Universidade do Minho
Escola de Engenharia

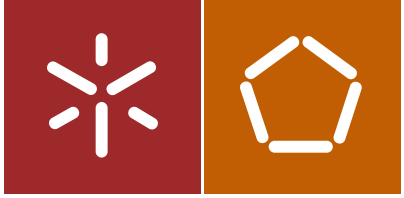
Lúcia do Carmo Fernandes de Araújo

Classificador de Literatura sobre Segurança de Sistemas de Informação

Classificador de Literatura sobre Segurança de Sistemas de Informação
Lúcia do Carmo Fernandes de Araújo

UMinho | 2014

Outubro de 2014



Universidade do Minho
Escola de Engenharia

Lúcia do Carmo Fernandes de Araújo

Classificador de Literatura sobre Segurança de Sistemas de Informação

Dissertação de Mestrado
Engenharia e Gestão de Sistemas de Informação

Trabalho efetuado sob a orientação do
Professor Doutor Filipe de Sá-Soares

Outubro de 2014

Declaração

É AUTORIZADA A REPRODUÇÃO INTEGRAL DESTA TESE/TRABALHO APENAS PARA EFEITOS DE INVESTIGAÇÃO E MEDIANTE DECLARAÇÃO ESCRITA DO AUTOR;

Universidade do Minho, 31 de Outubro de 2014

Assinatura: _____

Agradecimentos

Após um longo percurso e término deste projeto, que para mim, consistiu num extraordinário processo de aprendizagem, gostaria de registar o meu profundo apreço a todos quantos, de diferentes formas, me apoiaram na sua concretização. Agradeço a todos a cooperação, assistência, o saber e generosidade sem os quais não teria sido capaz de concretizar o mesmo, das quais gostaria de destacar:

Ao Prof. Dr. Filipe de Sá-Soares, meu orientador, pelo privilégio de comigo ter partilhado os seus conhecimentos e experiências, pelo interesse, ajuda e apoio que me prestou na resolução de dificuldades com que me ia deparando ao longo do projeto, pela disponibilidade para revisão do relatório, pela sua serenidade pelo fim deste projeto, entre outros aspetos.

Ao Professor Doutor Manuel Filipe Santos e Professor Doutor Paulo Cortez, pela forma segura e cuidadosa com que me guiaram na área do domínio de conhecimento de *Data Mining* e *Text Mining*.

Aos meus pais, José e Rosa, pelos valores incutidos, pela dedicação, apoio, compreensão e perseverança ao longo da elaboração deste projeto.

As minhas irmãs e tio João, pelo carinho e paciência que tiveram em todos os momentos da minha ausência.

Ao João Moreira pelo companheirismo, ânimo, chamadas de atenção, apoio constante, que foram fundamentais para a concretização da dissertação.

Para todos aqueles que me acompanharam, encorajaram e compreenderam os meus momentos de ausência, Muito Obrigada!

Orientador
Prof. Filipe de Sá-Soares

Autora
Lúcia Araújo

Classificador de Literatura sobre Segurança de Sistemas de Informação

Resumo

A quantidade de informação com que as organizações se deparam aumenta de dia para dia. A cada dia que passa as organizações estão mais dependentes e sobrecarregadas com informação. O modo como a organizar ainda se revela uma grande dificuldade para todos. Esta situação passa-se também com o conhecimento científico. A solução para este problema passa por classificar e organizar a informação de acordo com a semelhança dos conteúdos. O sucesso da classificação da informação passa sempre pela interação de um utilizador com a informação.

Para este estudo considerou-se como problema de investigação a ausência de um sistema de classificação abrangente para o conhecimento científico produzido sobre a Segurança de Sistemas de Informação (SSI).

Neste estudo intentou-se abranger a área da SSI como um todo e não apenas partes da sua área de investigação. Contudo, importa previamente destacar que não existe ainda um consenso claro entre os investigadores sobre o que deve ser incluído na área. A prévia compreensão de todos os aspetos referidos e de toda a sua envolvente julga-se permitir ao leitor perceber a grande importância deste estudo, assim como aos benefícios que do mesmo podem surgir. Contribuir para o preenchimento de um espaço de conhecimento em SSI insuficientemente explorado pelas investigações anteriores é o grande objetivo do presente estudo.

Orientador
Prof. Filipe de Sá-Soares

Autora
Lúcia Araújo

Classificador de Literatura sobre Segurança de Sistemas de Informação

Abstract

The amount of information that organizations face increases from day to day. Every passing day organizations are more dependent and overloaded with information. The way to organize information still proved a major difficulty for everyone. This also goes up with scientific knowledge. The solution to this problem is to sort and organize information according to the similarity of content. The success of the classification of information always involves the interaction of a user with information.

For this study was considered as research problem the lack of a comprehensive classification system for the scientific knowledge on the Information Systems Security (ISS).

In this study was attempt to cover the ISS area as a whole and not just parts of its research area. However, it is necessary first to point out that there is still no clear consensus among researchers about what should be included in the area. Prior understanding of all these aspects and all its surroundings its judged to allow the reader to realize the great importance of this study, including the benefits that the same may arise. Help to fill a space in SSI knowledge insufficiently explored by previous research is the main objective of the present study.

Índice

Índice de Figuras

Índice de Gráficos

Índice de Tabelas

Lista de Acrónimos

ACM – Association for Computing Machinery

BSI – British Standards Institution

CCS – Sistemas de Classificação de Computador

CIA – Confidencialidade, Integridade, Disponibilidade

CR – Computing Reviews

DM – Data Mining

DSSI – Desenvolvimento de Software de Sistemas de Informação

GR – Gestão de Risco

KDD – Knowledge Discovery in Databases

MISQ – Management Information Systems Ouarterly

NLP – processamento de linguagem natural

PLN – Processamento de Linguagem Natural

RI – Recuperação de Informação

SGBD – Sistemas de Gestão de Base de Dados/ DBMS - Database Management System

SGSI – Sistema de Gestão de Segurança da Informação

SI – Sistemas de Informação

SSI – Segurança de Sistemas de Informação

SVM - Support Vector Machines

TI – Tecnologias de Informações

TM – Text Minng

Classificador de Literatura sobre Segurança de Sistemas de Informação

“Thus, the task is not so much to see what no one yet has seen,
but to think what nobody yet has thought about that which everybody sees.”

Arthur Schopenhauer

Capítulo 1 – Introdução

1.1 Enquadramento

A era da informática tem vindo a revolucionar por completo o mundo do trabalho não havendo, na prática, nenhum sector da atividade humana que não tenha sido influenciado pelos computadores. O computador está a substituir cada vez mais o ser humano em todas as operações que podem ser codificadas de acordo com uma determinada sequência de operações simples e transformadas num programa. A título ilustrativo bastará considerar a gestão dos movimentos das contas de um banco, o controlo do aquecimento centralizado de um edifício ou ainda a supervisão de uma linha de montagem. Encontrou-se nos computadores instrumentos poderosíssimos para executar cálculos e processar informação em pouco tempo que, de outro modo, exigiriam anos de árduo trabalho. Pressupõem-se assim, que as tecnologias estão presentes em todas as organizações, que este é o meio utilizado para suprimir a informação, e que dado o elevado nível de globalização do sector tecnológico, a informação nas organizações terá de estar disponível em qualquer momento, sob a forma mais simples e correta possível, de fácil acesso.

Neste contexto, as organizações devem ser capazes de obter rapidamente a informação que apenas lhes interessa, no momento oportuno. Uma abordagem que pode auxiliar as organizações na obtenção da informação será a utilização de técnicas de categorização de informação capaz de selecionar apenas a informação pertinente, filtrando a informação que não é relevante. A informação, quando trabalhada e utilizada no momento certo pode revelar-se uma vantagem competitiva face meio em que está inserida num dado momento.

Porém, é importante que os utilizadores que recebem a informação saibam dar-lhe utilidade, transformando-a em ação, produção e produtividade para um entendimento comum. Do mesmo modo que a informação tem de ser trabalhada para um bom entendimento, a existência de consenso entre o que deve ou não pertencer a um domínio de conhecimento deve ser considerado e por isso a categorização de texto é a técnica de excelência do presente projeto para classificar um conjunto de documentos com uma ou

mais categorias existentes. No presente projeto esta técnica também é utilizada para organizar e filtrar informações.

Neste Capítulo introdutório identifica-se o âmbito do presente trabalho, os objetivos traçados para a sua execução, a abordagem de investigação mais apropriada e uma breve descrição da organização do estudo.

1.2 Motivação

A sociedade atual, o valor da informação e a importância das Tecnologias de Informação (TI) são, para este estudo, os aspetos chave. A sociedade atual está constantemente a ser "bombardeada" com informação. Essa informação pode ser de elevada importância de acordo com o âmbito em que se incluí. Na verdade as proporções com que a informação surge, por vezes, são bastante colossais e por isso as tecnologias de informação tentam dispor do meio mais adequado para limitar tamanha quantidade de informação.

Para responder a esta necessidade, a sociedade em cooperação com as tecnologias de informação dispõem de mecanismos de classificação que permitem reduzir de forma significativa a quantidade de informação que é transmitida ao utilizador final. Com base neste pressuposto pensou-se na importância da existência de um esquema classificativo para a área de investigação da presente investigação. Será explorado a viabilidade e existência de trabalhos já realizados neste âmbito, para assim justificar a base do presente estudo. Um dos problemas que está implícito no estudo de todo o conhecimento que engloba uma determinada área de investigação é o facto de esta potencialmente interligar grandes quantidades de informação.

Serão alvo de estudo técnicas que permitam destilar de forma rápida e eficiente grandes volumes de informação – área do *Text Mining* (TM). *Text Mining* é uma área de investigação que "tenta" resolver o problema de sobrecarga de informação usando técnicas do *Data Mining* (DM), aprendizagem máquina, processamento de linguagem natural (PLN), recuperação de informação (RI), e gestão do conhecimento" [Feldman e Sanger 2007a, p. X]. Por esta razão são criadas aplicações automatizadas que dão novos sentidos a palavras como classificar, arquivar e conservar, cujo objetivo é tornar racional e rentável o armazenamento e processamento de informação [Peixoto et al. 2003-2004]. A

classificação de documentos textuais é uma solução que permite separar a informação de acordo com o seu conteúdo, com o intuito de tornar a sua manipulação e auxilio nos processos de decisão mais rápidos, com uma margem de incerteza reduzida. A classificação de documentos surge com o propósito de tornar a informação uma mais-valia competitiva. No entanto, apesar de ser um sistema eficaz não dispensa a intervenção humana, uma vez que, dificilmente o computador está capacitado para lidar com todas as ambiguidades da linguagem humana. Classificar não é apenas uma tendência natural do conhecimento, mas também uma necessidade da inteligência humana. A razão da classificação de documentos facilita a recuperação da informação pretendida. De acordo com Jacobs [1992], a categorização de textos pode ser usada em três vertentes: na divulgação da informação (permite a distribuição de textos pelos utilizadores), na recuperação da informação (auxilia o utilizador a recuperar toda a informação que necessita e que está contida no sistema) e na navegação da estrutura do conhecimento (permite ao utilizador navegar por uma hierarquia até encontrar a informação que necessita).

Por fim, o propósito que encoraja a realização deste trabalho de investigação prende-se com, definir a área de estudo da Segurança de Sistemas de Informação (SSI) nas mais diversas temáticas, às já existentes na área. Para isso propor-se-á um conjunto de conceitos, retirados de estudos desenvolvidos pelos investigadores em anos anteriores, que depois de selecionados os mais aprimorados identifiquem com clareza cada temática da SSI. Este contributo permitirá reduzir a falta de acordo entre os investigadores da área sobre o que deve ou não estar incluído na lista de conceitos para descrever os conteúdos e conduz à existência de um universo conceitual para que seja facilmente compreensível os limites que são representados pelo conjunto de conceitos que o identifica.

Este estudo terá como foco contribuir para o preenchimento de um espaço de conhecimento em SSI ainda incompleto.

1.3 Objetivos

O presente trabalho de dissertação foi cuidadosamente planeado e por isso acredita-se na sua boa execução e no seu contributo inovador para a área de investigação. Reveste-se de um carácter teórico e aplicado. Teórico pois foi realizada uma revisão e

análise de literatura em relação aos SI, à SSI e ao TM. Aplicado pois foi desenvolvida uma solução que permitiu responder à lacuna identificada na presente área, já enumerada na Secção 1.2. Esta solução pode ser construída de duas formas, a primeira consiste na construção de raiz de uma aplicação que permita criar um esquema conceptual em SSI de forma semiautomática e a segunda possibilidade passará pelo uso de uma aplicação já existente que terá que ser configurada para o propósito em questão ou então acrescentado funcionalidades. Deste modo, a questão de investigação que preside este projeto é a seguinte: Como induzir semi-automaticamente o esquema conceptual temático da literatura em SSI? De modo a responder a esta questão foram formulados os seguintes objetivos.

Como primeiro ponto pretende-se obter as competências necessárias para o trabalho de dissertação, analisando e selecionando quais os métodos de investigação que serão mais adequados para o cumprimento do estudo. Este objetivo é de relevância acrescida uma vez que a escolha dos métodos determinará a consecução dos objetivos apresentados.

Em relação ao segundo objetivo, este consiste na revisão da literatura ao nível dos esquemas conceptuais temáticos para as áreas dos SI e da SSI, permitindo uma perspetiva abrangente dos esquemas conceptuais temáticos existentes que são mais relevantes, assim como a perceção das suas limitações e da relação que existe entre estes. A razão que leva ao estudo da área dos SI prende-se com o facto da área de SSI ser uma subárea dos SI. Este estudo compreende a possível relação que poderá existir entre esquemas conceptuais temáticos de ambas as áreas. Ainda neste objetivo é realizada uma revisão de literatura na área do *Text Mining*. Recorrendo às técnicas desta área de investigação será possível descobrir conhecimento através da análise de conjuntos de texto que retratem a literatura em SSI.

O terceiro objetivo visa a avaliação das ferramentas existentes que auxiliam a criação de esquemas de classificação para áreas de investigação. Este é um objetivo bastante importante uma vez que não é intenção deste trabalho classificar a literatura da área da SSI manualmente, e por isso é bastante relevante conhecer todas as aplicações que se debruçam sobre este aspeto.

No seguimento do ponto anterior tem-se um quarto objetivo que reside na aplicabilidade de uma ferramenta que permita o desenvolvimento *botom-up* do sistema de classificação para a literatura em SSI de forma semiautomática.

Para terminar, pretende-se que o último objetivo reporte todo o trabalho de dissertação realizado. Para isso serão analisados os resultados obtidos após a criação do esquema conceptual temático para a literatura em SSI, os aspetos críticos que estão subjacentes a este esquema, assim como o seu contributo para a área de investigação.

Espera-se que os objetivos apresentados sejam o meio para alcançar a resposta à questão de investigação definida no início deste capítulo. Como resultados do desenvolvimento deste projeto espera-se:

- Protótipo do classificador da literatura em segurança de sistemas de informação.
- Esquema conceptual da literatura em segurança de sistemas de informação.

1.4 Abordagem de Investigação

A presente secção prossegue com a apresentação da abordagem de investigação utilizada para o atual estudo.

O estudo retrata a análise e avaliação de trabalhos anteriormente realizados no âmbito de esquemas de classificação para a SI e a SSI. Este interesse foca-se na compreensão e exploração do que tem sido desenvolvido e no modo como responde às lacunas existentes nesse âmbito.

Um outro foco desta abordagem debruçou-se sobre a existência de aplicações que possam representar o conhecimento de uma área de investigação em que o requisito base é o texto não estruturado. Do diagnóstico realizado às aplicações existentes conclui-se que há uma ausência de *software* com os requisitos definidos para o estudo. Não é alvo deste estudo criar uma aplicação, mas, é alvo do estudo apresentar uma estrutura de conceitos bem definida recorrendo às aplicações existentes no mercado.

Para o projeto em causa é de salientar que este baseia-se numa revisão sistemática. Este tipo de revisão pretende responder a uma pergunta claramente formulada e utilizar métodos sistemáticos e explícitos para identificar, selecionar e avaliar criticamente as pesquisas relevantes e resultados devolvidos pela exploração dos dados. A revisão sistemática facilita a síntese de grandes quantidades de informação, reduz o risco

de erro, permite refazer todas as etapas de pesquisa, possuí um método estruturado e adaptável a diferentes contextos, os seus resultados são transpostos num formato fácil de ler e compreender, entre outros aspetos.

1.5 Organização da Dissertação

A presente secção descreve os capítulos que constituem este documento. No primeiro capítulo é enquadrado o projeto e apresentadas as razões que conduzem ao aprofundamento da área de conhecimento. O segundo capítulo descreve os fundamentos teóricos que alicerçam o estudo em causa, assim como a revisão de literatura no que concerne os aspetos relevantes da área. O terceiro apresenta toda a abordagem metodológica do projeto: problema e questão de investigação, objetivos do estudo e o método de investigação utilizado. No quarto capítulo é descrito em detalhe o trabalho realizado para alcançar o propósito deste estudo. O quinto capítulo apresenta as etapas e os resultados intermédios obtidos pela aplicação dos métodos de investigação. O sexto capítulo apresenta e discute os resultados finais obtidos neste projeto de investigação. O sétimo capítulo apresenta as contribuições, limitações, linhas de investigação futura que podem fazer com que este estudo prossiga para uma nova etapa de exploração, e conclusão final do estudo.

Capítulo 2 – Fundamentos do Estudo

2.1 Introdução

O presente capítulo será iniciado com a definição de terminologias específicas que envolvem a área em estudo. Outro aspeto que será abordado neste capítulo é a revisão de literatura pertinente para o problema de investigação discutido. Este processo tem um papel fundamental no que diz respeito ao âmbito em que este trabalho está inserido, contextualiza-o assim como "afunila" o estudo. Uma boa revisão de literatura é a base para a evolução do conhecimento [Webster e Watson 2002]. A secção 2.3 exibe detalhadamente o campo da revisão de literatura, sendo considerados trabalhos de investigação no âmbito da literatura da SI e da SSI. Um ponto também relevante, secção 2.4, apresenta os domínios da SSI já existentes e a secção 2.5 o "retrato" da literatura de "*Text Mining*". Tendo em conta que o objetivo do trabalho se insere na análise da literatura existente para a SSI, é importante recorrer a métodos que orientem a aquisição de informação qualificada a partir de texto de linguagem natural.

2.2 Definições Basilares

O projeto de dissertação retrata o campo de conhecimento da SSI. Será de elevada importância compreender o tema, assim como clarificar os conceitos que estejam relacionados com a área. Conceitos como organização, informação, sistemas de informação, gestão de sistemas de informação e segurança de sistemas de informação serão aqui aclarados com o objetivo de estabelecer uma base de entendimento comum entre o autor deste trabalho e a fácil leitura por parte do leitor.

Uma vez que a revisão de literatura mostra que diferentes investigadores apresentam diferentes entendimentos para os presentes conceitos, a presente secção tem por objetivo reduzir a incerteza da compreensão do leitor perante os conceitos apontados. De facto, será bastante visível que diferentes investigadores da área de estudo tenham um entendimento sobre determinado conceito bastante díspar dos seus pares.

2.2.1 Organização

O entendimento sobre o conceito organização tem ganho crescente abrangência quando considerado nas diferentes abordagens, variando apenas o foco para o qual este é apontado. Cada perspetiva ou abordagem salienta diferentes aspetos da organização. O entendimento proposto por [Schein 1972] concebe a organização como a coordenação racional de atividades de um conjunto de pessoas para a concretização de um objetivo ou intenção explícita e comum, devido à repartição do trabalho e funções, com hierarquia de poder e de responsabilidade. A organização pode então ser compreendida como um conjunto de recursos, quer sejam estes humanos ou não, organizados em função do cumprimento de uma missão [Rue e Holland 1986]. Segundo a perceção de Varajão [1998, p. 14], "organizações são entidades que surgem para satisfazer necessidades da sociedade que indivíduos isoladamente e outras organizações não conseguem (ou pelo menos não são capazes de o fazer igualmente bem), dependendo a sua sobrevivência e desenvolvimento da capacidade em afirmar a sua singularidade na satisfação das mesmas". Face à diversidade de entendimentos, Falkenberg et al. [1998] observam a organização como uma estrutura social com ações delimitadas e baseadas em informação, estabelecidas por metas, normas e regras de comportamento com características sistémicas.

Na construção de um entendimento para o conceito de organização é importante mencioná-la como um sistema social, aberto e dinâmico, no âmbito para o qual concretiza a sua atividade, em constante evolução e adaptação, com o intuito de garantir da sua existência e tendo como propósito os objetivos que deseja alcançar [Zorrinho 1991]. Autores como Dhillon e Backhouse [Dhillon e Backhouse 1994, 1996a, 1996b], Liebenau e Backhouse [1990], Galliers [1992b], Boland e Hirschheim [1987] consideram que a organização é vista sob três dimensões, como ilustrado na Figura 2.1.

A dimensão informal representa a cultura que carateriza a organização, onde se encontram definidas as regras, normas, compromissos e responsabilidades que governam as interações nas organizações. A dimensão formal é o conjunto de estruturas burocráticas e regras que regulam o comportamento dos elementos da organização. Já a dimensão técnica corresponde aos instrumentos que permitem mecanizar os elementos da dimensão formal. Tipicamente nas organizações modernas a dimensão técnica aponta para os sistemas informáticos como os seus principais elementos.

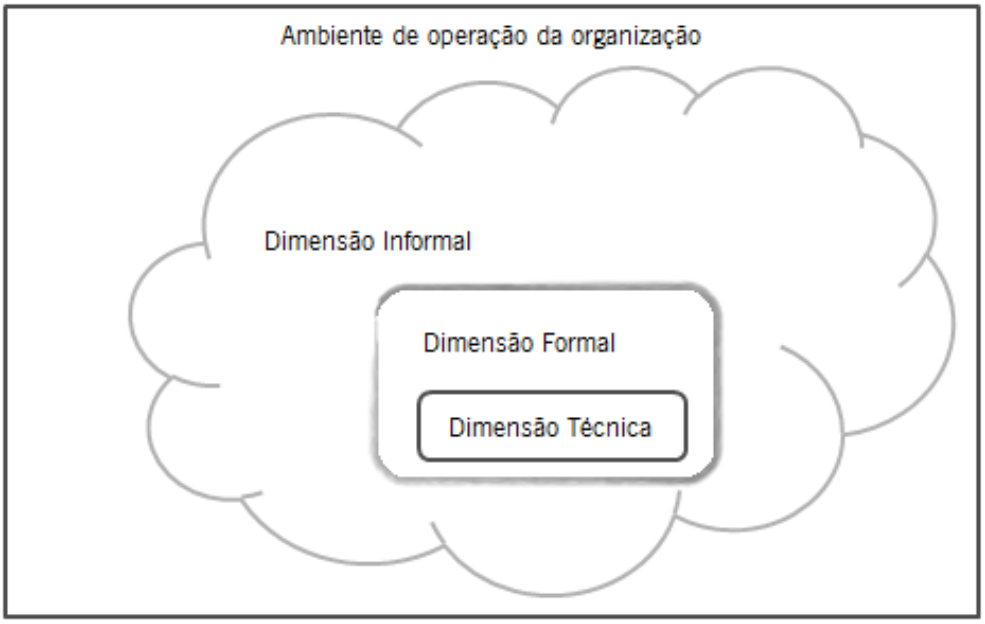


Figura 2.1: Dimensões de uma Organização/Sistema de Informação

Adaptado de Dhillon e Bachkouse [1994]

Alternativamente aos entendimentos já descritos subsistem várias outras opiniões sobre este conceito. Uma definição mais sistémica é a de Carvalho [2001], considerando que as organizações podem ser definidas como sistemas autónomos[1] viáveis[2]. O autor, ainda carateriza o conceito sob três perspetivas:

- Organizações como sistemas abertos (interação com o ambiente, em que existe a comunicação entre a organização e os seus clientes, os seus fornecedores e outras entidades que afetam o seu funcionamento);
- Organizações como sistemas cooperativos (cooperação entre agentes organizacionais para que a organização cumpra a sua missão executando as atividades que derivam da missão da organização));
- Organizações como sistemas viáveis (controlo e regulação – verifica se os objetivos definidos estão a ser cumpridos e quais as alterações que se estão a verificar no exterior).

Para o autor, os termos são designados por:

[1] Autónomos - existência independente.

[2] Viáveis - capacidade de adaptação às alterações no ambiente.

As formas mencionadas têm como resultado a comunicação que surge da interação e cooperação na organização. Paralelamente à perspetiva anteriormente mencionada, aponta-se um outro entendimento que atenta diferentes translações sobre este conceito. Morgan [2006] define que o conceito de organização é complexo e por isso aponta várias perspetivas de definição:

- Organização como uma máquina, com a mecanização a assumir o comando;
- Organização como organismos, baseado na natureza;
- Organização como cérebro, devido à aprendizagem e auto-organização;
- Organização como culturas, na presença da criação de realidade social;
- Organização como sistema político, com interesses, conflitos e poderes;
- Organização como prisões psíquicas, fruto da mentalidade dos colaboradores;
- Organização como fluxo de transformações, uniformidades de desdobramento da mudança;
- Organização como instrumentos do domínio, reconhecida como o rosto feio.

Segundo os termos apresentados, uma organização é formada por agentes que colaboram entre si com o objetivo de garantir a viabilidade da organização, cientes da fragilidade de alterações no seu ambiente, e que se baseia na satisfação dos seus clientes no âmbito da sua missão/finalidade, recorrendo a fornecedores para obter recursos que necessita para satisfazer os seus clientes. Abordagens simples sobre este conceito, definem organização como sistemas de natureza aberta e em constante ação com o meio envolvente.

Como se pode observar, o conceito de organização pode ser interpretado de várias formas. Por isso, considera-se importante esclarecer o entendimento adotado neste projeto. Assim, a organização é entendida, no âmbito deste trabalho, como um sistema complexo onde interatuam diferentes subsistemas inter-relacionados e interdependentes que interagem com o ambiente externo. A sua evolução é resultado da adaptação ao meio envolvente, da dinâmica interativa e evolução coerente dos subsistemas, não desprezando a constante atenção que a organização deve atender às alterações do ambiente para assim poder adaptar-se a novas mudanças, mantendo-se estável no mesmo e transformando-as em seu benefício [Varajão 1998]. Para concluir, é importante referir o elo inseparável que existe entre organização e sistema de informação, elucidando que o entendimento e compreensão de um

termo envolve o conhecimento e entendimento do outro, ou seja, para compreender um sistema de informação é essencial compreender o conceito de organização e vice-versa, isto porque ambos são inseparáveis. Assim como compreender o conceito de organização revela-se de elevada importância, é igualmente imprescindível ter uma compreensão correta não apenas do papel que a informação e as TI/SI desempenham na sociedade, como também a forma como são geridos os SI e quais os seus requisitos de segurança. Estes serão conceitos abordados nas próximas subsecções.

2.2.2 Informação

O conceito de informação é sem dúvida um conceito difícil de explicar. A informação permite comunicar conhecimento e aperfeiçoar o pensamento do ser humano. Intuitivamente, todo o ser humano reconhece a informação, todavia definir o conceito e identificar as suas fronteiras revela-se uma tarefa mais complexa.

Estando ciente do contexto deste projeto, o conceito de informação será perspetivado com foco na organização. Pretende-se que seja entendido desta forma uma vez que pela informação uma organização encontra-se em constante ritmo de mudança. A informação motiva todas as ações da organização no processo de tomada de decisão, no alcance da sua missão e objetivos que justificam e motivam a sua existência, com elevado nível de competitividade com o meio onde interage, possibilitando uma componente diferenciadora. Atualmente informação é encarada como uma "arma" estratégica para as organizações, que mediante a tomada de decisão por parte dos intervenientes é passível de adquirir vantagens competitivas para quem dela usufruí. Para Varajão [1998, p. 47] as "organizações têm consciência da informação como recurso estratégico que lhes permite competir de modo diferente através da criação de novos negócios, do estabelecimento de relações de força com clientes e fornecedores, e da sua própria renovação interna."

Alguns autores, como Liebenau e Backhouse [1990], entendem a informação como o resultado de um processo que considera que tudo o que se transmite pode ser compreendido, sendo concretizado por pessoas que interagem na organização. As organizações veem a informação como um recurso que faculta a interligação com o seu ambiente, permitindo compreender as suas necessidades, condicionantes e pressões, divulgar o mercado em que operam e serviços que prestam, competindo e adquirindo conhecimentos úteis para o seu funcionamento [Zorrinho 1995]. Um estudo mais lato sobre

este conceito define informação como o conjunto de "fenómenos [3] /objetos [4] /coisas [5] simbólicos, externos à mente humana, criados deliberadamente para serem usados no contexto de atividades humanas em operações que envolvem alguma forma de comunicação e/ou cognição" [Carvalho 2012, p. 2]. Por ultimo, Varajão [1998, p. 45] entende informação como sendo "um conjunto de dados, colocados num contexto útil e de grande significado que, quando fornecido atempadamente e de forma adequada a um determinado propósito, proporciona orientação, instrução e conhecimento ao seu recetor, ficando este mais habilitado para desenvolver determinada atividade ou decidir".

Cada vez mais as organizações têm consciência de que é a interligação entre a informação que possuem e os seus sistemas de informação permite-lhes maior eficiência e vantagem competitiva. Contudo, devido ao crescimento contínuo da informação e ao fluxo de informação, torna-se cada vez mais relevante gerir e proteger adequadamente essa informação.

2.2.3 Sistema de Informação

O papel que os SI desempenham nas organizações e na sociedade atual nem sempre é fácil de compreender, embora a sua utilização tem-se vindo a expandir e a surgir em novas áreas. Atualmente os SI são reconhecidos como uma área da ciência bastante madura e da qual as organizações muito dependem, mas, que ainda gera algum conflito quando se pretende contextualiza-la devido à diversidade de interpretações existentes.

Um possível esclarecimento para o termo sistemas de informação consiste no conjunto de pessoas, *hardware*, *software*, redes de comunicação, armazenamento de dados, e políticas e procedimentos que armazenam, recuperam, transformam e distribuem a informação numa organização [O´Brien e Marakas 2009]. Ainda nesta perspetiva, e validando o conceito anterior, sistemas de informação podem ser entendidos como um sistema que recolhe, processa, armazena e distribui informação pela organização, estando esta acessível a quem dela necessita [Buckingham et al. 1987]. Neste cenário, um sistema de informação pode ser suportado por computador que é usado num contexto organizacional [Alter 1996; Mentzas 1994; O'Brien 2000]. Visões mais tradicionais,

---

Carvalho [2012, p. 2] considera que o uso dos termos surge devido à enorme diversidade que os objetos informacionais podem assumir:

[3] "Fenómeno deve aqui ser entendido no sentido filosófico como correspondendo a algo que pode ser notado pelos sentidos ou pela mente humana; também no sentido normalmente usado em filosofia";

[4] "Objeto sugere algo com existência independente e externa à mente pensante";

[5] "Coisa sugere algo que não podemos ou queremos referir com designação mais específica, eventualmente por nos querermos referir a algo muito geral";

consideram que os sistemas de informação estavam relacionados com sistemas informáticos, mas com o passar do tempo esta ideia foi-se desmaterializando e atualmente considera-se que este conceito não está totalmente relacionado com o computador [O´Brien e Marakas 2009]. Já Bretschneider e Wittmer [1993] avançam com uma definição intermédia, em que aceitam que um sistema de informação possa não estar relacionado com o computador, mas observando o mundo real concluem que para uma organização é invulgar a existência de um sistema de informação que não esteja integrado com o computador.

Poder-se-á também dizer que um sistema de informação corresponde ao conjunto estruturado de procedimentos que, quando executados, geram informação para a organização. Porém, este termo pode ser visto com diversas interpretações para diversos grupos de pessoas, sendo difícil definir uma terminologia única para o domínio dos sistemas de informação [Falkenberg et al. 1998]. "De facto, a disciplina de Sistemas de Informação aparece como uma das consequências da proliferação massiva dos computadores nas organizações e na sociedade em geral, mas aparece claramente, não como uma extensão ou complemento da Informática, mas sim como um projeto próprio de investigação e ensino. " [Magalhães 1997, p. 53]

Pode observar-se que as opiniões entre os investigadores em relação ao conceito de SI são algo díspares. Autores como Ein-Dor e Segev [1993, p. 167] definem este conceito como "qualquer sistema computorizado com um utilizador ou operador de interface é um sistema de informação, desde que o computador não seja fisicamente embebido". Para Falkenberg et al. [1998, pp. 72-73] os sistemas de informação são vistos como "um subsistema de um sistema organizacional, compreendendo a conceção de como a comunicação – e informação – aspetos orientados de uma organização são compostos (...) e como esses operam, descrevendo assim as (...) ações orientadas à comunicação e fornecedoras de informação e acordos existentes dentro da organização". Estes autores ainda consideram que este conceito apresente três dimensões: a técnica, a social e a conceptual. Segundo o esclarecimento de Carvalho [2000, p. 6] face a este termo, sistemas de informação são "um objeto ativo que lida com (processos) informação" ou "um objeto cujo propósito é informar". Nesta linha, e para os objetos que o autor identifica, reconhece que estes podem ser observados pela ótica da organização, de sistema autónomo, de sistema baseado em computador, ou de atividade da organização, sem contudo se limitarem apenas a uma das óticas, isto é, um objeto pode estar presente apenas numa das visões ou em todas elas, como pode não pertencer a nenhuma delas. O objetivo do autor estava em definir os sistemas de informação segundo as quatro óticas/objetos apresentados, tendo por base as definições já mundialmente reconhecidas e agrupando-as segundo os quatro objetos

enumerados. Por fim, e na perspetiva de Livari [1992] os sistemas de informação são vistos como uma generalização dos objetos que a constituem, tentando não identificar as fronteiras entre sistemas de informação e organização, traduzindo o entendimento existente entre ambos.

Perante a discussão apresentada, e como se pode observar, existem diversos entendimentos para o conceito de SI. Uma possível justificação para esta realidade pode estar relacionada com o facto dos contributos para o crescimento desta área terem origem em distintas áreas como: ciências da computação, engenharia de computadores, ciências da organização e gestão, economia, sociologia, psicologia, conduzindo para que este conceito tenha a falta de consenso entre os investigadores dos SI [Falkenberg et al. 1998; McLean 1982; Westin et al. 1994]. Porém, e apesar dos autores citados poderem de forma simultânea apresentar definições muito semelhantes, todos eles identificam com clareza as componentes a que se referem para definir conceito, uma vez que ainda existem aspetos na sua definição que ainda não são claros e só desta forma é que se consegue um entendimento mais claro do conceito.

Ao examinar todo o conjunto de autores mencionados e com o objetivo de atribuir ao estudo em causa uma definição para o conceito de SI, considera-se a de Sá-Soares [2005, p. 27] que descreve SI como "um sistema social que tem por finalidade apoiar a significação e ação organizacionais através da síntese organizada de informação". Sabe-se que um SI é um sistema onde o foco de base é a informação, cuja finalidade é armazenar, tratar e fornecer informação para os processos e serviços das organizações. Convém realçar que este sistema é constituído pelo subsistema social (pessoas, processos, informação, documentos) e subsistema automatizado (máquinas, computadores, redes de comunicação) e que ambos estão interligados [Buckingham et al. 1987].

2.2.4 Gestão de Sistemas de Informação

Com a evolução das tecnologias de informação e o constante aumento do volume de informação que tem de ser processado, os SI tem assumido um crescente nível de importância nas organizações. Nesse sentido os SI da organização necessitam ser geridos de uma forma atenta, constante, profissional e de acordo com as melhores práticas disponíveis. Esta atividade é comumente designada por Gestão de Sistemas de Informação (GSI). Uma boa gestão de sistemas de informação permite benefícios notáveis, dentro os quais o melhoramento e facilitação do processo de tomada de decisão e do planeamento organizacional.

A natureza dinâmica da GSI é refletida como uma atividade que tem de estar em contínua atualização de modo a satisfazer as necessidades sociais da informação, sendo esta, única para cada organização. Por exemplo, a informação pode ser a mesma para duas organizações, mas terá um impacto diferente para as mesmas duas organizações, isto porque cada uma delas interpreta a informação retida de diferentes maneiras. O papel desempenhado pela atividade de GSI é a gestão da informação. Este regula, controla, processa e envia a informação necessária e útil aos destinatários da organização que necessitam da mesma para que de seguida possam identificar problemas, gerar informação, comunicação e auxiliar o processo de tomada das melhores decisões possíveis.

Tal como foi possível verificar nos conceitos anteriormente descritos existe uma diversidade de entendimentos. Para o presente conceito, gestão de sistemas de informação não será exceção. Para alguns autores, a "Gestão de Sistemas de Informação (GSI) é a atividade com preocupações que englobam as Tecnologias de Informação e Comunicação (TIC), o Sistema de Informação (SI) e a Gestão da Informação (GI)." [Gouveia e Ranito 1997, p. 46] Uma outra perspetiva defende que esta área de estudo consiste na conjugação das atividades de planeamento e desenvolvimento de sistemas de informação [Barki et al. 1993; Jordan 1993]. Paralelamente, Carvalho e Amaral [1993, p. 37] definem a gestão de sistemas de informação como "a atividade de gestão aplicada a uma das áreas funcionais das organizações cuja existência e importância tem vindo a ser reconhecida por diversos autores - a função de informação."

Alternativamente, Amaral [1994, p. 36] define este conceito como sendo a "Gestão do Recurso Informação e de todos os recursos envolvidos no planeamento, desenvolvimento, exploração e manutenção do SI." Um outro entendimento sobre este conceito, foca que a GSI permite às organizações no momento pretendido e para as pessoas desejadas, obterem um acesso a informação viável nos seus sistemas de TI [Reddy et al. 2009, p. 5].

Estando cientes de que os SI são sujeitos a períodos de excessos de informação, permanente estado de mudança da informação e formação de pequenos conjuntos de informação na organização é essencial perceber o conceito aqui apresentado. A informação bem "cuidada" pode ser a "arma" estratégica para a organização, defendendo-a das ameaças para o negócio, competindo de forma inteligente e transformando a organização, aumentando a sua interação.

Por isso, a GSI auxilia as atividades de uma organização, facilitando o suporte de decisão. Conclui-se também que uma boa gestão de sistemas de informação potencia o sucesso de uma organização. Será este o entendimento utilizado para este termo no presente projeto.

2.2.5 Segurança de Sistemas de Informação

A transformação dos SI em sistemas cada vez mais complexos e sofisticados, assim como a sua ascensão, provocou mudanças organizacionais bastante acentuadas. Anteriormente, dados eram preservados em formato papel e salvaguardados no interior da organização, mas com a criação dos computadores os dados começaram a ser controlados pelos repositórios de dados, que eram acedidos pelas diversas pessoas que compõem uma organização.

O mundo informático desenvolveu-se de tal forma que graças à utilização massiva da internet e dos seus serviços, pessoas externas à organização podem aceder com maior facilidade a esses dados, causando a destruição, a fraude, uso indevido da informação, e sem esquecer o acesso não autorizado. Daí, o número de ataques a que a informação está exposta é cada vez mais elevada, e consequentemente, a necessidade de proteger os sistemas de informação é crucial. Esta preocupação por parte das organizações guiou-as no sentido de implementarem sistemas de segurança no seu SI. Todavia, estabelecer os requisitos de segurança necessários para um SI numa organização veio a revelar-se uma tarefa com alguma dificuldade [Dhillon e Backhouse 2000a], sendo fundamental para uma organização reconhecer a segurança do SI como um potenciador do negócio, criar estratégias de SI, planear os benefícios resultantes do uso de segurança, conceber uma análise de risco, alocação de recursos e responsabilidades e, finalmente, identificar os controlos organizacionais apropriados.

Segundo Dhillon [1997] a segurança de sistemas de informação numa organização é vista sobre o prisma de sistema técnico, formal e informal. Este mesmo autor, defende ainda que a SSI quando focada nas atividades de manipulação de informação numa organização, é definida como um requisito de cuidado e proteção desta mesma informação, podendo ser influenciada pelas mudanças de contexto organizacional em que é compreendida e utilizada [Dhillon e Backhouse 2000a]. Para de Sá-Soares [2005, p. 29] a SSI é vista como "...um conjunto de meios através dos quais se garante a proteção do SI" ou Kim e Solomon [2010,

p. 8] que simplificam o entendimento da definição anterior, afirmando que a SSI protege os dados armazenados no SI. Dhillon e Backhouse [2001] compreendem a informação e a segurança de sistemas na perspetiva socio-organizacional. Para estes autores, estes sistemas auxiliam a colaboração entre indivíduos e grupos possibilitando a partilha de informação e processos de trabalho, e na realização de transações entre os parceiros. Para Gehrke et al. [1992] e, Herath e Rao [2009], a segurança é vista como um requisito das tecnologias de informação que previne das ameaças acidentais ou intencionais, depende das pessoas, dos processos e da tecnologia da organização.

À luz da discussão efetuada, considera-se que de forma a alcançar o bem-estar das organizações, será importante implementar variáveis de governação que tenham em atenção a confidencialidade, integridade e disponibilidade da informação nos seus sistemas de informação [McCumber 1991] com o intuito de prevenir, detetar e responder às ameaças a que estes sistemas estão expostos, e desta forma, proteger a informação:

- Confidencialidade – permite garantir que a informação está protegida da divulgação indevida, isto é, apenas colaboradores autorizados têm acesso à informação;
- Integridade – assegura que a informação que é manipulada mantenha uma correta consistência e representação física dos dados durante todo o seu ciclo de vida;
- Disponibilidade – garante aos utilizadores autorizados um constante acesso à informação para o seu uso.

A boa coordenação destes três princípios tradicionais – confidencialidade, integridade, disponibilidade (CIA) - permite que os objetivos definidos por uma organização sejam concluídos com elevado sucesso, uma vez que os seus sistemas de informação serão mais credíveis. Numa organização, a segurança proporciona a criação de novas oportunidades de negócio, uma vez que, o negócio está cada vez mais relacionado com a tecnologia e, por isso, subsiste a necessidade de refletir sobre o uso dos princípios básicos de segurança da informação como um meio de salvaguarda da informação. Os três princípios tradicionais apenas focam a solução de problemas ao nível dos dados que são armazenados nos sistemas computacionais [Dhillon e Backhouse 2000b]. Para estes autores, a SSI também deve ser

entendida como um impulsionador do negócio e, por isso, propõem quatro princípios – RITE – como complemento às variáveis de governo tradicionais de SSI:

- Responsabilidade – diz respeito à clara definição dos papéis desempenhados pelas pessoas das organizações, e o conhecimento dos mesmos, pelos próprios, e à sua responsabilização;
- Integridade – diz respeito à integridade das pessoas das organizações, quer na dimensão pessoal como profissional. Este traço deve ser uma característica de natureza mandatória;
- Confiança – nas organizações modernas a enfase numa relação eficiente entre as pessoas e a organização não deve ser o controlo excessivo. Assim, o auto-controlo e a responsabilidade das pessoas devem encontrar-se na base da atuação de todos. A organização poderá assim confiar que os seus colaboradores se regem pelas normas e regras por si definidas mantendo o seu bom desempenho e segurança;
- Ética – a organização define um conjunto de normas e regras para a eventual ocorrência de em determinadas situações os seus colaboradores as concretizarem. Contudo, podem ocorrer situações que não sejam previsíveis e por isso, a organização espera que os seus colaboradores ajam de acordo com os princípios éticos.

Mediante os diversos objetos apresentados pelos autores para definir o conceito de SSI, considera-se que o trabalho de literatura realizado por de Sá-Soares [2005] classifica este entendimento de modo bastante exato e abrangente, onde a SSI é vista como:

- Um estado – que traduz o nível de integridade da informação. Um estado de cuidado e proteção para com os SI;
- Um meio - através dos quais se garante a SSI como sendo de natureza estrutural, e o conjunto de produtos e recursos que o fazem são os controlos;
- Um processo – que evita ataques e ameaças acidentais e intencionais, garantindo a segurança dos computadores;
- Uma área de conhecimento – vocacionada para a investigação e para o ensino.

Porém, considera-se que ainda devem ser realizadas mais investigações para reconhecer a inclusão de variáveis alternativas, sem descurar, contudo, que deve ser a organização a definir a configuração de variáveis de governo que mais se adequa ao seu SI de modo a alcançar a sua segurança.

Após tudo o que foi referido, reconhece-se que a SSI inclui áreas como a gestão de segurança de informação, segurança de dados e computação (*computer and data security*) e segurança de redes [Whitman e Mattord 2011], concluindo-se ainda que garantir a segurança de sistemas de informação ainda é uma tarefa difícil e por vezes inacessível para os sistemas mais complexos, acrescentando também o valor dispendioso que é necessário para a sua implementação [Chaula et al. 2005].

2.3 Âmbito da Revisão da Literatura

2.3.1 Corpo de Conhecimento de SI

A mente humana classifica objetos de forma consciente ou inconsciente. Uma das expressões máximas desta capacidade consiste na sistematização do conhecimento que ao longo dos séculos os seres humanos foram realizando nas mais diversas áreas do saber. Classificar veio a revelar-se uma necessidade da inteligência humana proporcionando uma melhor organização do conhecimento para a mente humana. Segundo Brascher e Café [2008, p. 6], organização de conhecimento "(...) visa à construção de modelos de mundo que se constituem em abstrações da realidade."

O campo dos SI tem sido sujeito a grandes desenvolvimentos. Uma prova desta evolução está associada ao aumento do número de conferências, centros de investigação, publicações em revistas, entre outros aspetos que foram surgindo ao longo dos anos. O aumento do volume da produção científica e o crescimento do campo de SI conduz a maior especialização e divisão em subcampos. Investigadores têm revelado ter sérios problemas relacionados com as grandes quantidades de conhecimento acumulado nas áreas de investigação de SI [Barki et al. 1988].

O aumento exponencial de informação, a sua constante transformação e a dificuldade em saber o que deve ou não estar presente no domínio de uma determinada área, ganham cada vez mais dimensão ao longo do tempo. Confrontados com esta realidade, a necessidade

da criação de um sistema de classificação parece ser a solução adequada para organizar e mapear o território de uma área em estudo. Esquemas de classificação ou taxonomias são ferramentas importantes para representar o conhecimento de uma área da ciência, proporcionando um consenso geral sobre o conteúdo que está incluído nessa área, agrupando os conceitos com características comuns e segundo hierarquias, possibilitando uma clara compreensão dos conceitos e o relacionamento entre esses conceitos. Estes relacionamentos podem ser hierárquicos – conceitos organizados segundo a sua função e tipo de aplicação, ou sintático – conceitos podem pertencer a hierarquias diferentes, mas a representação é feita por classes compostas. Como resultado final obter-se-á uma rede ou estrutura de relacionamentos para a área de conhecimento em estudo.

O conjunto de características que cada esquema de classificação apresenta é único, permitindo maximizar a sua organização, gestão e recuperação da informação. A sua criação exige um vocabulário padrão, os termos selecionados devem apresentar critérios previamente identificados (coerência no tratamento dos dados) e o *software* a utilizar deve ser adequado às características do foco pretendido. Esta realidade das necessidades de informação possibilita uma navegação pelos termos que se apresentam de forma lógica (classes, sub-classes, sub-sub-classes) pelos diversos níveis de especificidade, facilitando a procura ao utilizador para o termo pretendido. É importante aclarar com rigor o propósito da classificação, de forma a conseguir eliminar as propriedades irrelevantes, apontando apenas as características que possibilitam a distinção entre os objetos.

Para Woods [2004] existem dois tipos de taxonomias: a taxonomia corporativa que facilita a representação de toda a informação existente na organização e a taxonomia clássica que considera que cada termo apenas e unicamente pode pertencer a um ramo da estrutura hierárquica. Em suma, "a organização das informações através do conceito de Taxonomia permite alocar, recuperar e comunicar informações dentro de um sistema de maneira lógica através de navegação" [Campos e Gomes 2007, p. 3].

As subsecções seguintes abordaram exemplos de organização de conhecimento para a área de SI, compreendendo de que modo estes podem contribuir para a realização de uma solução para a lacuna apresentada co Capítulo 1.

2.3.1.1 Esquema de Barki et al. [1988]

Em Junho de 1988, foi publicado na *Management Information Systems Ouarterly* (MISQ) o primeiro esquema de classificação para a literatura de SI proposto por Barki et al. [1988]. Este esquema tinha por base a análise de *keywords* da área, sendo composto por 1100 *keywords*, nove categorias no nível superior (disciplinas de referência, ambiente externo, ambiente tecnológico, ambiente organizacional, gestão dos SI, desenvolvimento e operações dos SI, uso dos SI, sistemas de informação e educação e investigação em SI) que incluem subcategorias, representado por uma estrutura em árvore onde os dois primeiros níveis são designados por letras dos abecedário e os restantes por números. O uso cuidado de *keywords* permite identificar com facilidade os conteúdos mais importantes de um documento de investigação, arquivar ou identificar diretorias com facilidade, sendo ferramentas valiosas na pesquisa de informação por meio do uso de computador. Um esquema de classificação por keyword considera a existência de quatro componentes: agendar (índice de vocabulário), notação (símbolo/identificados da categoria), índice (lista em ordem alfabética com todos os termos e assuntos incluídos no esquema e respetivas relações com a estrutura), e organização para a manutenção e revisão (mantém o esquema útil e utilizável) [Barki et al. 1988]. A Figura 2.2 exemplifica a estrutura do esquema proposto, podendo observar-se com facilidade a disposição das categorias e subcategorias.

E	IS Management
EA	Data Resource Management
	UF Data Administration
EB	Personnel Resource Management
	USE IS Staffing
EC	Administration of Computer Centers
ED	Hardware Resource Management
EE	Software Resource Management
EF	IS Project Management
EG	IS Planning
EH	Organizing IS
EI	IS Staffing

EJ	IS Evaluation
EK	IS Control
EL	IS Security
EM	IS Management Issues

Figura 2.2: Esquema Classificativo de SI ao Nível Superior de Barki et al.

Adaptado de Barki et al. [1988]

Cada uma das categorias apresentadas pode ainda incluir várias subcategorias. A Figura 2.3 exibe o que pode ser incluída na área de *IS Security* no esquema proposto por Barki et al. [1988].

EL	**IS Security**
EL01	DataSecurity
	UF Data Protection
EL02	Data Encryption
	EL0201 Public Key Encryption
EL03	Access Control
EL04	Physical Security
EL05	Authentication
EL06	Authorization
EL07	Passwords
EL08	Disaster Plans
	EL0801 Recovery

Figura 2.3: Sub-categoria do Esquema de SI de Barki et al.

Adaptado de Barki et al. [1988]

Os autores Barki et al. [1988], fundadores do esquema apresentado, consideram que o enquadramento de Ives et al. [1980] é a base para uma estrutura completa e adequada no desenvolvimento de um esquema de classificação para os SI. Para os autores este enquadramento é mundialmente aceite, já que incluí questões de SI ao mais alto nível, com adequada captação na estrutura de investigação em SI e averiguação da sua adaptação nos diferentes estudos de investigação. Esta *framework* foi avaliada como sendo a base para a primeira versão do esquema de classificação de SI, com o objetivo de analisar de que forma as *keywords* podem ser incorporadas a partir da literatura de SI. Para a criação do esquema, foram estudadas todas as *keywords* usadas em artigos publicados na MISQ, na *Information & Management* desde 1977, e a lista de *keywords* para artigos de SI na *Communications of the*

ACM, *Management Science* (desde Agosto de 1978) e *Design Science*, desde Janeiro de 1980. Analisadas e eliminadas as *keywords* repetidas da lista, totalizando 2000 *keywords*, foram analisadas e debatidas uma a uma sobre a classificação em que seriam incluídas ou a necessidade de criar subcategorias.

De forma a avaliar todo o trabalho efetuado pelos autores, foram levadas em consideração três questões: a primeira prende-se com a integridade do esquema, a segunda com a solidez da estrutura e por fim, a terceira com a potencial convergência e normalização. Estas questões foram distribuídas por 86 investigadores de SI do Canadá e Estados Unidos, tendo sido alcançado uma taxa de resposta de 55,8% [Barki et al. 1988]. Os resultados obtidos permitem sugerir que o esquema de classificação é fortemente satisfatório, porque apenas nove das 343 palavras propostas pelos investigadores não foram incluídas no esquema e quando solicitados a indicar a classe a que cada keyword pertencia no esquema a taxa média de assertividade foi de 68% [Barki et al. 1988]. "O nível de maturidade alcançado pelo campo de SI tornou o desenvolvimento de um esquema classificativo para que a indexação da sua literatura viável e desejável" [Barki et al. 1988, p. 309].

O esquema proposto pelos autores não é tido como final, pois ao longo do tempo aquele sofre alterações face à evolução pela qual passa o campo de investigação e pela contribuição que outros investigadores da área possam dar. Deste modo, em resposta a esta evolução Barki et al. [1993, p. 209] afirmam "o desenvolvimento deste esquema destinava-se a fornecer uma descrição da disciplina, introduzir uma linguagem comum, e permitir a pesquisa do desenvolvimento do campo".

Nesta linha de investigação, em 1993 os autores resolveram apresentar a evolução da disciplina de SI no que diz respeito aos temas e métodos de investigação. Foi então proposto um novo esquema de classificação baseado em artigos de *Information Systems Research* e *Organization Science* desde o seu início até o outono de 1992. Todos os artigos publicados entre 1987 e 1992 da MISQ, *Journal of MIS*, *Information and Management*, *Management Science* e *Communications of the ACM*, foram analisados. Para realizar todo este trabalho, Barki et al. [1993, p. 209] indicam que "o número total de *keywords* identificadas a partir deste exercício foi aproximadamente 2.000 *keywords*. Cada uma dessas *keywords* foi examinada por todos os três autores em relação ao esquema de classificação, e uma decisão foi feita com relação à sua inclusão. Em todos os casos os autores chegaram a acordo sobre se deve ou não incluir uma palavra-chave e quando no esquema colocá-lo".

Para este novo esquema de classificação são incluídas 1.300 *keywords*, revelando um acréscimo de mais de 175 novas palavras face à versão anterior, mas com 10 *keywords* excluídas. Ainda assim, a nova versão do esquema apresenta a mesma estrutura que o anterior, composta por categorias com quatro níveis onde os dois primeiros níveis são identificados por letras (do A até ao I) e os restantes níveis por números, conforme se mostra na Figura 2.4.

E	**IS Management**
EA	Data Resource Management
EB	Personnel Resource Management
EC	Hardware Resource Management
ED	Software Resource Management
EE	IS Project Management
EF	IS Planning
EG	Organizing IS
EH	IS Staffing
EI	IS Evaluation
EJ	IS Control
EK	IS Security
EL	IS Management Issues

Figura 2.4: Esquema Classificativo de SI ao Nível Superior de Barki et al.

Adapatado de Barki et al. [1993]

Porém, esta nova versão apresenta algumas diferenças quando comparada com a versão anterior: foi acrescentada uma nova categoria ao nível superior, assim como ao nível da subcategoria pode sofrer alterações, tal como se indica na Figura 2.5.

Ao identificar a mesma área de *IS Security* para exemplificar a estrutura do esquema mais atualizado, é possível ver a evolução da área. Já é notória a presença de mais subcategorias e com código de designação diferente.

É evidente que estes tipos de esquemas são bastante vantajosos para áreas de investigação. Nas palavras de Barki et al. [1993, p. 209], o esquema de classificação para a literatura dos SI deve "Primeiro, define o campo de SI em algum detalhe. Segundo, fornece

um vocabulário comum. Em terceiro lugar, fornece uma ferramenta com a qual a evolução da investigação pode ser estudada". Todavia, os dois esquemas anteriores não são os únicos esquemas desenvolvidos em torno da literatura de SI.

Vários investigadores pretendem que o seu conhecimento esteja organizado e que acompanhe as evoluções a que o campo está sujeito. É esta uma das perspetivas sobre a qual a ACM recaiu, na constante atualização do esquema classificativo.

EK	IS Security
EK01	DataSecurity
	UF Data Protection
EK02	Data Encryption
	EK0201 Public Key Encryption
EK03	Access Control
EK04	Physical Security
EK05	Authentication
EK08	Authorization
EK07	Passwords
EK08	Disaster Plans
	EK0801 Recovery
EK09	Computer Crime
EK10	Computer Viruses

Figura 2.5: Subcategoria do Esquema de SI de Barki et al.

Adaptado de Barki et al. [1993]

2.3.1.2 ACM Computing Classification System

O *ACM Computing Classification System* (CCS) é um sistema de classificação e de indexação da literatura por temas para a computação, desenvolvido pela *Association for Computing Machinery* (ACM) [ACM 1998]. O trabalho realizado pela ACM é apresentado em seis versões, a primeira em 1964, prosseguindo a evolução para 1982, 1983, 1987, 1991, 1998, sendo esta última a versão mais atualizada e a que se irá abordar. A última versão, 1998, apresentava a evolução ocorrida na natureza da computação. Este sistema é a base para a classificação de todos os documentos no ACM Guide to *Computing Literature* (Guide), uma bibliografia anual que lista mais de 20.000 novos itens e que indexou mais de 250.000 itens, desde 1982 [ACM 1998]. O CCS é hierarquicamente estruturado em quatro níveis[6],

[6] Três níveis codificados e um quarto não-codificado.

inclui termos gerais[7] e descritores de assunto implícito[8], sendo estes os três grandes conceitos da envolvente do CCS para 1998. Neste sistema, o primeiro nível é composto por 11 categorias com designação letra (A até K), onde cada uma destas categorias pode conter até quatro níveis. No segundo nível, apenas algumas categorias foram renomeadas. Em relação ao terceiro nível foram adicionadas 10 novas categorias, sendo estes dois níveis designados com a combinação de letra e de número. Por fim, o quarto nível fora adicionado com mais de 225 termos e mais de 150 termos foram reformulados [ACM 1998]. A Figura 2.6 ilustra uma pequena parte do esquema conceptual proposto pela ACM em 1998 onde são bem visíveis as alterações de que foram alvo os níveis.

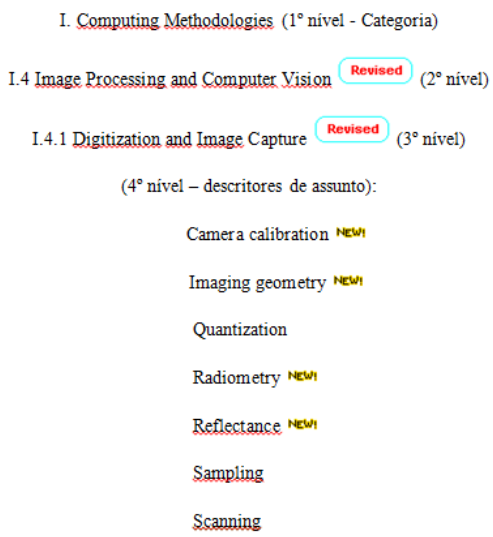


Figura 2.6: Esquema Conceptual da ACM

Adaptado de ACM [1998]

- A primeira categoria é para a literatura de computação em geral. Inclui subtemas para o material introdutório e pesquisa e para obras de referência;
- A segunda categoria é para *hardware* e tem subtópicos para estruturas de controlo e estruturas de micro-programação, estruturas de aritmética e lógica, estruturas de

[7] Conjunto fixo de 16 palavras que é aplico a todos os elementos da árvore que são relevantes (Algorithms, Design, Documentation, Economics, Experimentation, Human Factors, Languages, Legal Aspects, Management, Measurement, Performance, Reliability, Security, Standardization, Theory, Verification).

[8] Considera: nomes de propriedades de produtos, sistemas, linguagens, e pessoas de relevo no campo da computação, junto com o código da categoria em que se encontram classificados.

memória, entrada/saída de comunicações e dados de comunicações, registo de transferência de nível de implementação, design lógico, circuitos integrados, e desempenho e confiança;

- A terceira categoria está voltada para as preocupações na organização de sistemas de computador. Os subtópicos desta secção são arquiteturas de processadores, redes de comunicação por computador, sistemas para fins-especiais e sistemas com base em aplicativos, desempenho de sistemas e implementação de sistema de computador;
- A quarta secção versa *software*. Encontra-se subdividida em dois subtópicos para o desenvolvimento de *software*: técnicas de programação e engenharia de *software*. Linguagens de programação e sistemas operacionais são outros dois subtópicos para a infraestrutura de *software*;
- A quinta categoria são os dados. Os subtópicos estrutura de dados, representação de armazenamento de dados, criptografia de dados, codificação e teoria da informação e arquivos são incluídos nesta categoria;
- A sexta categoria é composta por diversos tópicos da teoria da computação, tais como: a computação por meio de dispositivos de resumo, a análise de algoritmos e complexidade do problema, lógicas e significados dos programas, lógica matemática e linguagens formais;
- A sétima categoria dirige-se à matemática da computação. Como subtópicos conta: análise numérica, matemática discreta, probabilidade e estatística, e *software* matemático;
- A oitava categoria é dos sistemas de informação, incluindo: modelos e princípios, gestão de base de dados, armazenamento e recuperação de informação, aplicação dos sistemas de informação, e interfaces e apresentação;
- A nona categoria, metodologias de computação, apresenta como subtópicos: a manipulação simbólica e algébrica, a inteligência artificial, a computação gráfica, o processamento de imagens e visão computacional, o reconhecimento de padrões, a simulação e modelação, documentos e processamento de texto;
- A décima categoria, abrange as aplicações informáticas. Os subtópicos são processamento de dados administrativos, ciências físicas e engenharia, vida e

ciências médicas, ciências sociais e comportamentais, artes e humanidades, engenharia assistida por computador, e computadores em outros sistemas;

- Por fim, a décima-primeira categoria, *Computing Milieux*, inclui o sector de informática, história da computação, informática e educação, computadores e sociedade, aspetos legais da computação, gestão da computação e sistemas de informação, a profissão da informática e computação pessoal.

Segundo Coulter et al. [1998], as fontes dos termos propostos para o CCS *Maintenance Committee* advêm de indexadores, texto livre de palavras e a frequência com que estas aparecem em base de dados bibliográficas relevantes, editores de *Computing Reviews* (CR) e público geral. O número de alterações sofridas pelo esquema em 1998 é bastante superior quando comparado com as versões anteriores, devendo-se este facto essencialmente à rápida evolução da computação. Estes mesmos autores consideram ainda que algumas questões não foram solucionadas: ao nível das categorias da estrutura principal, métodos e processos de suporte para a manutenção do CCS. Para o futuro do CSS referem que este terá que incorporar a descrição que preserva a integridade de arquivos, mantendo-se transparentes para o utilizador médio que faz uma pesquisa [Coulter et al. 1998]. Este sistema precisa de ser automático ou pelo menos dar apoio aos indexadores no processo de reclassificação da literatura existente num novo sistema. Seria interessante que o novo sistema a implementar em 2018, suportasse um processo contínuo de evolução para a classificação completa.

2.3.1.3 Síntese

Os esquemas de classificação apresentados anteriormente são muito semelhantes em relação a estrutura, mas diferem em relação ao objetivo de estudo. Os dois primeiros esquemas [Barki et al. 1988, 1993] incidem sobre na literatura dos SI, enquanto o ACM *Computing Classification System* retrata a ciência da computação [ACM 1998]. São esquemas a mencionar uma vez que permitem conhecer a área em que os SI estão incluídos, assim como reconhecer a relação que estes podem estabelecer com outras áreas que o circundam. Estudando estes esquemas é possível prever o que se pode analisar na literatura de SSI e futuramente relacionar estes esquemas com um esquema atual para a literatura de SSI.

2.3.2 Corpo de Conhecimento da SSI

"O valor da informação e a importância das Tecnologias da Informação (TI) no quotidiano individual, organizacional e social são hoje aceites como evidências." [de Sá-Soares 2005, p. 3]. Rápidas evoluções têm sido conhecidas ao nível das redes eletrónicas e dos sistemas de informação baseados em computadores, reconhecendo-se nas organizações o constante confronto com esta mesma evolução. "Mudança nas tecnologias de comunicação e informação e particularmente a sua confluência, levantou um conjunto de preocupações que se ligam com a proteção dos bens de informação organizacional. Alcançar um consenso sobre as salvaguardas para um sistema de informação entre os diferentes *stakeholders*[9] numa organização, tornou-se mais difícil do que resolver muitos problemas técnicos que possam surgir" [Dhillon e Backhouse 2000b, p. 125]. Opiniões antecedentes ao novo milénio, por parte de alguns utilizadores em relação ao uso generalizado de TI são de que estes não necessitavam de considerar questões relacionadas com segurança [Dhillon e Backhouse 2000b]. Porém, esta opinião está cada vez mais a ser refutada face aos imensos problemas de roubo de informação que as organizações estão sujeitas quer pelo meio externo, quer pelo meio interno. Como resolução deste problema, será necessário que os objetivos, os processos e os recursos das organizações sejam modificados, surgindo então o conceito de atividade organizacional de segurança de sistemas de informação.

Conforme observam Siponen e Oinas-kukkonen [2007, p. 60] "A importância da segurança da informação tem aumentado devido à maior utilização dos computadores e da Internet. Não de forma surpreendente, existem agora vários *journals*[10] e um largo número de conferências anuais e *workshops*[11] dedicados aos aspetos da segurança dos sistemas de informação (SI) e computação". Siponen e Oinas-kukkonen [2007] identificaram quatro questões como sendo as mais relevantes e abrangentes, relacionadas com a segurança de sistemas de informação - Acesso a Sistemas de Informação (SI), Comunicação Segura,

[9]Optou-se por manter o termo no seu idioma original, a língua inglesa, dada a elevada disseminação do mesmo. No entanto, querendo efetuar-se a tradução do termo para a língua portuguesa sugere-se o seguinte:
Stakeholoders - parte interessada;
Optou-se por manter o termo no seu idioma original, a língua inglesa, dada a elevada disseminação do mesmo. No entanto, querendo efetuar-se a tradução do termo para a língua portuguesa sugere-se o seguinte:

[10] Journals- revistas.

[11] Workshops – palestras.

Gestão de Segurança e Desenvolvimento de Sistemas de Informações Seguros - que devem ser solucionadas para que os sistemas de informação das organizações possam ser mais seguros. Associadas a estas questões estão implícitos quatro tipos de requisitos de segurança [Baskerville 1988; Parker 1998]: a informação não pode ser alterada por pessoas não-autorizadas (integridade), a informação deve estar disponível para as pessoas autorizadas e quando é necessária (disponibilidade), a divulgação não autorizada de informação deve ser detetada e impedida (confidencialidade) e a necessidade de assegurar que os contractos não possam ser negados a posterior (não-repúdio). Em relação à primeira questão em análise, acesso a SI, esta tem como objetivo garantir os requisitos de disponibilidade, integridade e confidencialidade (CIA), para que não seja comprometido o controlo de acesso para a leitura e escrita de informação realizados pelas pessoas ou pelos sistemas. No que diz respeito à comunicação segura, concede os três requisitos de segurança mencionados na primeira questão incluindo também o requisito não-repúdio. Esta segunda questão inclui os meios pelos quais as pessoas podem partilhar informação sem que terceiros saibam o que está a ser compartilhado. A gestão de segurança, terceira questão enumerada, tem como finalidade preservar em segurança o SI da organização incluindo o planeamento e evolução deste, sendo composta pelos quatro requisitos de segurança. Por fim, o desenvolvimento seguro do SI inclui o estudo dos diversos requisitos de segurança, como estes se manifestam, como são modelados e como garantem que um SI reflete os requisitos estabelecidos.

A SSI continua a revelar imensos desafios para todas as organizações. Muita da investigação realizada sobre esse aspeto tem natureza técnica, mas será de elevada importância que o conhecimento sobre a literatura em SSI esteja incutido nessas técnicas, uma vez que controlos técnicos não são suficientes para criar ambientes seguros na organização. Uma das possíveis razões que pode originar esta falta de ligação entre a natureza técnica e a natureza teórica deve-se ao facto de ainda não existir nenhum esquema conceptual temático que englobe todo o conhecimento que esta área abrange e por isso torna-se complicado saber o que esta realmente incluído na literatura da SSI tornando claro como é que a literatura pode auxiliar as questões técnicas. Daí cresce a necessidade da criação de um sistema de classificação para a área de SSI. Um sistema de classificação não só facilita compreender a diversidade de conceitos de uma determinada área de investigação, como permite que exista uma taxonomia que seja comum a todos os investigadores [Jacobs 1992].

As preocupações com a segurança da informação tornam-se cada mais fundamentais e por isso com o decorrer do tempo foram sendo criados modelos legislativos e referências gerais com instruções e preocupações básicas no âmbito tecnológico, tendo como objetivo padronizar o sistema de informação em geral. É fundamental compreender as ameaças a que uma organização está exposta, conhecendo todos os passos que devem ser tomados de forma a atenuar essas mesmas ameaças. Perante esta realidade, de seguida são apresentadas as certificações mais reconhecidas da SSI que foram sendo desenvolvidos ao longo do tempo e que hoje são mundialmente reconhecidos e utilizados por imensas organizações nas mais diversas áreas de atuação.

2.3.2.1 CISSP

A segurança da informação é, nos dias de hoje, considerada por algumas organizações, como uma parte integrante do negócio. Deste modo, todos os projetos de uma organização devem incorporar-se na segurança da informação e vice-versa, isto porque qualquer projeto de uma organização possui informação, informação essa que não pode ser perdida nem divulgada para o exterior. Para tal, existem várias organizações que têm por objetivo a criação de formalidades que visam a proteção da informação de uma organização.

A *International Information Systems Security Certification Consortium* ((ISC)²) é uma organização sem fins-lucrativos especializada na educação e certificações em segurança da informação [(ISC)² 1996]. Esta certificação internacionalmente reconhecida, surge em 1989 e certifica os profissionais do sector, de acordo com o padrão de segurança da informação. A certificação é composta por: *Certified Information Systems Security Professional* (CISSP) – cerificação que define a arquitetura, a gestão de risco e os controlos que garantem a segurança do negócio; *Certified Authorization Professional* (CAP) – certificação que auxilia os profissionais no processo de automatização e manutenção dos sistemas de informação; *Systems Security Certified Practitioner* (SSCP) – certificação responsável pela implementação e/ou atuação na infraestrutura de segurança de uma organização e *Certified Secure Software Lifecycle Professional* (CSSLP) – certificação que auxilia o desenvolvimento de um programa de segurança para o ciclo de vida do desenvolvimento de *software*. Para o caso de estudo, apenas se aprofundará a certificação CISSP.

A *Certified Information Systems Security Professional* (CISSP) é considerada mundialmente como a primeira credencial de segurança de informação [(ISC)² 1996]. Numa

organização, a CISSP debruça-se sobre a conceção e manutenção da infraestrutura de segurança, auxiliando os profissionais de *information assurance*[12] que definem a arquitetura, desenho, gestão e/ou controlos a garantir a segurança dos ambientes de negócio [Harris 2012]. Dada a perseverante evolução tecnológica e social é essencial que esta certificação esteja em constante atualização de modo a ser capaz de lidar com as novas tecnologias, novas ameaças e perfis de utilizador, implementando novas tecnologias de segurança que respondam às novas necessidades organizacionais. Para os profissionais de segurança da informação, tornou-se essencial a reflexão sobre o papel que cada individuo desempenha na conceção, desenvolvimento, implementação e manutenção de um programa de segurança de informação, assim como, averiguar se os objetivos vão de encontro aos requisitos de negócio da organização.

Esta certificação abrange uma variedade de conceitos de SSI, sendo eles agrupados de acordo com o assunto e reconhecidos em dez domínios, permitindo aos profissionais de SSI uma linguagem comum, facilitando a sua discussão e debate, resolvendo desta forma, as questões relacionadas com a área de estudo.

Seguidamente, descreve-se de modo sucinto cada domínio CISSP com base no que está descrito pela (ISC)[2] [1996]:

1. Controlo de Acesso (*Access Control*)

Os controlos de acesso são um conjunto de mecanismos que permitem a criação de uma arquitetura de segurança que protege os ativos do sistema de informação numa organização. Este domínio é composto pela infraestrutura tecnológica do sistema de processamento de informação, tipos de controlos, sensibilidade/criticidade dos dados e processos, que incluem o sistema, plataformas de rede e conexões de rede, e ambiente físico [Tipton 2009]. Através deste domínio é possível controlar os indivíduos que acedem ao sistema após a sua autorização, autenticação e monitorização, reduzindo as vulnerabilidades e ameaças a que o sistema está exposto. Ao alinhar pessoas, processos e tecnologia, as organizações podem estabelecer uma supervisão clara, reduzindo a exposição ao risco, construir a sua eficiência, e ter confiança nos seus ambientes de controlo.

[12] Utilizou-se o termo em língua inglesa dada a sua elevada disseminação. Na língua portuguesa o termo podia ser traduzido para: garantia da informação.

2. **Telecomunicações e Segurança da Rede (*Telecommunications and Network Security*)**

De forma a garantir a integridade, disponibilidade e confidencialidade da informação, este domínio engloba estruturas de rede, métodos de transmissão, protocolos e medidas de segurança, quer seja para transmissões em redes públicas e/ou privadas.

3. **Governação da Segurança da Informação e Gestão de Riscos (*Information Security Governance and Risk Management*)**

Este domínio baseia-se na execução correta das atividades de segurança numa organização, permitindo especificar o quadro de responsabilização e fornecer a supervisão de forma a garantir uma adequada mitigação dos riscos. Compreende estruturas e políticas, conceitos, princípios, estruturas e padrões que são utilizados para estabelecer os critérios necessários para a proteção dos ativos da informação na organização. A Governação de Segurança de Informação garante que as estratégias de segurança de informação estão alinhadas com os objetivos do negócio e regulamentos. A gestão de risco foca-se na minimização da perda dos ativos de informação, fazendo com que a organização conheça os riscos de gestão correntes, e as decisões a tomar com conhecimento de garantias de risco.

4. **Segurança no Desenvolvimento de Software (*Software Development Security*)**

O presente domínio é constituído pelas abordagens de desenvolvimento de software seguro, aplicações seguras e falhas de software, tendo por objetivo assegurar a confidencialidade, integridade e disponibilidade dos dados. O termo software é aqui entendido como sendo o software de sistema operativo (gere as operações de um sistema de computador) e software aplicacional (fornece funcionalidades para os utilizadores).

5. **Criptografia (*Cryptography*)**

O domínio Criptografia analisa os diversos tipos de criptografia, abordagens e tecnologias, incluindo algoritmos simétricos face a algoritmos assimétricos, infraestrutura de chave pública, funções *hash* e métodos de ataque, com o intuito de garantir a integridade, confidencialidade, autenticidade na transferência e armazenamento enquanto a informação se encontra oculta. A junção do uso de "cifra" (fórmula matemática) e de chave, permite a conversão de dados legíveis (texto simples) em formato não compreensível (texto cifrado) – criptografia.

6. **Arquitetura e Conceção de Segurança (*Security Architecture and Design*)**

O presente domínio analisa as melhores práticas para que o software seja projetado de forma segura, dando resposta às necessidades do negócio atual e futuro na organização. Concentra-se essencialmente nas políticas de segurança técnica da organização, implementação e execução dessas políticas. Versa claramente a conceção, implementação e operação dos controlos usados para executar os vários níveis de disponibilidade, integridade e confidencialidade de modo a garantir o funcionamento eficaz de uma organização. Do ponto de vista da conceção, descreve a um nível abstrato a relação entre os elementos-chave de *hardware*, sistema operacional, aplicativos, redes e outros componentes necessários para proteger os interesses da organização.

7. **Segurança Operacional (*Operations Security*)**

A Segurança Operacional permite identificar controlos sobre o pessoal, hardware, sistemas, auditoria e técnicas de monitoramento. Este domínio inclui preocupações com a proteção e controlo de recursos de processamento de informações em ambientes centralizados e distribuídos, sendo a disponibilidade o principal objetivo para a segurança operacional. Há uma série de processos e técnicas que podem ser implementados para garantir que um sistema mantenha a disponibilidade desejada quando confrontados com ameaças. Este domínio discute conceitos e técnicas que um profissional de segurança terá de implementar para satisfazer os requisitos de disponibilidade de um determinado sistema.

8. **Continuidade de Negócio e Recuperação de Desastres (*Business Continuity and Disaster Recovery*)**

Este foco consiste na preparação, nos processos e nas práticas necessárias para garantir a preservação da empresa face a grandes ruturas de operações normais. Envolve a identificação de riscos, seleção, implementação, teste e atualização dos processos, ações específicas necessárias para proteger de forma prudente os processos críticos de negócio dos efeitos do sistema principal, disrupções de rede e garantia de recuperação atempada das operações da empresa.

9. **Legalidade, Regulamentos, Observância e Investigação (*Legal, Regulations, Investigations, and Compliance*)**

Com este domínio analisa-se a criminalidade informática, leis e regulamentos. Por isso, inclui medidas e técnicas que devem ser usadas para determinar o incidente que ocorreu, bem como a recolha, análise e gestão de provas, caso existam. Aposta ainda, na forma de como desenvolver e implementar um programa de tratamento de incidentes.

10. **Segurança Física (Ambiental) (*Physical (Environmental) Security*)**

Por fim, este domínio analisa as ameaças, vulnerabilidades e contramedidas que podem ser utilizados para proteger fisicamente os recursos de uma empresa e informações confidenciais. Reflete sobre defensivas físicas, processuais, estratégias de recuperação, medidas a tomar e recursos disponíveis, incluindo a infraestrutura física da organização, políticas e procedimentos de segurança, ferramentas de segurança física e da equipa da organização.

2.3.2.2 CISM

No presente, a segurança de informação é um aspeto fulcral e preocupante para grande parte das organizações de todo o mundo. Falhas, desleixos e até mesmo a falta de conhecimento podem gerar prejuízos elevados para uma organização. Tais acontecimentos levam as organizações a pensarem nos benefícios da implementação de mecanismos de segurança de informação.

Implementar numa organização segurança de informação exige que profissionais estejam preparados com o objetivo de reduzir os riscos nas medidas de proteção que pretendem determinar para a organização, sem que para isso precisem de invadir a privacidade dos seus colaboradores e clientes, bastando apenas uma supervisão clara de todo o sistema de informação da organização. Neste sentido, a criação de certificações que englobem todas as normas, processos, tecnologias de informação apare como meio para a segurança de informação de uma organização.

A *Information Systems Audit and Control Association* (ISACA) surge em 1967, nos EUA, com o intuito de responder às necessidades das organizações em governação de tecnologias de informação [ISACA 2006]. A ISACA é uma organização sem fins lucrativos que auxilia os profissionais de governação de informação, segurança, controlo e auditoria de uma organização, garantindo a confiança e o valor dos sistemas de informação [ISACA 2006].

Certificações como: *Certified Information Systems Auditor* (CISA) – certifica quem controla, monitoriza e avalia as tecnologias de informação e sistemas de negócio; *Certified Information Security Manager* (CISM) – certificação focada na gestão de risco da segurança de informação; *Certified in the Governance of Enterprise* IT (CGEIT) – certifica profissionais de acordo com os conhecimentos e aplicações de princípios e práticas da organização na governação de TI; e *Certified in Risk and Information Systems Control* (CRISC) - certifica profissionais com experiência na identificação, avaliação e análise de risco, monitorização do risco, implementação e desenho de controlos de SI, monitorização e manutenção de controlos de SI; estão incluídas no conteúdo da ISACA [ISACA 2006]. Contudo, para este projeto apenas será abordada a certificação CISM.

A *Certified Information Security Manager* (CISM) é uma certificação da ISACA orientada para o negócio, nomeadamente para a gestão da segurança d informação. Define as principais competências e normas internacionais de desempenho, que os gestores de segurança da informação devem dominar [ISACA 2006]. A necessidade dos profissionais em terem uma certificação CISM surge da sua vontade em ter uma visão global na gestão de projetos, desenho, supervisionamento e avaliação da segurança de informação da organização. O reconhecimento desta certificação é a nível mundial e por isso um profissional CISM beneficiará de vantagem competitiva e evolução na carreira face a outros profissionais, rendimentos mais elevados e prestígio de conhecimento e experiência.

A certificação CISM é adequada para todos os gestores de segurança da informação (experientes) e para profissionais que têm responsabilidades de gestão de segurança da informação. De acordo com a ISACA [2006] a CISM organiza-se em cinco domínios, designadamente:

1. **Governação da Segurança de Informação (*Information Security Governance*)**

O presente domínio identifica e mantem a estrutura de administração da segurança de informação. Para que isso seja possível, é necessário que a estratégia de segurança da informação esteja alinhada com os objetivos de negócio da organização, assente na aplicação de leis e regulamentos. É necessário definir os papéis de governação de segurança da informação e as suas responsabilidades. Assegurar que os objetivos sejam atingidos, garantindo que os riscos são geridos de forma correta e verificando que os recursos da corporação sejam usados com responsabilidade.

2. **Gestão de Risco da Informação (*Information Risk Management*)**

Com o objetivo final de apoiar as organizações a manter a segurança dos seus SI, o domínio da gestão de risco de informação da CISM propõe às organizações o desenvolvimento de processos que visem definir e priorizar as estratégias de mitigação de risco necessárias para nesta matéria sejam garantidos níveis de segurança aceitáveis. Este domínio assenta no desenvolvimento de um processo de gestão de risco sistemático, analítico e contínuo.

3. **Desenvolvimento do Programa de Segurança da Informação (*Information Security Program Development*)**

Este domínio tem por objetivo especificar as atividades a serem executadas num programa de segurança da informação. As etapas identificadas vão desde a conceção, desenvolvimento e gestão do programa de segurança de informação com a finalidade de implementar uma estrutura de governação da segurança da informação. Também estão incluídas neste domínio atividades como: assegurar o desenvolvimento de arquiteturas de segurança da informação; estabelecer, comunicar e manter as políticas de segurança da informação que suportam a estratégia de segurança da informação; identificar os recursos internos e externos; e garantir o alinhamento entre o programa de sistema de informação e outras funções.

4. **Gestão do Programa de Segurança da Informação (*Information Security Program Management*)**

O presente domínio foca-se em dirigir, supervisionar e monitorar as atividades relacionadas com a segurança da informação na concretização do programa de segurança da informação. Para além do que já foi mencionado, certifica-se que os processos e procedimentos são realizados em conformidade com as políticas e normas das organizações e que a segurança da informação é uma parte integrante do processo de desenvolvimento de sistemas.

5. **Gestão de Incidentes e Resposta (*Incident Management and Response*)**

A gestão de incidentes e resposta permitem desenvolver e gerir a capacidade de responder e recuperar de eventos de segurança da informação perturbadores ou destrutivos.

Planear, estabelecer e gerir a capacidade de deteção de intrusão, investigar, responder; avaliar a continuidade dos negócios e planos de contingência; e realizar *forensics*, são os recursos utilizados para minimizar o impacto nos negócios e que compõem este domínio.

2.3.2.3 ISO/IEC 27000 Series

Nos dias que correm, os sistemas de informação necessitam de estar protegidos de forma a preservar o valor que a informação possui quer seja do ponto de vista organizacional ou individual, no setor público ou privado. A cooperação entre os fatores tecnologia, pessoas, processos, segurança, projetos de TI e o negócio de uma organização tem maior benefício quando alinhados com as melhores práticas de segurança da informação. Uma vez que a informação é o fator-chave da organização é necessário saber trabalhar com este ativo tão precioso.

Nesta perspetiva, para dar resposta às necessidades das organizações, surge a série de normas ISO/IEC 27000, composta por normas e regras que auxiliam da melhor forma possível a implementação e manutenção da segurança da informação em qualquer organização, pública ou privada, de grande ou pequena dimensão [(ISO) 2008]. A *International Organization for Standardization* (ISO) foi criada em 1947 e localiza-se em Genebra, Suíça. A sua finalidade consiste no desenvolvimento de padrões que suportam o comércio internacional. Para um cliente, fornecedor ou empresas parceiras, uma organização que possua uma ISO, garante que todas as informações que estão relacionadas com os seus produtos, possuem confidencialidade plena dos dados dos parceiros de negócio, assegurando a integridade e disponibilidade dessas informações [(ISO) 2008]. A junção da *International Electrotechnical Commission* (IEC), instituição criada em 1906 e localizada em Genebra, Suíça, cuja sua finalidade é desenvolver padrões para todos os tipos de electro tecnologias, facilitam o processo de desenvolvimento de padrões através de comités técnicos [(ISO) 2008].

O conjunto ISO/IEC 27000 foi desenvolvido com vista a ser entendido como um padrão global para a área dos SI e o seu objetivo passa por prover os fundamentos, princípios e conceitos necessários para que as organizações possam alcançar uma gestão dos ativos da informação sistematizada e robusta. O recurso à ISO/IEC2700 possibilita a aplicação de uma metodologia clara de Gestão de Segurança; reduz o risco de perda, roubo ou alteração da informação; promove confiança e o estabelecimento de regras claras para todos na organização; incrementa os níveis de segurança na gestão de processos; promove que os riscos e os seus controlos sejam constantemente verificados; garante qualidade e confidencialidade comercial; e a conformidade com a legislação vigente em matéria da segurança de informação pessoal, propriedade intelectual e etc. [(ISO) 2008].

Neste projeto apenas será apresentado o padrão ISO/IEC 27002 – guia de boas práticas que descreve os objetivos de controlo e os controlos recomendados para a Segurança da Informação, mas é importante não esquecer que existe a ISO/IEC 27001 – que especifica formalmente um sistema de gestão que se destina a trazer segurança da informação sob o controlo explícito de gestão; a ISO/IEC 27003 – que fornece uma abordagem de processos orientada para o sucesso da implementação de um Sistema de Gestão de Segurança da Informação (SGSI) de acordo com a norma ISO/IEC 27001; a ISO/IEC 27004 – que descreve as métricas e técnicas de medição aplicáveis para determinar a eficácia do SGSI, os objetivos de controlo e os controlos usados para implementar e gerir a Segurança da Informação; a ISO/IEC 27005 – que estabelece diretrizes para a gestão de risco em Segurança da Informação, fornecendo indicações para implementação, monitorização e melhoria contínua do sistema de controlos; e a ISO/IEC 27006 – que especifica requisitos e fornece orientações para os organismos que prestem serviços de auditoria e certificação de um SGSI.

Com o intuito de preservar o valor que a informação possuí surge o documento de referência, baseado nas melhores práticas de segurança da informação e que fornece um conjunto completo de controlos – ISO/IEC 17799 (origem no governo do Reino Unido). Em 1995 é publicada pelo *British Standards Institution* (BSI) sendo designada como BS7799 e em 2000, é republicada novamente pela ISO como ISO 17799, sendo-lhe acrescentadas algumas alterações [(ISO) 2008]. Em 2005 esta norma é submetida a revisão e é renomeada como ISO/IEC 17799:2005, reescrevendo, organizando e atualizando totalmente o seu padrão. Dois anos mais tarde, em 2007 é novamente republicada, passando a ser designada por ISO/IEC 27002:2005, sem que por isso, o seu conteúdo tenha sido alterado.

Esta subsecção retratará a norma ISO 27002 - Código de Práticas para a Gestão de Segurança da Informação (anteriormente conhecida como norma ISO 17799), assegurando que esta tem por objetivo o uso em simultâneo com a ISO 27001, e complementando-se. A ISO 27002 identifica para uma organização as diretrizes e princípios gerais para iniciar, implementar, manter e melhorar a gestão de segurança de informação. Em conjunto com a norma ISO 27001, a norma ISO 27002 traça os mecanismos de controlo e potenciais controlos, e orienta o desenvolvimento de padrões de segurança organizacional e práticas eficazes de gestão da segurança. Qualquer organização que pretende implementar medidas de gestão de segurança da informação ou até mesmo melhorar as que já possui pode fazê-lo, sem que para isso necessite de implementar a totalidade do código de práticas. O fator-chave para a decisão da implementação das práticas numa organização, serão os requisitos e riscos de segurança da informação.

Uma organização vê a informação como um ativo que deve ser protegido assim como toda a sua infraestrutura (redes, sistemas e funções de hardware e software). Uma vez que a norma ISO 27002 debruça-se sobre a informação, informação que pode existir sobre as mais diversas formas e através de diversos meios, estando sujeita a diversas ameaças e ataques (erro humano, danos com equipamento, etc.). Para facilitar a proteção da informação, as organizações precisam de desenvolver, implementar, monitorizar, avaliar e melhor os controlos de segurança. Deste modo, esta norma apresenta soluções que auxiliam a proteção da informação numa organização, estando organizada em 133 controlos estratégicos de segurança divididos por doze domínios. Porém, a organização deve primeiro identificar, selecionar e estabelecer os requisitos de segurança para os seus sistemas de informação para que o resultado sejam controlos apropriados ao seu contexto e ambiente de negócio.

Segundo a ISO/IEC [(ISO) 2008], a norma ISO 27002 é dividida em doze domínios sendo ela composta pelas melhores práticas de controlo de segurança de informação:

1. **Avaliação e Tratamento de Risco (*Risk Assessment and Treatment*)**

Para que este domínio seja concretizado, é necessário que uma organização identifique, quantifique e priorize os riscos, sendo que, de acordo com critérios e objetivos possa aceitar níveis de riscos consideráveis para a organização. O resultado será a mitigação desses mesmos riscos, consoante os objetivos e critérios de risco estabelecidos e adequados à segurança administrativa, técnica e física. O processo de avaliação de risco pode ser realizado inúmeras vezes até que seja possível avaliar toda a organização ou pequenas partes da organização, avaliar o sistema de informação individual ou componentes específicos do sistema/serviços. Esta avaliação pode incluir métodos sistemáticos de avaliação de riscos, métodos sistemáticos de comparação de riscos avaliados com base em critérios de risco, reavaliações periódicas para tratar alterações nos requisitos de segurança ou no ambiente de risco, e os meios de avaliação utilizados, e as relações com outras avaliações de risco. O tratamento dos riscos pode passar por aplicar controlos adequados que respondam à análise de risco elaborada, pode optar por transferir alguns riscos para terceiros (seguros) ou aceitar alguns riscos de forma consciente e objetiva.

2. **Política de Segurança (*Security Policy*)**

O nível de segurança dos sistemas informáticos de uma organização é em larga medida definido pela política de segurança adotada pela mesma. Assim, a definição por parte

das organizações de uma política de segurança robusta e adequada constitui-se determinante para a manutenção da integridade da mesma. O primeiro passo para o estabelecimento de uma política de segurança consiste em definir quais as metas de segurança que a organização pretende alcançar. Adicionalmente, a organização deverá comunicá-la a todos os seus colaboradores, analisá-la e revê-la em intervalos de tempo constantes ou até mesmo quando ocorram mudanças na organização.

Em termos práticos, o estabelecimento de uma política de segurança deve focar-se na criação, gestão e suporte de processos e procedimentos que visem garantir a proteção dos dados da organização. Para tal, deve garantir-se a observância das orientações providas por entidades de certificação, pelas normas de ética profissional, requisitos específicos de negócio e das leis e regulamentos aplicáveis.

3. Organização da Segurança da Informação (*Organization of Information Security*)

Este tópico retrata a gestão da segurança da informação dentro da organização, sendo, por isso, necessário estabelecer uma estrutura para geri-la. As atividades de segurança da informação são estruturadas mediante os representantes dos diversos departamentos da organização, com funções e papéis relevantes, cujo objetivo dita o início e o controlo da implementação de segurança da informação na organização.

A organização deve estabelecer acordos de confidencialidade com os seus colaboradores de modo a proteger a informação de carácter sigiloso, a informação acedida, comunicada, processada ou gerida por entidades externas (terceiros e clientes). O acesso por entidades externas às instalações de processamento de informação e de processamento e comunicação de informação das organizações deve ser controlado.

4. Gestão de Ativos (*Asset Management*)

A gestão de ativos tem um papel de elevado destaque numa organização, uma vez que esta deve compreender de forma clara quais os ativos de informação que contém (ativo é alguma coisa que tem valor acrescentado para a organização) e gerir a sua segurança de forma adequada. A informação e ativos devem possuir um nível adequado de proteção e para isso é necessário indicar a necessidade, prioridade e grau esperado de proteção para o seu manuseamento.

A organização deve possuir um inventário de todos os ativos de informação (hardware de TI, software, dados, documentação do sistema, meios de armazenamento, entre outros) e identificar o respetivo proprietário para todos os ativos.

5. Segurança dos Recursos Humanos (*Human Resources Security*)

O entendimento adequado de segurança da informação permite à organização gerir o acesso ao sistema de modo claro. Funcionários, fornecedores e terceiros devem compreender as suas responsabilidades, em especial as responsabilidades de segurança da informação (também elas presentes nos seus contratos) e devem concordar com o papel que desempenharão, com o intuito de reduzir o risco de roubo, fraude ou mau uso de recursos. Deste modo, as descrições de cargo e termos, e as condições de contratação devem ser explícitas, permitindo que estes facilitem o apoio às políticas de segurança da informação na organização.

É fundamental a educação e treino dos procedimentos de segurança da informação e o uso correto dos recursos de processamento da informação.

6. Segurança Física (*Physical Security*)

A segurança física ocupa-se da criação de medidas de segurança que devem ser implementadas por uma organização. O propósito base é impedir que pessoas não autorizadas acedam a todos e quaisquer elementos de uma organização (pode ser considerado o espaço físico, as instalações, os recursos ou informações armazenadas). Terá a organização de ser capaz de criar estruturas eficientes que respondam de forma adequada e rápida a estes atos potencialmente maliciosos.

7. Gestão das Operações e Comunicações (*Communications and Ops Management*)

Para uma organização, a identificação e a documentação dos procedimentos e responsabilidades de todos os recursos de processamento da informação é uma tarefa de gestão e operação com elevado relevo. Este domínio projeta garantir o correto e seguro funcionamento das instalações de processamento da informação, estabelecendo controlos de segurança e de gestão de redes.

O presente domínio recomenda, caso necessário, a utilização de segregação de funções para reduzir o risco do mau uso ou uso indevido dos sistemas. Planear e prever a

disponibilidade e os recursos dos sistemas permite reduzir o risco de falhas, calcular a capacidade futura dos sistemas, minimizando os riscos da sobrecarga. Outros aspetos a ter em atenção será a prevenção e deteção com a introdução de código malicioso, a troca de informação entre as organizações (baseadas em politicas formais) e a monitorização de atividades desautorizadas no processamento da informação.

8. Controlo de Acessos (*Access Control*)

O controlo de acessos, como o próprio nome indica, permite controlar o acesso à informação, ou seja, é a capacidade de autorizar ou negar o uso de um objeto[13] por um sujeito. Acesso aos sistemas de TI, redes e dados precisam de ser controlados com base nos requisitos de negócio e de segurança para evitar o acesso não autorizado, mediante o processo de autenticação (que sujeito pode aceder), de autorização (o que um sujeito pode fazer) e de auditoria (o que o sujeito fez).

Um utilizador deve consciencializar-se das suas responsabilidades, em especial o uso de senhas e segurança do equipamento dos utilizadores. Na organização os direitos de acesso dos utilizadores deve ser controlado por meio do registo dos utilizadores, restrições na alocação de privilégios, da gestão de senhas e revisão no acesso regular dos direitos.

9. Aquisição, Desenvolvimento e Manutenção de Sistemas de Informação (*Information Systems Acquisition, Development, Maintenance*)

Este domínio tem como principal objetivo garantir que a segurança é uma componente essencial para os sistemas de informação. Os requisitos de segurança de sistemas de informação devem ser identificados e justificados, acordados e documentados previamente ao seu desenvolvimento e implementação. O *software* gerado deve ser formalmente testado sob o ponto de vista da segurança e devem ser avaliados os problemas de risco.

10. Gestão de Incidentes de Segurança da Informação (*Information Security Incident Management*)

A comunicação e a gestão de eventos de segurança da informação, incidentes e pontos fracos, devem ser efetuadas com rapidez permitindo uma ação corretiva adequada. Responsabilidades e procedimentos são necessários para a gestão de incidentes. A forma

[13] O termo objeto é aqui entendido como uma entidade passiva, ou como um sistema ou arquivo.

consistente e eficaz permite implementar uma melhoria contínua e coletiva de evidências *forenses*.

11. Continuidade do Negócio (*Business Continuity*)

Uma organização é composta por várias atividades, sendo a continuidade do negócio uma atividade que permite a garantia da interrupção das atividades do negócio, protege os processos críticos de negócio dos efeitos de grandes falhas ou desastres dos sistemas de informação, e assegura a sua recuperação em tempo útil. Controlos permitem identificar e reduzir o impacto dos incidentes de segurança, auxiliando a recuperação rápida das operações essenciais, isto é, a continuidade do negócio deve ser implementada para minimizar o impacto sobre a organização e a recuperação de perdas de ativos de informação a um nível aceitável através da combinação de medidas de prevenção e recuperação.

A continuidade do negócio deve ser uma atividade realizada diariamente de forma a sustentar o serviço, a consistência e a capacidade de recuperação. A segurança da informação é considerada uma parte integrante do processo global de continuidade de negócio e de outros processos de gestão dentro da organização.

12. Conformidade (*Compliance*)

A conformidade é um domínio que deve garantir e evitar a violação de qualquer lei criminal ou civil, regulamentos, obrigações regulamentares ou contratuais e de qualquer requisito de segurança da informação. Contratar consultoria especializada pode ser a solução para analisar a conformidade da segurança dos sistemas de informação numa organização. A organização deve cumprir com a legislação adequada às suas necessidades (copyright, proteção de dados, restrições de criptografia, regras de provas, entre outros).

2.4 Fundamentos Teóricos

Esta secção tem como principal objetivo a elaboração da uma revisão de literatura na área dos esquemas de classificação para os SI e para a SSI, e uma breve revisão sobre o conceito de Text Mining. A revisão da literatura é uma característica fundamental dos projetos académicos. Um levantamento exaustivo sobre determinada área de conhecimento permite uma base sólida para o avanço do conhecimento, facilita o desenvolvimento da teoria,

estabelece um limite para a área que se pretende estudar e revela áreas onde é necessária a investigação [Webster e Watson 2002]. A revisão apresentada abrange literatura relevante para as áreas em questão e não se limita apenas a uma metodologia de pesquisa ou a documentos com origem de publicação comum. Uma boa revisão deve identificar as lacunas críticas do conhecimento, motivando investigadores a debruçarem-se sobre essa lacuna e solucioná-la.

Atualmente existem diversos estudos na área e são mais recentes quando comparados com o estudo de Claver et al. [Claver et al. 1999], estudo de base da presente secção. O conhecimento nesta área entretanto evoluiu e por isso começou a surgir o interesse de classificar todo este conhecimento para se conseguir obter uma melhor compreensão sobre o âmbito da área.

O presente estudo pretende rever os esquemas de classificação existentes na área de SSI e apresentar um novo sistema de classificação mais recente e com técnicas mais eficientes. Também será alvo de estudo esquemas de classificação dos SI, mas estes serão apenas os mais relevantes da área e só estudados na forma de revisão. A razão do estudo permitirá uma possível comparação que possa ser estabelecida entre os esquemas de ambas as áreas, mas sabendo desde já que a SSI faz parte de uma das temáticas dos SI.

Qualquer que seja a revisão de literatura deve ser sempre estudada à luz de um enquadramento conceptual, que permita avaliar o âmbito da área em questão. Nesta secção serão apresentados os seguintes enquadramentos: Baskerville [1993], Dhillon e Backhouse [2001], Siponen [2005a], Siponen [2005b] e Villarroel et al. [2005].

Por isso, Webster e Watson [2002] apresentam quatro critérios para que a revisão ao nível da literatura seja feita com rigor. Primeiro sugerem a seleção dos métodos de SSI, segundo consideram que deve-se incluir em cada classe de métodos de SSI representantes de diferentes regiões geográficas, terceiro necessita incluir em cada classe estudos que sugerem métodos alternativos em relação ao objetivo principal da classe, e quarto escolher os representantes de ambos os métodos antigos e recentes.

As organizações concedem crescentemente maior enfase e atenção na segurança dos seus SI. A segurança da informação é vista como um requisito importante que deve ser cuidadosamente ponderado, como um elemento presente em todas as fases do ciclo de desenvolvimento do sistema a partir da análise de requisitos para a implementação e manutenção, e não como um aspeto isolado. Violações de segurança não causam apenas

perdas às organizações, mas violações da segurança como a violação da privacidade (ex.: roubo de cartão de crédito ou informação) podem também ter impacto direto no bem-estar do utilizador comum de computadores. Por isso, é importante que sistemas de informação estejam devidamente seguros desde o seu início.

"Segurança deve ser considerada como um requisito fundamental no desenvolvimento de Sistemas de Informação (SI), e tem de ser tida em conta em todas as fases do desenvolvimento" [Fernández-Medina e Piattini 2003, p. 886].

2.4.1 Contextualização de Baskerville

Os investigadores têm presente a necessidade de ter um conhecimento bastante amplo sobre as diversas áreas de investigação. No entanto, estes não se podem esquecer que qualquer área de investigação evolui com o passar do tempo, assim como as técnicas de criação de recursos que lhe estão subjacentes. A segurança dos SI tem vindo a tornar-se uma preocupação cada vez maior devido à crescente disseminação da utilização de sistemas de computadores, o que tem, por outro lado, levado a que analistas de sistemas e designers tenham de desenvolver novos e melhorados métodos para a especificação de sistemas de segurança de informação [Baskerville 1993].

Neste âmbito, Baskerville [1993] realizou um estudo que tinha como principais objetivos: prover uma descrição analítica do progresso da análise da segurança dos SI e dos métodos de conceção, compreender as técnicas atuais de criação de recursos de computação segura e reconhecer os caminhos críticos para os novos métodos de desenvolvimento de sistemas em geral. Para que todos os analistas tivessem um consenso sobre o que realmente deve estar incluído em cada conhecimento, foi necessária a criação de uma taxonomia. A falta de consenso deve-se a que cada área de conhecimento opera dentro de um nível de abstração que é apropriado para lidar com um problema específico definido na evolução de desenvolvimento de sistemas [Baskerville 1993]. A contextualização deste estudo fundamenta-se numa taxonomia simples de "gerações" metodológicas que irá relacionar a evolução dos métodos da SSI com a perspetiva da ampla comunidade de desenvolvimento de SI. Utilizando as características genéricas dos métodos de análise e conceção dos SI, é possível comparar a evolução dos métodos de conceção de segurança com as gerações de métodos de desenvolvimento mais amplos dos sistemas de informação [Baskerville 1993]. Para este efeito, o autor utilizou um esquema de três gerações para

comparar e contrastar os métodos de segurança. A metáfora "geração" é particularmente útil na comparação de métodos de desenvolvimento de sistemas, porque uma nova geração evolui a partir da experiência da sua antecessora sem necessariamente implicar obsolescência da predecessora [Baskerville 1993].

"As características encontradas em três gerações de métodos gerais de conceção de sistemas de informação fornecem um enquadramento para comparar e compreender os métodos atuais de conceção de segurança. Estes métodos incluem abordagens que usam *checklists* de controlo, dividem requisitos funcionais em partições de engenharia e criam modelos abstratos dos problemas e soluções. Comparações e contrastes revelam que os avanços nos métodos de segurança ficam aquém dos avanços dos métodos gerais do desenvolvimento de sistemas. Esta análise também revela que métodos mais gerais falham na consideração rigorosa das especificações de segurança" [Baskerville 1993, p. 375]. Em relação aos métodos de conceção dos sistemas das gerações apontadas, a primeira geração – *Checkist Methods* – tem como principal objetivo a seleção dos vários componentes da solução, onde o meio para encontrar a solução ideal do sistema seria estudar todo o repositório de elementos do sistema disponível. Em relação à segunda geração – *Mechanistic Engineering Methods* – o particionamento de soluções de sistemas complexos será o principal objetivo, identificar e resolver cada particularidade do requisito funcional, sendo este o meio mais eficaz para descobrir a solução ideal do sistema. O grande desafio desta geração consiste no detalhe da organização num conjunto complexo de diversos elementos que devem ser integrados em sistemas com uma só solução. Por fim, a terceira geração – *Logical Transformational Methods* – sugere uma abstração do problema e soluções para o mesmo, e por isso utiliza o modelo dos atributos essenciais do problema de informação e a sua solução. A escolha correta dos atributos que devem ser atribuídos ao modelo é o desafio mais importante. Nas duas primeiras gerações de métodos de SSI, a análise e o design de tarefas está inteiramente enraizados nos elementos físicos do sistema, procurando determinar o que pode ser feito com a ajuda das soluções técnicas disponíveis, o que já não acontece com a terceira geração.

Baskerville [1993, p. 411] conclui:

- "Segurança é uma componente importante dos sistemas de informação. A análise de segurança e métodos de conceção evoluíram de forma semelhante

aos métodos de desenvolvimento de SI em geral. Compartilham características comuns como objetivos, meios, desafios e conceitos primários."

- "Os primeiros métodos de segurança focam checklists e análise de risco simples para apoio à decisão. Mais tarde, métodos de foco sobre o particionamento mecanicista da complexidade de um desejado sistema. Eles implicam uma pesquisa para controlos críticos que fornecem o mínimo de proteção satisfatória para todo o sistema de informação. Mais recentemente, o interesse desenvolve-se em métodos que se concentram em modelos abstratos como um meio para a compreensão das diversas necessidades de segurança no sistema de informação."

2.4.2 Contextualização de Dhillon e Backhouse

No seguimento do trabalho de Baskerville [1993], Dhillon e Backhouse [2001] também publicaram um estudo relacionado com a descrição e organização de toda a envolvente dos sistemas de informação atuais e investigação de segurança, isto é, reviram e avaliaram o corpo de conhecimento da SSI. Com este estudo, os autores pretendiam que a sua análise contribuísse para a tendência futura dos sistemas de informação e investigação em segurança, destacando as perspetivas sócio-organizacionais e avaliando as preponderâncias das soluções técnicas.

Como base para desenvolver este trabalho os autores utilizaram o enquadramento de Burrell e Morgan [1979] como um guia para estudarem as preocupações sócio-filosóficas dos trabalhos no campo dos SI e na área da SSI. Dhillon e Backhouse [2001, p. 128] afirmam " A fim de examinar a vasta literatura em sistemas de informação e segurança, precisamos de um enquadramento conceptual que nos ajude não só a classificar as obras, mas também a traçar as suas origens intelectuais. Teóricos como Burrell e Morgan [1979], Lane (1994) e Walsham (1993) afirmam que é importante para compreender os conceitos teóricos que formam a base de uma abordagem metodológica". Burrell e Morgan [Burrell e Morgan] apresentam um modelo bidimensional de análise das teorias sociais, conforme se ilustra na Figura 2.7, com o intuito de ajudar a classificar e compreender as teorias sociológicas existentes com base em quatro grandes paradigmas.

Figura 2.7: Esquema dos Quatro Paradigmas de Burrell e Morgan

Adaptado de Burrell e Morgan [1979]

Cada quadrante corresponde a um paradigma particular na sociologia. A primeira dimensão subjetividade/objetividade abrange pressupostos sobre as ciências sociais, podendo aqueles ser de natureza:

- Ontológica (nominalismo/realismo) – O nominalismo firma-se como uma realidade social relativa e o realismo supõe que o mundo real é rígido, tem estruturas intangíveis que existem independentemente das nossas atribuições de significados e significantes. O mundo social existe separado da perceção dos indivíduos do mesmo.
- Epistemológica (anti positivismo/positivismo) – os anti-positivistas rejeitam que a ciência social possa conceber conhecimento verdadeiramente objetivo. Já os positivistas admitem que se pode procurar explicar e prever o que acontece no mundo social através da busca por padrões e relacionamentos entre indivíduos. Mais ainda, admitem também que podem desenvolver-se e testar-se hipóteses e que o conhecimento consiste num processo cumulativo.
- Humana (voluntarismo/determinismo) – defende que são seres humanos limitados pelo seu meio ambiente ou então podem escolher livremente e de acordo com a sua vontade e consciência.
- Metodológica (ideográfica/nomotética) – Ideográfica concentra-se em saber tudo sobre determinado assunto e explorar os seus detalhes antecedentes e história de

vida. Nomotética pressupõe a utilização de testes quantitativos para a descoberta de relações causais entre as variáveis de interesse ao fenómeno.

Estes pressupostos compõem uma única dimensão de análise porque estão interconectados e não podem ser dissociados.

A segunda dimensão apresenta os pressupostos de natureza sociológica, que podem ser de ordem reguladora ou de natureza dinâmica (ordem/mudança) – abordagens que caracterizam os efeitos estabilizadores da ordem social, contra as abordagens mais focadas na mudança.

Ajustando estas duas dimensões, Burrell e Morgan [Burrell e Morgan] propõem quatro paradigmas, próximos mas distintos (funcionalismo, interpretativismo, humanismo radical e estruturalismo radical), que podem, no entanto, compartilhar algumas características com os seus vizinhos, sem que para isso percam a sua especificidade.

O paradigma funcionalista – tem sido o principal paradigma para o estudo organizacional, auxiliando explicações racionais sobre os assuntos humanos. Este pressupõe uma ação humana racional e acredita que se pode compreender o comportamento organizacional através do teste de hipóteses. Em oposição, tem-se o paradigma interpretativo, que procura explicar perspetivas do comportamento do indivíduo. Teóricos do paradigma humanista radical preocupam-se em especial com a liberação de restrições sociais que limitam potencialidades humanas. Por fim, o paradigma estruturalista radical caracteriza-se pelo facto dos teóricos verem inerentes conflitos estruturais na sociedade que geram mudanças constantes através de crises políticas e económicas.

Dhillon e Backhouse [2001, p. 130] realçam que – "Burrel e Morgan sustentam que os teóricos pertencentes a um determinado paradigma não podem reconhecer as visões alternativas da realidade que se encontram fora dos seus limites. Eles também afirmam que os quatro paradigmas são mutuamente exclusivos e contradizem-se uns aos outros, o que leva a considerarem que a teoria-social não deva pertencer a mais do que um paradigma, em determinado momento". No seu estudo, Dhillon e Backhouse [2001, p. 129] admitem que os "quatro paradigmas discutidos são definidos pelos pressupostos meta-teóricos que formam o quadro de referência e o modo de teorização":

- "O paradigma funcionalista representa uma perspetiva que está firmemente enraizada na regulamentação da sociologia e aborda o assunto de um ponto de vista objetivista, preocupado com a "regulação" e controlo de todos os assuntos organizacionais (...).

- O paradigma interpretativo preocupa-se com a compreensão subjetiva que os indivíduos atribuem a situações de integração social. Embora os interpretativistas concordem com os princípios reguladores dos funcionalistas, eles acreditam numa análise subjetiva do mundo social (...).
- O paradigma radical humanista opõe-se às teorias de regulamentação e sustenta uma mudança radical. Vendo a sociedade como anti-humana, os humanistas radicais salientam a emancipação dos seres humanos para que eles possam realizar todo o seu potencial (...).
- O paradigma estruturalista radical apresenta um ponto de vista que se opõe à visão de regulamentação da sociedade. Enquanto advogando uma mudança radical, estruturalistas radicais compartilham o ponto de vista objetivista dos funcionalistas (...).

Tendo por base o enquadramento de Burrell e Morgan [1979], os autores organizaram a literatura sobre SI e sobre SSI conforme se indica na Tabela 2.1.

Em relação à literatura de segurança, esta centra-se na proposta de *checklists*, na análise de risco e na avaliação. Esta parte da literatura preocupa-se com a parte formal automatizada do sistema de informação. O paradigma interpretativo permite avaliar as implicações dos SI baseados em computador. Nesta literatura, a organização social é analisada tal como é, e por isso é vista como um sistema emergente criado pelas pessoas em causa. Em relação à literatura da segurança, a utilidade das teorias interpretativas auxiliam na perceção das questões de segurança. A compreensão interpretativa facilita uma visão holística do domínio do problema. O paradigma humanista, no que se refere a literatura dos SI, usa conceitos assentes no presente paradigma, em que os sistemas simplificam uma discussão mais ampla de possíveis problemas organizacionais. No que toca à literatura de segurança, a estratégia de implementação em larga medida é vaga e imprecisa. Por fim, o paradigma estruturalista radical reflete os conflitos de interesses entre a gestão de topo e os utilizadores. Com base no estudo de Dhillon e Backhouse [2001] conclui-se que existe uma progressiva deceção com a conceção formal, racional e excessivamente mecânica na análise e conceção de controlos de SSI; grande parte da investigação em SSI tende a concentrar-se em estruturas formalizadas para a conceção da segurança; a confiança não é suficiente no *design* de abordagens e as avaliações recentes de segurança e métodos de conceção baseiam-se em conceções funcionalistas. Este estudo permitiu identificar as lacunas e as áreas problemáticas para as teorias e

abordagens de segurança de sistemas de informação, sugerindo que as perspetivas sócio-organizacionais devem ser igualmente ponderadas para alcançar a segurança de sistemas de informação.

Tabela 2.1: Classificação da Literatura sobre Sistemas de Informação e sobre Segurança de Sistemas de Informação

Adaptado de Dhillon e Backhouse [2001]

Paradigma	Teoria usada	Métodos de Sistemas de Informação e Trabalho Pioneiro	Métodos de Segurança e Trabalho Pioneiro
Funcionalista	Teoria dos Sistemas; Teoria da Contingência.	Sucesso dos SI (Ives et al., 1993); Identificação de Requisitos (Bailey & Pearson, 1983; Davis & Olson, 1984; Baroudi et al., 1986), Desenvolvimento de Sistemas (DeMarco, 1978)	Abordagens de Análise de Risco Tradicionais (Courtney, 1977; Parker, 1981; Fisher, 1984); Métodos de Avaliação de Segurança (Bell & La Padula, 1976; Van Der Veen et al., 1994)
Interpretativo	Teoria da Estruturação; Fenomenologia e Hermenêutica; Semiótica; Contextualismo.	Estratégia de sistemas de informação, desenho do sistema e implementação (Boland, 1985; Walsham, 1993) Uso de sinais no sistema de especificação (Liebenau & Backhouse, 1990)	Análise de risco e conteúdo comunicativo (Baskerville, 1991); Teoria dos atos de fala e Desenvolvimento da Segurança (Dobson, 1991); Considerações Pragmáticas e Segurança (Backhouse & Dhillon, 1996)
Radical Humanista	Teoria Crítica; Individualismo Anarquista.	Teoria dos Sistemas de Informação e Especificação do Sistema (Lyytinen & Klein, 1985)	Opções estratégicas para a segurança como descritas por Angell (1994); consideração teórica critica na análise de risco (Weber et al., 1992)
Estruturalista Radical	Teoria do Conflito.	Visão Contratual de Sistemas de Informação (Ciborra, 1987)	Não encontrada, exceto algumas semelhanças com o trabalho de Lane (1985)

2.4.3 Contextualização de Siponen

Com o passar dos anos foi-se observando o aparecimento de várias abordagens para os métodos de SI, contudo esses métodos de desenvolvimento de SI apresentavam falhas no endereçamento de preocupações com a segurança. Não obstante, todas essas abordagens acabaram por influenciar várias abordagens modernas da SSI. O facto de esses métodos serem desenvolvidos por investigadores de diferentes tradições de pesquisa e disciplinas de estudo distintas, contribuíram também para a controvérsia em volta desses métodos. Existem imensos métodos de SSI, no entanto, ainda não são reconhecidos todos eles. Comunidades de investigação em conjunto com o reconhecimento destes métodos pretendem compreender as diferenças que lhes estão subjacentes e perceber os inúmeros métodos de SSI existentes. Para isso são analisadas quatro características subjacentes: objetivos de pesquisa, papel organizacional da SSI, abordagem de investigação utilizada e aplicabilidade do desenvolvimento de SI. Utilizando o modelo de gerações de Baskerville [1993] juntamente com a informação encontrada, é possível identificar diretrizes normativas para a próxima geração de métodos de SSI. Os resultados do estudo têm implicações para a pesquisa de SSI, práticas de SSI e educação de SSI.

Siponen produziu duas publicações alusivas a este assunto. A primeira publicação pretendia apresentar as disciplinas e comunidades científicas que fundamentam as abordagens modernas de SSI, revelar os pressupostos explícitos que se escondem nas abordagens modernas de SSI e estabelecer uma classificação de cinco gerações para as abordagens modernas de SSI [Siponen 2005a], sucedendo esta publicação como uma evolução do trabalho efetuado por Baskerville [1993]. Comunidades científicas são compostas por diferentes conjuntos de investigadores que partilham uma determinada orientação, que pode ser em relação à educação ou experiência pessoa/trabalho. Segundo o autor, as comunidades de investigação e as disciplinas são um aval para a investigação das diferentes abordagens de SSI. São identificadas quatro comunidades de pesquisa para a segurança de informação: *IS/MIS security*, *Computer Security*, *Database Security* e *Cryptology/Communications Security*, que por sua vez investigam quatro disciplinas: *IS/Management Information Systems* (MIS), *Software Engineering*, *Computer Science* e *Mathematics*. A Figura 2.8 demonstra a relação existente entre as diversas abordagens modernas de SSI no que toca às origens e influências, com as quatro disciplinas apresentadas e com as comunidades científicas. As setas indicam a influência e as linhas pontilhadas indicam uma fraca influência. Autores que defendem as duas primeiras gerações de abordagens de *design* de SSI acabam por procurar inferir o dever (o que as organizações devem fazer)

a partir do ser (o que é possível fazer, ou o que existe). Em relação à terceira geração e gerações posteriores, defendem não tentar seguir essa abordagem.

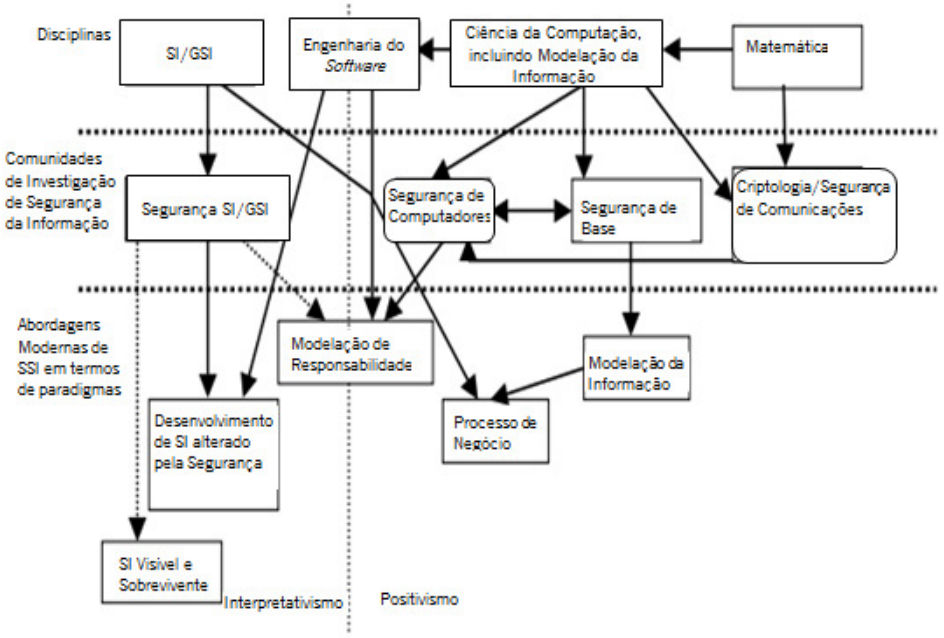


Figura 2.8: Comunidades Científicas vs. Disciplinas de Siponen

Adaptado de Siponen [2005a]

As gerações que são apresentadas segundo o estudo de Siponen [2005a] serão designadas por abordagens modernas de SSI, ou seja, os métodos de SSI que consideram o facto de que as organizações podem ter necessidades específicas de SSI e que o conhecimento de descobrir esses requisitos deve ser o ponto de partida para o desenvolvimento da SSI. Como se pode observar, a ilustração 2.8 mostra que as abordagens modernas de SSI são designadas por: *Viable and Survivable SI*, *Security-Modified SI Development*, *Responsibility Modeling*, *Business Process* e *Information Modeling*. Investigadores de SI nas suas estratégias de pesquisa tanto usam o positivismo como o interpretativismo. A primeira abordagem é composta por métodos de SSI de dois grupos de pesquisa, a sobrevivência dos SI (inclui a sobrevivência de ameaças futuras a SSI) [Karyda et al. 2001] e o modelo de sistema viável (estuda as diferentes ameaças/vulnerabilidades dos SI) [Hutchinson e Warren 2000], que é uma ferramenta que estuda de que forma ataques bem-sucedidos podem ser vistos à luz das cinco funções sistémicas: elementos operacionais, coordenação, controlo, inteligência e identidade. O objetivo é construir um sistema de informação viável e não robusto. Um sistema de informação viável

tem a capacidade de manter a sua existência através da gestão de risco, aplicando desta forma o modelo de sistemas viáveis proposto por Stafford Beer no início de 1970. Em relação a *Security-Modified SI Development* é um paradigma que inclui a abordagem de controlo lógico (melhora a conceção da SSI, oferecendo uma melhor forma de sistematização e controlo de *design* de SSI, incorporando no processo de desenvolvimento de SI estruturado) [Baskerville 1989], a abordagem espiral (garante que considerações de SSI são analisadas em todas as fases do desenvolvimento de SI) [Booysen e Eloff 1995], a metodologia virtual (consente um método para a construção de sistemas de segurança) [Hitchings 1995], a abordagem para o planeamento de SSI (mudanças no pensamento da SSI de modo a prestar mais atenção ao ser humano nesta dimensão da SSI) [Armstrong 2000; James 1996] e a metodologia de planeamento de SSI (propõe uma teoria baseada na prática a utilizar pelos gestores no modelo de SSI que reduz a minimização dos riscos de SSI) [Straub e Welke 1998], onde o trabalho efetuado na disciplina de segurança de SI/MIS é a principal influência das abordagens. As duas primeiras abordagens incidem sobre questões técnicas (ex.: modelação de processos), enquanto a metodologia virtual foca em questões organizacionais, culturais e humanas no desenvolvimento de SSI. A abordagem para o planeamento de SSI foca-se na componente humana de segurança, ao invés da metodologia de planeamento de SSI que é composta pelo modelo de planeamento de segurança de risco, um programa de sensibilização e uma matriz de contra-medidas. *Responsibility Modeling* é um paradigma composto pela abordagem de Strens e Dobson [1993] (descreve as responsabilidades do trabalho que ultrapassam as falhas de comunicação nas organizações), pela abordagem da análise semântica de responsabilidade (permite perceber os diversos requisitos de SSI nas organizações) [Dhillon e Backhouse 1996b], pela abordagem de casos de abuso (ajuda a compreender os requisitos de segurança e atividades nas organizações) [McDermott e Fox 1999] e pela abordagem de autorização baseada em tarefas (sugere um modelo aprimorado de autorização através da análise de responsabilidades nas organizações) [Thomas e Sandhu 1994]. Estas abordagens baseiam-se nos requisitos de segurança que podem ser descobertos analisando as responsabilidades do trabalho nas organizações e entendidas como linguagens de modelação que são usadas para propósitos de comunicação. O quarto paradigma, *Business Process*, é constituído pela abordagem de Hermann e Pernul [1999] (facilita a compreensão e a comunicação dos requisitos de segurança com as diversas pessoas envolvidas no desenvolvimento de SI) e pela abordagem de mercado eletrónico seguro e justo [Röhm e Pernul 2000]. Este paradigma permite o recurso a um padrão de modelação para a descrição das restrições de segurança existentes nos modelos de processos de negócio. Para finalizar, o último paradigma pondera o estudo da modelação de restrições de segurança (não é claro como o processo

de design de SSI é realizado) [Pernul 1992], a semântica de segurança orientada ao objeto (sugere uma notação que visa conceituar e sistematizar o desenvolvimento de SSI) [Ellmer et al. 1995], os dados e a semântica de segurança [Pernul e Quirchmayr 1994] e a modelação de DFD e ER (melhora a perceção do uso de modelos de controlo de acesso através de notação de modelação) [Pernul et al. 1998]. Este paradigma considera os aspetos de segurança da informação e a modelação de dados.

Vários enquadramentos têm sido desenvolvidos no seguimento dos quatro paradigmas de Burrel e Morgan [1979]. A razão disso deve-se ao facto desses paradigmas serem mutuamente exclusivos e por isso dados de SI ou métodos SSI só podem ser incluídos apenas num único paradigma. Enquadramentos mais recentes consideram os métodos de pesquisa, o papel organizacional dos SI, e objetivos de pesquisa, como paradigmas que não são mutuamente exclusivos e métodos de SI ou de SSI podem ser caracterizados sob diversas perspetivas. Siponen [2005a] utiliza um enquadramento composto por quatro pontos de vista: objetivo de pesquisa, papel organizacional da SSI, abordagens de investigação usadas e a sua aplicabilidade para o desenvolvimento dos SI, conforme se ilustra na Figura 2.9, relacionando estes com as cinco abordagens modernas de SSI apresentadas anteriormente. Os objetivos de pesquisa permitem perceber o papel das diversas abordagens de SSI, apresentando-se como críticos, interpretativos e orientados ao meio. Críticos porque ajudam a identificar possíveis falhas nas abordagens existentes, acreditam que nada pode ser tomado como garantido ou dogmaticamente aceite e são as melhores para o desenvolvimento de transformações profundas e alterações de paradigma, interpretativos uma vez que as pessoas estão cada vez mais cientes das suas ações tentando aumentar o profissionalismo e a compreensão de pesquisa dos fundamentos da SSI, e orientados ao meio pois são aplicados a prever e controlar processos naturais ou sociais, ou encontrar soluções para um problema real [Chua 1986].

O papel organizacional da SSI revela os pressupostos subjacentes aos métodos de SSI quanto à sua relação com a organização. Este papel pode ser técnico, sociotécnico ou social. Técnico quando a qualidade técnica é fraca, as preferências dos utilizadores não têm impacto no design de SSI e a resistência dos utilizadores é a origem dos problemas de SSI. Sociotécnico pois sistemas técnicos e organizacionais são importantes e a falta de ajuste entre estes é a principal causa de problemas da SSI.

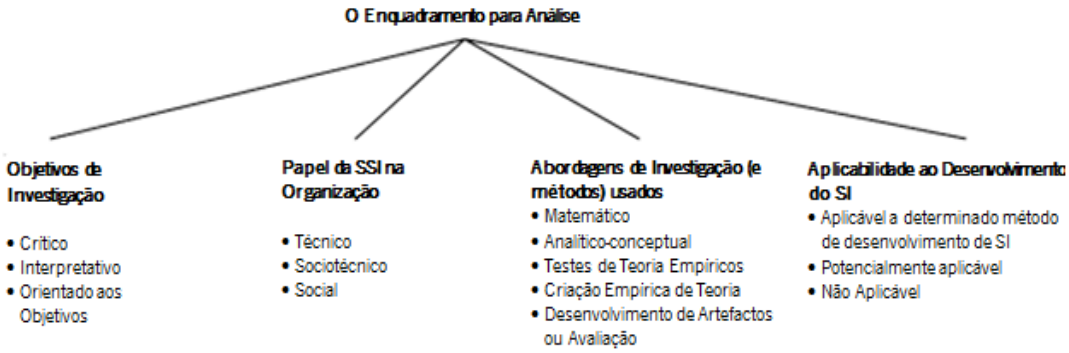


Figura 2.9: Enquadramento de análise de Siponen [2005a]

O social "...considera um sistema de informação principalmente como um sistema organizacional e social; um sistema de informação é visto como uma parte integrante e constitutiva da comunicação organizacional, controlo, coordenação, cooperação e contratos de trabalho e não apenas como um sistema de apoio separados para estas atividades organizacionais. " [Iivari e Hirschheim 1996, p. 553] Em relação às abordagens de investigação, existem diversos métodos de investigação com o propósito de classificar as abordagens de investigação ou estratégias. Para classificar a investigação o autor selecionou duas classificações, a primeira onde os métodos de investigação incluem matemática, conceitual-analítica, teste empírico da teoria, teoria empírica de criação e construção de artefactos ou avaliação de abordagens de pesquisa [Jarvinen 2000] e a segunda em que sempre que possível identificar o uso de métodos de pesquisa, incluindo a experiência em laboratório, experiência em campo, inquéritos, estudos de caso e simulações [Galliers 1992a] . Por fim, na aplicabilidade para o desenvolvimento dos SI as abordagens SSI podem ser integradas com os métodos de desenvolvimento do SI, caso contrário devem ser atividades contrárias e podem ser aplicáveis a um certo método ou processo de desenvolvimento de SI, aplicável potencialmente, ou não é aplicável a um método de desenvolvimento de SI [Baskerville 1988, 1992]. A Tabela 2.2 sintetiza os resultados da análise sobre o nível das abordagens individuais de SSI.

Tabela 2.1: Pressupostos Subjacentes aos Paradigmas e Abordagens de Desenvolvimento de SSI Modernos

Adaptado de Siponen [2005a]

Paradigmas e Abordagens		**Objetivos de Pesquisa**	**Papel Organizacional de SSI**	**Abordagens de Pesquisa**	**Suporte de aplicabilidade para o desenvolvimento do SI**
Segurança mudou o paradigma de desenvolvimento de SI	Abordagem de controlo lógico (Baskerville, 1988, 1989)	Orientado ao meio e crítico	Técnico	Análise conceptual, construção de artefacto e avaliação	Sim
	Metodologia virtual (Hitchings, 1995, 1996)	Orientado ao meio e interpretativo	Sócio-Técnico	Análise conceptual e construção de artefacto	Não
	Abordagem em espiral (Booysen and Eloff, 1995)	Orientado ao meio e interpretativo	Técnico	Análise conceptual e construção de artefacto	Sim
	Abordagem soft para o planeamento de SSI(James, 1996; Armstrong, 2000)	Orientado ao meio, interpretativo e crítico	Social	Análise conceptual, teoria de criação (por meio de levantamento quantitativo e qualitativo) e teoria teste (utilizando a pesquisa-acção)	Não
Paradigma de modelação de informação	Metodologia de planeamento de SSI (Straub and Welke, 1998)	Orientado ao meio e interpretativo	Entre técnico e socio-técnico	Análise Conceptual, testes de teoria da pesquisa-acção, utilizando como método de pesquisa	Não
	The security constraints modeling (Pernul, 1992)	Orientado ao meio	Entre técnico e socio-técnico	Construção de artefacto	Potencial
	Dados e semântica de segurança (Pernul and Quichmayr, 1994)	Orientado ao meio	Entre técnico e socio-técnico	Construção de artefacto e modelação matemática	Potencial

	Modelação DFD e ER (Pernul et al., 1998)	Orientado ao meio e interpretativo	Técnico	Construção de artefacto e modelação matemática	Potencial
	Semântica orientada a objetos de segurança (Ellmer et al., 1995)	Orientado ao meio	Técnico	Construção de artefacto e modelação matemática	Não
Paradigma de modelação de responsabilidade	Abordagem de Strens e Dobson (1993)	Interpretativo	Sócio-Técnico	Análise conceptual e construção de artefacto	Não
	Análise semântica de responsabilidade (Backhouse and Dhillon, 1996)	Interpretativo	Sócio-Técnico	Análise conceptual e construção de artefacto	Não
	Abordagem de autorização baseada em tarefas (Thomas and Sandhu, 1994)	Orientado ao meio e crítico	Técnico	Análise conceptual e construção de artefacto	Não
	Caso de abuso (McDermott and Fox, 1999)	Interpretativo	Sócio-Técnico	Análise conceptual	Sim
Paradigma de processos de negócio	Abordagem de Hermann e Pernul (1999)	Orientado ao meio e interpretativo	Técnico	Análise conceptual e construção de artefacto	Não
	Mercado eletrónico justo e seguro (Rohm and Pernul, 2000)	Orientado ao meio	Entre técnico e socio-técnico	Análise conceptual	Não
Paradigmas de SI viáveis e capazes de sobreviver	Sobrevivência de SI (Karya et al., 2001)	Orientado ao meio, crítico e interpretativo	Sócio-Técnico	Análise conceptual	Potencial
	Modelo de sistema viável (Hutchinson and Warren, 2000)	Interpretativo, Orientado ao meio	Técnico	Análise conceptual	Não

Características como orientado ao meio-fim e interpretativo são o objetivo de pesquisa mais predominante no estudo, sendo a característica crítica a menos utilizada. Em relação ao papel organizacional, o papel técnico é o mais usado, seguindo-se o sociotécnico e com apenas uma abordagem relacionada, o social. Ainda sob este ponto de vista, podem existir abordagens que se encontram entre o papel técnico e o sociotécnico. A análise conceptual foi o método mais influente em relação às abordagens de pesquisa. Testes da teoria empírica, testes da criação empírica, construção de artefactos e abordagens de pesquisas de avaliação podem ser usados sobre as diversas abordagens de SSI. Apenas três das abordagens apresentadas têm aplicabilidade para o desenvolvimento dos SI, mas existem outras quatro abordagens que com algumas modificações em determinadas partes podem ser métodos de desenvolvimento de SI.

Desta forma, se os utilizadores não sabem usar corretamente os sistemas de segurança, qualquer solução de segurança implementada numa organização nunca será a solução mais correta para solucionar os seus problemas. Contudo, métodos futuros de SSI devem estar mais direcionados para a integração das diferentes abordagens de SSI com o desenvolvimento de SI. Neste estudo ainda falta estabelecer a classificação de cinco gerações dos métodos de SSI a seguir. Para isso apenas se atualizou a classificação proposta por Baskerville [1993] onde se acrescentou uma quarta geração e que métodos futuros de SSI devem tornar-se uma quinta geração. Uma vez já descrita a investigação de Baskerville [1993] de seguida apenas será abordada a terceira, quarta e quinta geração. Assim, Siponen [2005a] considera que a terceira geração foca-se nos diversos meios de modelação de requisitos de SSI organizacionais, enquanto a quarta geração tem contributos das comunidades de SSI e adiciona aspetos de design sociotécnicos e sociais à geração anterior. Em relação à possível quinta geração, esta deve incluir técnicas sociais e métodos de SSI que sejam adaptáveis e desenvolvidos rigorosamente em conjunto com a prática. Cada uma destas gerações é composta por diversos paradigmas de SSI.

A Figura 2.10 permite perceber com clareza o que compõe cada uma das gerações. Com este estudo foi possível compreender as disciplinas e as comunidades de investigação que estão subjacentes aos paradigmas e abordagens modernas de SSI, estabelecendo uma comparação entre estes; descobrir os pressupostos explícitos que se escondem nas abordagens, sendo importante testar estas abordagens na prática com suporte em estudos empíricos rigorosos.

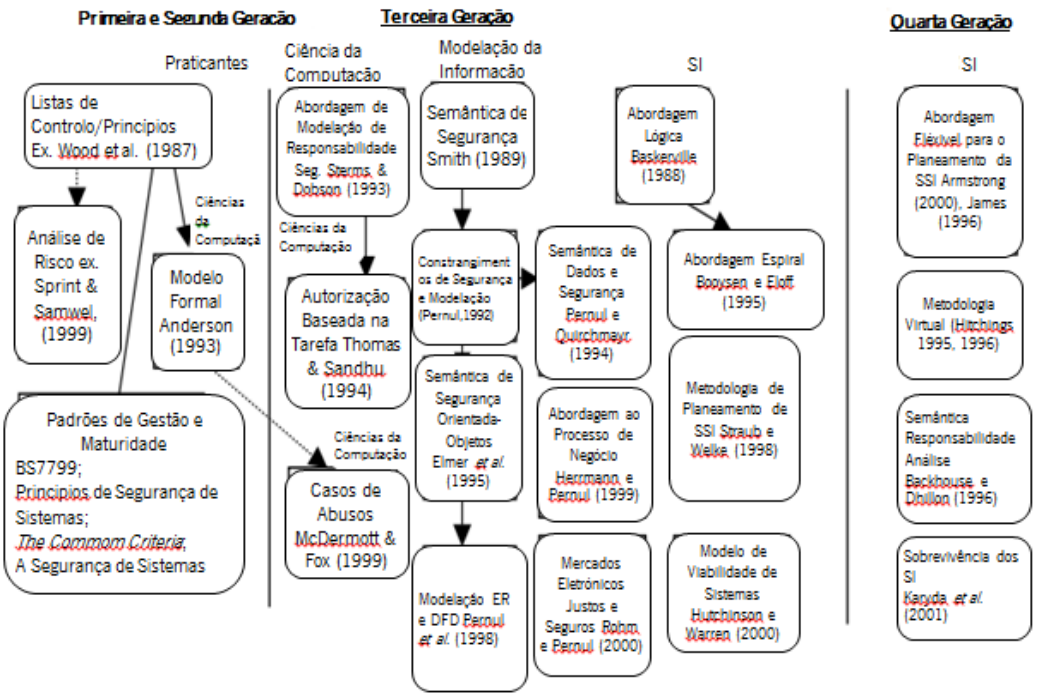


Figura 2.10: Classificação das Gerações de Siponen

Adaptado de Siponen [2005a]

A segunda publicação de Siponen tinha por objetivo comparar as principais suposições dos métodos tradicionais da SSI: *checklists*, normas de SSI, critérios de maturidade dos SSI, gestão de risco e métodos formais, sendo estes os métodos mais utilizados [Siponen 2005b]. Importa assim constatar os pressupostos subjacentes e as características dos métodos de SSI das gerações precoces sobre uma perspetiva crítica [Siponen 2005b]. O autor apresenta diversas características e pressupostos que devem ser tidas em consideração pelos investigadores que utilizam os métodos tradicionais da SSI, com o objetivo de proteger as organizações contra as ameaças. Classes de métodos tradicionais de SSI estão na base da primeira e segunda geração [Baskerville 1992; Siponen 2005a]. Investigadores classificam os métodos tradicionais de SSI em três [Baskerville 1988, 1992] ou cinco gerações [Siponen 2005a]. Existem diversos enquadramentos de análise para os diferentes métodos da literatura, no entanto no estudo de Siponen [2005b] o autor fundamenta a sua análise com o mesmo enquadramento usado no seu primeiro estudo [Siponen 2005a]. Este enquadramento é composto por objetivos de pesquisa [Chua 1986; Habermas 1984, 1987], papel organizacional da SSI [Iivari e Hirschheim 1996; Iivari e Kerola 1983], abordagens de investigação usadas [Jarvinen 2000] e a aplicabilidade para o desenvolvimento dos SI [Baskerville 1988, 1992] já descrita anteriormente será agora aplicada às cinco classes de métodos tradicionais

da SSI. A Tabela 2.2 mostra os resultados do estudo de Siponen [2005b], onde são apresentados os pressupostos subjacentes e as características dos métodos tradicionais de SSI à luz dos quatro pontos de vista identificados como sendo os mais eficientes para o presente estudo.

A primeira classe, listas (*checklists*) de SSI, é composta por quatro manuais de SSI: a AFIPS que permite criar pensamentos novos e construtivos sobre a SSI [Browne 1979]; a SAFE, composta por diversas áreas onde cada uma é formada por vários postos de fiscalização que aparecem na forma de perguntas concretas [Krauss 1972]; listas orientadas para o negócio onde os projetos permitem um rápido e fácil acesso a listas para os gestores [Moulton e Moulton 1996] e listas de controlo global que facilitam o uso de uma ferramenta para identificar controlos apropriados [Wood et al. 1987]. A segunda classe, normas SSI, permitem construir internacionalmente normas autoritárias e normas genéricas de SSI, sendo geralmente expressas em termos de objetivos ou imperativos [Siponen 2005b]. Fazem parte desta classe normas como a BS ISO/IEC 17799 [2000], a GASSP [Poore 1999] e a orientação inicial da SSI [Sanders et al. 1996]. A classe critérios de maturidade, caracterizada por uma escala objetiva de cinco níveis para classificar a maturidade da SSI [Siponen 2005b], foram selecionados três instrumentos: *System Security Engineering Capability Maturity Model* (CMM-SSE) [SSE-CMM 2003], *Software Security Metrics* [Murine e Carpenter 1984] e *Information Security Program Maturity Grid* [Stacey 1996]. A classe gestão de risco consiste na gestão e controlo do risco de SSI onde se selecionou para análise: a abordagem genérica de gestão de risco [Saltmarsh e Browne 1983], o método de negócio focado na gestão de risco [Halliday et al. 1996], o método X-ifying de gestão de risco [Frisinger 2001], o LRAM [Guarro 1987] e a abordagem de comunicação [Baskerville 1991]. Finalmente, a classe métodos formais com os trabalhos de Anderson [1993] e Barnes [1998], onde o desenvolvimento de SI deve basear-se em componentes validas de formalidade ou concretizas pelos métodos formais [Siponen 2005b].

A leitura pormenorizada da Tabela 2.2 permite observar que o objetivo de pesquisa mais utilizado é o orientado ao meio, estando também presente em alguns métodos o objetivo de pesquisa interpretativo e crítico, ou então a junção entre os três objetivos. Estudos interpretativos ajudam na resolução de problemas complexos de SSI, enquanto estudos críticos ajudam a identificar eventuais lacunas nos métodos principais, sendo estas indispensáveis [Siponen 2005b].

Tabela 2.2: - Resultados da análise de Siponen as principais suposições dos métodos tradicionais da SSI

Adaptado de Siponen [2005b]

Classes de métodos tradicionais de SSI					
	Métodos	**Objetivos de Pesquisa**	**Papel Organizacional de SSI**	**Abordagens de Pesquisa**	**Aplicabilidade para o DSI**
Checklists	AFIPS	Orientado ao meio e crítico	Técnico	Análise Conceptual	Não
	SAFE	Orientado ao meio	Sócio-Técnico	Análise Conceptual	Não
	Moulton and Moulton	Orientado ao meio	Entre o Técnico e Sócio-técnico	Análise Conceptual	Não
	Wood et al.	Orientado ao meio	Entre o Técnico e Sócio-técnico	Análise Conceptual	Não
Normas de SSI	BS ISSO/IEC17799 (2000)	Orientado ao meio	Técnico	Análise Conceptual	Não
	GASSP	Orientado ao meio e interpretativo	Sócio-Técnico	Análise Conceptual	Não
	Sanders et al.	Orientado ao meio	Técnico	Análise Conceptual	Não
Critérios de Maturidade	SSE-CMM	Orientado ao meio	Técnico	Análise Conceptual	Não
	Grid de maturidade de SSI	Orientado ao meio e interpretativo	Sócio-Técnico	Análise Conceptual	Não
	Métricas de SSI	Orientado ao meio e crítico	Técnico	Análise Conceptual	Potencialmente aplicável
Gestão de Risco	Métodos Genéricos de RM	Orientado ao meio	Entre Técnico e Sócio-Técnico	Análise Conceptual	Não
	Métodos de negócio focados na RM	Orientado ao meio e crítico	Técnico	Análise Conceptual	Não
	The X-ifying RM method	Orientado ao meio e crítico	Sócio-Técnico	Análise Conceptual, Teoria da criação e testes de pesquisa empírica; pesquisa	Não
	LRAM	Orientado ao meio e interpretativo	Técnico	Análise Conceptual, modelação matemática	Não
	Abordagem de comunicação	Orientado ao meio e crítico	Sócio-Técnico	Análise Conceptual	Não
Método Formal	Barnes	Interpretativo	Técnico	Modelação matemática	Não
	Anderson	Orientado ao meio e crítico	Técnico	Modelação matemática	Não

O papel organizacional da SSI mais predominante é o técnico, seguindo-se o sociotécnico. Também existem casos em que aparecem ambos os papéis para identificar o mesmo método, ou então também se vê que o ponto de vista social não se identifica com nenhum método. No que se refere à abordagem de investigação utilizada, grande parte dos métodos identifica-se por uma abordagem de análise conceptual, de seguida pela abordagem de modelação matemática e apenas um dos métodos utilizada as duas abordagens anteriores, adicionando a teoria da criação e teste de pesquisa empírica, e avaliação. Finalmente, em termos da aplicabilidade dos métodos no desenvolvimento de SI, apenas um dos métodos é que pode ser potencialmente aplicável (Grid de maturidade de SSI). Os métodos *checklists* e as normas da SSI asseguram as melhores práticas, as normas de maturidade da SSI têm como principal objetivo expor o nível de maturidade da segurança de todos os SI baseando-se em estudos empíricos de forma a testar e aperfeiçoar na prática os métodos da SSI [Siponen 2005b]. "A conclusão será de que os métodos tradicionais de SSI denotam certas características e pressupostos, os quais devem ser tidos em conta pelos desenvolvedores dos métodos, investigadores, profissionais/praticantes utilizando os métodos tradicionais de SSI" [Siponen 2005b, p. 314].

2.4.4 Contextualização de Villarroel et al.

Com o passar dos anos o conceito de segurança foi evoluindo assim como a noção por parte das organizações da necessidade de segurança nos seus sistemas. Ainda muito recente se considera que soluções de segurança baseiam-se na implementação de *firewalls*, *routers*, servidores de configuração, *passwords* e criptografia, quando o ideal seria resolver os verdadeiros problemas de fundo dos sistemas de informação e implementar uma conceção de software adequada para os sistemas em questão [Villarroel et al. 2005]. Como resposta a este pressuposto, têm sido desenvolvidos métodos com o objetivo de incluir a segurança nos processos de desenvolvimento. A segurança deve influenciar todos os aspetos de conceção, implementação e teste de *software*.

Villarroel et al. [2005] publicaram um artigo de forma a reunir todas as metodologias desenvolvidas com o propósito descrito anteriormente e estabeleceram a comparação que existe entre as mesmas. Afirmam, no entanto, que "(...) deixamos bem claro que os aspetos de segurança não podem ser completamente especificados por esses métodos já que eles têm uma série de limitações que nós temos que ter em conta. Ao mesmo tempo, cada um destes métodos engloba aspetos muito importantes em matéria de segurança que podem ser usados como base para novos

métodos ou extensões que podem ser desenvolvidos". Neste estudo, os autores identificaram onze métodos:

- *MOMT (Multilevel Object Modeling Technique)*

Marks et al. [1996] asseguram que este método permite desenvolver vários níveis para bases de dados seguras e estabelecem regras de interação entre os elementos do modelo. Este modelo é composto por três fases: análise (analisa futuras vulnerabilidades do sistema tendo em conta o modelo objeto multinível, o modelo dinâmico multinível e o modelo funcional multinível), sistema de *design* (design de base de dados a vários níveis, alto-nível, estrutura de sistemas e bases de dados multiníveis) e objeto de *design* (*design* de módulos em detalhe de sistemas automatizados). Contudo, este método apenas explica a fase de análise não apontando soluções válidas para as atuais situações onde a tecnologia utilizada e as necessidades de segurança se alteram.

- *Business Process-driven Framework for Security Engineering*

Vivas et al. [2003] sugerem um enquadramento orientado ao processo de negócio, onde as decisões tecnológicas são dirigidas pelo modelo de negócio. Baseia-se em UML (*Unified Modeling Language*) e integra requisitos de segurança no modelo de processo de negócio do sistema. Esta abordagem tem um fim experimental e exploratório, sendo focada na identificação e discussão de problemas e não na oferta de soluções.

- *UMLSec: Secure Systems Development Methodology using UML*

"O objetivo é auxiliar a difícil tarefa de desenvolvimento de sistemas de segurança-crítica numa abordagem baseada na notação da Linguagem de Modelação Unificada. Apresentamos a extensão UMLsec do UML que permite expressar com segurança informações relevantes dentro dos diagramas de especificação do sistema" [Jurjens 2002, p. 412]. Segurança multinível e controlo de acesso obrigatório são os modelos de segurança com maior relevo na atual proposta. Este método integra a versão de modelo UMLSec definido para a introdução de padrões no processo de conceção, tendo em conta requisitos de segurança como confidencialidade e integridade. Contudo, não considera o *design* de base de dados segura de acordo com aspetos conceptuais, lógicos e físicos.

- *Secure Database Design Methodology*

Consiste num método de criação e implementação de base de dados seguras, que consideram as limitações em matéria de informação sensível na fase de requisitos. Os modelos e as linguagens auxiliam as ferramentas a especificar os constrangimentos e a classificar a informação em diversos níveis de segurança, especificando o papel que os utilizadores devem desempenhar para aceder à informação [Fernández-Medina e Piattini 2003]. No entanto, este método não é o mais apropriado para o desenvolvimento de SI seguros.

- *Security and Privacy Requirements Analysis Methodology within a Social Setting*

Este enquadramento trata os requisitos de segurança e privacidade de uma estrutura de modelação orientada a agentes sob uma perspetiva *top-down* ou *bottom-up*. Suporta um conjunto de técnicas de análise: análise do atacante (ajuda a identificar possíveis invasores do sistema e as suas intenções maliciosas), análise de dependência das vulnerabilidades (ajuda a identificar vulnerabilidades ao nível das relações organizacionais entre os *stakeholders*), análise de contramedidas (motivação do invasor, vulnerabilidade do sistema e as suas capacidades para atacar são fatores de risco) e análise de controlo de acesso (preenche a lacuna entre modelos de requisitos de segurança e modelos de implementação de segurança) [Liu et al. 2003]. No entanto, este enquadramento não tem em conta o processamento de base de dados, assim como, as fases de desenvolvimento de SI.

- *A Paradigm for Adding Security into IS Development Methods*

Este paradigma ajuda os desenvolvedores a usar e modificar os métodos de SSI existentes, conforme necessário. De acordo com Siponen [2002, p. 52] e para este paradigma, "Primeiro, olhamos transversalmente para o Desenvolvimento de Software de Sistemas de Informação (DSSI) e as metodologias de desenvolvimento de segurança de SI a fim de encontrar conceitos fundamentais comuns (sujeitos e objetos). De seguida mostramos os padrões dos métodos de SSI existentes resultando quatro conceitos adicionais (restrições de segurança, classificações de segurança, sujeitos de abuso e cenários de abuso, e políticas de segurança). Finalmente, nos consultamos um painel de praticantes para obtermos comentários sobre os padrões". Foram identificadas seis dimensões que permitiram capturar o maior número de padrões no desenvolvimento de sistemas de segurança: assuntos de segurança (entidades que têm uma dependência de segurança para com os ativos da organização), objetos de segurança (ativos da

organização que são pertinentes em termos de segurança de informação), restrições de segurança (incluem acesso de leitura, acesso de escrita, etc.), classificações de segurança (necessidade de classificar objetos de segurança e assuntos de acordo com as informações sobre sensibilidade de segurança), sujeitos de abuso (refere-se aos objetos que podem ser alvo de uma violação de segurança) e cenários de abuso, e política de segurança (conjunto de leis, regras e práticas que regulam como uma organização gere, protege e distribui informações confidenciais). O problema desta abordagem é ao nível dos modelos gráficos e linguagem utilizada para sustentar a proposta ou prover maior formalidade [Villarroel et al. 2005].

- *CoSMo: An Approach Towards Conceptual Security Modeling*

Modelação conceitual deve abranger requisitos de segurança e mecanismos de segurança de alto nível. Este método compreende a necessidade de integrar considerações de segurança nos processos de modelação de software com base em UML. Autenticação (garantir a autenticidade), integridade/coerência (mantém a informação única e protege que esta seja modificada por indivíduos estranhos; normas que regem a atualização e validação dos dados), sigilo/confidencialidade (divulgação não-autorizada da informação; normas que regem a classificação de dados e acesso a ela), privacidade (organizações devem manter a informação dos indivíduos confidencial), autorização (conjunto de regras sobre "quem tem que tipo de acesso e a que informação), controlo de acesso (processos que controlam autorização, limitando o acesso aos dados armazenados apenas para utilizadores autorizados), disponibilidade (atender a indivíduos autorizados, com informações sempre que solicitado), responsabilidade (capta os requisitos de que as pessoas podem ser responsabilizadas por atividades de segurança relevantes), auditoria (recolha e organização dos dados e análise dos dados para descobrir violações de segurança), não-repúdio (indivíduo não consegue negar ter realizado determinada ação relacionada com dados), anonimato (ausência de identidade), validade (validade de contratos e assinaturas digitais), criptografia (mecanismos: algoritmos de criptografia, assinaturas digitais e funções de verificação de integridade), certificados digitais (mostram e provam a autenticidade dos participantes de comunicação envolvidos) e confiança a terceiros (mercado eletrónico em que outros participantes são obrigado a confiar, porque eles realizam tarefas sensíveis) são todos os requisitos e mecanismos de segurança que podem ser vistos na literatura [Artelsmair et al. 2002].

- *Using Aspects to Design a Secure System*

Este método, orientado a aspetos de modelação, permite aos desenvolvedores incluir preocupações de *design*, facilitando a criação de um design sistémico e consistente. Aspetos são abordados como padrões de *design*, e têm como benefícios: permitir a compreensão e comunicação das preocupações com segurança, maior facilidade em modelar e perceber o seu comportamento. Adicionalmente, os aspetos são reutilizáveis em diversos sistemas, as alterações às inquietações de segurança são executadas em locais (aspetos), e o impacto das preocupações de segurança nas unidades de *design* pode ser observado através da criação de aspetos de segurança em modelos primários e na avaliação dos modelos resultante [Georg et al. 2001].

- *A Methodology for Secure Software Design*

Fernandez [2004] propõem "apresentar uma metodologia para criar *software* seguro. Esta abordagem faz uso do princípio básico da segurança e desenvolvimento orientado ao objeto. Consideramos *design* orientado-objeto, o *Unified Modeling Language* (UML), e padrões (1,12) como o essencial na criação de software bem concebidos. (...) A ideia principal do método proposto é que os princípios de segurança devem ser aplicados a todos os fase de desenvolvimento e que cada estágio pode ser testado para o cumprimento desses princípios". O desenvolvimento deste método é realizado em quatro fases: requisitos (casos de uso explicam as interações necessárias com o sistema), análise (padrões de análise podem ser usados para implementar o modelo conceitual de forma mais confiável e mais eficiente), *design* (casos de uso devem estar representados na interface) e implementação (reflete no código das regras de segurança definidas para a aplicação). Em todas estas fases devem ser aplicados os princípios de segurança e realizadas auditorias de forma a validar a instituição de política que foi implementada. A combinação de diferentes tipos de padrões facilita uma funcionalidade diferente e de qualidade, mas carece ao nível do modelo gráfico e da utilidade de linguagem que apoia a metodologia ou que proporciona uma maior formalidade.

- *ADAPTed UML: A Pragmatic Approach to Conceptual Modeling of OLAP Security*

Este método e linguagem conceitual para a modelação de segurança OLAP baseia-se em UML ("*ADAPTed UML*"), dá suporte à modelação e ao *design*, e supõe uma política de segurança central que permite sincronizar os modelos conceptuais e lógicos com o modelo físico da ferramenta OLAP. Este modelo é composto por cubos, medidas, níveis de dimensões e atributos da dimensão. O modelo contempla a conceptualização de restrições de permissão, as quais podem

consistir num aspeto positivo (subsídios explícitos) ou negativo (negações explícitas). Os requisitos devem ser realizados e ser rigorosos, a fim de suportar as fases de design subsequentes. Uma outra funcionalidade disponível é a MDSCL, a qual consiste numa linguagem de restrições de segurança multidimensional para qualquer sistema alvo, baseada em MDX e que depende do OLAP conceitual e do modelo de segurança pretendido. Para expressar as restrições de autorização (negativa), sugere-se um conjunto de demonstrações HIDE (*Hide Cube*, *Hide Measure*, *Hide Slice*, *Hide Level*, *Hide Level Where*, *Hide Measure Where*) [Priebe e Pernul 2001].

- A *UML Extension for Secure Multidimensional Conceptual Modeling*

Modelação multidimensional (MD) é a base para uma *Data Warehouses* (DW), Base de Dados MD e OLAP, permitindo um forte mecanismo de descoberta de informações de negócio, cruciais para os processos de tomada de decisão estratégica [Fernández-Medina et al. 2004]. Esta abordagem é uma extensão UML que permite representar as principais informações de segurança e limitações multidimensionais da modelação ao nível conceptual. A sua importância prende-se com o facto de que em todas as fases do processo de desenvolvimento considera os aspetos de segurança com foco nas operações de leitura (operação mais comum por parte dos utilizadores). A implementação desta abordagem proporciona modelos multidimensionais seguros como qualquer *Database Management System* (DBMS) comercial que seja apto de implementar base de dados a vários níveis, como OLS ou DB2 Base de Dados Universal (UDB). De forma a apresentar uma abordagem para a extensão UML são consideradas três fases [Fernández-Medina et al. 2004, p. 220]: em primeiro lugar "Definir com precisão na organização os utilizadores que terão acesso ao sistema multidimensional"; em segundo "Classificar as informações para o modelo MD", e em terceiro "Cumprir obrigatoriamente o controlo de acesso". Os elementos que compõem o esquema são: descrição (pequena descrição da a extensão em linguagem natural), extensões pré-requisitos (atual extensão precisa da extensão anterior), estereótipos/valores de *tagged* (definição dos estereótipos e/ou valores etiquetados), regras de boa formação (a semântica estática da meta classes são definidas tanto em linguagem natural como num conjunto de invariantes expressa por meio de expressões OCL) e comentários (adicionar qualquer comentário, decisão).

Fernandes-Medina et al. [2004], para compararem os onze métodos apresentados anteriormente, basearam-se no enquadramento de Khwaja e Urban [2002a] que apresenta a diferença clara de conceitos de especificação e técnicas de especificação. De acordo com Khwaja e Urban [2002b, p. 581] são vários os critérios de avaliação disponíveis para os desenvolvedores

utilizarem nas fases de especificação, assim como existem diversas técnicas de especificação. Os mesmos autores esclarecem que um critério é entendido como uma declaração usada para identificar as propriedades, isto é, valores. (...) As especificações podem ou não realizar uma propriedade" [Khwaja e Urban 2002b, p. 582]. Khwaja e Urban explicam ainda que os critérios de avaliação permitem identificar o conjunto de propriedades pretendidas nas especificações e nas técnicas de especificações para a avaliação da sua qualidade. Os critérios técnicos são o nível de formalidade, por isso, um elevado nível de formalidade em especificação técnica auxilia no alcance de definições precisas, ambíguas, consistentes e completas, e especificações verificáveis.

Um elemento de um critério de especificação pode ser: compreensível (especificação do sistemas deve ser um modelo cognitivo, compreensível), adequado (funcionalidade separada da implementação), inequívoco (falta de precisão), completo, consistente, correto, verificável, testável, modificável (manutenção, adaptável), rastejável, mínimo (permite economia de expressão). Em relação aos critérios técnicos de especificação que foram identificados, a Tabela 2.2 reflete os critérios técnicos que existem e qual o tipo de critério de especificação que lhe está subjacente. Uma vez que cada critério de especificação técnica pode ser relacionado com um ou mais critérios de especificação, sendo que a resposta de cada metodologia será relacionada com o cumprimento da técnica com relação a um critério de especificação. A concretização de um critério técnico deve ter como resultado o cumprimento de todos os critérios de especificações relacionados a esse critério e a especificação deve ser compreensível, adequada e mínima. Os critérios apresentados completam os aspetos formais, no entanto, cada uma das propostas apresentadas apresenta alguns pontos fracos. Villarroel et al. [2005, p. 318] afirmam que "É muito difícil desenvolver um método que preencha todos os critérios e compreenda todas as restrições de segurança em termos de confidencialidade, integridade e disponibilidade. Se esse método fosse desenvolvido, a sua complexidade evitaria o seu sucesso. Portanto, a solução seria uma solução mais completa em que as técnicas e modelos definidos pelas normas do modelo mais aceite fossem utilizadas".

A razão que leva à escolha dos métodos anteriormente descritos deve-se ao facto destes na generalidade tentam resolver problemas de segurança (especialmente relacionados com confidencialidade), realçam aspetos de modelação da segurança e usam linguagens de modelação que facilitam o processo de *design* de segurança. Os métodos apresentados incluem aspetos de segurança muito importantes, que vão desde as primeiras fases de desenvolvimento até as fases finais, e que podem ser usados como suporte para o aperfeiçoamento dos métodos atuais.

Tabela 2.2: Comparação usando Critérios de Avaliação para Especificações de Software e Técnicas de Especificação

Adaptado de Villarroel et al. [2005]

Critérios Técnicos	Critérios de Especificação	Metodologias										
		Marks et al. (1996)	Vivas et al (2003)	Jurjens (2002)	Fernández-Medina and Piattini (2003)	Liu et al. (2003)	Siponen (2002)	Artelsmair et al. (2002)	Georg et al. (2002)	Fenández (2004)	Priebe and Pemul (2004)	Fenández Medina et al. (2004)
Adequação Expressiva	Compreensível	X	X	X	X	X	X	X	X	X	X	X
	Adquado	(x)	X	(x)	X	(x)	X	X	X	X	X	X
	Mínimo	X	X	X	X	(x)	(x)	X	X	X	X	X
Construtibilidade specifications		X	X	X	X	(X)	(X)	(X)	(X)	(X)	X	X
Âmbito das Especificações	Completo	(x)	X	(x)	(x)	(x)		(x)	(x)	(x)		X
Nível de Formalidade	Inequíveco	X	(x)	X	X	X	(x)	(x)	X	(x)	X	X
	Consistente	X	(x)	X	X	X	(x)	(x)	X	(x)	X	X
	Completo	X	(X)	X	(x)	X	(x)	(x)	X	(x)	X	X
	Verificável	X	(X)	X	X	X	(x)	(x)	X	(x)	X	X
	Válido	X	(X)	X	X	X	(x)	(x)	X	(x)	X	X
Base Formal	Inequívoco	X	(x)	X	X	X	X	X	X	X	X	X
	Consistente	X	X	X	X	X	X	X	X	X	X	X
	Completo	X	(x)	X	(x)	X	X	X	X	X	X	X
	Verificável	X	X	X	X	X	X	X	X	X	X	X
	Válido	X	X	X	X	X	X	X	X	X	X	X
Extensão da aplicabilidade		(x)	X	(x)	X	(x)	X	X	X	X	X	X
Fácil de usar		X	X	X	X	(x)	X	(x)	(x)	(x)	X	X
Ajuda			(x)	(x)	(x)				(x)		(x)	(x)

Ambiente Integrado e Suporte de Ferramentas			(x)	(x)	(x)							
Especificação de suporte à organização	Compreensível	X	(x)	X	X	X	(X)		(x)			(x)
	Modificável	X	X	X	X	X	(X)		(x)			(x)
Suporte para Manutenção	Modificável	X	(x)	X	X	X						
	Rastreáveis	(x)	(x)	X	(x)	X						
Executável	Compreensível	(x)	(x)		X	(x)	(x)				(x)	
	Inequíveco	(x)	(x)		X	X	(x)				(x)	
	Consistente	(x)	(x)		X	X	(x)				(x)	
	Completo	(x)	(x)		(x)	(x)	(x)				(x)	
	Correcto	(x)	(x)		X	X	(x)				(x)	
	Verificável	(x)	(x)		X	X	(x)				(x)	X
	Válido	(x)	(x)		X	(x)	(x)				(x)	X
Tolerância para a incompletude	Verificável	X	X	X	(x)	X	X		(x)			
	Válido	X	X	X	(x)	X	X		(x)			
Múltiplas visualizações	Compreensível	X	(x)	X	X	X		(x)	(x)	(x)	X	
Flexibilidade e Simplicidade de Notação	Compreensível	X	X	X	X	(x)	(x)	(x)	(x)	X	X	X
Verificação Interna de Apoio	Inequívoco											
	Completo											
	Consistente											
	Verificável											
Validação Externa de apoio	Correcto	X	(x)	(x)	(x)	(x)						
	Válido	(x)	(x)	(x)	(x)	(x)						
Suporte para outros estágios de desenvolvimento	Rastreáveis			(x)	(x)							
Apoio à Geração de Documentação	Compreensível	(X)		(X)	(X)			(X)	(X)			

2.4.5 Conclusão

Na secção 2.4 foram apresentados vários enquadramentos sobre a literatura da SSI. Cada autor apresenta a sua visão de como desenvolver um SI seguro, mas também alertam para a importância de que cada um dos métodos de SSI tenha em consideração determinados pressupostos que podem ser únicos. Uma vez que cada um destes métodos de SSI foi analisado isoladamente, nesta secção será feito um estudo comparativo entre os diferentes métodos já explorados no que toca às principais características, limitações, lacunas existentes, semelhanças, terminando com uma breve reflexão da importância da análise dos métodos, esquematizado na Tabela 2.3. Para complementar o estudo, também foram obtidas conclusões sobre os esquemas de classificação sugeridos pelos investigadores, assim como apresentados os seus benefícios.

Após este estudo, foi possível concluir que o enquadramento para classificar a literatura em SSI mais referenciado pelos estudiosos da área é o de Burrell e Morgan [1979]. Os investigadores que se debruçam sobre a literatura de SSI consideração diferentes modos de classificar a informação quando comparados com os métodos utilizados na literatura de SI. Também é evidente que pesquisas em SSI detêm um atraso quando comparadas com pesquisas na área dos SI, existindo uma divergência entre os métodos gerais de desenvolvimento e os métodos de desenvolvimento de segurança [Baskerville 1993]. A ausência de estudos empíricos na literatura em SSI revela que o campo ainda está muito aquém da sua evolução.

O primeiro aspeto a ter bem presente do estudo realizado ao longo da secção 2.4 é que apenas se refere a áreas específicas da literatura de SSI, não abrangendo a área no global. Em relação aos estudos apresentados, pode- concluir que estes ainda não conseguem resolver todos os problemas com que se depara a literatura em SSI, isto porque não se consegue incluir todo o conhecimento que envolve a SSI nos métodos apresentados. A importância da revisão de literatura em SSI permite analisar todo o trabalho que tem sido realizado no âmbito da classificação da literatura em SSI considerando os aspetos positivos e negativos de cada um dos métodos, e obtendo uma conhecimento vasto sobre os diversos métodos já utilizados e sobre o que está dentro de cada um deles. Como possível "guia" para traçar a envolvente da SSI, o trabalho realizado por Siponen [2005a, 2005b] será o que melhor se adequa à literatura em SSI, uma vez que abrange um vasto conjunto de métodos que são analisados à vista dos principais aspetos que são tidos como os mais importantes para a SSI.

Tabela 2.3 - Comparação dos Esquemas de Classificação da Literatura de SSI

Esquema	Características	Limitações	Lacunas	Reflexão
Baskerville [1993]	Métodos focados para *checklists*, análise de risco simples de apoio à decisão, mecanismo de particionamento da complexidade do sistema desejado; Fornece uma descrição analítica da evolução dos SI, uma análise de segurança e métodos de conceção; Procuram controlos críticos que fornecem a proteção mínima para os SI.	Falta de integração dos métodos apresentados com os métodos gerais de desenvolvimento de SI, não permite grandes progressos para os métodos de segurança; Preocupação ainda para a segurança e para a segurança na análise de sistemas.	Desenvolvimento de métodos que foquem em modelos abstratos como meio para a compreensão das várias necessidades de segurança no SI; Métodos gerais não consideram especificações de segurança com rigor.	A análise de segurança e conceção de métodos evoluíram de forma idêntica aos métodos de desenvolvimento de SSI; A sinergia entre métodos de segurança e métodos de desenvolvimento de SI pode ser a resolução de alguns dos problemas existentes.
Dhillon e Backhouse [2001]	Mapeamento do território dos SI e da pesquisa de segurança; Analisa as preocupações sócio-filosóficas das diversas abordagens de segurança e SI; Sistemas baseados em computador interagem dinamicamente com os ambientes formais e informais que são usados.	Paradigmas mutuamente exclusivos que se contradizem uns aos outros; Avaliações em segurança e métodos de conceção baseiam-se em conceções funcionalistas; Grande parte da investigação em SSI tende a concentrar-se em estruturas formalizadas para conceção da segurança.	Confiança não é suficiente na conceção de abordagens; Crescente desilusão com a conceção formal, racional e excessivamente mecanicista na análise e conceção de SI.	Considerando as perspetivas sócio organizacionais e avaliações das preponderâncias das soluções técnicas, será o caminho mais estável para alcançar a SSI.
Siponen [2005a]	Análise dos objetivos de pesquisa, o papel organizacional da SSI, a abordagem de investigação e aplicabilidade do	Apenas três dos métodos de SSI podem ser integrados em métodos de SI; Métodos de SSI existentes pertencem	Deverá existir uma quinta geração sociável e adaptável a métodos de SSI;	Abordagens apresentadas pretendem melhorar a SSI; Implicações para a pesquisa da SSI, para

	desenvolvimento de SI são os pontos de vista analisados; Apresentação de disciplinas e comunidades científicas, revelação dos pressupostos explícitos, classificação de cinco gerações para as abordagens modernas de SSI.	apenas às quatro primeiras gerações identificadas.	Ausência de estudos empíricos.	práticas de SSI e educação de SSI; Futuros métodos de SSI devem estar mais direcionados para a integração das diferentes abordagens de SSI com o desenvolvimento de SI.
Siponen [2005b]	Comparação dos principais pressupostos ao nível de *checklists*, normas de SSI, critérios de maturidade dos SSI, gestão de risco e métodos formais.	Resultados não são comprovados; Métodos tradicionais de SSI oferecem pouca evidência sobre a sua utilidade na prática.	Devido a dimensões sociais dos SI é necessário existirem métodos de interpretação; Prestar atenção a questões sociais; Necessidade de estudos empíricos.	Pesquisas em SSI não é apenas a resolução de problemas concretos com a introdução de outros métodos ou ferramentas; Futuros métodos tradicionais de SSI devem fornecer orientações sobre como esses métodos se integram com o desenvolvimento de SI; Métodos tradicionais de SSI extravasam certas características e pressupostos que devem ser tidos em conta pelos desenvolvedores dos métodos, investigadores, profissionais.
Villarroel et al. [2005]	Dicas para construir sistemas corretos com base me técnicas formais de representação de requisitos e análise de correções em cada etapa do processo; Apresentação de onze metodologias.	Aspetos de segurança não podem ser completamente especificados pelas metodologias apresentadas. Quanto mais complexa for uma metodologia, menor será o seu sucesso.	Incapacidade de desenvolver metodologias que preencham todos os critérios e compreenda todas as restrições de segurança em termos de confidencialidade, integridade e disponibilidade;	Segurança deve influenciar todos os aspetos de conceção, implementação e teste de *software*; Metodologias analisadas geralmente tentam resolver problemas de segurança, realçar aspetos de modelação da segurança e usam linguagens de modelação.

2.5 Text Mining

Nos últimos anos assistiu-se a uma rápida evolução dos recursos computacionais, que permitiram a geração de grandes volumes de dados. De acordo com Dilly [1999], a quantidade de informação dobra em cada 20 meses, comprovando a taxa explosiva do crescimento de informação. Com o passar do tempo o armazenamento de dados ficou mais acessível, fruto da constante redução de custo de armazenamento e processamento de dados. Este custo está particularmente relacionado com a capacidade de extrair informação de alto nível que está subjacente a esses dados, isto é, a informação útil que serve o suporte de decisão, exploração e compreensão dos dados gerados. Ao analisar os dados gerados, pode ser percetível a existência de padrões ou tendências com elevada importância, proporcionando maior competitividade ao ambiental organizacional, otimização do processo de negócio, análises de mercado fascinantes, entre outros aspetos.

Segundo Hearst [1999], a lógica subjacente à análise de dados em *Data Mining*[14] é o sinônimo de "descoberta de conhecimento em base de dados" *(knowledge discovery in databases* - KDD), sendo um processo automatizado com metas bem definidas cujo objetivo é a captura e análise de grandes conjuntos de dados para extrair um significado. Han et al. [2011] consideram DM como sendo o processo de extração ou mineração de conhecimento de grandes quantidades de informação. Para Linoff e Berry [1997], DM é a exploração e análise de dados por meios automáticos ou semiautomáticos em grandes quantidades de dados, com o objetivo de descobrir regras ou padrões interessantes.

Para a compreensão e a realização do presente projeto não tem elevada importância conhecer em detalhe todas as características de DM, mas sim, compreender que para este projeto recorre-se a características que têm inspiração em DM, como é o caso da subárea *Text Mining*[15]. Recorrendo ao TM pode extrair-se informação relevante de grandes bases de texto, sem precisarem de uma leitura prévia. A diferença entre estas duas abordagens é que no DM as técnicas de mineração são aplicadas sobre dados estruturados (base de dados), enquanto que o TM aplica as técnicas de mineração de dados em dados não estruturados (formato de texto) [Feldman e Sanger 2007b].

[14] Termo usado para referir mineração de dados mas que dada a elevada disseminação do termo na língua original, inglês, neste trabalho será usado o termo original.

[15] Termo usado para referir mineração de texto, respetivamente, mas que dada a elevada disseminação do termo na língua original, inglês, neste trabalho será usado o termo original.

Ao longo desta secção será apresentada alguma literatura em TM e os respetivos pormenores que podem revelar-se de elevada importância para o desenvolvimento do projeto de dissertação. O objetivo desta secção não é detalhar toda a revisão de literatura sobre a temática de TM, mas sim, apresentá-la de forma genérica para que o leitor compreenda a razão de na realização deste projeto recorrer-se a técnicas desta temática.

O TM tem vindo cada vez mais a adquirir elevada importância face ao rápido crescimento da quantidade de informação textual disponível, tanto na Internet como em mecanismos de procura. Neste contexto, TM constitui-se uma técnica de grande importância, permitindo que elevadas quantidades de informação textual possa ser acedida, analisada e transformada numa fração de segundo. Através deste processo pode-se extrair informação relevante de base de textos, identificando e explorando padrões interessantes, classificando-os, sem que para isso seja necessária a sua leitura prévia.

O *Text Mining* é uma área interdisciplinar que não interage somente com *Data Mining*, aprendizagem máquina e estatística, mas também com linguagem computacional ou processamento da linguagem natural [Waegel 2006]. Uma vez que TM recorre a uma base de dados textual composta por um conjunto de documentos de texto de diversos domínios de conhecimento e categorias para extrair informação, será essencial recorrer a um esquema de classificação de conteúdos [Yang e Lee 2005]. Os autores, Feldman e Sanger [2007b], afirmam que um documento pode apresentar-se com as seguintes componentes: ter carácter (letras, números, caracteres especiais, espaços), palavras (conjunto de carateres com um dado significado), termos (palavras ou conjunto de palavras que podem caraterizar o assunto do documento) e conceitos (características geradas por um documento pelo seu significado). Sendo assim, qualquer documento possui um determinado contexto, podendo este fazer parte de uma coleção de documentos.

Subsistem diversas definições da literatura de TM, mas nem todas apresentam o mesmo significado, sendo que existe alguma divergência entre autores sobre o que deve ou não ser considerado. "*Text mining*, também conhecido como *text data mining* ou *knowledge discovery from textual databases*, refere-se ao processo de extrair padrões interessantes e não triviais ou conhecimento dos documentos de texto" [Tan 1999, p. 65]. Mais recente ainda, Feldman e Sanger [2007a, p. 1] descrevem que "*Text Mining* pode ser definido como um processo de conhecimento intensivo no qual um utilizador interage com uma coleção de documentos ao longo do tempo usando um conjunto de ferramentas de análise". Este processo tem semelhanças superficiais com

DM devido às técnicas que ambos utilizam, como o pré-processamento, a autoaprendizagem, os algoritmos de descoberta de padrões, a gestão de conhecimento e os elementos da camada de representação (ferramentas que permitem ver e explorar os resultados obtidos). Além das diferenças já apontadas anteriormente, o foco do DM incide sobre o pré-processamento ao nível da limpeza e padronização de dados, onde os dados já estão armazenados num formato estruturado, enquanto que no caso do TM o foco está no pré-processamento ao nível da identificação e extração de características representativas em documentos de linguagem natural [Feldman e Sanger 2007b].

Idealmente, um sistema de TM deverá apresentar características dinâmicas. A interface deve receber um conjunto de documentos por via de um ciclo iterativo de entrada e saída com o utilizador final, onde este seleciona as opções de critérios, os dados são processados e apresentados os vários tipos de padrões, mapas ou tendências [Feldman e Sanger 2007b].

A Figura 2.11 apresenta a arquitetura de um sistema de TM: 1ª - Tarefa de pré-processamento; 2ª - Operações de Mineração; 3ª - Camada de Apresentação, Componentes e Funcionalidade do Browser; e 4ª - Refinamento dos Processos.

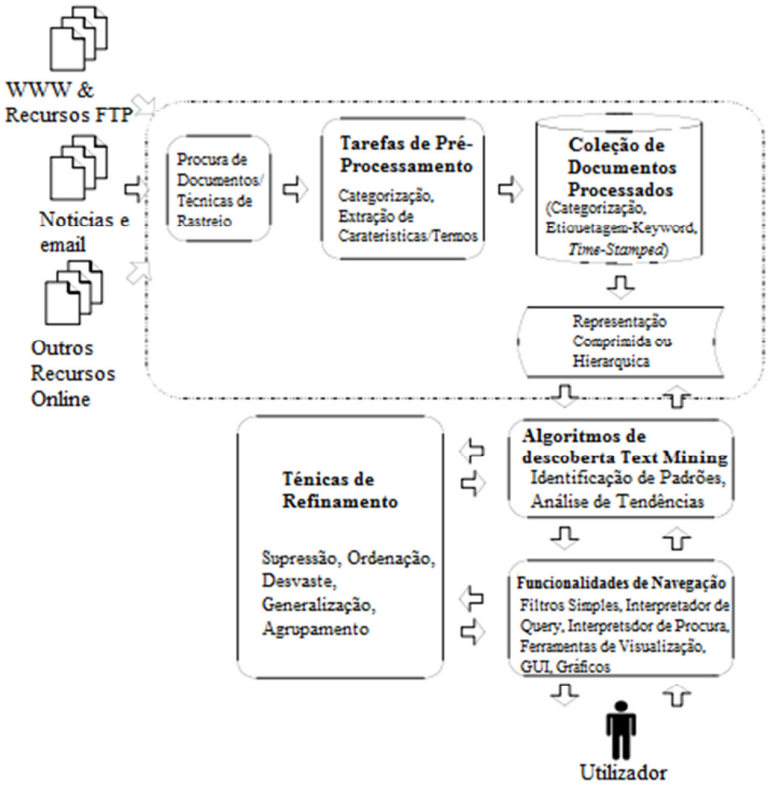


Figura 2.11: Arquitetura de um Sistema de Text Mining Genérico

Adaptado de Feldman e Sanger [2007b]

A primeira tarefa incluí todos os processos necessários para converter os dados originais de documentos num formato adequado para posterior aplicação dos métodos de extração de conhecimento; o próximo passo é onde se descobrem padrões, analisa-se a tendência via cálculo de frequência de termos e aplicam-se algoritmos de descoberta de conhecimento no texto. A fase seguinte exibe uma interfase gráfica com o utilizador que apresenta os padrões extraídos e uma linguagem que interpreta os pedidos dos utilizadores. Por fim, a quarta etapa, tem o objetivo de aperfeiçoar os resultados alcançados através da aplicação de novos métodos de preparação dos dados, da aplicação de algoritmos e da análise de resultados.

A figura 2.11 permite compreender com detalhe todas as fases que facilitam o processo de TM na extração de informação contida num documento de texto, com o intuito de permitir que esta seja processada por computadores reduzindo o erro do ser humano como um intermediário. Contudo, a utilização de "*text mining*", sendo que esta utiliza informação textual representada por palavras, pode gerar problemas de compreensão (exemplo: sinônimos, polissemia), sem que haja a preocupação com a padronização ou com a estruturação de dados. Por essa razão é importante realizar todas as fases deste processo (Pré-processamento, Processamento e Pós-processamento) com a máxima exigência.

O Pré-processamento remove os dados desnecessários para entender o texto e extrair o conhecimento [Monteiro et al. 2006]. O processo de preparação de documentos inclui: correção ortográfica, remoção de *stopwords*, lematização, definição de *n-grams*, cálculo de pesos e seleção de palavras-chave.

A correção ortográfica visa eliminar possíveis erros ortográficos em palavras candidatas a serem termos a indexar (eliminar hífens, pontuação, acentos, palavras em maiúsculas e minúsculas), recorrendo a um corretor ortográfico em conjunto com um dicionário de línguas onde se compara a palavra do texto com as do dicionário. A remoção de *stopwords* consiste na remoção de palavras que não demonstram a mínima relevância, não possuem representatividade alguma, que se repetem muitas vezes num texto (exemplo, vogais) e não acrescentam nem retiram informação relevante ao documento. A lematização ou *stemming* composto pela remoção de variações de palavras (plural, gerúndio, prefixos, sufixos, género e número) de modo que a palavra fique só com a *stem* (raiz) [Monteiro et al. 2006]. Desta forma é possível reduzir o número de palavras sem perder detalhe e precisão [Elberrichi e Aljohar 2007].

O Processamento, segunda etapa, incluí técnicas como o cálculo de peso, seleção de palavras-chave, extração de termos semelhantes e representação textual. O cálculo de pesos é

utilizado para descrever documentos no modelo de espaço vetorial, em que, quanto mais uma palavra aparece num documento (isto é, quanto maior a sua frequência), maior é o seu contributo para a importância desse documento. Para o caso da seleção de palavras-chave que são as palavras importantes do texto que resumem e descrevem o conteúdo de um documento, são ignorados símbolos e caracteres de controlo de arquivo de formatação. Para uma correta determinação das *keywords* (palavras-chave) é imprescindível que sejam removidas as *stopwords*. Um dos recursos utilizados para descobrir a importância dessas palavras é calcular a frequência com que elas aparecem no texto. A extração de termos semelhantes tipicamente são palavras similares que são utilizadas num mesmo contexto. Para a sua extração existem diversos métodos. A representação textual é composta pela sumarização, agrupamento e categorização.

O objetivo da sumarização é produzir versões mais curtas e sumárias de um documento original, permitindo reduzir o tamanho e nível de detalhe de um texto, mantendo a sua informação chave [Fan e Wallace 2005].

O agrupamento (*document clustering*) é uma técnica de processamento textual que permite agrupar um conjunto de documentos desorganizados em grupos similares, isto é, são analisados os documentos e criam-se grupos similares de acordo com o conteúdo retratado nos mesmos [Passarin 2005]. O fator mais importante num algoritmo de *clustering* é a medida de similaridade. Neste processo, utilizam-se algoritmos como *K-Mean*, agrupamento glomerativo hierárquico, entre outros. A categorização consiste na atribuição de documentos de linguagem natural em categorias pré-definidas de acordo com os seus conteúdos [Sebastiani 2002]. É uma forma eficiente de categorizar documentos com a máxima probabilidade de correção, sem a atribuição de demasiadas categorias erradas e com um custo computacional aceitável. Muitos dos métodos utilizados para a categorização de documentos são da área da aprendizagem máquina [Sebastiani 2002]. Para categorizar os documentos utilizam-se classificadores probabilísticos, regressão lógica Bayesiana, árvores de decisão e regras de associação, redes neuronais, *support vector machines*, entre outros [Feldman e Sanger 2007b].

A última etapa do processamento de texto, o Pós-processamento, é a avaliação e validação dos resultados obtidos na fase de análise por meio de medidas de qualidade, com o intuito de melhorar a compreensão de conhecimento descoberto pelo algoritmo utilizado na etapa do processamento. Terminada esta etapa, o utilizador têm ao seu dispor um conjunto de informação que lhe permite melhorar e otimizar o foco para a qual esta será utilizada.

Para concluir, o TM é uma tecnologia que tem evoluído a ritmo exponencial contribuindo para isso o desenvolvimento dos SI organizacionais assim como as necessidades a que as organizações estão expostas e exigências de mercado. São notórios alguns problemas técnicos uma vez que as tarefas de mineração de texto são extremamente sensíveis, assim como ainda não é possível automatizar por completo o processo de mineração de texto, facilitando aos utilizadores que não têm conhecimento de um domínio específico autonomamente interpretar e aplicar as etapas necessárias para a aplicabilidade desta técnica.

Capítulo 3 – Abordagem Metodológica

3.1 Introdução

No capítulo precedente à abordagem metodológica reviu-se a literatura pertinente ao projeto de dissertação no que se refere à segurança dos sistemas de informação, particularmente aos esquemas conceptuais temáticos existentes para a área, terminando com a revisão na área de *Text Mining*.

O atual capítulo apresenta a abordagem metodológica aplicada ao trabalho de dissertação considerando-a como a mais pertinente e adequada para alcançar os objetivos apresentados. Através dos contributos do capítulo anterior, objetiva-se descrever e justificar o problema e a questão de investigação para os quais se procura obter resposta. Neste mesmo capítulo, ainda são expostos os objetivos a alcançar e métodos utilizados para que desta forma sejam alcançados os objetivos pretendidos do trabalho.

3.2 Problema de Investigação

Constantemente o ser humano é bombardeado com situações inesperadas e face às quais tem que, de forma rápida e eficaz, descobrir meios que o auxiliem a enfrentar tais situações. Tudo o que rodeia o ser humano tem informação, tendo o ser humano de ter a capacidade de saber interpretar de forma rápida e adequada tudo o que o envolve incessantemente (pessoas, objetos do mundo). Só assim será capaz de agir e aperfeiçoar o seu conhecimento e comportamento. Na prática, o que foi mencionado anteriormente tem bastante importância, mas por si só não é suficiente. A informação ganha proporções imensas e mediante o foco em que se insere pode influenciar diversas atividades e ações. Para dar resposta a este tipo de situações, pode-se contar com a ajuda de alguns recursos, como por exemplo, os esquemas de classificação de informação. Os esquemas são ferramentas que auxiliam o ser humano na organização de ideias e conhecimento. Ao se esquematizar a informação (vista em texto, meios de comunicação, etc.) será mais clara a sua compreensão e entendimento universal. Segundo Wadsworth e Rovai [1992], esquemas são estruturas mentais ou cognitivas, pelas quais os indivíduos intelectualmente se adaptam e organizam o meio, isto é, os esquemas são o conjunto de processos dentro do sistema

nervoso. Assim como uma criança quando nasce não apresenta qualquer esquema maciço, sendo este desenvolvido ao longo do seu crescimento, no caso dos esquemas de conhecimento, com o decorrer do tempo e a crescente utilização, vão-se tornando cada vez mais em estruturas complexas e independentes. Os esquemas (sistemas para organização do conhecimento) são ferramentas eficazes na sociedade e existem já desde os tempos remotos, estando presentes em todas as áreas de conhecimento humano, desde o modo simples ao mais complexo. Estes facilitam diversas situações do dia-a-dia sem que para isso o ser humano tenha que fazer grandes esforços mentais (organizam novas perceções no esquema).

A revisão de literatura realizada no capítulo anterior reflete que investigadores de diversas áreas de conhecimento dedicam bastante atenção e pesquisam as melhores práticas para representar a informação e organização/representação do conhecimento, reconhecendo a classificação como meio útil de ordenação de conhecimento. A literatura retrata que a classificação está presente na base de sistemas e atividades que se ocupam da organização de conhecimento nas mais diversas formas e não apenas nos sistemas que visam a recuperação de informação. Compreende-se, deste modo, a elevada importância de representar a informação por meio de sistemas/esquemas de classificação. Dessa mesma revisão de literatura, e reconhecendo que todas as ações do ser humano no mundo são envolvidas por atos classificatórios, estando por toda a parte, encontrando-se na vida social de forma onipresente, destacam-se diversos tipos de sistemas de classificação que podem ser mais ou menos complexos.

Olhando ao que já está redigido e do que se pode concluir, ainda existe um vazio quanto à organização de conhecimento na área deste projeto, Segurança de Sistemas de Informação. De toda a revisão de literatura, compreende-se os esforços que têm sido realizados para que o resultado final seja a classificação da área. Pela revisão de literatura, comprova-se que alguns autores e organizações estruturaram a área segundo a perspetiva que mais se adequa ao fim pretendido, não considerando todas as temáticas que a envolvem. É este o ponto de partida para o presente trabalho. Os sistemas de classificação tendem a descrever de forma ordenada, todas as esferas de uma área de conhecimento, de acordo com os critérios ou características pretendidas para o fim desejado.

Pela revisão de literatura, pode verificar-se que já existem sistemas de classificação reconhecidos para a literatura dos Sistemas de Informação e para as Ciências da Computação, mas ainda existe um vazio em relação à organização da SSI. Para melhor caraterizar o problema de investigação, importa referir que à segurança dos SI está a ser reconhecida a sua crescente

importância, garantindo que a concretização de um protótipo para os sistemas de classificação na área de SSI seja bastante revelador.

Concluída a elucidação do problema de investigação que motiva este trabalho, a secção seguinte expõe e aclara a questão de investigação que se associou ao problema enunciado.

3.3 Questão de Investigação

Apresentado o problema de investigação, são levantadas várias questões de investigação que norteiam a realização desta dissertação. Segundo a revisão de literatura efetuada foi fácil perceber o quão importante é classificar informação e a importância que organizar o conhecimento por meio de esquemas de classificação nos dias atuais.

Sabe-se que a conceção de um esquema de classificação é bastante prolongada no tempo e implica bastantes recursos, mas os benefícios com a conceção de um esquema de classificação transpõem as desvantagens que daí resultam. O entendimento universal de uma área de conhecimento e o tempo despendido para a criação desse esquema, por si só já são o fator revelador deste trabalho.

O esquema que é proposto neste trabalho tem como elo de criação o conceito de "semiautomático", isto porque, dada a quantidade de informação que é produzida diariamente, e à evolução da tecnologia, torna-se inconcebível não fazer uso dos utilitários que facilitam a vida dos seres humanos e recorrer à tecnologia já existente.

Toda a pesquisa bibliográfica foi de elevada relevância e com importância para a compreensão de todo o problema ao qual se procura dar resposta, apresentando como questão de investigação: "Como induzir semi-automaticamente o esquema conceptual temático da literatura em SSI?"

Acredita-se que esta questão de investigação corporiza de forma correta e satisfatória os propósitos deste estudo. Para alcançar a resposta à questão de investigação apresentada procura-se construir uma visão clara da realidade que envolve toda a área de SSI, recorrendo ao esquema de classificação.

A próxima secção identifica e caracteriza cada um dos objetivos definidos para este projeto.

3.4 Objetivos do Estudo

Pela literatura revista foi reconhecida a importância que os esquemas de classificação conceptuais temáticos facultam como meio para a organização e categorização da informação e do conhecimento. Por outro lado, também foi possível observar as dificuldades terminológicas e de categorização na representação de conceitos quando se pretende estruturar e representar os domínios de conhecimento. Segundo Hjerland [2003] existe a necessidade de criar bases metodológicas apropriadas para a elaboração deste tipo de sistemas para estruturar o conhecimento. Portanto, este trabalho propõe apresentar os aspetos apontados pela literatura da área, que fundamentam a criação de um esquema de classificação para a área de SSI pela ampla utilização na organização da informação.

Para realizar com sucesso o propósito deste trabalho, definiu-se um conjunto de objetivos, apresentados e discutidos nesta secção, que auxiliam a compreensão das várias etapas de todo o trabalho. Para o desenvolvimento desta investigação e considerando o seu âmbito, foi considerado o uso de métodos de investigação, detalhados na próxima secção, que permitem compreender o conjunto de passos que devem ser executados para obtenção do produto final.

O primeiro objetivo definido para este trabalho realiza-se com a familiarização da temática em análise e a escolha dos métodos de investigação a empregar. Para efetuar toda esta tarefa foi feita uma revisão aos fundamentos e literatura dos SI e SSI, na procura da fundamentação teórica em que os autores apoiam-se para a construção de sistemas de classificação.

Como segundo objetivo definiu-se a identificação de aplicações, investigando a sua aplicabilidade por meio da experimentação. A ferramenta teria que permitir o desenvolvimento *bottom-up* de um sistema de classificação para a SSI e respeitar um conjunto de requisitos que proporcionem a criação do esquema de classificação.

A construção de um repositório de dados composto por distintas publicações em *journals* da temática de SSI realizadas por diversos autores no intervalo de 2000-2011, será aqui apresentado como sendo o terceiro objetivo;

O quarto objetivo será o mais revelador: é a criação do esquema classificativo para a SSI, com base nos resultados obtidos pelo objetivo anterior. Justificar e validar a natureza do esquema de classificação para a SSI, também será abordado neste objetivo;

Por último, apresenta-se o quinto objetivo. Com o desfecho do quarto objetivo e com base no seu resultado, propor-se-á um conjunto de interpretações, ilações, vantagens e contributos, que,

primeiramente, auxiliam compreender o esquema de classificação criado, e, posteriormente, sejam a base a partir da qual se atualize e produzam esquemas de classificação cada vez mais aperfeiçoados e abrangentes para a SSI.

Pelo desenvolvimento deste trabalho e através da revisão de literatura, observam-se progressos científicos na área de estudo da SSI. Pretende-se identificar a base para elaborar um sistema de classificação da SSI e o método para estruturar esse modelo conceptual. O contributo deste trabalho irá preencher um espaço de conceção de conhecimento em SSI ainda insuficiente.

A próxima secção retratará o posicionamento científico deste trabalho, apresentando e explicando os métodos de investigação que regem os procedimentos da investigação científica.

3.5 Método de Investigação

As secções anteriores descreveram o problema, a questão e os objetivos que norteiam este estudo. A presente secção aduz à compreensão do leitor em relação ao método de investigação do presente estudo, e com o qual se procura responder aos desafios que este tenta solucionar. Sendo este um trabalho de investigação, e como qualquer trabalho nesse âmbito, pretende encontrar soluções. O seu ponto de partida inicia-se com o enquadramento do problema e identificação da questão de investigação, para a qual a resposta ainda é desconhecida. Neste sentido, a procura para a explicação de um facto (fenómeno de reconhecimento indiscutível) é tal e qual a apresentação do enunciado de um problema para o qual deve ser encontrada uma solução.

A metodologia científica consiste no estudo dos métodos de "conhecer", "procura" e "conhecimento". Recorrer a um método de investigação é fundamental, uma vez que este é o meio para atingir um fim. Pelo conjunto de procedimentos técnicos, visa a homogeneização de padrões para a execução e apresentação do trabalho de investigação.

Segundo Barros e Lehfeld [1986], a metodologia não procura soluções, mas escolhe a forma de encontrá-las, integrando os conhecimentos por meio dos métodos em vigor nas diferentes disciplinas científicas ou filosóficas. Apesar das diversas formas de classificar, as abordagens de investigação são normalmente distinguidas entre abordagens quantitativas e qualitativas [Myers 1997a]. A investigação quantitativa aplica-se a estudos que procuram compreender fenómenos exprimíveis em termos de quantidade e envolvem a geração quantitativa de dados que são sujeitos a análises rígidas e formais [Kothari 2004]. A investigação qualitativa, por seu lado, envolve o "estudo, uso e recolha de uma variedade de materiais empíricos" [Denzin e Lincoln 2005, p. 3] e

visa auxiliar os "investigadores a compreender as pessoas e os contextos culturais dentro dos quais elas vivem" [Myers 1997b, p. 3]. É do conhecimento humano que a escolha do método deve ser feita em função da natureza do problema a estudar. Neste sentido, considerou-se pertinente seguir um método de investigação quantitativo (investigação científica tradicional), baseado no pensamento racional positivista, isto é, pelas observações empíricas constroem-se teorias (expressas na forma dedutiva) que tentam explicar o que é observado.

O método escolhido permitirá trilhar o percurso através do qual se pretende atingir o resultado do trabalho de investigação e através deste fazer "ciência". A interpretação mais clara deste conceito define método como o caminho, o meio, o percurso, o projeto de raciocínio previamente determinado para abordar e compreender os fenómenos/problemas. Considera-se a definição de método como sendo um procedimento passível de ser repetido de forma a atingir um objetivo tangível ou intangível a que este se propõe. Mediante o método utilizado, o seu rigor vai permitir distinguir o conhecimento científico de outro tipo de conhecimento. Conhecer o método de investigação utilizado neste trabalho permitirá ao leitor uma compreensão ampla da importância do desenrolar deste trabalho, ou seja, neste contexto o método é responsável pela transparência e pela objetividade da investigação.

Os métodos de investigação, tema de foco desta secção, podem ser classificados segundo as suas características. Frente a tal cenário, é possível afirmar que existem diferentes tipos de métodos que correspondem a cada ramo da ciência e a cada tema a ser investigado. No contexto deste projeto apenas será abordado o método científico (experimental), primeiro método selecionado para este projeto, sendo apresentado ao longo desta secção um outro método complementar. O método investigação preditiva é o método mais adequado a este projeto, retratando apenas as suas características, sem que para isso seja necessária a explicação dos restantes tipos de métodos existentes. Este método científico permite descobrir novo conhecimento através de uma aprendizagem cuidada, estando dividido por quatro etapas: a observação (natural ou científica), a formulação de hipóteses, a experimentação (provar a hipótese), e por fim, a elaboração de resultados e respetiva interpretação.

Procedendo à caracterização de cada uma das fases do método identificado, a primeira etapa é a observação/revisão sistemática/científica onde inicialmente definiu-se e fixou-se as condições em que o trabalho se vai realizar, assim como o principal recurso: o tempo. Este tipo de revisão baseia-se num estudo secundário, isto é, um estudo que revê todos os estudos primários (estudo empírico que investiga uma questão de pesquisa específica) relacionados com uma questão

específica de pesquisa e que atende aos critérios de legibilidade, com objetivo de integrar e sintetizar as evidências relacionadas com a questão de investigação específica [Antman et al. 1992]. Pelo uso de métodos explícitos e sistemáticos, e com uma seleção baseada em critérios aplicados de forma uniforme, pretende-se alcançar resultados confiáveis através de uma avaliação criteriosa, rigorosa, reprodutível e aditável, a partir dos quais se possam tirar conclusões e tomar decisões. Será uma síntese quantitativa com inferências baseadas nos resultados da pesquisa. Para elucidar esta primeira etapa do método de investigação sobre a revisão sistemática, conclui-se a necessidade de identificar, avaliar e interpretar todas as pesquisas disponíveis face ao tema do projeto, sumarizar as evidências existentes sobre a tecnologia, identificar as lacunas da atual pesquisa e fornecer a estrutura para posicionar esta nova pesquisa.

A segunda etapa, formulação de hipóteses, surge como sendo a resposta possível à questão colocada. Depois de observar o fenómeno e de reunir a documentação suficiente sobre observações já efetuadas por outros autores, o investigador deve procurar uma argumentação que permita explicar e justificar cada uma das características de tal fenómeno. A validade da hipótese só se revela quando esta é posta à prova por meio de experimentação, devendo ser uma hipótese verificável e verificada. As hipóteses podem ser induzidas (origem na observação dos fenómenos) ou deduzidas (a partir de teorias existentes).

Como terceira etapa do método identificado, surge a experimentação. A sua utilização permite verificar a relação que existe entre dois fenómenos ou ordens de factos. O objetivo primordial desta etapa é provar a veracidade da hipótese levantada. Uma vez formulada a hipótese, o investigador deve comprovar que esta é válida em todos os casos e, para tal, deve realizar experiências (terceira etapa do método científico) nas quais se reproduzam o mais fielmente possível as condições nas quais se verifique o fenómeno estudado. Se sob tais condições o fenómeno acontece, a hipótese terá validade, ou seja, será uma proposição verdadeira nas condições estipuladas. Caso contrário, a hipótese não terá qualquer importância.

Para finalizar toda a sequência do método, temos a elaboração e conceptualização de resultados, que consiste na interpretação e enquadramento dos dados recolhidos na sua total tradução recorrendo a gráficos, quadros, tratamentos estatísticos, etc.

Em cooperação com o método científico descrito anteriormente, ainda foi reconhecido o paradigma *Design Science* ou *Constructive Research* que pretende contribuir para o desenvolvimento do esquema conceptual temático para a literatura em SSI.

A *Design Science* surge na engenharia e ciências artificiais [Simon 1996], sendo considerada uma investigação prescritiva que procura alargar os limites das capacidades humanas e organizacionais, criando novos e inovadores artefactos, constatando no ambiente em estudo a existência da interação entre as pessoas, a organização e a tecnologia [Hevner et al. 2004]. Pela construção e aplicação do artefacto, é possível obter o conhecimento e a compreensão do domínio do problema, descobrindo a solução que é pretendida e respondendo aos objetivos já definidos. Segundo Hevner et al. [2004], Benbasat e Zmud [1999], Winograd [1996] a pragmática da *design science* desempenha um papel bastante importante na literatura dos SI devido a sua aplicabilidade na conceção. Em paralelo com a opinião dos autores citados, considera-se a *Design Science* uma abordagem bastante apropriada a este projeto, isto porque, o resultado será a construção de representações, modelos, métodos e instâncias que são originadas através da investigação de design, com uma avaliação por meio de critérios de valor ou utilidade [March e Smith 1995].

O avanço das áreas científicas formula a conceção de artefactos cada vez mais complexos, com maior diversidade quanto à sua aplicabilidade e que são destinados à resolução de problemas organizacionais identificados. Para avaliar a qualidade e eficácia dos artefactos, podem ser utilizados métodos computacionais e matemáticos, ou recorrer a técnicas empíricas [Hevner et al. 2004]. Na opinião de March e Smith [1995] a *Design Science* cria objetos que auxiliam os propósitos humanos, sendo divididos por quatro tipos de produtos inovadores e valiosos: as construções, os modelos, os métodos e as instanciações. Aqueles autores ainda defendem que este paradigma com elevada relevância para as TI, *Design Science*, restringe-se apenas a duas atividades: construir (construir o artefacto com uma finalidade específica) e avaliar (determina o uso e competência pretendida para o artefacto), que estão condicionadas pelo desempenho do artefacto face ao ambiente em que atua. Pela avaliação recorrendo à teoria, deverá compreender-se a razão que resultou no funcionamento do artefacto ou, caso contrário, o que levou à sua falha, concluindo todo o processo com a justificação da teoria. O desempenho do artefacto está relacionado com o ambiente em que atua, onde o seu desentendimento pode originar artefactos com *design* inapropriado. Para reduzir resultados indesejados pela construção de um artefacto, é essencial antecipar os efeitos do seu possível uso.

A essência do *Design Science* é a criação de uma clara utilidade de artefactos de TI que permitem às organizações enfrentar tarefas relacionadas com informações importantes. Cientes de como os autores já citados traçam a *Design Science*, Hevner et al. [2004, p. 82] resumem esta pragmática de forma clara, focando sete requisitos:

Requisito 1: A criação de um artefacto com propósito inovador sob a forma de construção, modelo, método ou instanciação;

Requisito 2: O seu objetivo consiste no desenvolvimento de soluções de base tecnológica para problemas de negócio importantes e relevantes;

Requisito 3: A avaliação do artefacto é crucial. A sua utilidade, qualidade e eficácia deve ser rigorosamente demonstrada;

Requisito 4: O artefacto deve ser inovador, fornecendo contribuições claras e verificáveis para o problema enunciado e até então sem solução, ou até mesmo resolver um problema conhecido de forma mais eficaz e eficiente;

Requisito 5: O artefacto deve ser rigorosamente definido, formalmente representado, coerente e internamente consistente;

Requisito 6: A procura pelo artefacto eficaz requer a utilização de meios disponíveis para atingir os fins pretendidos no ambiente do problema;

Requisito 7: Os resultados devem ser comunicados de forma eficaz, para que diferentes leitores compreendam a razão do artefacto.

Uma avaliação rigorosa é a base para aceitar as novas contribuições do novo artefacto. Mediante o objetivo para o qual um artefacto é concebido, caso os artefactos já existentes sejam adequados, a criação de um novo artefacto será desnecessária. Outro prisma será o caso do novo artefacto não compreender com rigor o mundo real ou não resolver o problema, e por isso a sua utilidade será nula [Hevner et al. 2004].

Pela junção dos dois métodos apesentados anteriormente, julga-se assim, concluir com exatidão o objetivo principal deste trabalho. Olhando ao âmbito em que todo o projeto se realiza, é importante não esquecer que o TM será um bom processo para a extração de informação de documentos de texto não estruturado, sem que para isso seja necessário lê-los previamente. O projeto em questão passa pelas principais atividades: obter documentos, pré-processar os documentos, extrair conhecimento ou classificá-lo, e avaliar o conhecimento, sendo que estas coincidem com as atividades da mineração de texto, contribuindo para a gestão do conhecimento.

O desenvolvimento do esquema de classificação consiste na descrição da literatura de SSI. Ao contrário do que a revisão de literatura apresentou, o esquema que será proposto irá abranger o maior número de temáticas, que os já referidos na literatura, que apenas se debruçaram por áreas especificas. O contributo para a identificação de uma linguagem comum para referir os aspetos da

SSI e a observação do relacionamento entre as diversas temáticas traduz a eficácia pretendida do esquema proposto inicialmente. O estudo ainda poderá ser complementado recorrendo à qualidade de dados e de informação que pode estar na origem do esquema e daí surgirem novas avaliações e propostas futuras para a SSI.

Para melhor compreensão de todo o desenrolar do trabalho, o próximo capítulo, Capítulo 4, irá descrever as atividades e tarefas realizadas que proporcionam a criação do esquema temático.

Capítulo 4 – Descrição do Estudo

4.1 Introdução

O presente capítulo descreve o trabalho realizado para a criação do esquema conceptual temático na SSI, com foco nos diversos pontos expostos no Capítulo 3 deste documento, abordagem metodológica.

Pretende-se que o leitor, após a leitura do presente capítulo, tenha uma visão global e pormenorizada de todas as fases e tarefas subjacentes deste estudo. Para simplificar a compreensão deste capítulo, julga-se que com a apresentação da Figura 4.1, que realça as principais fases deste projeto, o nível de entendimento de cada fase e respetivas tarefas será claro. A cada uma destas fases é associada uma numeração para facilitar a sua identificação e descrição detalhada ao longo das seguintes secções.

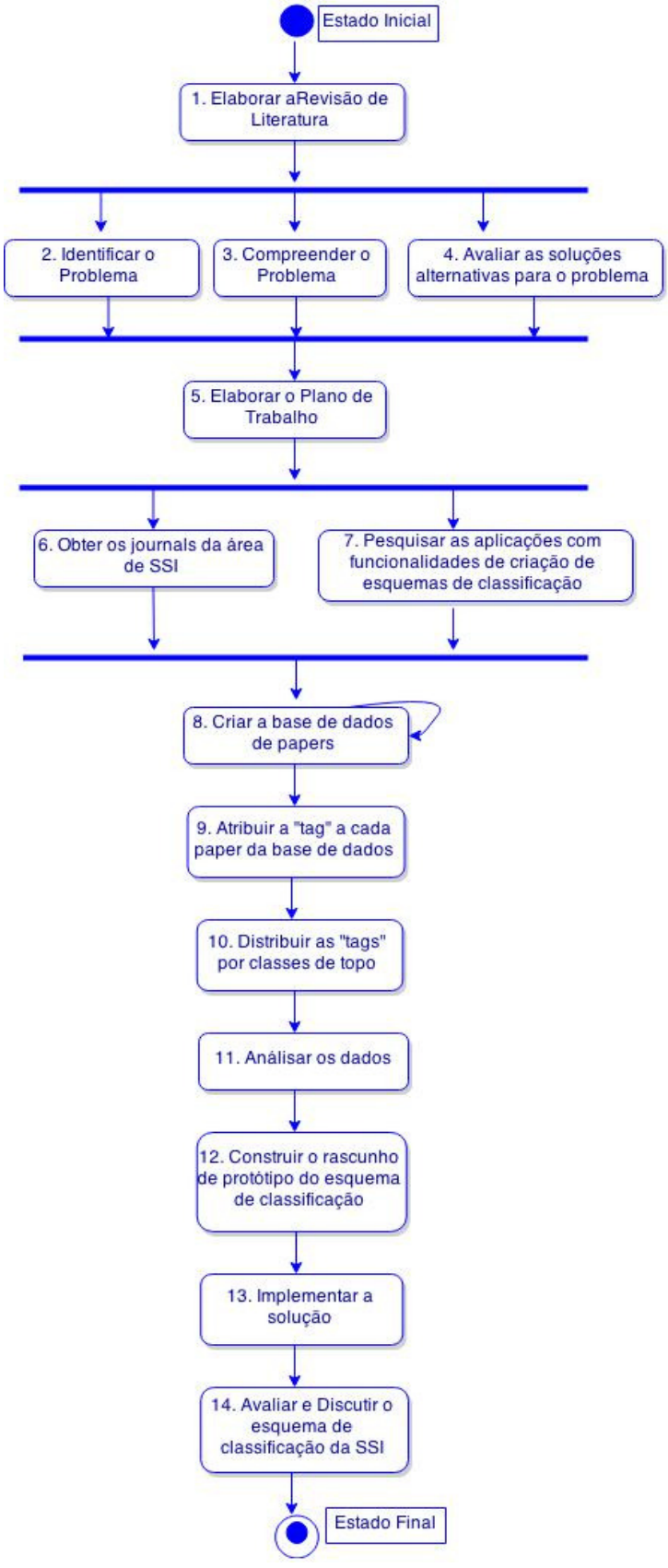


Figura 4.1: Principais Fases do Estudo

4.2 Posicionamento do Estudo

O trabalho de dissertação iniciou-se com o ponto número um – "Fazer Revisão de Literatura" (Figura 4.1). Esta fase foi fundamental para este projeto, uma vez que simplificou o processo cumulativo de aquisição de conhecimento, situando-o dentro da área de pesquisa na qual se insere e assim contextualizando-o. Ao contextualizar este trabalho, o investigador reconheceu quais os investigadores que já abordaram esta temática e quais os trabalhos desenvolvidos neste âmbito. Para o leitor deste documento, proporciona a compreensão da linha teórica em que o projeto está inserido, tendo como base os autores identificados na revisão de literatura. Este foi um momento de compreensão e interpretação do conhecimento existente, já que o ponto de partida desta investigação inicia-se através de estudos prévios neste âmbito. A partir dos estudos já efetuados por outras pesquisas, foi alvo do presente estudo acrescentar um elo adicional de conhecimento sobre o que já subsiste neste âmbito, iniciando este ciclo com uma leitura e compreensão do conjunto de conhecimento já existente sobre o âmbito de pesquisa.

Com este exercício conseguiu-se focar aspetos bastante relevantes para este trabalho. O primeiro passo consistiu no entendimento de um conjunto de definições que estão ligadas com a área em estudo, possibilitando um entendimento claro e único desses conceitos. O segundo passo abrangeu o reconhecimento verdadeiro da importância que os esquemas de classificação desempenham para uma área de conhecimento: facilita a descrição detalhada da área, a existência de vocabulário comum e a rápida procura de conceitos. Esta apreciação faz-se com o primeiro esquema de classificação na área dos SI [Barki et al. 1988] e com o ACM *Computing Classification System* (CCS), esquema de classificação para as ciências da computação. Com estes dois estudos compreendeu-se as boas práticas para a construção de esquemas e as limitações que os autores apontaram nesses trabalhos. De seguida, terceiro aspeto, considerou-se que sendo a SSI o foco deste trabalho, seria fundamental compreender todo o corpo de conhecimento que a circundava. A análise deste ponto passou por explorar as normas e padrões já existentes e reconhecidos mundialmente, que estruturaram o conhecimento da SSI. Um outro aspeto explorado que preencheu esta fase de revisão de literatura salientou a análise de trabalhos científicos publicados na área e cujo principal objetivo focava a análise e mapeamento do domínio de investigação da SSI. Estas publicações faziam comparações ao *design* das metodologias de SSI apontadas pelos autores e analisavam a tendência futura dos SI e pesquisas em segurança. O último aspeto a ser observado comportou o estudo sobre *Text Mining*, que consiste na extração de informação de dados não

estruturados ou semiestruturados (foco chave deste trabalho). Neste estudo compreendeu-se o conceito, as técnicas, os algoritmos, as metodologias e as estruturas de dados utilizadas pelo *Text Mining*.

Todos os documentos analisados e que foram alvo de referência de informação ao longo do projeto de dissertação, foram sendo organizados recorrendo ao *software* EndNote, facilitando deste modo a escrita deste documento. Este *software* permite pesquisar, armazenar e organizar as referências bibliográficas de publicações de artigos científicos.

A revisão de literatura efetuada neste projeto permitiu reconhecer e dar crédito às investigações de outros autores, compreender e familiarizar com a área de conhecimento em estudo, e impor num campo de conhecimento já moderadamente estabelecido a possibilidade de receber novas pesquisas que possam contribuir para a organização da temática de SSI.

4.3 Apresentação do Problema de Investigação

A presente secção retrata o ponto dois, três e quatro ilustrado na Figura 4.1. A formulação do problema de investigação iniciou-se com uma leitura e conhecimento prévio de pesquisas chave sobre o assunto. Se o número de pesquisas que é lido sobre o assunto é elevado, mais condições o investigador tem para formular um problema de pesquisa interessante, pertinente e original. Ao saber o que já está feito neste âmbito, as condições para determinar o próximo passo a ser dado para aumentar o conhecimento, torna-se mais elevado.

Pela revisão de literatura, e com o capítulo que retrata o método de investigação, é deixado bem claro como é que o problema identificado se insere na área de conhecimento da SSI e em que ponto se encontra o estado de conhecimento sobre o assunto. A compreensão do problema e avaliação de soluções alternativas para o problema apontado são fases também elas de elevada importância. Com estas fases foi já explicado o contexto teórico no qual o problema está inserido, ou seja, quais os modelos teóricos que foram desenvolvidos para explicar o fenômeno. Também foi explicado o estado de conhecimento empírico e experimental que foi acumulado até então sobre este problema, ou seja, o quanto e o que se sabe sobre o assunto. Desta forma, estão incluídos os conhecimentos já estabelecidos sobre o problema, os aspetos que ainda não foram investigados, as contradições que existem sobre os resultados obtidos e as dúvidas sobre a sua explicação.

Com a primeira tarefa deste projeto, revisão de literatura, foi possível justificar e precisar o problema, assim como as hipóteses a serem formuladas. Com a realização das tarefas seguintes,

identifica-se e compreende-se o problema de investigação, terminando com a avaliação de soluções alternativas que podiam ser viáveis para o projeto. Para responder a este desafio realizaram-se pesquisas *web*, procurando esquemas conceptuais temáticos que podiam ajustar-se à literatura de SSI ou que estão incompletos.

4.4 Plano de Trabalho

A elaboração do plano de trabalho, fase número cinco da Figura 4.1, permitiu o planeamento e organização de ações que eram necessárias desenvolver, viabilizando a formalização e acompanhamento de todo o processo de implementação do projeto de investigação. Nesta fase foi também possível definir uma adequada questão de investigação, uma estratégia de pesquisa pertinente para investigar o problema apresentado, apresentar o método do estudo mais apropriado e identificar os objetivos cruciais para o presente trabalho.

A documentação elaborada para o plano de trabalho refletiu integralmente o que foi desenvolvido e o que está direcionado para as próximas fases. A elaboração do plano de trabalho serviu como um instrumento de controlo geral, comtemplando as tarefas necessárias para se obter o resultado final com referências claras aos prazos estimados para a sua execução e os respetivos recursos necessários.

Através de uma breve observação ao plano de trabalho é possível verificar que o presente trabalho está dividido em duas partes: inicialmente uma parte teórica de compreensão dos conceitos e conhecimentos que envolvem as áreas de estudo sobre as quais o estudo se debruça e uma segunda parte que consiste na criação de um esquema concetual temático que utiliza os conhecimentos assimilados na primeira parte do plano de trabalho descrito.

4.5 Recolha de *Journals*

Esta secção espelha todo o trabalho realizado ao nível da pesquisa de artigos científicos relacionados com a SSI, fase número seis da Figura 4.1. O objetivo desta fase baseou-se na aquisição de artigos científicos para mais tarde popular a base de dados criada, fase número oito da Figura 4.1. No mercado editorial existem diversas formas de apresentar as publicações. Para o presente trabalho bastará mencionar os tipos de publicações que foram alvo de análise: os periódicos científicos e os artigos científicos, que divulgam os conhecimentos adquiridos através de métodos científicos.

Os periódicos científicos permitem uma aproximação percetível sobre o progresso da ciência e sobre o reconhecimento dos seus benefícios, geralmente transmitindo novas pesquisas. Através dos periódicos publicados na SSI é constante o conhecimento do impulsionar de alterações na forma como a SSI é aplicada, executada em diversas situações, reconhecida a sua utilização pelas organizações, entre outros aspetos. Este tipo de publicação é organizada por catálogos, onde existem vários outros periódicos alocados por várias bibliotecas. A indexação de periódicos científicos facilita a disponibilidade de agrupar e a fonte de procura dos artigos em periódicos é facilitada.

No caso dos artigos científicos, a sua finalidade passa por apresentar resultados concisos de uma pesquisa, seja ela experimental, quase experimental ou documental, realizada de acordo com o método científico aceite por um grupo de investigadores. Normalmente é abordado o problema ou objetivo de investigação, o conjunto de hipóteses, as possíveis soluções do problema ou modos de se alcançar o objetivo, uma descrição dos métodos e técnicas utilizados, uma análise dos resultados obtidos e uma conclusão que aponta qual a hipótese que foi verificada experimentalmente.

Geralmente, a estrutura lógica dos periódicos científicos e dos artigos científicos começa com um *abstract*, que é um resumo do *paper* composto, habitualmente por um a quatro parágrafos. De seguida é apresentada a "introdução", que faz uma descrição geral da pesquisa, incluindo uma revisão de pesquisas similares. A próxima componente de um artigo científico é a secção de "material e métodos ou metodologia", que detalha os procedimentos utilizados na pesquisa. Depois tem-se a secção "resultados e discussão" que descreve os resultados obtidos e discute as suas implicações. Por fim, é descrita a "conclusão" onde é contextualizada a investigação, sugeridas possíveis aplicações e a realização de pesquisas complementares. Eventualmente, pode existir uma secção de notícias que descreve desenvolvimentos científicos (envolve questões politicas), ou até mesmo um editorial e ou, cartas ao editor. Uma das diferenças entre um periódico científico e um artigo científico, é que no primeiro caso este proporciona às comunidades científicas um meio formal de comunicação e disseminação da produção técnico-científica, enquanto que no segundo caso é um trabalho de base acadêmica que apresenta resultados concisos de uma pesquisa realizada de acordo com um método científico reconhecido por parte da comunidade de investigadores. Assim, pela divulgação e debate do conhecimento elaborado em pesquisas, um artigo científico é o meio fundamental para a divulgação e desenvolvimento da SSI.

Tabela 4.1: Journals Identificados

Journal	Objetivo do *Journal*
Communications of the Association for Information Systems (CAIS)	Destina-se em promover o livre fluxo de ideias dentro da comunidade de Sistemas de Informação.
Journal of Association for Computing Machinery (JACM)	Publica trabalhos significativos sobre os princípios da ciência da computação. Debruça-se nos limites de subdisciplinas de informática e nas fronteiras entre ciência da computação e outras áreas.
Information Systems Frontiers (ISF)	*Journal* de pesquisa e desenvolvimento nos interfaces em sistemas de informação e tecnologia de informação com valor académico e industrial.
Information Technology and People (IT&P)	Representa a inovação contínua de estratégias sociais e organizacionais na conceção e uso de tecnologia da informação, a partir da rede para o *desktop*.
European Journal of Information Systems (EJIS)	Revista científica interdisciplinar que pretende oferecer uma perspetiva europeia diferenciada na teoria e prática de sistemas de informação. Tecnologia, desenvolvimento, implementação, estratégia, gestão e políticas relacionadas com a teoria e prática de sistemas de informação são os tópicos abordados.
Information Systems Research (ISR)	Revista acadêmica que cobre investigação nas áreas de sistemas de informação e tecnologia da informação.
Journal of Strategic Information Systems (JSIS)	Foca-se em questões de gestão, negócios e organizacionais associadas à introdução e utilização de sistemas de informação como uma ferramenta estratégica, e considera estas questões num contexto global.
Information Systems Journal (ISJ)	Publica artigos sobre a conceção e implementação de linguagens, modelos de dados, modelos de processos, algoritmos, *software* e hardware para sistemas de informação.

Journal of Computer Information Systems (JCIS)	Fórum que apresenta ideias e pesquisas dos membros da Associação Internacional de Informática de Sistemas de Informação e outros sistemas de informação e profissionais de negócios.
Information Resources Management Journal (IRMJ)	Aborda um vasto conjunto de questões relacionadas com o uso, o fracasso, o sucesso, as políticas, estratégias e aplicações da tecnologia da informação nas organizações. Apresenta as mais recentes descobertas e opiniões de especialistas de gestores líderes de tecnologia da informação em diversos setores, focando em aspetos de gestão e organizacionais da gestão de recursos de tecnologia da informação.
Information and Management (I&M)	Recolhe e divulga informação sobre novos desenvolvimentos e avanços na área de SI, incentivando o uso de políticas de gestão e estratégias para o negócio.
Journal of Management Information Systems (JMIS)	Revista acadêmica que publica artigos originais de pesquisa nas áreas de sistemas de informação e tecnologia da informação.
Journal of the Association for Information Systems (JAIS)	Revista científica que cobre investigação nas áreas de sistemas de informação e tecnologia.
Management Information Systems Quarterly (MISQ)	Retrata a valorização e comunicação de conhecimentos sobre o desenvolvimento de serviços baseados em TI, a gestão dos recursos de TI, bem como a utilização, impacto e economia de TI, com implicações de gestão, organizacionais e sociais.
Journal of Database Management (JDM)	Publica pesquisas em todos os aspetos da gestão de base de dados, análise de sistemas e *design* e engenharia de software.
International Journal of Secure Software Engineering (IJSSE)	Publica pesquisas originais sobre as preocupações de segurança que devem ser consideradas no desenvolvimento de *software*. Inclui todos os aspetos de segurança de *software* para o desenvolvimento, implementação e processos de gestão de sistemas de software.

Computers & Security (C&S)	*Journal* representativo na área de segurança de TI. Destina-se a profissionais envolvidos com a segurança, auditoria, controlo e integridade de dados de computador em todos os setores - indústria, comércio e acadêmico.
Journal of Information Technology Theory and Application (JITTA)	Revista de pesquisa que acolhe artigos de um amplo espectro de abordagens de investigação (trabalhos de investigação de agenda, documentos interpretativos ou exploratórios, investigação especulativa, comentários *state-of-research*, ou trabalhos de pesquisa completos), ensaios de investigação, e trabalhos de aplicação (documentos normativos ou estudos de caso).
Information Management & Computer Security (IMCS)	Contribui para o avanço do conhecimento sobre a segurança e garantia de informações e sistemas de informação. Foca na interação de aspetos técnicos e humanos.
Journal of Information Security and Privacy (JISP)	Centra-se em questões de privacidade de informação e segurança, tanto para académicos como profissionais.
Journal of Information System Security (JISSec)	Publica pesquisas académicas em segurança de sistemas de informação.
International Journal of Computer Science and Security (IJCSS)	Revista on-line para a publicação de trabalhos em ciência da computação e tecnologias de segurança de computadores. Os assuntos abordados incluem: controlo de acesso, segurança de computadores, criptografia, comunicação e segurança de dados, base de dados, comércio eletrónico, multimídia, bioinformática, processamento de sinal e processamento de imagem, etc.
Internacional Journal of Information Security (IJIS)	Periódico com pesquisa em segurança da informação que oferece a publicação imediata do trabalho técnico importante, seja teórico, aplicável ou relacionado com a implementação.
International Journal of Information Technology and Management (IJITM)	Revista que aborda a tecnologia da informação, a sua evolução e perspetivas futuras. Aborda aspetos tecnológicos, gestão, políticos, económicos e organizacionais da aplicação de TI.

Após uma breve descrição do tipo de publicações investigadas, prossegue-se para uma descrição detalhada das etapas realizadas para a seleção e análise dessas mesmas publicações. Esta análise iniciou-se com uma pesquisa ostentada em todos os *journals* que publicam nas áreas de SI e SSI. Considerou-se que numa primeira fase era necessário ter uma base do que existia ao nível de *journals* que pudessem contribuir para uma boa análise e concretização do trabalho. A Tabela 4.1 mostra todos os *journals* analisados e respetiva descrição, para que o leitor melhor compreenda a seleção que foi realizada. Com esta pesquisa e mediante o foco do trabalho, apenas foram considerados os *journals* que estavam estreitamente relacionados com o campo de literatura da SSI. Para proceder à sua seleção, compreendeu-se qual o foco de cada *journal* e assim foi possível limitar o número de *journals* que seriam explorados, passando de 24 para 14 *journals* (Gráfico 1).

O próximo passo surgiu com a identificação dos periódicos e artigos científicos que compreendiam o intervalo temporal entre 2000 e Maio de 2011, inclusivamente. Com a definição deste período temporal, procedeu-se à obtenção de todos os *papers* que respeitassem o intervalo temporal definido e que estavam incluídos na lista de 14 *journals*. O Gráfico 4.1 representa para cada *journal* o número total de artigos científicos que foram obtidos entre 2000 e 2011, contabilizando um total de 4643 *papers*. A designação de cada *journal* poderá ser consultada recorrendo à Tabela 4.1.

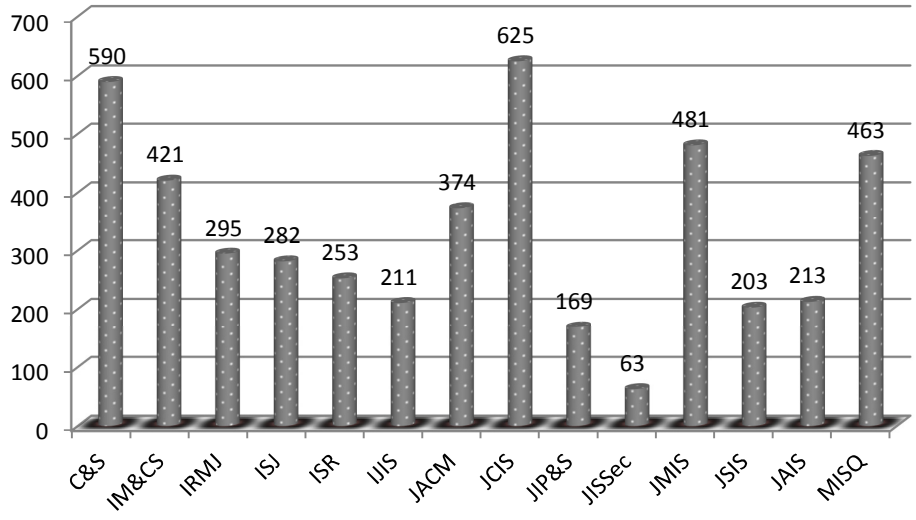


Gráfico 4.1: Número de Papers obtidos por Journal

Concluída a obtenção de documentação, a próxima fase consistiu na seleção refinada de todos os *papers* de cada *journal*, isto é, era o momento em que eram analisados todos os *paper* (um a um) e selecionada a sua inclusão ou exclusão na base de dados, criada a posteriori. Note-se que para esta seleção foram consideradas as seguintes componentes de cada *paper*: o título, as palavras-chave e, casualmente, o *abstract* e/ou conclusão. Após esta triagem, constatou-se que dos 4643 *papers* trabalhar-se-ia sobre 1000 *papers* que abordavam conteúdos em SSI relevantes (Gráfico 4.2).

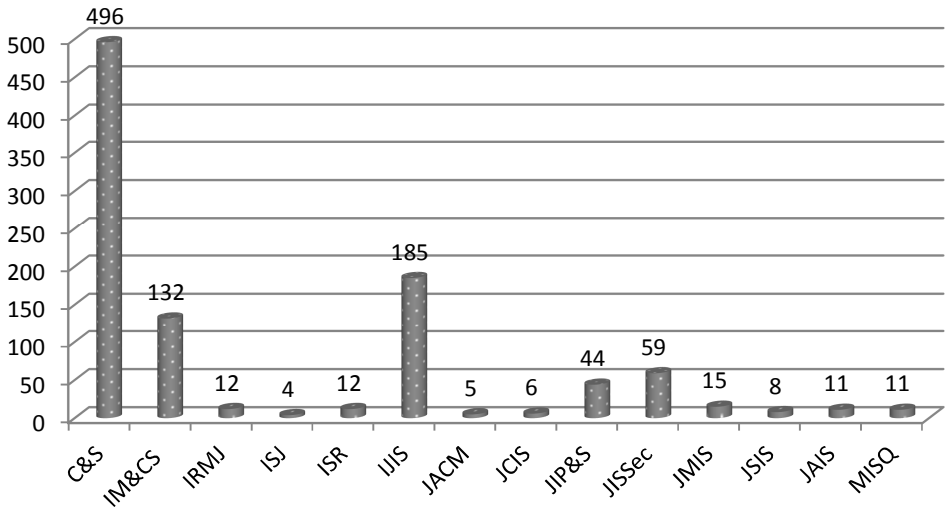


Gráfico 4.2: Número de *Papers* por cada *Journal* (Seleção Final)

Identificados todos os *papers*, elementos que iriam compor a base de dados, a próxima secção debruça-se sobre o estudo de aplicações de *software* que poderiam ser úteis no tratamento e "trituração" de toda a informação que compõe um dado *paper*. Com exceção do corpo de conteúdo de cada *paper*, todos as outras características de um *paper* foram importadas para uma base de dados, tema descrito na secção 4.8.

4.6 Aplicações de *Text Mining*

Na secção anterior foi descrita a forma como se procedeu à identificação e seleção de "artigos/periódicos científicos" a tratar neste projeto. Evidentemente, a compreensão clara deste processo permite ao leitor uma perceção lúcida do tipo de informação que está a ser trabalhada e as necessidades futuras do projeto.

Assim, a secção que agora se inicia, identificada como a fase número sete da Figura 4.1, procede à apresentação e descrição das principais características de aplicações de *software* que existem no mercado, com particular destaque para a identificação de aplicações de código aberto (*open source*) que podem contribuir para a criação do objetivo principal do presente trabalho, com o intuito de possibilitar a aplicação das várias técnicas de processamento textual e respetivos algoritmos.

4.6.1 Seleção de Aplicações

Num processo de descoberta de conhecimento existem várias etapas de processamento de dados e ferramentas aplicacionais que podem auxiliar essas mesmas etapas. Dado o foco deste projeto, a escolha de uma ferramenta de *Text Mining* é sem dúvida uma etapa com elevada importância dada a diversidade de ferramentas que existem no mercado.

A aplicação escolhida terá que dispor de certos requisitos para poder auxiliar a criação do esquema conceptual temático: terá que ser capaz de extrair a informação considerada essencial para analisar o tema de um "artigo/periódico científico", armazená-la, facilitar a exploração da informação recorrendo às funcionalidades disponibilizadas por tal aplicação de modo a conseguir identificar a palavra-chave/etiqueta que nomeie o assunto do "artigo/periódico científico" em questão e devolver dados que facilitem a discussão e conclusão de um resultado final que apoie a construção do esquema concetual temático, tendo como base as "palavras-chave/etiqueta" identificadas em todo o conjunto de "artigos/periódicos científicos".

A discussão sobre os aspetos e características relacionados com a análise de aplicações de *software* é uma tarefa extensa. A abordagem desta secção passou por considerar todas as aplicações que pudessem trabalhar ao nível do processamento e análise de informação em documentos de texto, com o intuito de providenciar o auxílio apropriado ao estudo do problema apresentado anteriormente. A aplicação nomeada terá de ser capaz de proporcionar uma lista estruturada de conceitos/termos do domínio da SSI, onde os termos estejam organizados hierarquicamente; seja um instrumento de organização intelectual, atuando como um mapa conceitual dos tópicos explorados em SSI; e evidencie um modelo conceptual do domínio de SSI. O uso adequado dos recursos computacionais já existentes será um importante aliado facilitando toda a análise, experimentação e avaliação de informação que surja em resposta do processamento da base de dados, e recorrendo a *software*.

O objetivo traçado para a aplicação identificada passa por representar o conhecimento existente e explicitado em SSI, assente numa estrutura de documentos agregados. O primeiro passo consistiu na identificação de todas as aplicações que poderiam, com alguma característica iminente, auxiliar a organização e tratamento de informação. Posteriormente, analisou-se individualmente cada aplicação, atentando o seu contributo para o problema identificado. Para a seleção das ferramentas usou-se, como base, os portais http://www.predictiveanalyticstoday.com/top-30-software-for-text-analysis-text-mining-text-analytics/ e http://www.kdnuggets.com/software/text.html.

A lista que se segue, Tabela 4.2, exibe uma pequena parte de ferramentas de TM existentes. Com base nos critérios propostos Cruz [2007] foram identificadas ferramentas que apresentavam características que pudessem materializar o objetivo deste trabalho: ser uma plataforma, preferencialmente serem *open source* (a lista apresenta todas as que foram ponderadas mesmo que não sejam *open source*), possuir interface gráfica, ter conectividade a base de texto em diferentes formatos (txt, pdf, Web e email), possibilidade de integração com outras ferramentas de processamento da linguagem natural e apresentar a capacidade de suportar as etapas da descoberta do conhecimento (KDD).

Tabela 4.2: Lista de Ferramentas de Text Mining

Ferramenta	URL
PolyAnalyst	www.megaputer.com/
IBM Text Analytics	www-01.ibm.com/software/ebusiness/jstart/textanalytics/
Lexalytics Text Analytics	lexalytics.com/software
Smart Logic 5	www.smartlogic.com
Ai-One	www.ai-one.com
WordStat	provalisresearch.com
SAP Text Analytics	www.sap.com
SAS Enterprise Content Categorization	www.sas.com/en_us/software/analytics/enterprise-content-categorization.html
SAS Text Miner	www.sas.com/technologies/analytics/datamining/textminer/
Attensity	www.attensity.com/
Content Analyst	www.contentanalyst.com

Cogito	www.expertsystem.net/
Angoss Text Analytics	www.angoss.com/
WEKA	www.cs.waikato.ac.nz/ml/weka/
Knowledge Server	www.autonomy.com/
OpenText	www.opentext.com/
AeroText	www.rocketsoftware.com/
Textalytics	textalytics.com/home
Luxid Information Analytics	www.temis.com/home
Intelligent Miner for Text	www-01.ibm.com/common/ssi/cgi-bin/ssialias?infotype=OC&subtype=NA&htmlfid=897/ENUS11L3691&appname=System%20Storage
Clementine Data Mining System	www-304.ibm.com/partnerworld/gsd/solutiondetails.do?solution=10387&expand=true
OneCalais	www.clearforest.com
Oracle Data Mining	www.oracle.com
RapidMiner	rapidminer.com
Saplo	saplo.com/products
Muscat Structure	http://www.smartlogic.com/home/products/products-overview
Readware	http://www.readware.org/
Information Intelligence	http://www.recommind.com/
Statistica Text Miner	http://www.statsoft.com/

A lista que se segue, caracteriza cada uma destas aplicações. Não é descrita a aplicação na sua totalidade, apenas as características que melhor se evidenciam. O utilizador tem ao seu dispor na Tabela 4.2 os locais de endereço web que poderá explorar para um melhor entendimento de cada uma das aplicações.

1. PolyAnalyst – Megaputer Intelligence

PolyAnalyst é um pacote de mineração de dados comerciais da *Megaputer*. Abrangendo todo o ciclo de análise de dados completo de carregamento de dados e integração de modelação e

elaboração de relatórios, oferece uma ampla seleção de algoritmos para análise automática de texto e dados estruturados. Pode ler a partir de uma variedade de base de dados, pacotes estatísticos, ler o texto de HTML, Word e PDF. O motor OLAP permite que os dados sejam agregados ou "cortados e cortados" antes de aplicar algoritmos de *Data Mining*. Destacam-se ferramentas como árvores de decisão, redes neurais, algoritmos genéticos, lógica *fuzzy*, raciocínio baseado em casos e categorização de texto.

O sistema permite aos utilizadores executar inúmeras operações de descoberta de conhecimento: Categorização, *Clustering*, Previsão, Análise de ligação, Palavras-chave e extração de entidade, Descoberta de Padrões, Deteção de Anomalias.

2. IBM Text Analytics

A solução dispõem de um ambiente de captura de conhecimento em dicionário e regras semânticas para reutilização, extração de informação costumizável e entidades reconhecimento de entidades relacionamento recorrendo ao treino, regras e análise da linguagem natural.

3. Lexalytics Text Analytics

Permite transformar texto não estruturado sobre entidades (pessoas, lugares, empresas, produtos), sentimentos, citações, opiniões e temas de texto em conhecimento organizado. Está integrado em sistemas de pesquisa de mercado, monitoramento das redes sociais, pesquisa/análise da voz do cliente, pesquisa corporativa e política pública.

4. Smart Logic 5 – Knowledge management

Plataforma de análise de conteúdo de texto, processamento de linguagem natural, classificação baseada em regras, modelação de ontologia/taxonomia e visualização da informação.

5. Nathan, Ai-One

É uma aplicação que permite construir inteligência artificial em dados não estruturados, semiestruturados ou estruturados, detetando o significado contextual dos dados por meio de padrões. Apresenta uma interface bastante *user-friendly* para o utilizador, autónoma na sua aprendizagem, multilinguística e transparente na sua aprendizagem.

6. WordStat – Provalis Research

Facilita a mineração de texto para grandes quantidades de texto. Possuí ferramentas de análise estatística de dados e de dados qualitativos.

7. SAP Text Analytics

Software de análise de texto a partir de SAP, com origem na Inxight Software, com funcionalidade de integração de dados, qualidade, perfil e análise de texto que auxiliam a tomada de decisão.

8. SAS Enterprise Content Categorization

Alimentada pela tecnologia *Teragram*, aplica-se ao processamento de linguagem natural (NLP) e técnicas linguísticas avançadas para categorizar automaticamente grandes volumes de conteúdo multi-língua que é adquirido, gerado ou existente num repositório. Analisa os conceitos gramaticalmente e o conteúdo de entidades e eventos, os quais são depois usados para criar meta dados, desenvolver taxonomias, gerar as regras das categorias e definições dos conceitos que podem ser aplicados a grandes volumes de documentos e os quais desencadeiam os processos de negócio.

9. SAS Text Miner

SAS Text Miner disponibiliza ao utilizador um conjunto de ferramentas para descobrir e extrair inteligência de coleções de documentos de grandes dimensões. Ajuda a identificar tendências e negócios, e gera oportunidades significativas para a compreensão de questões chave do negócio mais eficiente e com menos riscos. A utilização desta ferramenta permite classificar documentos pré-definidos, encontrar relações explícitas ou associações entre documentos, e incorporar dados textuais não-estruturados. Esta solução permite agilizar o processo de mineração de dados, para criar modelos preditivos e descritivos com base na análise de grandes quantidades de dados. O seu desempenho facilita a análise recorrendo a capacidades como *stemming*, reconhecimento automático de termos *multi-word*, normalização de várias entidades, tais como datas, moeda, percentagem, ano, *part-of-speech tagging*, extração de, apoio para os sinónimos, e linguagem de análises específicas.

10. Content Analyst

A empresa dispõe de uma plataforma de integração CAAT que possuí um conjunto de soluções de análise de texto que reduzem o tempo necessário para discernir as informações relevantes a partir de dados não estruturados. Providencia soluções de pesquisa de conceitos, categorização com base no conceito, *clustering* dinâmico, *email threading*, *near-duplicate detection*, *language analytics* e sumarização para uma larga escala de dados não estruturados.

11. Cogito – Expert System

Expert System é uma empresa com foco em tecnologias com elevada análise semântica, permitindo uma gestão mais eficiente do volume de informação, transformando-a em conhecimento acessível e utilizável para as organizações, no desenvolvimento de produtos, inteligência competitiva, gestão na interação com o cliente, entre outros. Incluí a pesquisa semântica e de linguagem natural, análise de texto, desenvolvimento e gestão de taxonomias e ontologias, categorização automática, extração de dados e meta dados e processamento de linguagem natural. A informação pode provir de documentos, *email*, *intranet*, *internet*, conteúdos *mobile* e *information streams*, apresentando soluções de pesquisa semântica, análise de texto, inteligência e segurança, categorização automática e processamento de linguagem natural.

12. Angoss Text Analytics

A presente solução, com a combinação de resultados de *Text Mining* com dados não estruturados apresenta resultados de *Data Mining* satisfatórios, permitindo ao utilizador extrair entidades, categorizar e visualizar relações entre os temos. A informação pode ser analisada com a descoberta visual, análise de comparação, análise de tendências, associação entre conceitos, documento de exploração, análise preditiva, suporte de processamento de linguagem natural e multi-língua e modelos de implementação.

13. WEKA

A ferramenta *Weka* é um programa em Java de código aberto que tem como principal objetivo a mineração de dados. Os seus recursos facilitam o processamento de texto, a classificação, a regressão, o *clustering*, as regras de associação e de visualização. *Weka* tem como objetivo agregar algoritmos provenientes de diferentes abordagens/paradigmas na subárea da inteligência artificial dedicada ao estudo da aprendizagem por parte de máquinas. *Weka* procede à

análise computacional e estatística dos dados fornecidos recorrendo a técnicas de *Data Mining* tentando, indutivamente, a partir dos padrões encontrados gerar hipóteses para soluções.

14. Knowledge Server – Autonomy

Solução automatizada que facilita o processo de mineração de texto, *clustering*, consulta, sumarização, categorização e atribuição de *ranking* de relevância de grandes quantidades de informação internas e externas. Cria mapas de conceitos e tópicos agrupados sem intervenção humana, usando a probabilidade *Bayesiana* e reconhecimento de padrões.

15. OpenText

Possui soluções com foco na gestão de informação estruturada e não estruturada. Permite a captura, gestão de documentos e de registos, fluxo, pesquisa e armazenamento, aplicações e *add-ons* (ex. *email*) que podem ser analisados para obtenção de informação eficiente e rápida para o negócio da organização.

16. AeroText – Rocket

Conjunto de aplicações de mineração de texto para análise de informação não estruturada, com soluções de ambiguidade. Este motor de análise de texto apresenta funcionalidades como descobrir entidades e a relação entre estas, as suas características, entre outras, podendo ser configurado para reconhecer outras aplicações e terminologias relacionadas com o objetivo da pesquisa em causa.

17. Textalytics

Conjunto completo de APIs multilinguísticas de análise de texto para classificação automática, extração de entidades e conceitos, entre outros aspetos. Esta solução é otimizada para os diferentes cenários de aplicabilidade de APIs específicos para o negócio em questão.

18. Luxid Information Analytics

Plataforma com soluções para estruturação, gestão e exploração de conteúdos não estruturados das organizações. Permite a captura de conceitos chave, visualizações gráficas, entre outros aspetos.

19. Intelligent Miner for Text – IBM

A aplicação integra um conjunto de soluções de análise de texto. Inclui análise linguística, agrupamento automático e/ou atribuição de classificação em categorias definidas, sumarização de documentos e identificação de conceitos-chave recorrendo à extração.

20. Clementine Data Mining System

Sistema de baseado em programação visual, que inclui diversas técnicas de aprendizagem automática, como Indução de Regras, Redes Neuronais e Árvores de Decisão. Disponibiliza ainda ferramentas que permitem manipular, explorar, visualizar e construir modelos sobre os dados. Todos os passos do processo de descoberta de conhecimento podem ser realizados pelo Clementine. Possui uma interface de programação visual que facilita a construção de modelos de *Data Mining* para o processo de descoberta de conhecimento. A ferramenta oferece facilidades para a exploração e manipulação de dados, além de várias técnicas de modelação e recursos gráficos, para a visualização de dados. As operações são representadas em uma área de trabalho por nós (nodes) que, conectados, formam o fluxo de dados, chamado de *streams*.

21. ClearForest – OneCalais

A solução é composta por um motor de pesquisa com acesso a base de dados de variadas fontes de dados. Esta solução de processamento de linguagem natural analisa os resultados e extrai significado da informação não-estruturada, recorrendo a padrões via palavras-chave, apresentando a informação de forma gráfica e intuitiva.

22. Oracle Data Mining

A presente solução possui uma base de dados de análise preditiva com algoritmos de *Data Mining*. Os utilizadores podem aceder a interfaces gráficas para explorar os dados, encontrar padrões, relacionamentos e conhecimentos ocultos.

23. RapidMiner

A aplicação desenvolvida em java é uma ferramenta de ambiente integrado para o *Data Mining*, *Text Mining*, análise de dados e *Business Intelligence*. Possuí uma utilização versátil em qualquer sistema operativo e ambiente de trabalho, assim como, uma interface gráfica intuitiva e flexível para o desenho de um processo de análise.

24. Saplo

Permite extrair informação em grandes quantidades de texto por meio de uma API.

25. Muscat Structure – Smartlogik Discover

Categorização automática em tempo real, usando uma ferramenta de construção de regras para especificar os documentos que fazem ou não parte de uma categoria específica de uma taxonomia.

26. Readware

Sofisticado sistema de classificação com um *ConceptBase* de estruturas fundamentais de conhecimento, correlacionado às consultas e documentos. Codifica um modelo de significado em vez de um modelo de um documento ou um texto como outro *software* faz. *Readware* é capaz de usar operações abstratas e identificar objetos abstratos, processos e relações temáticas de PNL, *Text Mining* e de *software* de máquina de aprendizagem que não pode, possivelmente ter acesso.

27. Information Intelligence

Usando uma classificação ou taxonomia existente, este *software* pode classificar automaticamente documentos com base em semântica e análise probabilística. O motor *MindServer Categorizer* realiza a análise sobre o corpo dos documentos, criando taxonomias e categorias de conteúdo em tempo real.

28. Statistica Text Miner

Recorrendo a técnicas de modelação preditiva, *clustering*, classificação e procedimentos exploratórios, a Statistica fornece um conjunto abrangente de análise de dados, gestão e visualização de dados, assim como soluções de mineração de dados.

4.7 Criação da Base de Dados

Nesta secção abordam-se os aspetos relacionados com a criação da base de dados. Este processo foi um meio laborioso para conseguir alcançar a resposta para o objetivo do projeto. O conceito de base de dados está em constante interação nas atividades do ser humano, mesmo por vezes de forma não explícita (por exemplo, quando fazemos compras num hipermercado, quando

procuramos um livro na biblioteca, uma agenda com as moradas de pessoas conhecidas, entre outros exemplos). Emerge como um meio para os sistemas de informação que necessitam de receber, armazenar, processar e apresentar a informação onde esta é necessária e na forma pretendida pelo utilizador final. Para Damas [2005, p. 33] "uma base de dados consiste numa coleção de dados estruturados, organizados e armazenados de forma persistente por uma aplicação informática".

O uso de uma base de dados resolve problemas como o alto nível de redundância, incoerência da informação, inflexibilidade, acessos concorrentes, isolamento e integridade dos dados, elevados custos de manutenção, entre outros aspetos. O objetivo de criar e manter uma base de dados permite que um programa possa ser modificado, alterando a forma de utilização dos dados, sem que isso implique alterações nos restantes programas que utilizam esses mesmos dados. Uma das vantagens mais significativas do uso de base de dados consiste na redução de custos de manutenção através da separação entre a forma como os dados são percebidos pelo programador e a forma como esses dados são armazenados fisicamente.

Para facilitar o acesso, o uso e obtenção de grandes quantidades de dados, um utilizador recorre a programas que lhe permitem desempenhar tarefas de armazenamento e manipulação de dados, proporcionando aos programadores e utilizadores finais os dados tal como eles são pedidos – Sistemas de Gestão de Base de Dados (SGBD). Segundo Damas [2005, p. 15] um SGBD é uma "aplicação informática (...) que fornece a interface entre os dados que são armazenados fisicamente na base de dados e o utilizador". Estes sistemas de gestão e armazenamento de dados servem de intermediário entre o nível aplicacional e a base de dados, impedindo a manipulação direta por parte de cada aplicação cliente, garantindo a segurança, integridade e validade dos dados armazenados. Com a utilização de um SGBD é fornecido um conjunto completo de serviços de acesso aos dados, recorrendo-se a uma linguagem de pesquisa, linguagem SQL, que possibilita a um utilizador interrogar ou operar sobre a base de dados, de modo a realizar as tarefas pretendidas.

Para a realização do presente trabalho foi identificada a importância da criação de uma base de dados que armazene toda a informação e dados relevantes, alusivos aos periódicos e artigos científicos identificados na fase número seis da Figura 4.1. A atual secção retratará todo o trabalho elaborado na fase número oito da Figura 4.1, isto é, descreverá todo o processo que envolveu a criação da base de dados. A próxima subsecção descreve o trabalho elaborado com recurso à linguagem de programação Java. Esta secção surge em resposta à necessidade eminente

de tratamento de dados para execução das técnicas de *Text Mining* sobre a base de artigos científicos.

4.7.1 Semelhança entre as Aplicações

Dada a lista apresentada na secção anterior, a presente secção compara as ferramentas identificadas com um conjunto de características específicas que as ferramentas de *Text Mining* normalmente apresentam. A escolha das ferramentas está relacionada com as principais etapas de processamento de descoberta de conhecimento que são necessárias realizar para elaborar o esquema conceptual temático. Estas etapas de descoberta de conhecimento estão relacionadas com tarefas como: classificação (procura e junção de termos semelhantes num conjunto de rótulos categóricos predefinidos), regras de associação (procura de termos que ocorrem frequentemente de forma simultânea nos dados), *clustering* (agrupar documentos com conteúdos semelhantes), sequência (descobre alterações na quantidade de frequência de conteúdo), deteção de desvios (procura de termos com características distintas dos padrões que normalmente são identificados), regressão (pesquisa de termos no mapeamento de conteúdos de dados), sumarização (identifica características comuns entre conjunto de dados) e associação (associação de palavras recorrendo a um dicionário).

A Tabela 4.3 apresenta de modo sucinto as ferramenta anteriormente apresentadas, relacionando-as com as tarefas que caracterizam a temática *Text Mining* no processamento de descoberta de conhecimento. A primeira coluna da tabela identifica as tarefas, à exceção das duas últimas linhas, e as restantes colunas correspondem às ferramentas identificadas. A numeração das colunas está de acordo com a numeração das ferramentas na secção anterior.

A lista elaborada na secção 4.6.1 facilitou a caraterização de ambientes integrados de aplicações que permitem a descoberta de conhecimento. O levantamento efetuado de ferramentas de *Text Mining* quer elas sejam comerciais ou de código aberto que permitem a aplicação de várias técnicas de pré-processamento, aplicação de algoritmos e extração de informação, auxiliou na criação de um modelo que destaca as características mais evidentes de cada aplicação para realizar a descoberta de conhecimento e facilitar a comparação entre as mesmas.

Apesar de se estar a generalizar as tarefas de descoberta de conhecimento, nada impede que esta seja a base para relacionar a implementação com outras especificidades e de outras aplicações. Para este caso prático e da análise elaborada conclui-se que nenhuma das aplicações apresentadas responde a todos os requisitos necessários para a criação da solução pretendida do

presente estudo, assim como, não existe nenhuma forma de as integrar. Para o desenvolvimento do esquema de classificação da literatura de SSI, um dos principais requisitos era identificar uma aplicação *open source* capaz de representar o esquema com clareza e interface apelativa. Dos resultados obtidos da subsecção anterior apenas foram encontradas duas, de um total de 28, aplicações *open source*, das quais não satisfazem as necessidades requeridas e apresentadas anteriormente. O facto de grande parte das aplicações não serem de utilização livre impossibilitou a sua experimentação e adequação ao presente projeto. Grande parte das aplicações tem tarefas de descoberta de conhecimento como a classificação, *clustering*, regressão e sumarização.

Também foi importante analisar a forma como a informação seria apresentada. Uma área de visualização "amiga do utilizador" facilita e torna mais acessível a interpretação da informação, dedicando-se o utilizador ao problema de pesquisa e não à compreensão de como funciona a aplicação. É importante salientar que não existem técnicas que satisfaçam todas as expectativas de visualização de informação, mas é recomendável que existam combinações entre as técnicas de visualização para facilitar a interpretação do resultado.

Tabela 4.3: Comparação das Aplicações de *Text Mining*

Tarefas KDD	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28
Classificação	X	X	X	X		X	X	X	X	X	X	X	X	X		X	X	X	X	X	X	X	X		X	X	X	X
Regras de Associação	X	X			X			X	X			X	X	X	X	X	X	X	X	X				X			X	X
Clustering	X	X	X		X	X	X		X	X	X	X	X	X	X				X	X	X	X	X		X		X	X
Sequência			X																X	X								X
Deteção de Desvios	X																			X				X				X
Regressão	X			X	X	X	X	X	X	X	X	X	X	X	X	X	X	X				X	X	X	X	X	X	X
Sumarização	X		X	X	X	X		X	X	X	X	X		X			X	X	X		X			X	X	X		X
Associação		X	X	X				X	X			X									X	X	X					
Licença de Uso Comercial	X	X	X	X	X	X	X	X	X	X	X	X		X	X	X	X	X	X	X		X		X	X	X	X	X
Licença de Uso Gratuito													X										X					

4.7.2 Programa Java

A presente secção descreve de forma detalhada o que foi desenvolvido recorrendo à linguagem de programação Java. Esta secção, de acordo com a Figura 4.1, que descreve as principais fases do projeto de dissertação, encontra-se inserida dentro do ponto 8 sendo uma sub-tarefa desse ponto. Com este exercício, a compreensão e acompanhamento de toda a criação da base de dados de papers será simplificada, permitindo uma abordagem clara do trabalho. Das diversas formas existentes para proceder ao tratamento de informação, optou-se neste trabalho de investigação pela realização de um Programa em código Java. Java é uma linguagem de programação orientada a objetos – objetos são instâncias de classes que determinam o comportamento do objeto.

As razões que levaram à escolha desta linguagem de programação para auxiliar o presente trabalho tem origem numa tecnologia de elevada capacidade que possibilita um maior controlo sobre as tarefas, maior rapidez, facilita a implementação de mecanismos de boa execução, maior versatilidade na execução das tarefas, segura, confiável e simplifica a replicação de tarefas. Entendeu-se que recorrendo a uma aplicação para o auxílio do processamento e análise de informação seria uma mais-valia, uma vez que existe um maior controlo sobre as tarefas a executar. O programa auxilia o carregamento de dados para a base de dados, valida a exatidão dos dados inseridos quando comparados com o ficheiro original, facilita a execução de operações sobre os dados presentes na base de dados para a geração aleatória de números, tratamento de *stopwords*, contagem de palavras ao nível do título e *abstract* de cada *paper*, entre outras operações.

Dada a complexidade e dimensão dos dados tratados, é importante despender de um maior detalhe de todas as tarefas implementadas. O Programa Java implementado é composto por seis classes, uma classe responsável por exibir toda a informação tratada por meio de uma interface, duas classes que trabalham sobre a informação já introduzida na base de dados e as restantes três classes fazem o processamento de informação para inserir a informação na base de dados. Prosseguindo para a explicação aprofundada de cada uma das seis classes, toda a implementação começa com a criação de uma ligação entre a base de dados e o Programa Java, facilitando a sua comunicação, designando-se a classe por "*BDConnector*". Esta classe, ao estabelecer uma ligação com a base de dados criada, auxiliou a passagem de informação contida nos ficheiros de texto para a base de dados, permitindo deste modo, todo o tipo de pesquisa e análise sobre os dados. Toda a informação sobre as publicações foi colocada em vários ficheiros ".txt", apenas agrupada pelos 14

journals (tema abordado em maior detalhe na próxima secção) considerando-se que este seria o modo mais apropriado para lidar com a elevada complexidade da informação selecionada.

Ao nível da introdução dos dados na base de dados trabalhou-se nas classes: "*SourceDirFrame*" e "*PropertiesHandler*". No primeiro caso, "*SourceDirFrame*", a classe apresenta comportamentos tais como: averigua a existência de *papers* repetidos nos ficheiros de texto recorrendo à análise e comparação dos diferentes títulos de acordo com o *journal*, verifica se o número de secções/componentes introduzidas por cada *paper* corresponde à estrutura definida de 8 elementos de dados para definir as características básicas de um *paper*, remove todos os parágrafos e espaços duplicados no título e *abstract* dos *papers*, remove todos os espaços duplicados no início, meio e final de cada keyword, facilita a leitura do ficheiro de texto que contém toda a informação relativa aos *papers*, e finalmente carrega/injeta todos os dados dos ficheiros de texto relativos de cada *paper* para a base de dados. No caso da classe "*PropertiesHandler*" permite o carregamento do ficheiro de texto que contém todas as configurações necessárias para o Programa Java poder aceder às várias bases de dados criadas (facilita a leitura das configurações da BD, *username* e *pass*).

Para que os dados executados nos diversos comportamentos das classes fossem possíveis de visualizar, foi criada a classe "*LogWindow*". Esta classe apresenta, se solicitado, uma interface que é moldada de acordo com o que está a ser executado no momento, permitindo mostrar os resultados de todas as interações e ações que estão a ser realizadas no momento pelo Programa Java.

A segunda parte desta secção foca nas classes que facultaram, mediante os comportamentos definidos na mesma, a observação e tratamento de informação da base de dados. Foram criadas duas classes: a "*OtherOperations*" e a "*DatabaseOperations*". No caso da primeira classe, o programa permitiu o carregamento de todos os dados contidos num ficheiro ".txt" com uma lista de *stopwords*, verifica se a lista de *stopwords* continha elementos repetidos, e gera 100 números aleatórios distintos compreendidos entre 1 e 1000, confirmando que cada número gerado ainda não se encontra na lista de números já gerados. Os números gerados referenciam o "idpaper" de um dado *paper* que será alvo de análise detalhada do rigor da informação presente na base de dados. A classe "*DatabaseOperations*" é a classe responsável por todas as operações feitas sobre a base de dados. Nesta classe foi implementado um conjunto de regras que identificam a ordem de ocorrência de cada *keyword* num dado *paper*, introduzindo e guardando essa valor no campo na base de dados. A próxima etapa cruzou a lista de *stopwords* com as *keywords* de cada

"título" e *abstract* de cada *paper*. Uma outra tarefa surge com a necessidade de saber exatamente "que palavras e quantas palavras" estão presentes nos "títulos" e *abstract* de cada *paper*, mas já com a eliminação de *stopwords*, ou seja, saber quantas palavras de cada tipo estavam a ser repetidas nestes dois campos de dados. Assim, foi contabilizado quantas palavras eram encontradas repetidas e colocado o seu valor total num campo criado na base de dados, sendo que, a presente contagem apenas funciona para palavras "únicas" e nunca para expressões. Por fim, nesta fase, aproveitou-se para limpar todos os parágrafos e espaços extras, vírgulas, pontos, dois pontos, parênteses e ponto-vírgula.

Em anexo, no Apêndice A é possível ver um pequeno extrato do código desenvolvido.

Concluída a explicação da presente secção, crê-se com este esclarecimento um entendimento simplificado para a próxima secção que retrata a criação da base de dados.

4.7.3 Desenvolvimento da Estrutura de Dados

A presente subsecção descreve em detalhe todos passos que foram necessários realizar para a criação da estrutura de dados. Todavia, antecedente a esta mesma etapa realizou-se a seleção de todas as palavras contidas na secção *keywords* dos "artigos científicos", isto é, esta etapa consistiu na seleção de palavras *keywords* que seriam incluídas e que definiam claramente a área de SSI. Com estas mesmas palavras *keywords* foi criado um documento "Excel", composto por duas folhas (*keywords* de SSI e *keywords* que não pertenciam a SSI) onde eram expostas todas as *keywords* dos "artigos/periódicos científicos" selecionados. Uma vez que era interesse e análise do projeto debruçar-se apenas sobre as *keywords* de SSI, as *keywords* resultantes da seleção foram organizadas de acordo com a relação temática que podia ser estabelecida entre elas, criando-se um esquema, esboço, da literatura de SSI tendo apenas como base os estudos que tem sido desenvolvidos na área. Desta organização, resultaram 85 grandes grupos de *keywords* de SSI, organizadas pelo seu grau de importância e relação (ver Apêndice C). Estes 85 grupos abrangem uma variedade de *keywords* que estão relacionadas tematicamente entre elas (tema abordado na próxima secção).

Para o desenvolvimento da estrutura de dados apenas serão observadas as *keywords* da temática de SSI, que será uma tabela integrante da base de dados e cuja denominação nesta primeira fase será - "arvorekeywords". O desenvolvimento da base de dados fez-se recorrendo a uma tecnologia de armazenamento de informação e de dados do tipo base de dados relacional. Este tipo de estrutura garante que cada dado tem um, e apenas um lugar na base de dados,

armazenando toda a informação em tabelas, sendo uma ou mais colunas o índex da tabela (chave pela qual a informação é procurada).

Inicialmente começou-se a utilizar a aplicação "*phpMyAdmin*" para auxiliar o controlo da base de dados *MySql* no que se refere às etapas de criação, cópia, eliminação, renomeação e alteração de tabelas, permitindo a manutenção "cuidada" das mesmas. Apagar, editar, adicionar campos, exportar ou importar uma base de dados, entre outras funcionalidades são características desta aplicação que deu suporte a todo o tratamento de dados. Perante a dimensão deste projeto foram consideradas várias etapas no desenvolvimento da base de dados, criando-se várias versões que foram "batizadas" como "pilotov0", "pilotov1", "pilotov2", e "pilotov3". Após a criação da base de dados foi iniciado o processo para criação de tabelas/entidades e respetivos campos/atributos, assim como de tabelas auxiliares designadas por "teste" (eliminadas mais tarde).

O primeiro passo realizado para a criação da base de dados consistiu na elaboração de um esquema visual que representasse a estrutura a implementar de entidades e atributos, de forma transparente e que esteja em conformidade com as boas práticas de criação de uma estrutura de base de dados, o qual é apresentado na Figura 4.4. Sendo "entidade" um conceito abstrato, pode ser definido como a identificação dos elementos e do conjunto de atributos comuns do mundo real que estão a ser analisados. Para "atributo" entende-se o que representa os dados da entidade com um determinado domínio de valores, podendo ser um atributo opcional ou obrigatório. Os atributos existentes na definição da entidade não têm todos o mesmo grau de importância e por isso, alguns atributos conseguem identificar univocamente cada uma das ocorrências da entidade e são escolhidos como identificadores/chaves.

O passo seguinte passou por estabelecer uma relação entre as entidades de forma a relacionar os dados de uma com os dados da outra entidade, onde a linha que une as entidades é o que permite estabelecer a relação entre elas. Usualmente, existindo uma relação direta entre o modelo relacional e a implementação física da base de dados desse mesmo modelo, é natural que se opte por este modelo para fazer a modelização dos dados, uma vez que a sua implementação numa base de dados relacional é quase direta.

Perante o problema descrito e mediante os conteúdos teóricos apresentados foram consideradas, numa primeira etapa, oito entidades para descrever o problema: "arvorekeywords", "autor", "keyword", "keywords_idarvore", "paper", "relacoesarvorekeys", "tag" e "tipoarvorekeys". A entidade "arvorekeywords" versa todas a *keywords* de SSI que foram encontradas em todos os "artigos/periódicos científicos" respeitando o seu grau de descendência definido. A entidade "autor"

apresenta segundo cada "artigo científico" os autores que o conceberam. A entidade "keyword" mostra, mediante o "artigo científico", as *keywords* que foram definidas pelos autores para identificar essa mesma publicação. A entidade "keywords_idarvore" é constituída por todas as *keywords* que identificam os "artigo/periódico científico", mas com uma particularidade bastante distinta da entidade "arvorekeywords", apenas as "keywords" de SSI são inequivocamente identificadas com um identificador que permitirá estabelecer uma ligação com a entidade "arvorekeywords". A entidade "paper" é a tabela que apresenta todas as componentes que constituem um "artigo científico" com a exceção do "corpo de texto" deste mesmo *paper*; a entidade "relacoesarvorekeys" contribui para estabelecer a relação entre as diversas *keywords* de SSI com base na tabela "arvorekeywords". A entidade "tag" é uma tabela que sofreu várias alterações ao longo da evolução do projeto dado os vários casos de estudo considerados para a seleção da *keyword* que melhor identificava um *paper*. Por fim, a entidade "tipoarvorekeys" é uma tabela auxiliar que apresenta os três tipos de *keywords* ("keywords de SSI", "keywords não SSI" e "keywords vazias") que são abordadas em todos os "artigos/periódicos científicos" selecionados. A Figura 4.2 apresenta a primeira estrutura de dados inserida na base de dados.

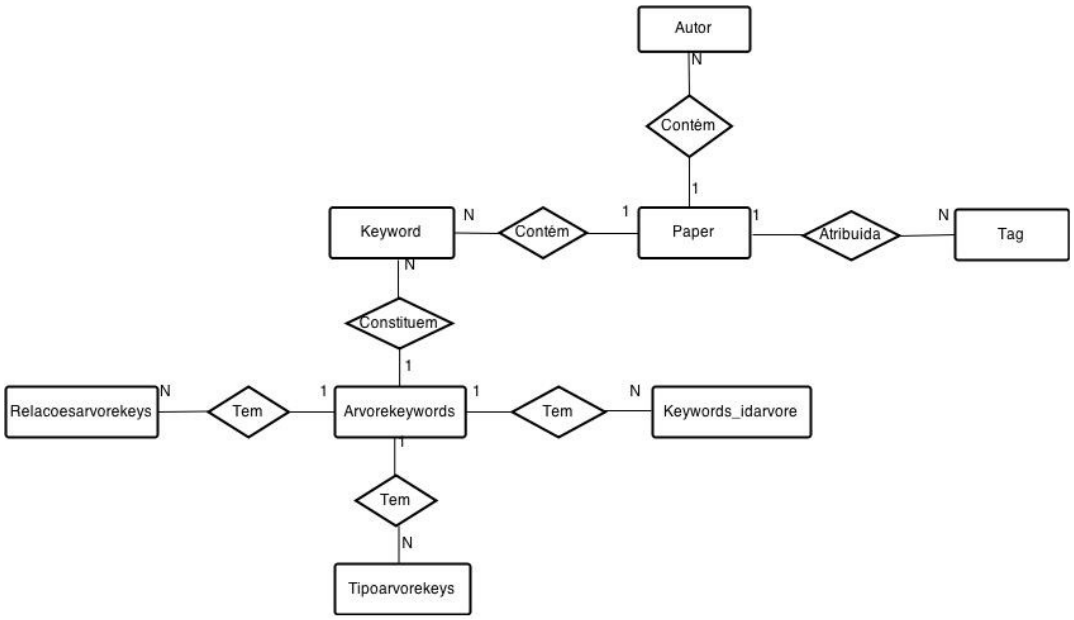


Figura 4.2: Diagrama Entidade-Relação (versão 0)

Neste momento já está compreendida a estrutura teórica que suportou a construção da primeira versão do diagrama de entidade-relação. Com a especificação do primeiro modelo de dados, passou-se para a implementação física da base de dados de acordo com a estrutura

especificada no modelo entidade-relação e pelo modelo de dados respetivo. No esboço representativo da base de dados é possível ver a estrutura das tabelas que permitem armazenar dados, especificando os seus atributos. A Figura 4.3 exibe os atributos alusivos de cada entidade e a sua relação. Os atributos que se apresentam sublinhados são os identificadores/chaves de cada tabela.

A descrição de cada uma das oito entidades permitirá ao leitor uma melhor compreensão futura face à estrutura de dados construída.

Começando pela entidade "paper", constituída por um "id" único, nome do "journal" (secção 4.5), "titulo", "ano" de publicação, "volume", "número", "página de inicio e final", e "abstract". O atributo "abstract" foi analisado porque em algumas situações era questionável o assunto do "paper" quando apenas se focava no "título e keywords".

A segunda entidade é a "keyword", composta por um "idkeyword" irrepetível, um "idpaper" que permite relacionar as "keywords" com o respetivo "paper" e a "keyword" que indica a descrição da "keyword".

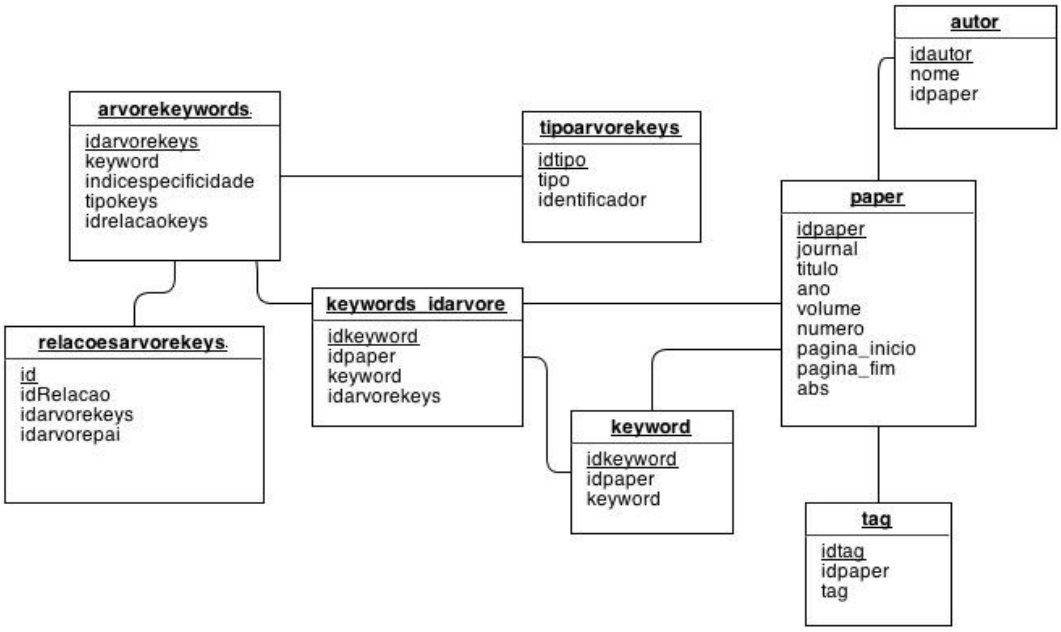


Figura 4.3: Atributos do Diagrama Entidade-Relação (versão 0)

Uma terceira entidade, "autor", constituída pelo identificador único "idautor", "nome" do autor, e "idpaper" que permite estabelecer a relação com o "paper".

Uma quarta entidade foi criada, a "keywords_idarvore", com o atributo "idkeyword", "idpaper" para relacionar-se com a entidade "paper", "keyword" que nomeia esta mesma, e "idarvorekeys" que estabelece a ligação à entidade "arvorekeywords".

A quinta entidade é a "arvorekeywords", que tem presente um "idarvorekeys", a descrição da "keyword", o atributo "indicespecificidade" que atribui a cada *keyword* o nível de ascendência em que se posiciona na "arvorekeywords" (é considerado valor fraco em nível 1 e valor excelente em nível 6, que é o máximo número de filhos que um grupo de "keywords" possui). Quando foi realizado o esboço do esquema da literatura de SSI mencionado anteriormente, todas as *keyword* obedeceram a uma estrutura que considerava a semelhança temática entre elas e que era organizada desde a *keyword* mais genérica à mais detalhada. Cada coluna no ficheiro correspondia a um valor de "indicespecificidade" e por isso separou-se a informação coluna a coluna colocando cada coluna numa folha diferente do Excel e acrescentando-se o campo "indicespecificidade" e restivo valor. De seguida, pegou-se na informação das várias folhas e colocou-se tudo numa única folha importando-se novamente para a base de dados. O atributo "tipokeys", identifica se a "keyword" pertence ao domínio da SSI, sendo esta uma árvore apenas constituída por *keywords* de SSI não resulta nenhuma dúvida em relação à sua inclusão e tipo, e a "idrelacaokeys" que facilita a integração com a entidade "relacoesarvorekeys".

A entidade "relacoesarvorekeys" composta por quatro atributos, o "id", o "idRelacao", o "idarvorekeys" que facilita a ligação à tabela "arvorekeywords", o "idarvorepai" que atribuí a todas as *keywords* com a mesma descendência o mesmo identificador e facilita na compreensão do grau de parentesco.

A sétima entidade, "tipoarvorekeys" com atributos como o "idtipo", o "tipo" que conceitua se a "keyword" está incluída/excluída no domínio da SSI, e o "identificador" que corresponde a um valor que define a "keyword" quanto ao seu tipo.

Por fim, a entidade "tag" constituída pelo "idtag", "idpaper", e a "tag" que designa a "tag" atribuída a um dado *paper*, onde um *paper* pode conter mais do que uma "tag" para o identificar.

As oito entidades apresentadas anteriormente constituem a estrutura relacional que é repercutida no "phpMyAdmin", sendo relacionadas por meio dos identificadores/chaves. Toda a estrutura é iniciada com a tabela "paper", que dispõe toda a informação importante sobre um determinado "paper". Nesta fase foram realizados testes à base de dados para verificar a integridade e veracidade dos dados, mas apenas serão abordados no Capítulo 5.

Apesar da aplicação "*phpMyAdmin*" ter elevado potencial de análise, não respondem satisfatoriamente a todas as exigências de manuseamento. A criação da base de dados foi feita através das funcionalidades disponibilizadas pelo "*phpMyAdmin*". Após uma compreensão profunda do modelo, conclui-se que seria benéfico encontra uma aplicação que possibilitasse um manuseamento e integração eficiente e prática, facilitando todas as alterações que poderiam surgir ao longo de toda a evolução da estrutura de dados, passando-se a trabalhar com o *MySQL Workbench*. Para facilitar a execução deste longo procedimento, exportaram-se após a versão "pilotov0" todos os dados para o *MySQL Workbench*, sendo a versão "pilotov1" e posteriores versões inteiramente criada por esta nova aplicação.

4.7.3.1 Versão Zero

A primeira versão criada de base de dados (pilotov0) concentrou-se na inserção de dados sobre os periódicos/artigos científicos da área de SSI. Esta versão foi criada com base nas entidades e atributos expostos na Figura 4.3. Uma vez que os SGBD não trabalham em documento com o formato ".pdf" de selecionar partes de texto que sejam importantes de analisar e tendo em conta que toda informação dos periódicos estava neste tipo de extensão de documento, toda esta etapa fez aumentar o tempo despendido para a tarefa.

Perante o contexto deste projeto identificar características como o "journal, ano, volume, título, keywords e abstract" de um dado artigo não é uma tarefa automática e por isso todas as componentes relevantes para o trabalho em questão foram copiadas para um documento cujo seu formato era ".txt". A escolha deste formato resulta do facto de um SGBD permitir que seja importado documentos com o formato ".txt". Neste ponto, observou-se que nem todos os "periódicos/artigos científicos" eram constituídos por "keywords" e "abstract" e em situações deste género optou-se por colocar o campo com valor "NULL", sendo representado no SGBD por hífen, "-". Como a tabela "tag" foi criada nesta primeira versão e nesta versão ainda não são atribuídas "tags" aos "papers", a todos os campos da tabela atribuiu-se o valor "NULL", representado por hífen, "-". Todo este processo segue uma estrutura definida e facilmente interpretada pelo código java (secção 4.7) que permite a inserção de informação no SGBD.

A estrutura de inserção dos "periódicos/artigos científicos" nos ficheiros ".txt" segue a nomenclatura identificada de seguida, onde cardinal, "#", faz a separação entre as diversas tabelas.

NomeJournal<TítuloJournal<Ano<Volume<Número<PáginaInício<PáginaFim<Abstract#NomeAutor1<NomeAutor2<...#keyword1<keyword2<...#tag1<tag2<...#

Nesta primeira etapa cada registo de um determinado *paper* é composto por quatro tabelas, "paper", "autor", "keyword", "tag". Para que seja facilmente compreensível a estrutura utilizada nos ficheiros com extensão ".txt", existem quatro "#" que fazem separar as quatro tabelas que representam um dado *paper*, isto é, reflete quando inicia e termina a informação relativa a um dado *paper*. É importante referir que inicialmente apenas foram introduzidas estas quatro tabelas pela importação dos documentos "txt", sendo as restantes tabelas criadas por meio das funcionalidades disponibilizadas do SGBD.

Ao longo desta subsecção serão apresentadas várias tarefas e análises da precisão dos dados introduzidos na base de dados. É sem dúvida um processo bastante prolongado e de elevada relevância já que os dados a trabalhar terão de ser verdadeiramente exatos quando comparados com o documentos original.

Para assegurar que o número de campos inseridos na base de dados se encontrava de acordo com a estrutura definida, ou seja, que não estava a faltar qualquer tipo de campo ou que os dados não tinham sido inseridos no atributo errado na base de dados sobre um dado "paper", recorreu-se a funcionalidades disponibilizadas pela linguagem Java. Com a implementação de instruções Java foi possível a criação de métodos para assegurar previamente nos documentos ".txt" se a introdução e importação de todos os dados de cada *paper* cumpria a estrutura definida: oito campos de valores para identificar o *paper*, campos com dados sobre o nome de autores, campos com dados das "keywords" do paper e campos para dados das "tags", que seriam "NULL" nesta primeira fase. Esta verificação e correção permitiu identificar que alguns *papers* contidos no documento "txt" não estavam totalmente corretos devido a ausência de campos. Também foi possível verificar que para alguns casos, os campos que caracterizavam um "periódico/artigo científico" estavam sendo apontados erros como um sinal de pontuação que era representado pelo sinal gráfico "?" em palavras e/ou ao longo de frases, causando-lhes erros ortográficos. Este erro foi originado devido ao processo de copia de texto que era realizada entre os ficheiros ".pdf" ou "html" para o documento "txt" e onde algumas palavras sofreram pequenas alterações na sua percetibilidade e entendimento. Com as funcionalidades disponibilizadas pelos arquivos ".txt" foi facilmente corrigido este erro, onde foram identificados todos o locais onde o sinal "?" estava presente e corrigida a ortografia de todas as palavras que o continham. A próxima tarefa passou por

limpar os parágrafos repetidos e espaços repetidos entre duas palavras em cada *paper*, analisando o seu "título" e "abstract", assim como verificar a existência de repetição de *papers*. As tarefas mencionadas foram bem-sucedidas, valendo mais uma vez a execução de código Java.

Porém, a introdução correta dos dados relativos a cada paper na base de dados ainda não estava concluída, uma vez que sinais gráficos como por exemplo, " / ", " " ", " ' ", " < " identificados ao longo do texto não eram interpretados corretamente ao carregar a informação para a base de dados. Para ultrapassar este obstáculo e utilizando a linguagem de programação Java, executou-se código que transformava estes sinais gráficos de informação contida no "txt" em conhecimento facilmente interpretado pela base de dados (Ver Tabela 4.4). Após carregar toda a informação para a base de dados e recorrendo a comandos SQL converteram-se novamente os carateres para os sinais originais. Era espectável que toda a informação contida na base de dados representava exatamente o conteúdo de informação de todos os *papers* selecionados e introduzidos.

Sinal Gráfico do Paper	Carater inserido para BD
/	*
"	T1
'	T2
<	T3

Tabela 4.4 - Nomenclatura de conversão de sinais

Para garantir que a informação que era importada para a base de dados estava correta, recorreu-se ao aplicativo de linguagem de programação Java (descrito na subsecção 4.7.1). Foi criado um método que gerasse 100 números aleatórios com intervalo compreendido de 0 até 1000, onde estes números representam a identificação de um *paper* (idpaper). Os 100 números gerados permitiram identificar os *papers*, observar e comparar todos os elementos que o caraterizam, analisando a exatidão dos dados. Desta pequena amostra foram encontrados 28 *papers* cujos seus dados não correspondiam exatamente ao exposto no periódico ou artigo científico. Dos 28 papers, 22 apresentavam erros ortográficos ou falta de espaço entre duas palavras no seu "abstract", três continham o nome do "autor" mal escrito e três foram identificados com o "volume" incorreto. Com esta amostra de 100 *papers*, depreende-se que o número de *papers* com erros aumentará se a amostra passar a ser os 100 *papers*, onde o número de *papers* incorretos poderia passa de 28 para 280. Esta dúvida foi apurada, e por isso, de forma ligeira foram analisados os restantes 900 *papers*,

comparando o conteúdo presente no ".txt" com o formato original desse mesmo *paper*. Pelo resultado desta análise consta-se que ao nível dos "títulos" do *paper* foram encontradas nove palavras com erros ortográficos e sete ocorrências de ausência de espaço entre duas palavras. Pela análise do *abstract* e não esquecendo a dimensão de texto que este pode alcançar, identificaram-se 283 ocorrências de falta de espaçamento entre as palavras e 110 palavras com erros ortográficos. Para concluir a observação, olhou-se à conformidade do "nome dos autores" de todos os *papers* e foram encontrados dois nomes errados. Deste processo, foi possível concluir que 29% dos *papers* apresentam repetição de espaços (TotalEspaços/TotalPapers= 290/1000) e que 7,65% das palavras contêm erro (TotalPalavrasErradas/TotalPalavras= 121/158127) onde 8861 é o total de palavras contidas nos "títulos" dos 1000 *papers* e 149266 é o total de palavras contidas no *abstract* de todos os *papers*.

Após todas estas verificações, foram importados todos os dados exatos contidos no "txt" alusivo a cada journal, denominando esta primeira versão por "pilotov0" e utilizando a aplicação "phpMyAdmin".

4.7.3.2 Versão Um

A versão "pilotov1" é a etapa que sucedeu a versão "pilotov0". Nesta versão foi adicionado o atributo "posicaokey" na tabela "keywords_idarvore" que corresponde à posição que cada "keyword" ocupa num artigo, isto é, representa a ordenação pela qual as *keywords* são apresentadas pelos autores de um "periódico/artigo científico, podendo compreender valores entre 1 e 6. Considera-se que o nível 1 é a melhor posição em que a "keyword" pode estar colocada num *paper*, diminuindo o seu índice de riqueza semântica à medida que a posição da "keyword" progride para o nível 6.

Outro aspeto que permite expor a evolução desta versão é a presença do atributo "calculovalor" identificado na tabela "keywords_idarvore". Este atributo apenas considera as *keywords* que estão relacionadas com a temática da SSI e permite identificar por meio da fórmula: ((1+(1/posicaokey))*indicespecificidade) um primeiro resultado experimental para a seleção da "tag" para os *papers* que compõem a base de dados e que apresentam o campo da SSI. Nesta primeira fase de seleção da "tag" reconhece-se que o resultado máximo do "calculovalor" retornado via fórmula de um dado *paper* será a "tag" que irá identificar esse mesmo paper. Para avaliar o cálculo do atributo "calculovalor" fez-se uma *query* que retorna valores de atributos como: "idkeyword, idkeywordniveis, keyword, posicaokey, indicespecificidade". Após o resultado deste

cálculo foram importados para a base de dados todos os atributos e respetivos valores, complementando deste modo a entidade "keywords_idarvore". O atributo "indicespecifidade" pertence à entidade "arvorekeywords" já presente na versão "pilotov0".

Para além destes dois atributos, ainda nesta versão, foram consideradas as mesmas entidades e atributos que compuseram a versão "pilotov0". Na Figura 4.4 é possível visualizar uma parte da estrutura da versão "pilotov1".

4.7.3.3 Versão Dois

Um olhar atento sobre o desenvolvimento da versão anterior delineia que esta ainda era bastante simples e que ainda se fazia apontar a ausência de informação útil para concluir-se resultados mais ricos. Com o propósito de atingir resultados viáveis e assertivos, foi criada uma nova versão, designada por "pilotov2". Esta versão é composta por todas as entidades e atributos que caraterizavam as versões anteriores, com a adição de novos componentes de análise.

Para a presente versão é de salientar a criação de uma nova tabela, a "classekeywords", composta por todas as *keywords* pertencentes ao domínio da SSI com base nos documentos recolhidos, totalizando 1239 *keywords*, organizadas segundo as classes temáticas de SSI, e definidas inicialmente aquando a criação do esquema da "arvorekeywords". Os atributos que a compõem vão desde o "id" que é o identificador único da tabela, "idclasse" que representa a classe a que determinada "keyword" pertence na "arvorekeywords", o "nível" que representa o "indicespecificidade" da "keyword", o "idkeyword" que estabelece a ligação com a tabela "arvorekeywords", a "keyword", e o "idpainivel" que permite identificar perante os seis níveis da "arvorekeywords" e mediante cada keyword qual é o seu antecessor.

Após a criação desta versão e observando os resultados obtidos pôde-se concluir que existem imensas *keywords* que pertencem ao mesmo pai. Com as características da presente versão de base de dados ainda não era suficiente para aclarar a "tag" que irá representar um determinado "artigo/periódico científico". Para responder a este desafio, a próxima subsecção é dedicada à descrição de uma versão posterior à "pilotov2" e que comtemplará todas as evoluções e resultados alcançados.

4.7.3.4 Versão Três

A versão três, designada por "pilotov3", foi a última versão a ser desenvolvida e é a que apresenta maior contributo para a solução do projeto em causa. Esta versão final apresenta todas as características das versões "pilotov0", "pilotov1" e "pilotov2", com a adição de todas as entidades e atributos considerados relevantes e que criam a "pilotov3". A Figura 4.4 apresenta toda a estrutura de dados final, seguindo-se a explicação de todos os elementos. Serão analisadas vinte e duas entidades, não esquecendo que nove destas entidades são já adquiridas da versão "pilotov2", sofrendo apenas alterações pontuais em algumas delas.

Tal como já mencionado anteriormente, a base de dados é composta por 1000 *papers*, dos quais 880 contêm *keywords* e os restantes 120 não exibem nenhuma keyword. Perante este facto, foi imprescindível tratar a informação sob duas formas: *papers* com *keywords* e *papers* sem *keywords*, assim como criar entidades específicas.

Iniciando a análise da estrutura final da base de dados pode-se dividir esta análise em duas partes. A primeira parte referente às entidades já apresentadas nas versões anteriores e a segunda parte com exposição para as novas entidades e atributos.

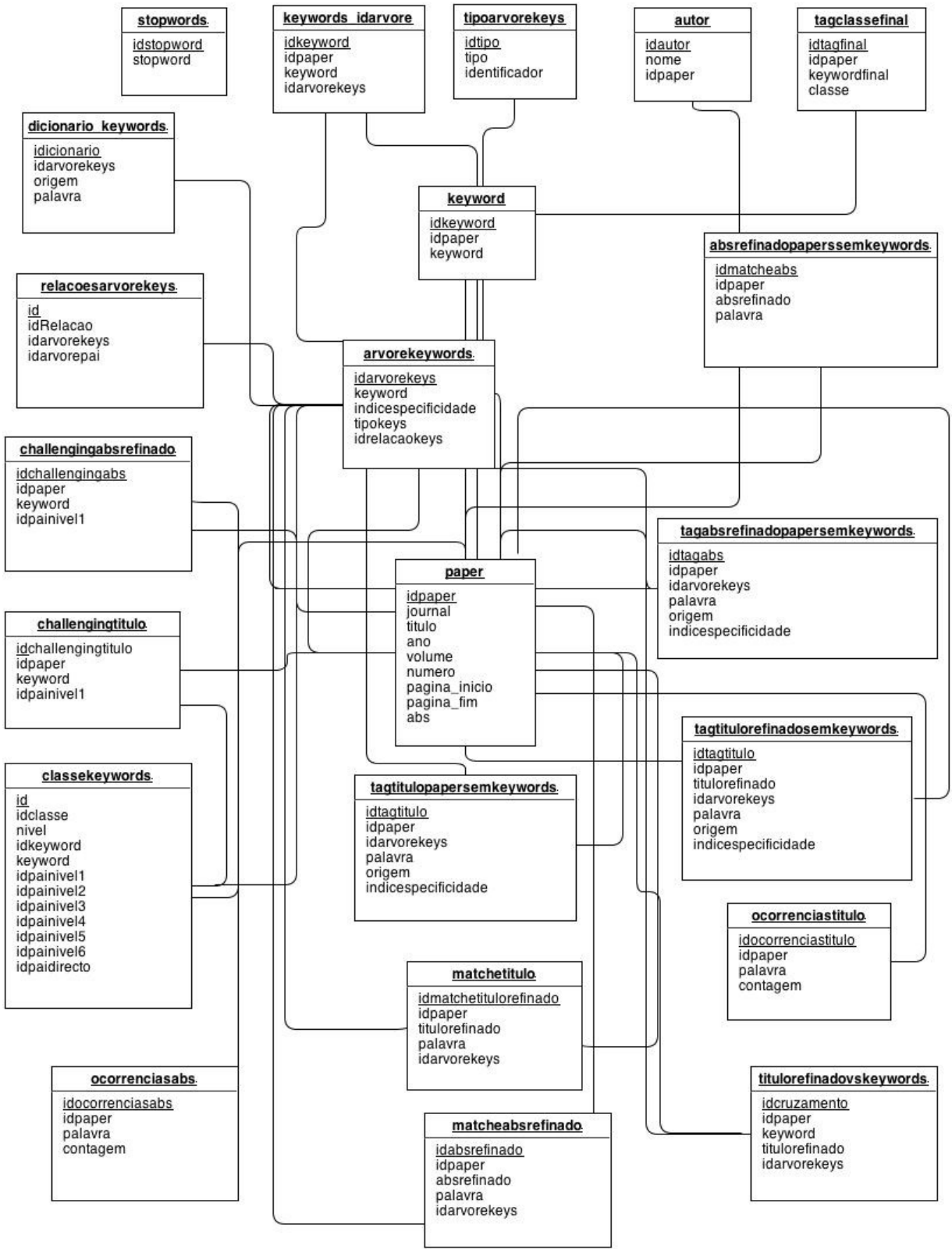


Figura 4.4: Estrutura da Base de Dados - "pilotov3"

Em relação à primeira parte e começando com a entidade "paper", é pretendido que esta retrate a informação pertinente de um determinado artigo/periódico científico", sendo composto pelos atributos: "idpaper" – identificador único da tabela, "journal" – nome do periódico publicado, "titulo" – título/assunto do "paper", "ano" - ano de publicação, "volume" – secção da publicação, "numero" – número do "paper" publicado em determinado volume, "pagina_inicio", "pagina_fim", "abs" (abstract) – breve resumo do "paper", "titulorefinado" – é o título do "paper" com a eliminação das "stopwords", e "absrefinado" – é o "abstract" do "paper" com a aplicação de uma "query" que removeu todas as "stopwords". Prossegue-se para a entidade "arvorekeywords", composta pelo "idarvorekeys" – id da entidade, a "keyword" – palavra-chave do "paper", o "indicespecificidade" – posição que a "keyword" se encontra na árvore, o "tipokeys" – determina se a keyword pertence ao âmbito da SSI e a "idrelacaokeys" – identificador que permite estabelecer a relação entre as diversas *keywords*. A próxima entidade, "keyword", mostra todas as *keywords* (pertencentes ou não ao âmbito de SSI) que estão presentes num dado "paper", onde "idkeyword", "idpaper" e "keyword" são os atributos que compõem esta entidade. Como em qualquer tipo de "paper" existe o autor que o concebeu, sendo a entidade "autor" composta pelos atributos "idautor" – identificador da tabela, "nome" – nome do autor/autores e "idpaper" – identificador que permite estabelecer a relação com a entidade "paper". A entidade "tipoarvorekeys" composta pelos atributos "idtipo" – identificador da entidade, "tipo" – representa os três tipos de *keywords* que podem existir num determinado "paper" e "identificador" – que aponta os três tipos de *keywords* que podem ser identificadas (security – *keywords* da literatura de SSI, non-security – *keywords* genéricas e em nada referência a SSI, *void* – *keywords* genéricas, como por exemplo, "*security*"). Com o objetivo de focar apenas as *keywords* que pertencem à SSI criou-se a entidade "keywords_idarvore" que apresenta todas as "keywords" que definem o assunto de determinado paper, mas atributos como "idarvorekeys", "posicaokey" (posição ocupada pela keyword na sequência de keywords apontadas no paper) e "calculovalor" (valor obtido pela fórmula já identificada na versão "pilotov1") apenas são trabalhados para as "keywords" do tipo *security*, e atributos como o "idkeyword", "idpaper", "keyword" são os atributos comuns a ambos os tipos de *keywords*. A entidade "relacoesarvorekeys" permite estabelecer a relação entre as diversas *keywords* de SSI identificadas ao longo de todos os papers e que estão presentes na entidade "arvorekeywords". Nesta fase é estabelecida uma relação entre "keyword-pai" e "keyword-filho". Esta relação é baseada no pormenor e abrangência de cada keyword, podendo incluir outras keyword mais especificas dentro da área a que se refere. O "id", "idrelacao", "idarvorekeys" e

"idarvorepai" são os atributos da entidade "relacoesarvorekeys". Para a entidade "classekeywords" com os atributos "id", "idclasse", "nível", "idkeyword", "keyword", "idpainivel1", "idpainivel2", "idpainivel3", "idpainivel4", "idpainivel5", "idpainivel6", "idpainiveldirecto" pretende-se apresentar para cada keyword de SSI a classe que a inclui, olhando-a perante os diferentes níveis que se encontra pela "arvorekeyords". Para concluir a primeira parte da apresentação da estrutura da base de dados, falta referir a entidade "tagclassefinal" que inicialmente era designada por "tag" e nesta versão final, "pilotov3", foi renomeada para melhor caracterizar os atributos da entidade. É composta pelos atributos "idtagfinal", "idpaper", "keywordfinal" (tag(s) que caracteriza o *paper*), "classe" (classe de topo-família do nível a que pertence a *tag*) e apresenta para todos os 1000 *papers* a respetiva *tag* e classe a que pertence a *tag*.

Para a descrição da segunda parte da estrutura da base de dados, são apontadas 12 entidades. Começando pela entidade "stopwords" - entidade auxiliar que facilitou todo o processo de remoção de palavras dispensáveis ao nível do atributo "titulo" e "abs" de cada paper, sendo composta pelos atributos "idstopword" e "stopword". Em anexo pode ser visualizada toda a lista de "stopwords" que foi incluída nesta entidade. Esta lista surge de uma pesquisa efetuada via Internet, que após a análise de vários resultados foi selecionada a lista *default* proposta pela equipa do "MySQL". Porém, ao analisar esta lista concluiu-se que seria útil acrescentar mais exemplos de palavras que poderiam mais tarde ser trabalhadas nos atributos do "titulo" e "abs" dos *papers*, assim como excluir exemplos de palavras que estavam contidas na lista (totalizando 47 palavras), porque iriam influenciar negativamente a sua aplicabilidade uma vez que algumas *keywords* seriam eliminadas quando se cruzava a informação. Para introduzir automaticamente a lista de "stopwords" na base de dados foi implementado um método em Java. A entidade "dicionário_keywords" constituída por todas as *keywords* do âmbito da SSI encontradas nos "artigos científicos" e que podem ser encontradas nas diversas categorias gramaticais. Esta entidade pretende responder a todas as eventuais formas de exibição da *keywords* num *paper*, quer esta se encontre no singular, plural, com e sem hífen, entre outros situações. A entidade é composta pelo "idicionario", "idarvorekeys", "origem" – atributo que refere se a keyword tem origem no *paper* ou se a keyword foi acrescentada à lista do dicionário identificando-se por 1 e 2 respetivamente, e a "palavra" – identifica a keyword.

Para expor a perspetiva dos *papers* que apresentavam keywords como um dos elementos que caracterizava o *paper*, foram criadas sete entidades. A entidade "matchetitulo" composta pelo "idmatchetitulorefinado" (identificador único da tabela), "idpaper" (identificador do "paper"),

"titulorefinado" (título com a remoção de "stopwords"), "palavra" (keyword de SSI encontrada ao longo do título), "idarvorekeys" (identificador único na entidade "arvorekeywords" e que associa a keyword encontra as *keywords* presentes na árvore). O objetivo desta entidade passa pelo cruzamento dos títulos, já sem *stopwords*, com as *keywords* exibidas em toda a tabela "arvorekeywords"/"dicionario_keywords" para averiguar a semelhança das palavras e eleger uma tag. A próxima entidade "matcheabs", espelha o *abstract* de cada *paper* já sem "stopwords" e cruza-o com as *keywords* que compõem a tabela "arvorekeywords"/dicionário_keywords. É uma entidade composta pelo "idabsrefinado" (identificador único da tabela), "idpaper" (identificador do paper), "absrefinado" (abstract com a remoção das stopwords), "palavra" (keywords de SSI encontradas ao longo do absrefinado), e "idarvorekeys" (identificador da keyword para estabelecer a relação com a entidade "arvorekeywords"/"dicionário_keywords"). A entidade "challengingtitulo" é o resultado em keyword do cruzamento entre a entidade "matchetitulo" e o máximo valor do atributo "calculovalor" de cada paper na entidade "keywords_idarvore", ou seja, é o cruzamento entre o "matchetitulo" de cada "paper" e a "tag" eleita por cada "paper" via fórmula (versão "pilotov1"). Entidade é composta pelo "idchallengingtitulo" (identificador único da tabela), "idpaper" (identificador do "paper"), "keyword" (keyword resultante do atual cruzamento), e "idpainivel1" (identificador de raiz da atual keyword). Na mesma linha de trabalho encontra-se a entidade "challengingabs", que foca o "abstract" de cada paper e não o título, como é o caso da entidade descrita anteriormente. O "idchallengingabs" (identificador único da tabela), "idaper" (identificador da tabela), "keyword" (keyword resultante do atual cruzamento), "idpainivel1" (identificador de raiz da atual keyword) são atributos que compõem a entidade "challengingabsrefinado".

Para aperfeiçoar a seleção da "tag" final ainda foi ponderado o número de vezes em que ao longo do "título" e "abstract" cada palavra ia aparecendo, ou seja, separando palavra a palavra de uma determinada frase, quantas vezes ela está presente ao longo de todo o texto. Para responder a este desafio, adaptou-se a estrutura da base de dados adicionando-se a entidade "ocorrenciastitulo" e entidade "ocorrenciasabs". Os seus atributos são o "idocorrenciastitulo" (identificador único da tabela referente ao titulo do *paper*), "idpaper" (identificador do paper), "palavra" (palavra encontrada ao longo do título), "contagem" (número total de vezes que a palavra aparece ao longo do título) e "idocorrenciasabs" (identificador único da tabela referente ao *abstract*), "idpaper" (identificador do *paper*), "palavra" (palavra encontrada ao longo do *abstract*) e "contagem" (número total de vezes que a palavra aparece ao longo do *abstract*), respetivamente.

Para terminar, surge uma segunda perspetiva de análise aprimorada – observar os 120 *papers* que não apresentam na sua estrutura *keywords* e que foram designados por "paperssemkeywords". A decisão de separar e tratar os dados de modo distinto nos dois tipos de papers é de elevada importância, uma vez que a informação e as experiências são desenvolvidas de forma distinta. Para examinar resultados válidos para os "paperssemkeywords" foram criadas duas entidades: a "tagtitulorefinadosemkeywords" e a "absrefinadopaperssemkeywords" que retratam a informação ao nível do título e *abstract* de cada *paper*. A entidade "tagtitulorefinadosemkeywords" surge do cruzamento entre o atributo "titulo" de cada um dos 120 *papers* e de todas as "keywords" que se encontram na tabela "dicionário_keywords", fazendo referência às "keywords" que sejam comuns. É uma entidade composta pelo "idtagtitulo" (identificador único da tabela), "idpaper" (identificador do paper), "titulorefinado" (título com a remoção de stopwords), "idarvorekeys", "palavra" (keyword encontrada pelo cruzamento), "origem" (identifica se a keyword é a original presente na tabela "arvorekeywords" ou se é uma keyword adicionada ao "dicionário_keywords" olhando à sua exibição no singular e plural), "indicespecificidade" (nível em que a keyword se encontra na "arvorekeywords"). Por fim, a entidade "absrefinadopaperssemkeywords" deriva do cruzamento entre o atributo "absrefinado" e a entidade "dicionário_palavras" resultando deste cruzamento as keywords que sejam comuns. Esta entidade tem como atributos o "idmatcheabs"(identificador da tabela), o "idpaper" (identificador do "paper"), o "absrefinado" ("abstract" com a remoção de "stopwords") e a "palavra" ("keyword" resultante do cruzamento).

Terminada a descrição da estrutura de base de dados, a próxima secção retrata o processo realizado para agrupar as *keywords* resultantes das opções tomadas sobre o conteúdo da base de dados.

4.7.4 Atribuição de Classes

O dia-a-dia do ser humano é um constante confronto com informação. Eventualmente seria difícil gerir e organizar toda a informação se não existissem regras que a norteassem como um meio de expressão e entendimento entre os seres humanos. Esta premissa permite concluir que a conceção de regras orais e escritas para a linguagem é o meio essencial para o bom funcionamento e entendimento social. Neste sentido, a presente secção vem expor e descrever o trabalho realizado na obtenção de uma estrutura real de conceitos que se inter-relacionam, sendo a base da estrutura apenas as *keywords* retiradas dos artigos científicos explorados ao longo deste projeto e com foco na temática da SSI. Todas as *keywords* exploradas foram agrupadas de acordo com a sua

semelhança temática, designando-se este agrupamento por classes de palavras. É importante referir que a base desta etapa prestou-se dos conhecimentos obtidos pela informação explorada na Secção 2.5, possibilitando uma perceção clara da distribuição dos conceitos pelas várias temáticas de investigação.

Para realizar esta tarefa começou-se por selecionar o corpo de conhecimento da SSI que melhor se enquadrava na lista de *keywords* de SSI obtida pela exploração dos artigos científicos selecionados. A tentativa deste enquadramento não foi bem-sucedida uma vez que, as *keywords* identificadas pertenciam a diversos contextos e por isso não podiam ser fragmentadas por apenas 10 temáticas genéricas para assim conduzir à perda de significado. As 10 temáticas aqui referidas correspondem aos 10 dominós definidos na certificação de CISSP. Considerou-se esta certificação como a base para o desenvolvimento do esquema de classificação uma vez que, a sua abrangência temática é consideravelmente bem estruturada. O passo seguinte, um pouco mais demorado mas também maís assertivo, consistiu em analisar as *keywords* que se relacionavam entre si e criar pequenos grupos dessas mesmas *keywords*. Ao observar estes pequenos grupos concluiu-se que era possível existir pelo menos uma keyword que identificava o grupo e a área de investigação que retratava. Neste momento existia uma keyword global e um conjunto de "keywords filhas" que se inter-relacionavam. A próxima tarefa consistia em pegar nas "keywords filhas" e reorganizá-las. Daqui resultou uma estrutura hierárquica composta pelas diferentes *kywords* de SSI e de acordo com a classe a que pertencem, não esquecendo o seu índice de riqueza semântica. A disposição das *keywords* mostra-as distribuídas por 85 áreas de investigação e podem estar localizadas desde a raiz até um nível mais detalhado (intervalo compreendido entre 1 a 6).

A Figura 4.5 mostra um pequeno extrato de uma classe composta pela "keyword pai" (*Secuirty Principles*) e pelas que a ela lhe estão associadas.

Security Principles	*Integrity*	*Information integrity*
		Storage integrity
		File integrity
		systemic integrity
	Confidentiality	*Inferential Disclosure*
		Information disclosure
		Disclosure inference
	Availability	
	Non repudiation *Non-repudiation* *Nonrepudiation*	*Fair non-repudiation*

Figura 4.5: Estrutura da Classe *Security Principles*

Cada coluna da figura representa um nível em que quando maior o nível em que a keyword está localizada, maior será o seu nível de detalhe. Para uma visão global de toda a estrutura criada deve ser consultado o Apêndice C.

A classe *Security Principles* é composta apenas por três níveis, mas o nível de detalhe de outras classe poderá estar compreendido entre o intervalo de 1 até 6. Nesta estrutura (Apêndice E) é visível que as grandes áreas de investigação que a revisão sugere foram levadas em consideração, melhorando a expressividade deste modelo de *keywords*.

A partir deste momento existia uma configuração na base de dados designada por "arvore de keywords" que funcionou como alicerce das próximas etapas para identificação da melhor *tag* para definir o *paper* estudado no momento.

4.8 Identificação da "*Tag*"

Para terminar este capítulo de descrição do estudo, a presente secção atravessa todo o conjunto de atividades já descritas como um meio para atingir e descrever o objetivo do trabalho de dissertação. Perspetivando todas as tarefas realizadas nas secções anteriores, esta secção representa, através de várias fases as possíveis conclusões para eleição da tag final que identifica a temática de um paper. O processo de seleção da tag é descrito através da caraterização de casos de exploração que apresentam todo o procedimento conduzido para eleger a tag de acordo os requisitos que estão a ser considerados no caso. Para a análise de dados recorreu-se a funcionalidades disponibilizadas pelo SQL e Excel.

Uma vez que os dados iniciais foram afinados com o auxílio do Programa Java e a informação facultada com recurso a uma base de dados, foi possível obter resultados com boas opções para identificar a *tag* final de um dado *paper*.

Caso 1 # Exploração da *tag* segundo o índice de riqueza semântica

A primeira fase de análise foi consideravelmente vasta, dado que o objetivo era identificar para cada uma das *keywords* que compunham um dado *paper* o seu "indicespecificidade", termo utilizado para descrever o "índice de riqueza semântica" da keyword. Este valor é alcançado de acordo com a disposição que a keyword apresenta na "arvorekeywords" (conteúdo já descrito anteriormente).

O valor que é concedido ao atributo "indicespecificidade" compreende uma escala entre 1 e 6, que representa os vários níveis em que a keyword pode estar colocada na entidade "arvorekeywords". A posição ocupada apresenta: 1 corresponde a um índice de riqueza semântica "muito fraco", 2 ao valor "fraco", 3 "médio", 4 "médio-bom", 5 "bom" e 6 a um índice de riqueza semântica "muito bom". Para associar às keywords o valor do "indicespecificidade" foi criada uma *query* que cruzava as "keywords" do tipo "security" presente em cada *paper* com as *keywords* da entidade "arvorekeywords" sendo o seu *output* o valor que correspondia ao índice de riqueza semântica da keyword em causa.

Pela análise de uma pequena amostra foi possível observar que não seria eficaz escolher a *tag* de um dado *paper* segundo o seu índice de riqueza semântica, isto porque em muito casos apesar do índice de riqueza semântica ser "muito bom" não era assinalado pelos autores como a primeira keyword que identificava o *paper*, da mesma forma que esta podia não ser a única keyword do tipo *security* associada ao *paper* em questão.

Caso 2 # Exploração da *tag* segundo o cálculo entre valores de atributos

No caso anterior, Caso 1, a análise de dados efetuou-se sobre toda a informação que compunha a base de dados, trabalhando assim sobre os 1000 *papers*. Para o presente Caso 2 teve-se como objetivo principal restringir a quantidade de informação que era tratada. Para isso pensou-se numa forma possível de começar a trabalhar com menos informação, concluindo-se que cruzando informação de várias fontes de atributos seria o primeiro passo para tal limitação.

Neste sentido, inicialmente, calculou-se recorrendo a uma fórmula testada por meio de uma pequena amostra, qual seria a *tag* mais apropriada para um *paper* com foco no "indicespecificidade" e "posicaokey", uma vez que os valores destes atributos são muito assertivos e precisos. O resultado devolvido por esta fórmula era assertivo, passando-se a trabalhar com todos os valores destes dois atributos na fórmula criada.

A fórmula utilizada foi: (1+(1/posicaokey))*indicespecificidade. Cada um destes atributos já estão descritos na secção anterior (Secção 4.8). O resultado devolvido segundo a aplicação da fórmula é bastante íntegro, uma vez que é o resultado máximo e ótimo entre todas as *keywords* da SSI que o *paper* identifica. Pela aplicação da fórmula, foi devolvido resultado para 880 *papers* e para 120 *papers* não lhes foi atribuída qualquer valor pois não apresentam qualquer tipo de *keywords* do tipo *security*, não sendo possível aplicar a fórmula neste pequeno conjunto de *papers*.

No caso dos 880 *papers*, verificou-se que para alguns casos foi-lhes atribuída mais do que uma *tag*. No Apêndice F é possível ver o resultado da aplicação da fórmula.

Caso 3 # Exploração da *tag* segundo o cruzamento de valores

Para o Caso 3 voltou-se a trabalhar novamente com os 1000 *papers* que compunham a base de dados. A este processo designou-se "noção de matches" uma vez que nesta etapa era objetivo cruzar o texto contido no atributo "titulo" e "abstract" de cada *paper*, separadamente, com todas as *keywords* que constituíam a entidade "arvorekeywords". O resultado final passou por encontrar as *keywords* semelhantes entre estes dois elementos cruzados e compreender quais poderiam ser eleitas a ou as *tags* vencedoras que mais tarde classificavam determinado *paper*. Neste momento, o conteúdo de cada atributo "titulo" e "abstract" de cada *paper* já está submetido à triagem da entidade "stopwords", ficando apto para comparar as diversas *keywords* da "arvorekeywords" com as *keywords* semelhantes dos títulos, tendo sempre em conta os singulares/plurais das *keywords*, assim como abreviaturas e palavras com/sem hífens. Da aplicabilidade deste caso resultaram *keywords* para *papers* que anteriormente não tinham sido identificadas pelos casos um e dois.

Caso 4 # Exploração da *tag* segundo os casos anteriores

O presente caso versa sobre os resultados obtidos nos casos anteriormente descritos. A este caso atribuiu-se a designação "noção de *challenging*", cujo resultado foi proveniente do cruzamento entre o resultado da "noção de matching" ao nível do "titulo" e "abstract", Caso 3, com o resultado proveniente da identificação de uma tag viável segundo o seu maior índice de riqueza semântica, Caso 1. Do desafio entre os dois casos apenas resultaram valores que correspondiam aos *papers* que apresentam *keywords* do tipo *security*, uma vez que o Caso 1 apenas explora os *papers* que são compostos por *keywords* do tipo *security*.

A segunda parte da análise deste Caso 4 apoia-se no resultado obtido pela exploração do Caso 3, "noção de matches", com o resultado obtido pela aplicação do Caso 2, aplicação do cálculo entre valores de atributos.

Este desafio auxiliou na verificação de semelhança entre o resultado obtido pela primeira experiência (caso 1) e o resultado retornado pela segunda exploração (caso 2), sendo possível uma melhor precisão na eleição da(as) *tags* que melhor identificam um *paper*. Esta análise permite

identificar uma maior robustez para as *tags* resultantes deste cruzamento uma vez que cada um dos casos já está a comtemplar outras definições de requisitos de identificação da *tag*.

Caso 5# Exploração da *tag* para *papers* sem *keywords*

De todo o conjunto de 1000 *papers* que compunham a base de dados, 120 *papers* não eram compostos por qualquer tipo de keyword. Para este conjunto foram adotadas outras formas de eleger a tag de modo a que estes *papers* também fossem classificados.

Inicialmente, começou-se por identificar no "titulo" as keywords que estavam presentes na estrutura da "arvorekeywords". Do resultado desta exploração identificaram-se como tags as *keywords* que apresentavam maior "indicespecificidade" no "titulo". Para complementar a escolha da tag verificou-se se esta também estava presente no "abstract". Se estes dois casos eram verificados, adotou-se a "tag" reconhecida pelo cruzamento para identificar o *paper*. Caso não se verificasse semelhança entre as *keywords* seria identificada a que apresentava maior "indiceespecificidade".

No Caso 3, exploração da *tag* segundo o cruzamento de valores – "noção de matches", foi o caso em que conjunto foi explorado juntamente com os restantes "papers".

A terceira tentativa passou por pegar no resultado obtido pela "noção de matches" ao nível do "titulo" e cruzar este resultado com o resultado obtido pela "noção de matches" ao nível do "abstract". O valor retornado destes dois cruzamentos permitiu ver as keywords que eram semelhantes e atribuir uma possível tag aos papers da amostra. Para as keywords resultantes deste cruzamento foi identificado o seu valor de "indicespecificidade", onde a tag passou a ser o valor mais alto identificado no paper em questão.

Para os papers que não foram identificadas qualquer tipo de keyword por meio dos métodos descritos anteriormente, optou-se por escolher a keyword do "titulo" ou "abstract" que apresentava melhor valor para o conceito de "indicespecificidade" da "arvorekeywords" juntamente com o número de vezes que a mesma aparecia no texto. Como esta solução é de reduzida certeza e ainda foram encontrados 39 papers em que as tags identificadas apresentavam o mesmo índice de especificidade, foi necessário ler o "titulo" e "abstract" para eleger a melhor "tag".

Capítulo 5 – Análise dos Resultados

5.1 Introdução

No capítulo anterior foram descritas todas as etapas intermédias julgadas essenciais para obter resposta à questão de investigação enunciada.

O objetivo deste capítulo é, precisamente, proceder à apresentação dos resultados da análise dos contributos recolhidos após a realização das etapas do capítulo anterior e, consequentemente, obter as respostas finais para a questão de investigação. Estruturalmente, o capítulo encontra-se organizado em duas secções. Após esta introdução, apresenta-se na Secção 5.2 a análise dos resultados. Esta análise é observada em duas perspetivas, os resultados obtidos com base em keyword dos *papers* e os resultados obtidos sem recurso a *keyword*. Na última secção resumem-se as principais contribuições deste capítulo.

5.2 Análise

Os resultados obtidos nas etapas descritas nas secções anteriores serão analisados de seguida. De todos os *papers* recolhidos, serão apresentados resultados de 1000 publicações (Gráfico 5.1), das quais 120 (22%) não possuem qualquer lista de *keywords* inicial, sendo os restantes 880 (78%) *papers* os que apresentam keyword definidas pelos autores.

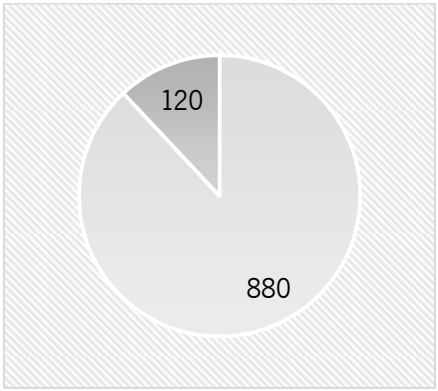


Gráfico 5.1: Distribuição *Papers* com/sem *Keywords*

Uma vez que a análise é focada na literatura da SSI para um período temporal bem definido, e para toda a amostra inicialmente analisada, apenas foram identificados 1000 *papers* que respondiam aos requisitos pretendidos, o gráfico 5.2 apresenta para cada ano o número de *papers* que foram analisados. É importante recordar que no decorrer do processo de seleção de fontes para a elaboração do projeto descartaram-se deliberadamente todos os *papers* que no seu "título" ou "abstract" não apresentavam qualquer tipo de keyword que pudesse estar relacionada com a área ou que a sua análise em termos da frequência da keyword é reduzida apenas ao aparecimento no "título". Não é surpreendente o facto de inicialmente serem analisados imensos artigos, uma vez que foram explorados 12 *journals* e o resultado final apenas ser restringido a 1000 *papers*. Está técnica de seleção restringiu consideravelmente a quantidade de *papers* para preencher a base de dados, mas por outro lado, proporcionou um resultado bastante minucioso.

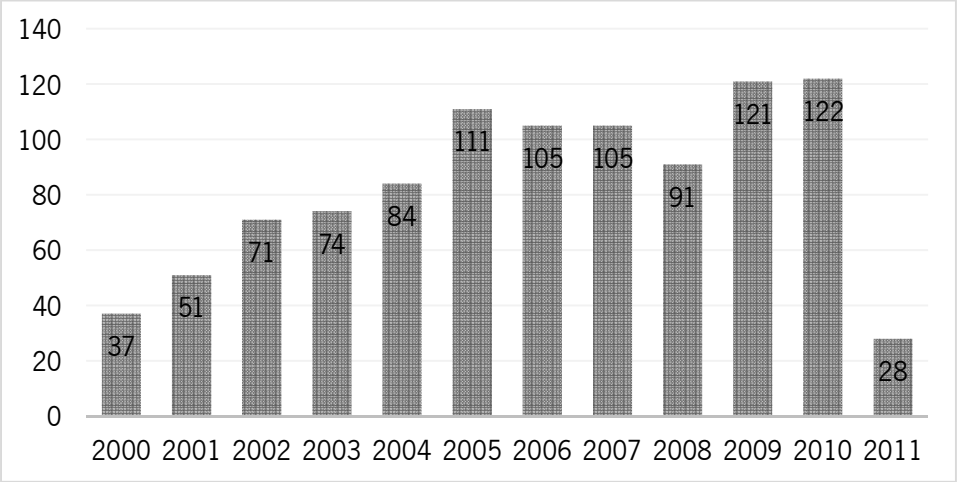


Gráfico 5.2: Número Total de *Papers* Distribuídos pelo Ano de Publicação

O Gráfico 5.2 mostra a visão global de como a área de SSI tem evoluído durante o período de 2000 e início de 2011. Poder-se-á assumir que o número de publicações aumenta com o desenvolvimento da própria área. É de notar que em 2005 a temática de SSI começa a ter maior número de "artigos científicos", mas é a partir de 2009 que se faz notar o aumento de publicações na área. É importante referir que o número de publicações para 2011 é reduzido porque apenas foram analisadas as publicações até Maio e apenas para os *journals* que já tinham publicado até a data.

O Gráfico 5.3 ilustra para cada um dos 14 *journals* identificados, o número de *papers* que foram analisados. Tal como em análises anteriores, nesta também é possível verificar que o

"Journal Computers & Security" é o que mais se distingue, procedendo-se o "International Journal of Information Security" e assim por diante, sendo o "Information Systems Journal" o que menor percentagem apresenta para análise de informação.

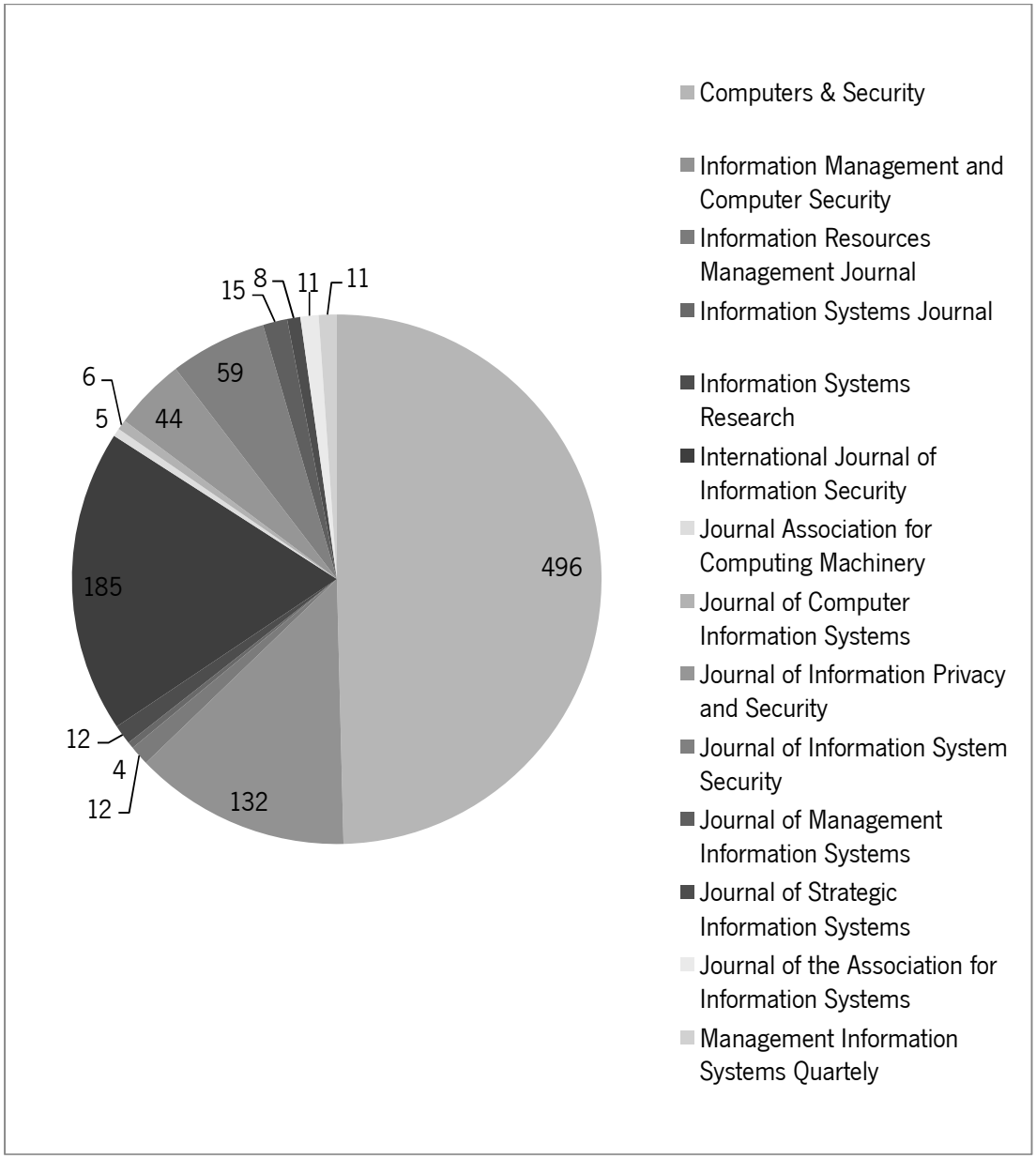


Gráfico 5.3: Distribuição do Número de *Papers* Explorados

Sem dúvida que a área da SSI começa a ter elevada importância para os investigadores e sociedade em geral uma vez que o meio informático está constantemente a ser fragilizado com os atos da sociedade acarretando graves problemas pessoais, organizacionais e sociais.

De seguida são apresentados os resultados da exploração dos 1000 *papers*. Os resultados são divididos por dois casos: *papers* que eram compostos por *keywords* e *papers* que não eram constituídos por *keywords*. Esta separação resulta das razões já apresentadas na Secção 4.10, isto porque, o tratamento de dados realizou-se em práticas diferentes.

Caso #1 – Papers com Keywords

Nesta parte serão analisados os resultados mais relevantes que se conseguiram apurar para os *papers* que apresentavam *keywords* de SSI como um elemento participante do corpo do artigo, totalizando 880 *papers*. Será possível observar diversas análises e distintas interpretações, sendo que na secção de "Apêndices" também serão apresentados alguns resultados mais extensos.

Número de *Keywords* Totais vs. *Papers*

Uma vez que existem 880 *papers* com *keywords*, considera-se importante averiguar de acordo com o número de *keywords* que podem compor um *paper*, qual é a distribuição do número de *papers* pelo número total de *keyword*. O gráfico seguinte ilustra o resultado desta análise, mas é importante referir que a análise total de *keywords* é sobre as *keywords* que pertencem e não pertencem à temática da SSI.

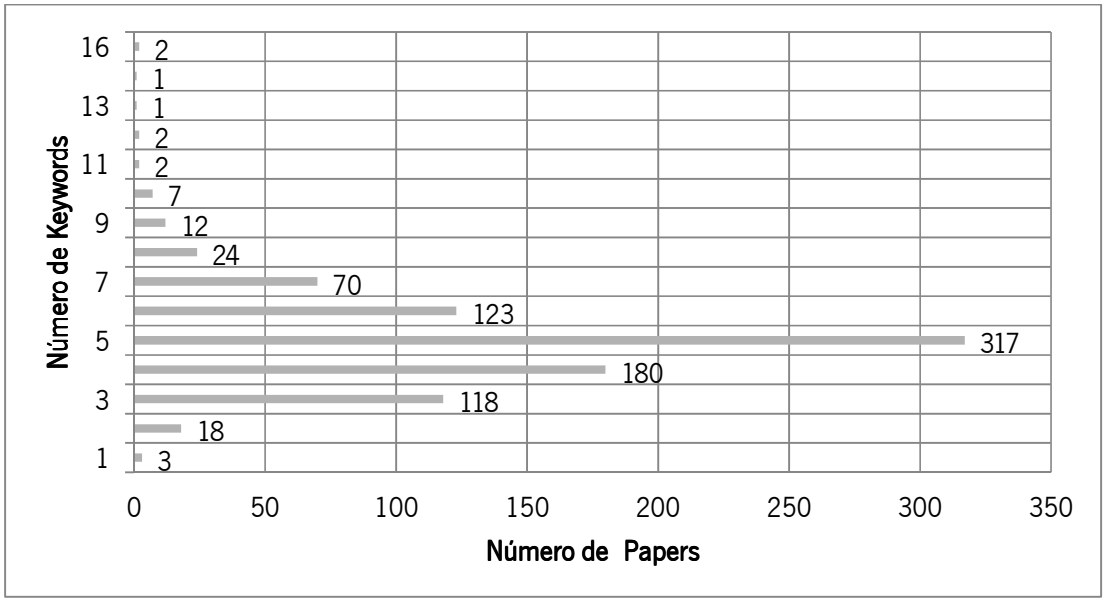


Gráfico 5.4: Número de *Papers* vs. Número de *Keywords*

Da visualização do gráfico compreende-se que grande parte dos papers são constituídos por cinco *keywords*. Ainda segundo esta análise é possível concluir que existem *papers* que podem ser compostos por 16 *keywords*, que nenhum *paper* é composto por 15 *keywords* e que em média, cada *paper* possuí entre três a seis *keywords* para o tipificar.

Número de *Keywords* vs. *Papers*

A próxima análise é de elevada relevância uma vez que transmite em termos globais o número de *keywords* de SSI pelo número total de *papers*, isto é, permite perceber quantos *papers* possuem x número de *keywords* (por exemplo, existem 180 *papers* com 1 keyword).

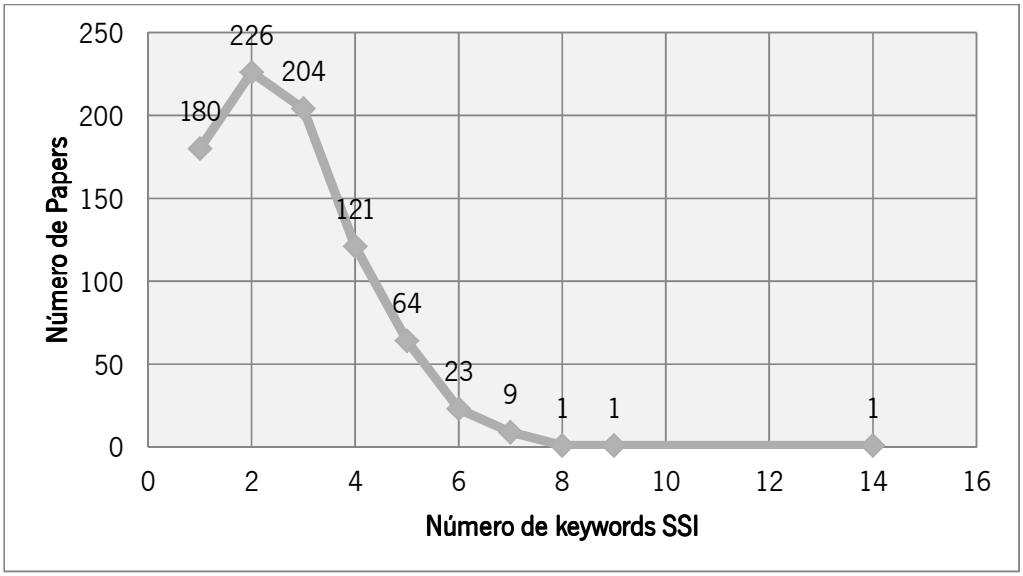


Gráfico 5.5: Número de *Keywords* de SSI vs. Número de *Papers*

Ao observar o Gráfico 5.5 conclui-se que grande parte dos *papers* pode ser caracterizados por 2, 3, 1, ou 4 *keywords*, tornando a sua categorização mais eficiente. É também notável que há medida que o número de *keywords* aumenta, existe uma redução acentuada do número de *papers* e existe um único *paper* que possuí na sua lista 14 *keywords* do tipo *security*.

Número de *Keywords* vs. Posição

Depois de todo o trabalho concretizado para excluir as *keywords* que não pertencem à temática de SSI, avaliou-se o impacto que cada keyword de SSI podia representar de acordo com a sua posição no *paper*. Esta análise pressupõe que os autores dos *papers* dispõem as *keywords* de

acordo com o seu grau de relevância para o *paper*. O gráfico seguinte mostra para a amostra de 880 *papers*, a existência de *papers* que podem possuir 16 *keywords* para o classificar. Estas *keywords* apenas retratam a temática da SSI e desta análise relaciona-se a posição com que aparecem na lista de *keywords* com o número de *keywords* em cada uma das posições.

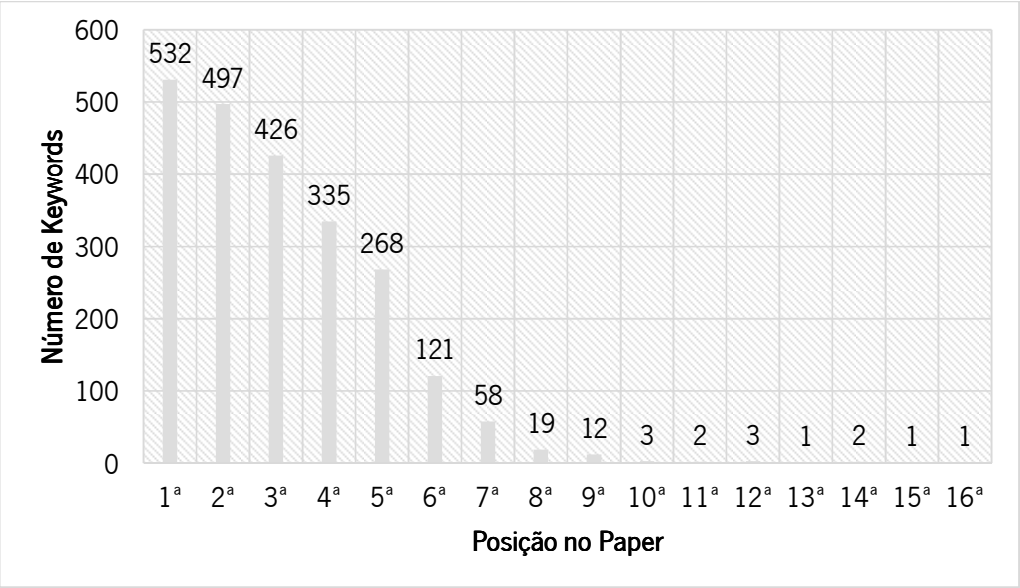


Gráfico 5.6: Número de *Keywords* vs. Posição no *Paper*

Pela observação, comprova-se que efetivamente os autores colocam as *keywords* mais expressivas nas primeiras posições da lista de *keywords* de um *paper*.

Número de *Papers* vs. *Abstract*

Para além de refletir sobre os dados de *keywords*, também é importante observar outras componentes do *paper*. De toda a exploração e validação da informação na base de dados, é evidente que todos os *papers* são compostos pelo "título", mas o mesmo não se pode afirmar da componente "abstract". A próxima análise identifica os *journals* que na sua estrutura de documento não apresentam "abstract" e o número de papers que se registam nesta situação.

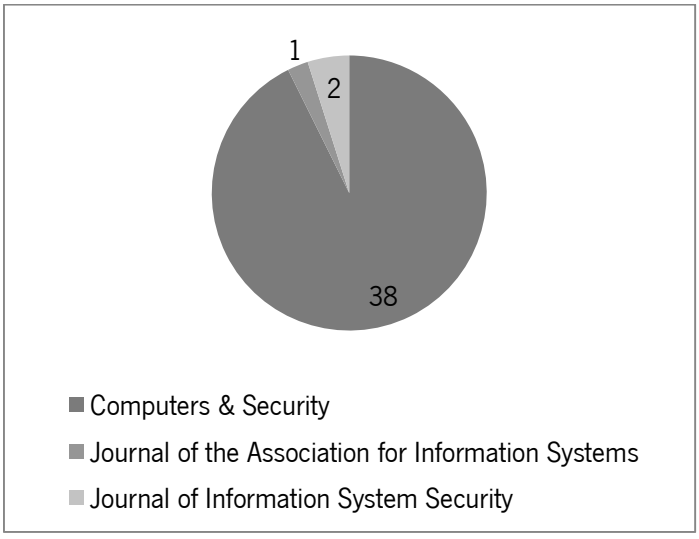


Gráfico 5.7: *Papers* sem *Abstract*

Da observação desta análise pode-se concluir que o número total de *papers* que não têm "abstract" totaliza 41, estando a sua distribuição restrita a três *journals* diferentes. Poder-se-á afirmar que os restantes *journals* são todos dotados de todos os elementos que caraterizam o paper. Ainda sobre esta análise e porque é importante observar os resultados até ao detalhe máximo, a próxima pesquisa detalha para cada um dos três *journals* apresentados anteriormente, o número total de *papers* que não têm "abstract" pelo nome do *journal* e período temporal em que se verifica esta asserção.

Um olhar atento, concluí que o "*Journal Computer & Security*" é o que apresenta maior número de *papers* sem "abstract". Este resultado é afetado pelo facto de grande parte dos "artigos científicos" pertencerem a esse *journal* , levando a crer que por isso existe uma maior percentagem de informação com erro.

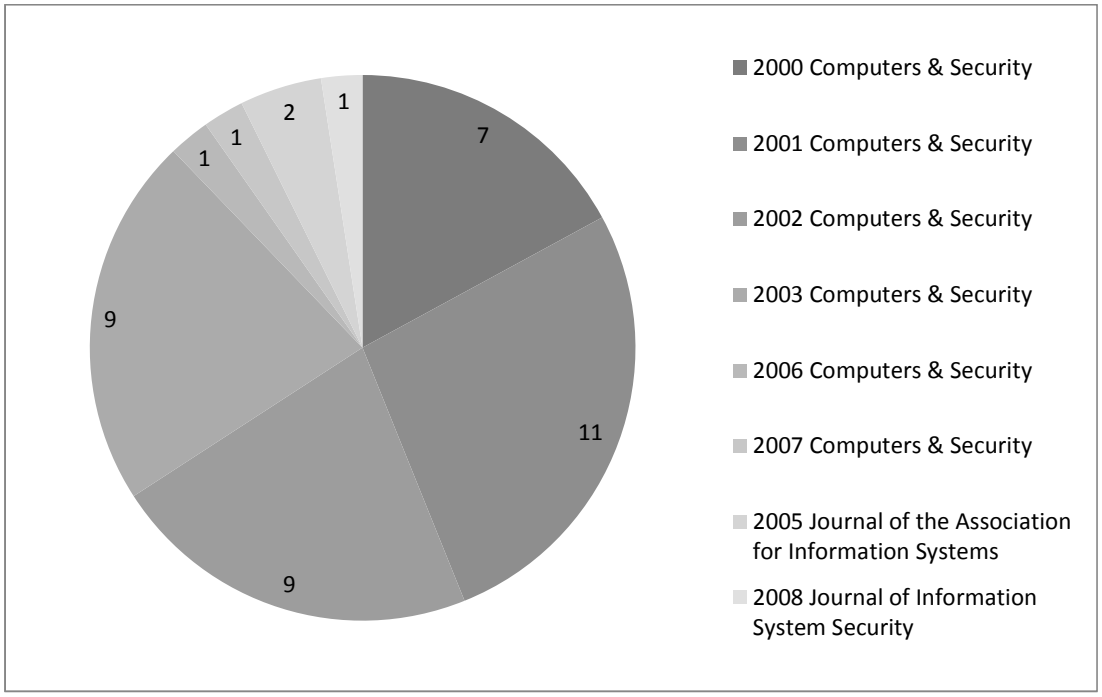


Gráfico 5.8: Distribuição *Papers* sem *Abstract* vs. *Journal*

Lista de Classes vs. Anos

A análise que se segue apresenta em formato tabular (Tabela 5.1) o cruzamento obtido quando se pretende saber para cada ano quais as classes de quê é que lhe está associado assim como o número de vezes em que isso ocorreu. A contagem feita considera que, se para determinado ano, a um dado paper lhe foi identificado duas ou mais *keywords* pertencentes à mesma classe pai, será apenas contabilizado uma única vez a classe e não acumulado o número de vezes da classe.

Ao observar a tabela é possível concluir que a literatura da SSI teve um crescimento com um ritmo bastante acentuado a partir de 2004 com maior concentração em 2006, 2007 e 2008. As classes aqui representadas podem ser consultadas no Apêndice C, já referenciado na subsecção 4.7.3. Desta análise, ainda é possível observar que grande parte das publicações realizadas estão concentradas nas classes 7 (*Attacks*), 8 (*Cryptology*), 11 (*Identification*), 12 (*Authentication*), 13 (*Access*), 14 (*Computer Network Security*), 15 (*Internet Security*), 17 (*Detection*), 25 (*Data security*), 31 (*Security Risk*), 35 (*Security Management*), 36 (*Security Policy*), 37 (*Information Security Governance*), 41 (*Security Compliance*), 52 (*Computer Crime*), 54 (*Security and Privacy*) e 55 (*Secure Voting*), concluindo-se que os âmbitos mais focados pelos autores são sem dúvida os temas mais críticos da temática da SSI e por isso o seu foco fazer-se notar pelo intervalo identificado.

Lista de Classes vs. *Journals*

A Tabela 5.2 representa para cada um dos 14 *journals* trabalhados as classes que os representam e quantos *papers* estão associados. Facilmente observa-se que o *journal* C&S é o que abrange maior número de classes, resultado proveniente do facto de este *journal* ser o que representa maior número de *papers* em exploração. A segunda posição é associada ao jornal de IJIS que também está representado por diversas classes. Nesta análise ainda é possível concluir que, tal como na análise anterior, as classes que mais se destacam são a 7 (*Attacks*), 8 (*Cryptology*), 9 (*Defences*), 11 (*Identification*), 12 (*Authentication*), 13 (Access), 14 (*Computer network security*), 15 (*Internet Security*), 17 (*Detection*), 25 (*Data Security*), 31 (*Security Risk*), 35 (*Security Management*), 36 (*Security Policy*), 37 (*Information Security Governance*), 41 (*Security Compliance*), 52 (*Computer crime*), 54 (*Security and Privacy*) e 55 (*Secure Voting*). Esta análise reforça as temáticas mais trabalhadas pelos autores de cada *journal*, do mesmo modo que permite estabelecer uma relação em termos de foco dos 14 *journals*.

Tabela 5.1: Classes de *Keywords* vs. Anos

Classes /Ano	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011
1						2	1		1	2	1	
2						1						
3		1	1	2	4	4	2	2	2	2	5	
4			1		1	1	1	3	3	6	4	
5		1	1	1	2		1	2	5	2	4	1
6		2		2	1	1	4	3		1	2	1
7		6	3	6	6	14	24	14	14	29	20	6
8	5	7	12	17	28	25	29	11	15	26	22	6
9	4	1		3	4	4	2	9	3	9	9	2
10					1	1	1	3		1	1	
11			1	1	5	4	4	4	6	6	2	
12	5	1	7	9	16	15	10	13	9	19	15	2
13	1	1	5	5	7	6	8	5	6	11	12	4
14	4	1	1	4	10	20	12	18	11	17	22	
15			3	1	6	5	4	7	5	10	6	2
16						1	1		1			
17		2	2	5	7	11	8	15	7	10	10	2
18					1	1	1	1	2			
19				1		2	1	2	1	3	1	
20			1		2		1	2		1	1	
21		1	1		1		5	1		3	3	
22										1		1
23				1								1
24	1			1	1	1	1	1	1	4		2

25		2	3	2	10	17	8	16	13	10	16	
26			1	1	5	1	3	4	4	4	1	1
27			1						1		1	
28								1	1			
29				1	1	1			1	2		
30						1						
31	1	1	2	4	7	11	5	9	8	6	8	2
32		1	2			2		2				
33					1	2	1		3	1		1
34					1	2	1		2	2	2	
35	3	2		2	2	6	7	4	2	3	4	
36	1	1	3		4	2	4	4	6	4	9	
37	2	1	1		2	3	5		3	1	3	
38						1	2	3	1	1	2	
39							1	1	2	1	2	
40							1					
41	1	2	1	1	7	4	8	5	2	5	8	1
42					1	1		2	2	2	2	1
43					1			1	2			
44								1		1	1	
45						1	1	5	6	2	4	
46	3	2			2					1	1	
47				1	1	2		1	3	3	2	
48							1		1			
49					1			1			2	
50		1	1	1		1		1	1	1		

51						1						
52		3		3	4	6	9	7	3	6	5	1
53				2							1	
54		2	6	4	13	14	9	13	14	11	14	2
55		3	1	3	2	2	2	1		3	2	1
56			1				1					
57								1		1		
58								1				1
59						1						
60			1									
61											1	
62						1						
63							1					
64											1	
65										1		
66											1	
67												1
68			1									
69										1		
70											1	
71												1
72								1				
73										1		
74			1									
75										1		
76		1										

77										1		
78							1					
79									1			
80										1		
81								2				
82								2		1		
83		2										
84								3	1		1	
85						1	1	2	2	1		

Tabela 5.2: Classes Vs. *Journals*

Classes / *Journals*	C&S	IM&CS	IRMJ	ISJ	ISR	IJIS	JACM	JCIS	JIP&S	JISS	JMIS	JSIS	JAIS	MISQ
1	3					1		1	1		1			
2	1									1				
3	14	1	1		1	6		1		1				
4	13								2	2			1	2
5	16			1	1	1				1				
6	15					2								
7	92	8	1	1	1	20		2	3	6	5		2	1
8	103	7				88	3			2				
9	23	3			1	8			5	4	2	1	1	2
10	5					1			1		1			
11	18	3				7			2	2		1		
12	71	11				30	1		5	2			1	
13	40	1				22	2		2	2			2	

14	73	10				28			3	5		1		
15	36		2		1	4			2	3	1			
16	2									1				
17	55	3			1	9	1	2	2	4	2			
18	5									1				
19	7					2				1			1	
20	7					1								
21	9				1	1			1	3				
22	1					1								
23	1					1								
24	5					6	1			1				
25	11	74	2		1	3			1	4	1			
26	12			1	1	8			1	2				
27					1	2								
28	1					1								
29	3					3								
30							1							
31	34	15	2		1				1	6	2	1		2
32	6					1								
33	3				2					1	2			1
34	3				2				2	1	2			
35	14	6	1		3	2			1	6				2
36	18	2	1		1	7	1		2	2		1		3
37	14	2				1				2				2
38	6				1				1	1				1
39	3	1				1					1	1		

40		1												
41	26	4			1	4			1	4	1	1		3
42	5								3	2				1
43	1								2	1				
44	1								1					1
45	12			1		1			2	1			1	1
46	8					1								
47	3	1	1							8				
48	2													
49	4													
50	3				1	1			1	1				
51	1													
52	23	6	1		4	3		1	3	3	2	1		
53	2										1			
54	27	17	4		2	20		2	13	7	7	5	1	
55	8	7		1		3							1	
56	2													
57					1					1				
58	2													
59						1								
60	1													
61											1			
62					1									
63						1								
64	1													
65	1													

66						1								
67						1								
68					1									
69									1					
70										1				
71	1													
72	1													
73	1													
74						1								
75	1													
76				1										
77	1													
78				1										
79													1	
80									1					
81	1									1				
82	1					2								
83						1	1							
84								1	1	3				
85	4									1	1			

Número de Classes Distintas vs. Total *Papers*

As próximas análises irão apresentar em maior detalhe a informação extraída ao longo das várias etapas descritas no Capítulo 4. Grande parte das análises são fruto das várias interpretações apuradas entre o cruzamento da informação contida na tabela "arvorekeywords" e restantes tabelas. A próxima Tabela 5.3 permite observar o número de classes distintas que podem estar presentes em cada *paper*, apresentando o total de *papers*. O número de classes distintas é representado pelas keyword que compõem um *paper*.

Tabela 5.3: Número Classes vs. Total de *Papers*

NúmeroClasses	TotalPapers
0	169
1	182
2	228
3	202
4	124
5	63
6	21
7	8
8	1
9	1
10	0
11	0
12	0
13	0
14	1
15	0
16	0

Da análise, pode-se concluir que grande parte dos *papers* apresentam uma, duas ou três áreas distintas para classificar um *paper*, mas, também existe um *paper* que apresenta 14 classes de *keywords* diferentes para o categorizar. No Apêndice D poderá consultar-se estes resultados em maior detalhe uma vez que esta análise olha a todas as classes existentes e não apenas à classe pai.

Número de *Papers* vs. Lista de Classes

O Gráfico 5.9 apresenta pelas diferentes classes pai o número de *papers* que foram associados a tag eleita. Tal como já abordado anteriormente, existem classes que ressaltam as

conclusões da análise: 7 (*Attacks*), 8 (*Cryptology*), 9 (*Defences*), 12 (*Authentication*), 13 (Access), 14 (*Computer network security*), 17 (*Detection*) e 54 (*Security and Privacy*). Uma nota importante deve ser mencionada para esta análise: todas as *tags* eleitas mencionam a classe pai sem por isso renunciar a acumulação de uma mesma *tag* no mesmo *paper* para a mesma classe.

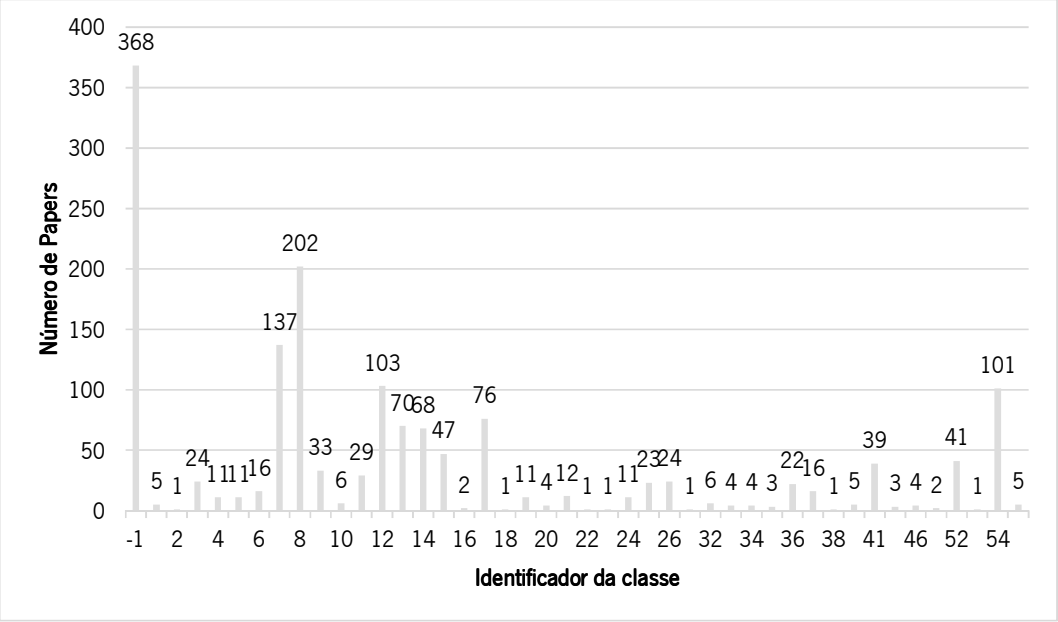


Gráfico 5.9: Número *Papers* vs. Classes

Com facilidade se nota que existem classes que não estão apresentadas no gráfico (total de 35 classes). Este resultado provém do facto de nestes *papers* as *tags* que o identificam não pertencerem a classes pai. Pela observação do gráfico conclui-se, ainda, que existem 368 *keywords* que foram eleitas *tag* e que a sua classe pertence à classe pai. Nesta análise não é possível detalhar este grupo de *keywords*, podendo consultar-se o Apêndice C se conhecer em maior detalhe.

Outras Análises

Além destes resultados, também é possível consultar o Apêndice F que apresenta a *tag* atribuída a cada *paper* segundo a fórmula do "calculovalor", apresentada na secção 4.10. Esta fórmula relaciona a posição da keyword no *paper* com o índice de especificidade definido na "arvorekeywords".

Caso #2 – Papers sem Keywords

Anteriormente apresentaram-se os resultados obtidos pela exploração e análise das publicações que na sua estrutura eram compostas por lista de *keywords*. Inversamente à parte anterior, a presente secção debruça-se sobre as temáticas da SSI com a nuance de resultados e conclusões relativas a periódicos e artigos científicos que não possuíam na sua estrutura de documento qualquer tipo de lista de *keywords*, ou seja, os autores das publicações não lhes atribuíram *keywords*.

Nestas condições foram identificados 120 *papers*, o equivalente a 12% da amostra de 1000 *papers*. Este conjunto de *papers* mereceu um manuseamento distinto quando comparado com o caso anterior, uma vez que não puderam ser compridas todas as etapas de exploração de dados definidas na secção 4.10.

Triagem entre Título e "arvore de keywords/dicionáriostopwords"

Para a presente experiência, recorreram-se aos dados contidos na tabela de "arvore de keywords/dicionáriostopwords" e compararam-se os termos semelhantes entre o "titulo" de cada paper com os termos desta tabela. A exploração da *tag* segundo o cruzamento de valores - "noção de matches", foi o caso em que esta pequena amostra foi explorada em conjunto com os restantes *papers*.

Ao analisar os dados do título foi possível concluir que apenas 79 *papers* distintos tinham no título *keywords* com a temática de SSI, ou seja, 41 dos *papers* ainda estavam sem uma primeira análise na identificação de uma possível tag. Nesta análise identificaram-se 41 *papers* com mais do que uma *keyword* de SSI associada, com resultado final obtido de 153 *keywords* identificadas no conceito de "*matche*" para o "titulo". Nesta análise ainda é verificado que um *paper* pode ser associado a várias *keywords* que podem ser todas referentes à mesma classe.

Número de *Paper* vs. *Abstract*

Ao trabalhar sobre esta pequena amostra foi verificado que, apesar de todos os *papers* serem compostos na sua estrutura pelo "título", o mesmo não acontecia com a presença do "abstract". Nesta situação de formatação incompleta do *paper* encontravam-se 30 *papers* (3%), 27 dos quais pertencem ao *Journal Computer & Security*, dois ao *Journal of Information System Security* e um ao *Journal of the Association for Information Systems*.

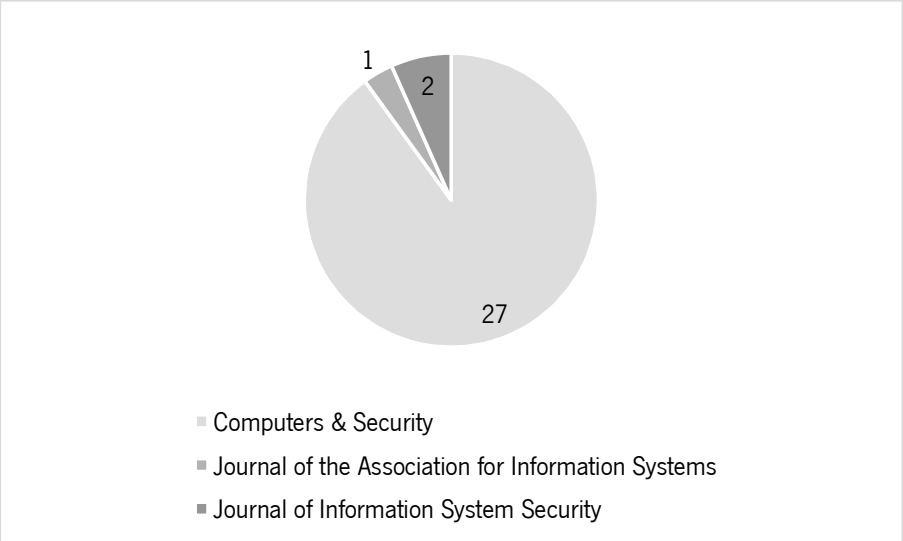


Gráfico 5.10: Número de *Papers* vs. *Abstract*

A restrição apresentada anteriormente conduziu a análise para foco do "título" e por consequente diminuiu o leque de exploração da amostra.

Triagem entre Título e *Abstract*

A presente análise estuda os *papers* sem *keywords* com o objetivo de cruzar as *keywords* de SSI encontradas no "titulo" e as do "abstract". Dentro da amostra definida são obtidos 46 resultados para uma possível *tag*. As *keywords* que eram comuns a estas duas perspetivas passaram a ser a *tag* do *paper*. Para as *keywords* resultantes deste cruzamento foi identificado o seu valor de "indicespecificidade", onde a tag passou a ser o valor mais alto identificado no *paper* em questão.

Seleção *tag* – *Papers* sem resultados para pesquisas anteriores

Para o presente Caso 2 não são apresentados resultados para seleção da *tag* segundo a fórmula do "calculovalor". A razão que conduziu a esta situação tem como base o facto de esta amostra de 120 *papers* não estar composta por lista de *keywords* e por isso não se obtém a posição da keyword.

Para todos os *papers* que não foi alcançada *tag* por meio do cruzamento entre "titulo" e "abstract" foi necessária uma leitura ligeira do documento para conseguir associar-lhe uma temática de SSI com base nas *keywords* já identificadas. Nestas condições estavam contabilizados 38 *papers*.

5.3 Conclusão

Neste espaço será apresentado o resultado obtido da análise elaborada aos resultados obtidos no presente capítulo. Todos os resultados apresentados ou intermédios foram explorados para que fosse possível tirar o máximo proveito da informação proveniente de toda a base de *papers*. Os resultados são provenientes de todos os casos explorados na secção 4.10 e para cada *paper* é apresentada a(s) keyword(s) que maior robustez mostra quando constituída como um elemento de todas as etapas da seleção da *tag*.

Uma *tag* é considerada robusta se o resultado é proveniente do "calculovalor", "*matche* titulo", "*matche* abstract" e o cruzamento entre estas duas condições devolve a mesma *tag*. Contudo, caso a keyword apenas se encontre na lista de *keywords* com maior "indicespecificidade" e esteja no "titulo" será considerada *tag* uma vez que a premissa inicial do presente trabalho era apenas investigar a informação dos *papers* para a lista de *keywords* e "título". Sem ignorar os *papers* que não apresentavam lista de *keywords*, o rigor de seleção da *tag* também esteve presente. Para estes casos foi considerada uma *tag* robusta as *keywords* cujo "indicespecificidade" apresentava maior valor em conjunto com os matches encontrados ao nível do "titulo" e "abstract".

Para os casos em que foram consideradas a *tag* com exposição análoga ao "calculovalor", "titulo" e "abstract" foram identificados 182 resultados. Neste âmbito inicial ainda foram identificados 230 resultados em que a *tag* escolhida estava presente no "calculovalor" e "titulo". É importante referir que os resultados para um dado paper podem retornar várias *tag* e que este processo de identificação deve ser replicado por cada ciclo de tags trabalhadas. Na segunda etapa de avaliação da *tag* com verificação para o "titulo" e "abstract" foram identificados 129 resultados e 161 resultados em que a *tag* é *tag* via "calculovalor" e está presente no "titulo".

O processo de verificar se a keyword é *tag* via fórmula "calculovalor", ou via "titulo", ou via "abstract" foi preservado até que o número de *keywords* identificadas pelo "calculovalor" com foco pelo "indicespecificidade" fosse concluído.

A *tag* final para cada *paper* pode ser consultada no Apêndice G. Este resultado é fruto de todas as hipóteses testadas até que já não existissem mais *keywords* provenientes do "calculovalor" para cada *paper*. Para os *papers* que não eram compostos por lista de *keywords* a tag base considerada foi a keyword pertencente à temática de SSI que estava presente no "titulo" seguindo-se para a validação ao "abstract". Neste resultado foi selecionada a keyword que apresentava maior "indicespecificidade". Ainda para este reduzido conjunto de *papers* e para os quais não foi possível

definir nenhuma *tag* foi necessário avaliar as *keywords* de SSI semelhantes encontradas no "titulo" e "abstract" que apresentavam melhor "indicespecificidade" cruzando-a com a informação do *papers*. Para este conjunto de *papers* e para os casos em que não foi identificada nenhuma keyword no "titulo" e "abstract foi necessário ler o corpo do *paper* para conseguir identificar uma *tag* para o *paper*.

Por fim, conclui-se que o resultado final de todo o trabalho não relacionou todas as classes de keywords definidas com a *tag* final de cada paper. Apesar de todas as classes terem sido o ponto de partida do projeto, o resultado final restringiu a oferta que foi alcançado por todas as etapas já descritas anteriormente. A titulo de exemplo, foram consideradas para cada *paper* apenas a contagem de classe pai, caso as duas *keywords* eleitas pertencessem a mesma classe pai. Na eventualidade de ambas as *keywords* robustecerem nas mesmas proporções é adotada a *keyword* mais específica. Todos os resultados têm origem nas várias etapas de eleição de *tag*.

Capítulo 6 – Síntese

6.1 Contextualização

O presente capítulo apresenta a solução final do projeto de dissertação. Dada a dimensão do projeto considerou-se conveniente apresentar os resultados em formato ilustrativo e por isso, para obter esse resultado recorreu-se a duas aplicações de código-aberto: *Pajek e VOS Viewer*.

O Pajek, primeira aplicação utilizada, é um programa que permite a análise e visualização de grandes redes de informação. De acordo com a página web oficial desta aplicação, ela fornece ferramentas para análise e visualização de redes como redes de colaboração, redes genealógicas, redes de Internet, redes de citações, redes de difusão (SIDA, notícias, inovações), *Data Mining* (redes de 2-Mode), entre outras. Para este caso prático foi criado um ficheiro de texto que continha a informação necessária para a criação de uma rede. A rede de informação foi criada manualmente com recurso ao Excel que continha todas as "*keywords-tags*" De seguida foram definidos os identificadores da *keyword* - vértices (ponto de ligação entre dois *edges* - *keyword*) e os *edges* (linhas que representam a ligação entre dois vértices e a sua distância). Após terminar a criação da rede foi possível carregar a mesma na aplicação e analisar os resultados devolvidos.

A próxima etapa incide sobre a utilização da aplicação *VOS Viewer*. O VOS é um programa que permite construir mapas baseados nos dados da rede. Os mapas são construídos utilizando as técnicas de mapeamento de agrupar do VOS e permite visualizar e explorar os mapas. O programa apresenta os mapas de diferentes maneiras, cada qual foca um aspeto diferente do mapa. Para uma análise mais detalhada ainda é possível fazer *zoom*, pesquisa de texto, etc. Para alcançar um resultado viável foi necessário criar um mapa com toda a informação necessária à construção do esquema para que depois de importado na aplicação ela representasse esse mesmo mapa. O mapa da aplicação VOS obedece à estrutura de: "id,label,x,y,weight,cluster" onde id representa o id que faz com que o mapa se consiga relacionar com a rede, a *label* que define a keyword, x e y são as posições que ocupa a *keyword*, *weight* - tamanho da representação total do objeto *tags* que foram identificadas pelos diversas etapas descritas anteriormente e cluster - cor diferenciadora das várias temáticas.

Após uma breve contextualização de como estas duas aplicações funcionam, a secção seguinte apresenta os resultados obtidos para o esquema temático da literatura de SSI.

6.2 Esquema de Classificação da SSI

Encerrados todos os passos intermédios para a obtenção do esquema final de classificação sobre a temática da SSI, a presente secção descreve as etapas realizadas para a obtenção de um esquema de classificação figurativo por meio de aplicações ilustradas. A base deste esquema são todas as *tags* finais identificadas para cada *paper* (Apêndice G). Apesar de terem sido identificadas pelo menos 1000 *tags* para os 1000 *papers*, o resultado final do número de *tags* únicas identificadas foi de 534 *keywords*. Desta análise pode concluir-se que 53,4% das *tags* associadas aos 1000 *papers* são únicas.

Com todas as *tags* finais definidas foi criada uma árvore de *keywords* com as 534 *keywords*. Esta árvore obedeceu aos mesmos critérios de criação da árvore produzida inicialmente para este projeto. Era composta pelos identificadores únicos da keyword e pela descrição da mesma. Para se conseguir estabelecer a relação entre as *keywords* foi colocado num ficheiro Excel a keyword origem, a keyword filha e o seu índice de especificidade quando comparado entre as duas *keywords*. Nesta fase já existe uma primeira versão de todo o esquema conceptual temático para a SSI, que foi melhorado recorrendo ao *software Pajek*. Contudo, uma vez que o *Pajek* só aceita a leitura de ficheiros ".txt" foi necessário replicar a informação da arvore presente no Excel para o formato aceite pelo *Pajek*.

A partir deste momento trabalhou-se toda a informação de modo a que o *Pajek* conseguisse processa-la. Numa primeira fase foi necessário estabelecer um código composto pelo "idkeyword" e a sua descrição. Cada "id" representava um vértice na árvore para o esquema final. A próxima etapa consistiu na definição de *edges* que representa o "idkeyword" considerada pai na arvore, pelo "idkeyword" considerada filha na árvore, pelo número de níveis que separa os dois "identificadores". Para estabelecer os níveis entre as duas "idkeyword" foi necessário inverter a ordem que é espectável nestas situações, isto é, se a diferença entre os dois "idkeyword" for de apenas um nível, esta diferença será invertida para 6 (número máximo de "indicespecificidade" na tabela "arvorekeywords"), seguindo-se a mesma lógica para os restantes níveis (diferença de níveis = leitura *Pajek*: 1=6, 2=5, 3=4, 4=3, 5=2, 6=1). Após a elaboração do ficheiro com os códigos da rede aceites pelo *Pajek*, carregou-se a informação do ".txt" para a aplicação. As figuras que se seguem exibem apresentações da informação nos diversos *layouts* disponibilizados pela aplicação *Pajek*. A Figura 6.1 mostra a relação estabelecida entre as diversas *tags* identificadas nas várias

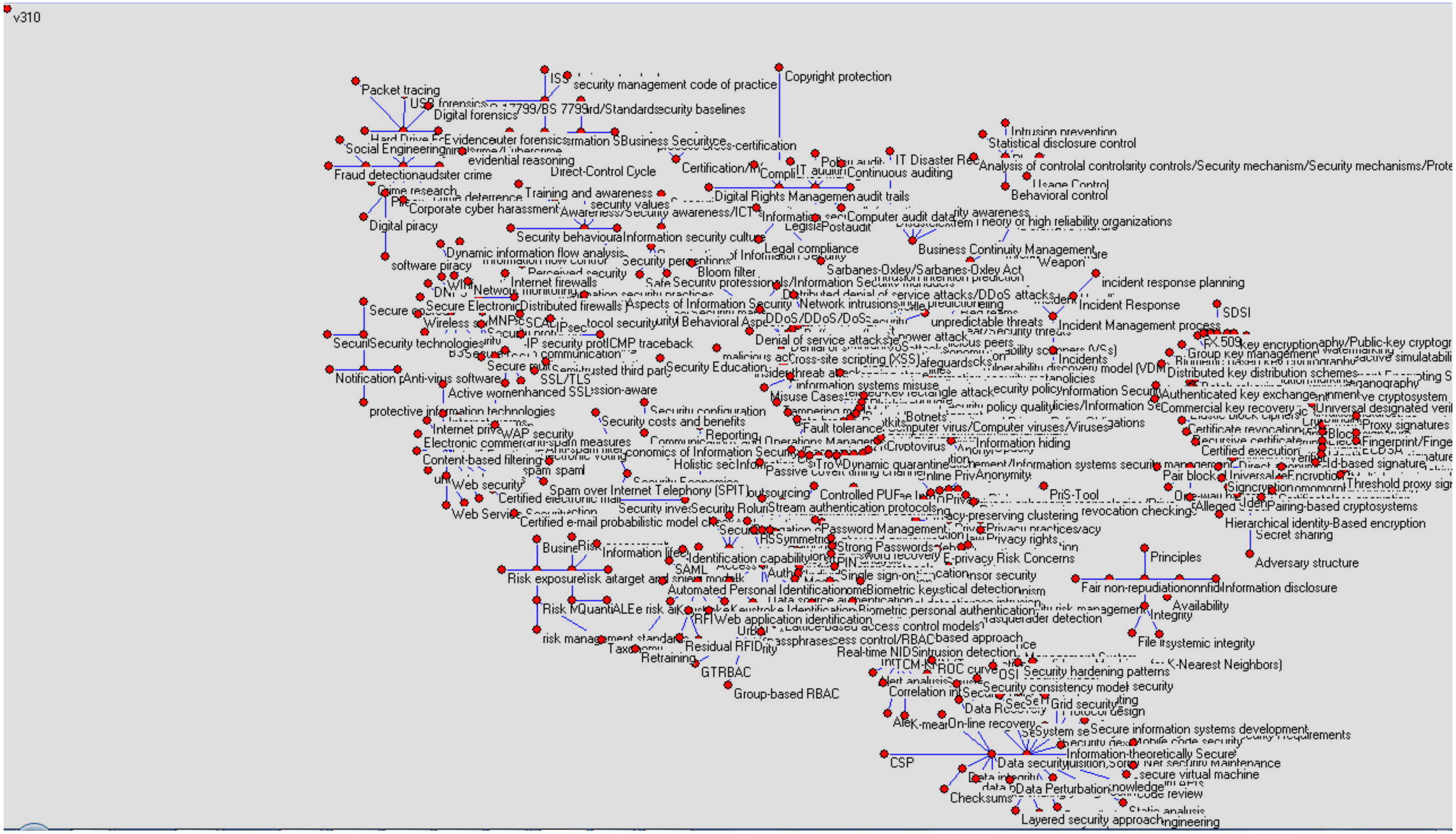


Figura 6.1: Esquema em *Layout Energy* - *Fruchterman Reingold*

etapas descritas ao longo da secção 5.2. De acordo com *layout* escolhido, sendo o que mais se adequa para representar uma primeira versão do esquema de classificação, é facilmente percetível a relação que existe entre as várias temáticas da SSI assim como a dimensão que cada área começa a alcançar. Ao observar a figura é possível identificar para alguns casos, os grupos criados pelas *tags*.

Nesta fase já é notável a abrangência da temática da SSI, mesmo que seja apenas com referência as *keywords* que foram identificadas como *tags* de *papers*. A representação ilustrada na Figura 6.1 sobre a temática da SSI por meio da aplicação *Pajek* ainda não está totalmente concluída. Nesta abordagem está em falta o tamanho que um vértice pode adquirir. Para realizar este processo foi contabilizado o número de vezes que uma keyword aparece na lista total de *papers* com *tag*. Este é o ponto de partida para melhorar a exibição do esquema temático e por isso recorreu-se à aplicação VOS.

A aplicação de colaboração científica *Vos Viewer* permite apresentar o resultado perfeito do esquema conceptual temático para SSI. Esta aplicação recebe como *input* um ficheiro ".txt" que representa a rede, isto é, o ficheiro é composto por códigos que estabelecem a relação entre as várias *keywords* e a sua distância em "indicespecificidade" (ficheiro *Pajek*) e recebe um outro ficheiro ".txt" que representa o mapa de toda a temática de SSI, composto pelo identificador que se relaciona com o identificador do ficheiro *Pajek*, pela descrição da keyword, a posição em termos de eixo de x e y que localizam a disposição da keyword na rede, o tamanho do vértice que representa o número de vezes que uma keyword foi identificada como *tag*, e a cor que irá representar o conjunto de *keywords* que compõem uma área, conforme se ilustra na Figura 6.2.

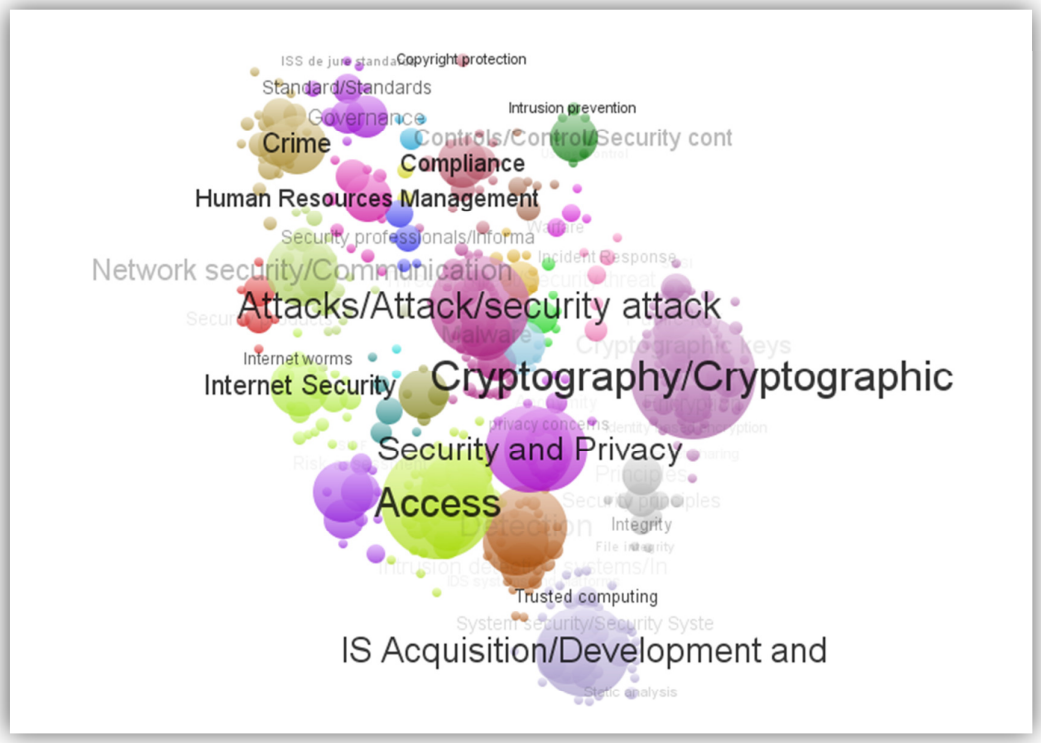


Figura 6.2: Esquema na Perspetiva de *Label View*

Uma vista simples sobre o esquema e o seu detalhe em termos de perfeição da distância entre as várias *keywords* e respetivos níveis, é visível na representação da Figura 6.4. Nesta exibição – *Scatter View*, pode observar-se a proporção ocupada pelos grupos mais abrangentes, assim como compreender os limites das várias áreas quando comparado com a Figura 6.2.

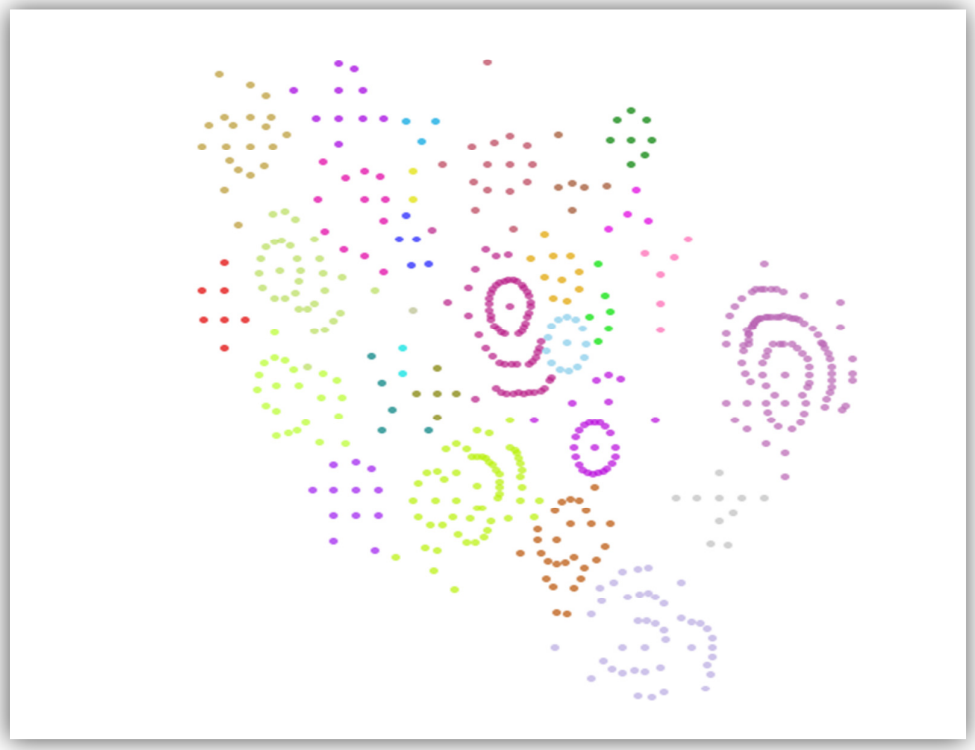


Figura 6.3: Esquema em *Scatter View*

Uma análise em termos da densidade de cor que cada área de SSI ocupa – *Cluster Density View* – pode ser visualizada na Figura 6.4. Com esta análise é facilmente identificado as grandes áreas da SSI uma vez que é comtemplado a abrangência total de todas as *keywords* pela temática de topo.

Figura 6.4: Esquema em *Cluster Density View*

A representação que melhor ilustra o esquema de classificação da literatura de SSI está contemplada na Figura 6.5. Este diagrama apresenta a visão global de todo o esquema criado e onde é possível observar as áreas que mais se destacam (cor vermelha/amarela), a densidade que ocupam e a sua relação temática em termos de disposição com as áreas vizinhas.

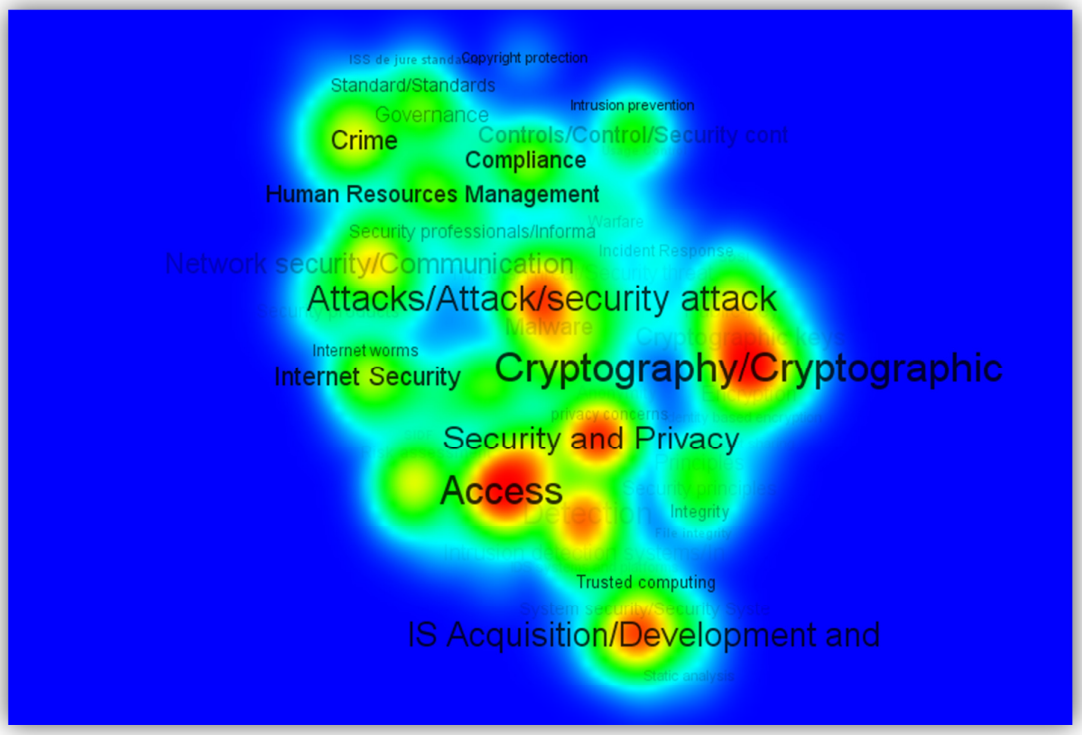


Figura 6.5: Esquema em *Density View*

Por fim, a Figura 6.6 exibe uma pequena parte em maior detalhe das áreas da SSI. Com esta visão é possível estabelecer o nível de detalhe em que este trabalho foi realizado sem descurar do que era o foco principal. Contudo, dada à dimensão do esquema criado é possível ver no Apêndice H cada uma das áreas em detalhe máximo.

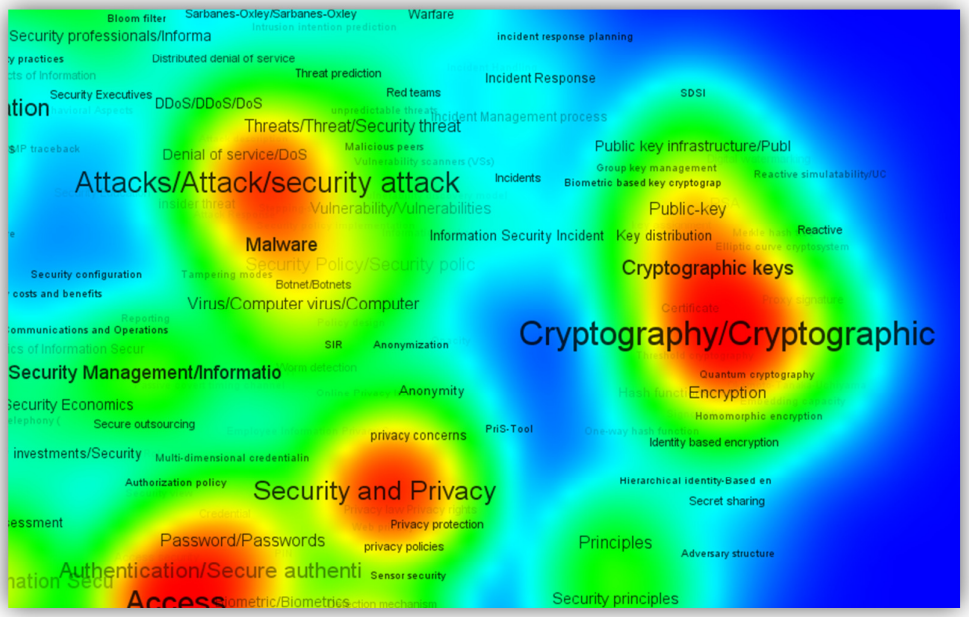


Figura 6.6 - Esquema em detalhe

6.3 Discussão dos Resultados

O tema deste projeto vem dar resposta a uma necessidade eminente de consenso no que diz respeito ao que deve estar incluído em termos de conceitos dentro de um determinado âmbito na temática da SSI. No capítulo anterior e nas secções anteriores a este, foram apresentados diversos resultados que propõem uma metodologia de consenso entre os investigadores para a temática da SSI.

Desde os primeiros resultados foram logo identificados os grandes grupos de conceitos, permanecendo-se em termos de extensão quase inalterável. Apesar do esquema final ser composto apenas por 534 keywords, é notória a fluidez que existe entre os vários conceitos e a forma como se relacionam. São evidenciadas sete grandes áreas: *Attacks/Security Attack*, *Cryptography/Cryptographic*, *Security and Privacy*, *Data Security*, *Internet Security*, *Access* e *Computer Network Security*, que estão em expansão, fruto das circunstâncias com que a sociedade se depara na sua rotina habitual. Os investigadores têm sem dúvidas explorado bastante estas temáticas, para que se consiga combater problemas como fraude de informação, invasão de acessos, etc. Um pormenor importante e que coincide com a reviravolta na sociedade no que toca a

informação e aos sistemas de informação, é que estas temáticas começam, de modo geral, a revelar-se a partir de 2004.

Os gráficos seguintes ilustram a evolução do tratamento das temáticas mais significativas (sete temáticas) ao longo dos anos, pelo número de papers publicados (intervalo de análise definido para este estudo). O Gráfico 6.1 exibe a evolução do número de publicações para a temática Attacks/Security Attack. A leitura do gráfico permite concluir um decrescimento do número de publicações em 2004, assim como um aumento acentuado em 2009.

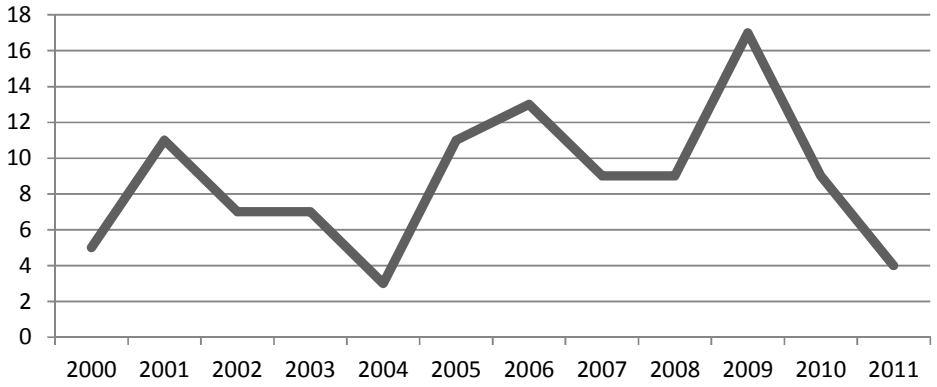


Gráfico 6.1: Evolução do Número de Publicações sobre a temática Attacks/Security Attack

A seguinte ilustração, Gráfico 6.2, mostra a evolução do número de publicações sobre a temática Cryptography/Cryptographic. Para esta temática é notório um crescimento moderado entre o ano 2000 e 2004 e uma descida acentuada entre 2006 e 2008.

Gráfico 6.2: Evolução do Número de Publicações sobre a temática Cryptography/Cryptographic

O gráfico 6.3 apresenta de que modo aumenta, diminuí ou se mantém constante o número de publicações para a temática Access. Para esta temática são salientadas as publicações para os anos 2002, 2006, 2009 e 2010.

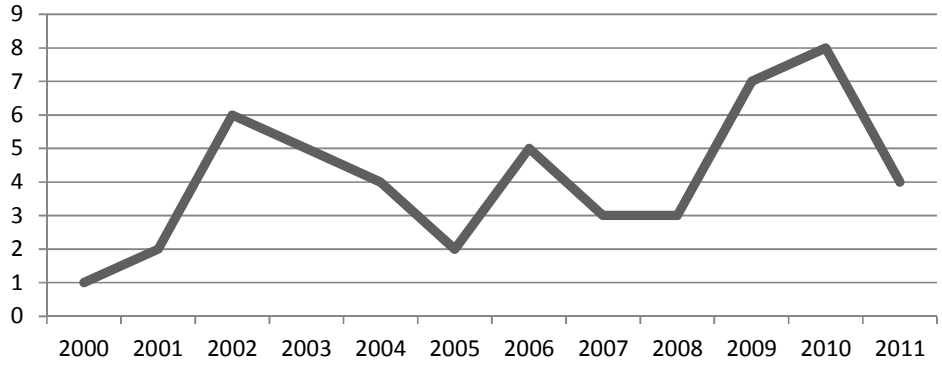


Gráfico 6.3: Evolução do Número de Publicações sobre a temática Access

A evolução do número de publicações para a temática Computer Network Security é exibida no Gráfico 6.4. Nesta temática salienta-se a diferença para o número de publicações no ano 2004 quando comparado com o ano 2010.

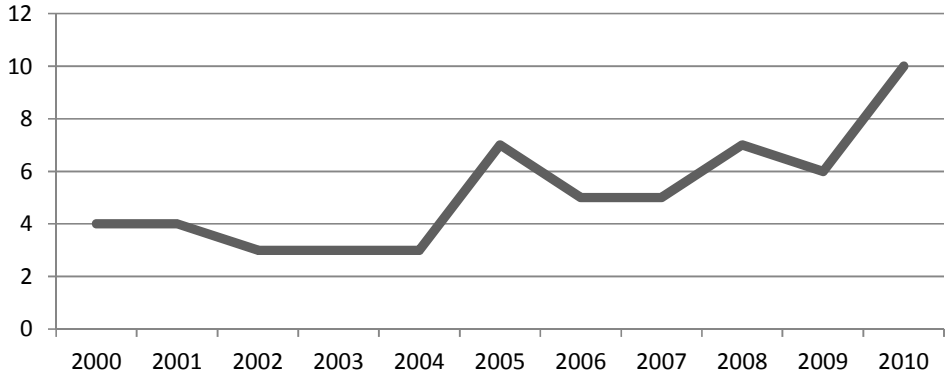


Gráfico 6.4: Evolução do Número de Publicações sobre a temática Computer Network Security

De seguida, Gráfico 6.5, é exibida a evolução do número de publicações para a temática Internet Security. Salienta-se o ponto de viragem no ano 2006 com um crescimento acentuado até 2008, decrescendo a partir deste ano até ao final do intervalo de análise.

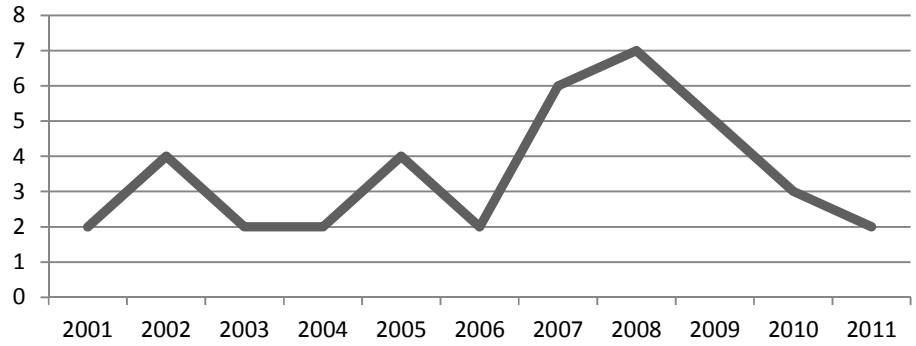


Gráfico 6.5: Evolução do Número de Publicações sobre a temática Internet Security

O Gráfico 6.6, que traduz a evolução do número de publicações para a temática Data Security mostra o crescimento moderado do número de publicações desde o ano 2002 até 2005, fazendo-se sobressair novamente no ano 2007 e 2010. Esta é uma temática já madura para a área de SSI.

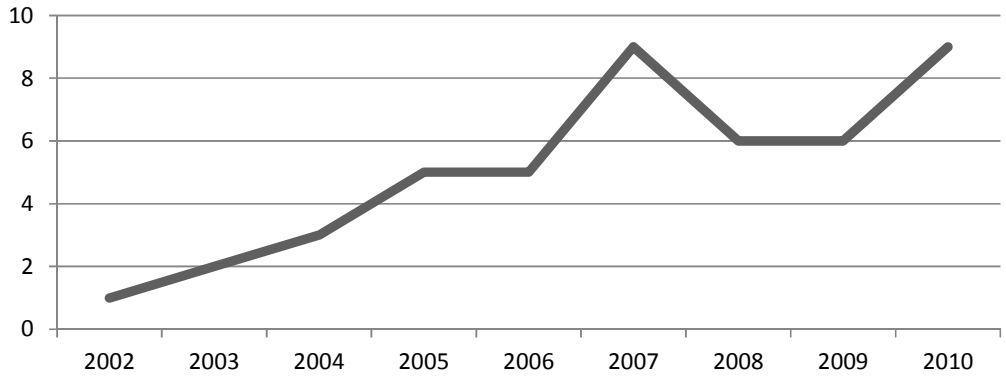


Gráfico 6.6: Evolução do Número de Publicações sobre a temática Data Security

Por fim, o Gráfico 6.7, apresenta a evolução do número de publicações para a temática Security and Privacy. Este gráfico mostra que esta temática nunca se encontra com número contante de publicações, atingindo o seu máximo, de acordo com o intervalo de análise, no ano 2010.

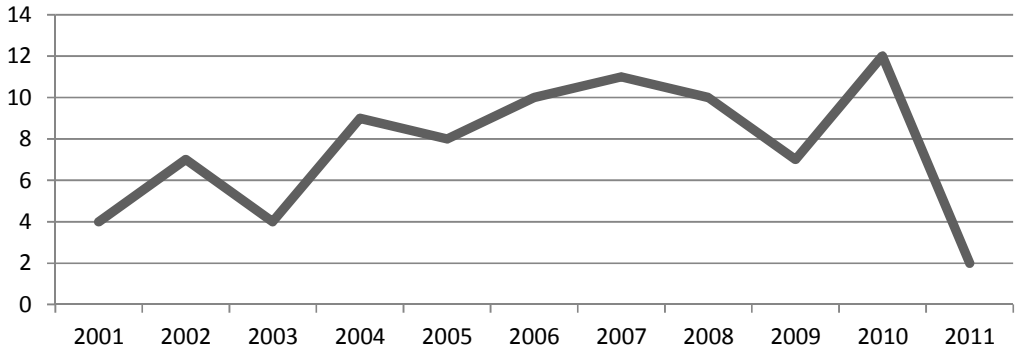


Gráfico 6.7: Evolução do Número de Publicações sobre a temática Security and Privacy

Apesar destas sete grandes áreas se distinguirem no meio das restantes, é viável pela observação dos resultados na secção anterior, que mesmo as áreas que não estão tão sobressaídas em termos de dimensão se relacionam bastante bem com as restantes sete apontadas.

O próximo Capitulo, Conclusão, apresenta todo o conjunto de observações que foram possíveis de concluir como as contribuições do presente estudo, as limitações/dificuldades que foram consideradas no estudo, reflete linhas de investigação futura e aponta as considerações finais do estudo.

Capítulo 7 – Conclusão

O presente capítulo aponta as contribuições relevantes da realização deste estudo para o tema deste projeto, sendo composto por quatro secções. Inicialmente apresenta as contribuições associadas à presente investigação e depois aponta as limitações do projeto. São apresentadas as propostas de investigação futura que apontam para novas perspetivas no seguimento deste projeto e por fim são apresentadas as considerações finais do estudo.

7.1 Contribuições

Nesta secção são apresentadas as principais contribuições deste estudo para o domínio da segurança de sistemas de informação.

O aspeto mais importante deste estudo consiste na confirmação da possibilidade de desenvolvimento de um esquema conceptual temático para a literatura de SSI. Este esquema não se foca apenas numa área de investigação e facilita a promoção de um maior nível de consenso sobre as áreas\temáticas pertencentes à SSI.

O segundo aspeto também ele de relevo baseia-se no facto de este esquema ter sido construído de modo "semiautomático". Com base na exploração da *tag* segundo o índice de riqueza semântica, a exploração da *tag* segundo o cálculo entre valores de atributos e a exploração da *tag* segundo o cruzamento de valores, etapas mais relevantes do processo de seleção da *tag* (Secção 4.10), a validação dos resultados obtidos foi bem sucedida e a criação do esquema de classificação foi realizada de alguma forma por um processo semiautomático.

Por fim, e com todos os conceitos identificados, é possível analisar a área em termos evolutivos, apurar quais as temáticas que têm maior exploração por parte dos investigadores, quais os autores que mais exploram a SSI, entre outros aspetos. Esta contribuição também permite concluir quais as áreas de maior fragilidade ao longo dos tempos e qual a sua continuidade como uma área em expansão.

7.2 Limitações/Dificuldades

O projeto de investigação ao longo do seu desenvolvimento decorreu com certas limitações. A atual secção descreve sucintamente as limitações do estudo.

A primeira limitação encontrada refere-se ao período de análise dos *papers*. Uma vez que o período selecionado é compreendido entre o ano 2000 e o ano 2011 poder-se-á concluir que não são abrangidos conceitos que anteriormente foram investigados.

O facto do esquema criado não ser atualizado automaticamente com os novos estudos, comtemplando recentes conceitos, conduz ao desfasamento da utilização deste esquema para anos futuros. O esquema criado apenas servirá de base para futuros trabalhos de investigação. Além desta limitação, ainda é apontada uma restrição à utilização do número de *keywords* final. As *tags* finais, consideradas, resultam de uma "trituração" máxima ao conteúdo de toda a base de dados.

Por fim, a cooperação entre investigadores para avaliar a veracidade do esquema final na atribuição de *keywords* a um dado conjunto de *papers*, ou seja, não foi conseguido obter participação por parte de pessoas externas para avaliar a consolidação e atribuição de *keywords* dada ao conjunto total de *papers* investigados (este processo implicaria a leitura do artigo).

7.3 Linhas de Investigação Futura

Após finalizar o projeto admite-se que diversos caminhos poderiam ter sido considerados para que se verificassem outras oportunidades de exploração da informação com vantagens e desvantagens associadas as opções de escolha, mas que serão relevantes considerar para trabalho futuro.

Uma primeira linha foca os critérios para julgar as melhorias e a qualidade do esquema conceptual produzido que passaria por analisar de que forma a solução classifica eficientemente e eficazmente determinado documento. Também seria importante reconhecer se o tempo de processamento de documentos e os recursos são os adequados, se determinado documento não está a ser classificado averiguar a razão de tal sucedido. Idealmente era existir uma aplicação capaz de responder a todas estas linhas de pensamento assim como, a aplicação sugerir ao utilizador a inclusão de outras categorias que estão de alguma forma relacionadas, ponderar a sugestão de criação de uma nova temática que poderá ser recente para a área da SSI, ou então a exclusão de tal documento do repositório de investigação.

Outros aspetos a considerar prendem-se com: elaborar um esquema classificativo para a SSI tendo como *input* todas as publicações que foram sendo elaboradas até à presente data, dotar o sistema classificativo desenvolvido de capacidades de atualização e desenvolver um esquema para a SSI totalmente automático.

Idealmente, uma possibilidade inovadora e vantajosa, seria a aplicação possuir elevada capacidade de aprendizagem, isto é, melhorar o seu desempenho por quantas mais vezes é executada, melhorando os resultados devolvidos, permitindo ao longo do tempo ter esquemas conceptuais temáticos mais aperfeiçoados e com conclusões cada vez mais enriquecedoras, como por exemplo, que autores estudam determinada temática da SSI; em que ano determinada temática teve maior evolução e sobre que tema; que relação existe entre as diversas temáticas, entre outras conclusões que foram investigadas mas com processos ainda semiautomáticos.

Poderá também ser vantajoso apresentar o esquema conceptual na forma de "*living graph*" facilitando assim a sua exploração. Para melhorar o seu resultado, seriam apresentadas várias perspetivas sobre as hipóteses testadas em relação ao sucesso ou insucesso da criação do esquema conceptual, com os pontos de vista que melhor respondem as necessidades do esquema.

Finalmente, reconhece-se que a temática de SSI é uma temática em evolução, em parte resultado do ambiente globalizado em que se vive e por isso a criação de um esquema conceptual temático seria a reposta à ausência do entendimento entre os investigadores da área.

7.4 Considerações Finais

O grande desafio do presente projeto de dissertação centrou-se na representação de conhecimento da temática de SSI por meio de um sistemas de organização de conhecimento que permita a sua representação. Os sistemas de organização de conhecimento são sistemas de conceitos semanticamente estruturados que incluem termos, definições, relacionamentos e propriedades de conceitos [Dahlberg 2006]. Segundo Hodge [2000] são destacadas duas categorias classificatórias para os sistemas de organização de conhecimento: a classificação e categorização, e os modelos de relacionamento. A primeira categoria engloba os sistemas de categorização, sistemas de classificação bibliográfica, listas de itens de assunto, e taxonomias, enquanto que a segunda inclui as ontologias, redes semânticas e dicionário de palavras.

No presente estudo e com base nas categorias referidas por Hodge [2000], afirma-se que o esquema de organização de conhecimento elaborado para a literatura de SSI pertence à primeira categoria identificada pelo autor. Este esquema obedece a regras que foram já especificadas anteriormente e que contribuem para a sequência lógica do esquema.

Neste projeto foram expostos diversos problemas e soluções relacionadas com a aquisição e análise de informação para a temática da SSI. As grandes quantidades de informação estão a dificultar cada vez mais a análise de temas e a obtenção da informação que realmente importa pesquisar, assim como, que a pesquisa pretendida seja realizada em tempo útil. A criação de um esquema concetual temático vem dar resposta a esta necessidade e permitir um consenso entre o que inclui cada área e de que forma se encontra.

Espera-se que todo o tempo disponibilizado na elaboração deste estudo permita de forma positiva auxiliar os investigadores para novos trabalhos de investigação.

Apêndice A - Código Java (classe *OtherOperations*)

Neste Apêndice é apresentado uma classe desenvolvida na linguagem Java que auxiliou no carregamento de dados na base de dados criada.

```
/**
 * @(#)OtherOperations.java
  * @author lucia
 * @version 1.00 2011/9/22
 */
import java.util.ArrayList;
import java.util.Random;
import javax.swing.JFrame;
import javax.swing.UIManager;
public class OtherOperations
{
        private LogWindow logw;
        private String stopwordFile;

    public OtherOperations(LogWindow logw)
    {
        this.logw=logw;
    }

    //gera X numero aleatorios
    //X= numInNeed
    //retorna um arraylist com os numeros gerados
    public ArrayList generateXNumber(int numInNeed)
    {
                logw.setLog("Gerador de 100 números:\n");
```

```
            ArrayList numbers= new ArrayList();


        //cada volta do ciclo gera um numero
        for (int i=0; i<numInNeed; i++)
        {
        while(true) //ciclo infinito até gerar um numero aleatorio sem ser repetido
                {
                        int randomInt= this.generateRandomNumber();
                        //verifica se o numero já existia
                                boolean
        isrepeted=this.checkRepeatedGeneratedNumbs(randomInt, numbers);
                        //se ainda não existia
                        if(!isrepeted)
                        {
                                //adiciona o numero repetido
                                numbers.add(randomInt);
                                //quebra o ciclo infinito e vai gerar outro numero
                                break;
                        }
                }
        }
        //imprime o numeros na janela
        for(int i=0;i<numbers.size();i++)
        {
                logw.setLog((i+1)+"º numero:"+numbers.get(i));
        }
        //retorna arraylist com todos os numeros gerados
        return numbers;
}

//gera um numero aleatorio de 1 a 1000
private int generateRandomNumber()
```

```
        {
            Random randomGenerator = new Random();
            int randomInt = randomGenerator.nextInt(1000)+1;
            return randomInt;
        }

        //verifica se o numero gerado já existe na lista de numero já gerados
        private boolean checkRepeatedGeneratedNumbs(int numbTocheck, ArrayList
numbers)
        {
            //se o array está vazio retorna logo false!
            if(numbers.size()==0)
                return false;
            //percorre o array à procura do numero a verificar se existe repetido
            for(int i=0;i<numbers.size();i++)
            {
                if(numbTocheck==(Integer)numbers.get(i))
                    return true; //encontrou o numero logo é repetido
                else
                    continue; //se não é igual continua o ciclo
            }
            //se chegou aqui significa que percorreu a lista de numeros
            // já gerados e não encontrou o numero que estava a verificar. Logo não é
repetido!
            return false;
        }
      //Stopwords
//lê o ficheiro de stop words e devolve uma string com o conteudo ou um erro de leitura
        private String getStopWordsFromTxtFile()
        {
            logw.setLog("A importar listas de stopwords!");
           // ler ficheiro de texto com as stopwords
```

```
            PropertiesHandler filereader= new PropertiesHandler();
            String dados=filereader.readallFile(this.stopwordFile);
            logw.setLog("Dados (stopwords) importados para a aplicação, com
sucesso!");
          return dados;
        }

      //recebe a lista de stopwords e devolve a array de stopwords sem quaisquer
repetição de palavras
        public ArrayList verificaStopwordsrepetidos(String lista)
        {
            String [] temp=lista.split("\n");
            for(int i=0;i<temp.length;i++)
            {
                    for(int t=i+1;t<temp.length; t++)
                    {
                            String stopword=temp[i];
      //                    if()
                    }
            }
            return null;
        }
          public static void main(String  [] args)
        {
            JFrame.setDefaultLookAndFeelDecorated(true);
                    try
                    {
            UIManager.setLookAndFeel("org.jvnet.substance.skin.SubstanceRavenGraph
iteGlassLookAndFeel");
            }
                    catch (Exception e)
            {
```

```
                System.out.println("Substance Raven Graphite failed to initialize");
            }
            OtherOperations oo=new OtherOperations(new LogWindow());
            oo.generateXNumber(100);
        }
    }
```

Apêndice B – Lista de *Papers* Analisados

idpaper	titulo	ano
1	Safe Contexts for Interorganizational Collaborations Among Homeland Security Professionals	2010
2	Overcoming Online Information Privacy Concerns: An Information-Processing Theory Approach	2007
3	The Influence of Experiential and Dispositional Factors in Phishing: An Empirical Investigation of the Deceived	2010
4	The Deterrent and Displacement Effects of Information Security Enforcement: International Evidence	2008
5	Interoperability of E-Government Information Systems: Issues of Identification and Data Sharing	2007
6	Gaining Trust Through Online Privacy Protection: Self-Regulation, Mandatory Standards, or Caveat Emptor	2008
7	Information Risk of Inadvertent Disclosure: An Analysis of File-Sharing Risk in the Financial Supply Chain	2008
8	Privacy Concerns Versus Desire for Interpersonal Awareness in Driving the Use of Self-Disclosure Technologies: The Case of Instant Messaging in Two Cultures	2011
9	Intrusion Prevention in Information Systems: Reactive and Proactive Responses	2007
10	Understanding the Value of Countermeasure Portfolios in Information Systems Security	2008
11	Decision-Theoretic and Game-Theoretic Approaches to IT Security Investment	2008
12	Information Security: Facilitating User Precautions Vis-à -Vis Enforcement Against Attackers	2009
13	Risks and Benefits of Signaling Information System Characteristics to Strategic Attackers	2010
14	Investments in Information Security: A Real Options Perspective with Bayesian Postaudit	2009
15	An Information Systems Security Risk Assessment Model Under the Dempster-Shafer Theory of Belief Functions	2006
16	Incremental Information Security Certification	2001
17	Violation of Safeguards by Trusted Personnel and Understanding Related Information Security Concerns	2001
18	Security with unfortunate side effects	2003
19	Corporate Governance and Information Security	2001
20	Methods for Protecting Password Transmission	2000
21	Trends in academic research: vulnerabilities analysis and intrusion detection	2002
22	Host intrusion prevention: Part of the operating system or on top of the operating system?	2005

23	Technology evolution drives need for greater information technology security	2005
24	Cyber crime: the backdrop to the Council of Europe Convention	2001
25	Technology and Electronic Communications Act 2000	2002
26	Fighting Against the Invisible Enemy. Methods for detecting an unknown virus	2001
27	Computer forensics	2003
28	The Corporate Information Assurance Officer (CIAO)	2001
29	Protecting The Web Server And Applications	2001
30	The implications of immunology for secure systems design	2004
31	From information security to... business security?	2005
32	Safety and Security in Multiagent Systems: Report on the 2nd SASEMAS workshop (SASEMASâ€™05)	2005
33	Crimes And Misdemeanours: How to Protect Corporate Information in the Internet Age	2000
34	From Risk Analysis to Security Requirements	2001
35	Life Was Simple Then	2000
36	Security policy update	2002
37	Some trends in research in cryptography and security mechanisms	2003
38	Multiparty Biometric-Based Authentication	2000
39	Smurfing, Swamping, Spamming, Spoofing, Squatting, Slandering, Surfing, Scamming and Other Mischiefs of the World Wide Web	2000
40	Cryptanalysis of a user friendly remote authentication scheme with smart cards	2004
41	Global Interoperability for Key Recovery	2000
42	Vulnerabilities categories for intrusion detection systems	2002
43	Compsec 2002: the complete security circle	2002
44	Information Security Governance e Compliance management vs operational management	2005
45	Controlling Risks of E-commerce Content	2001
46	The Weakest Link	2001
47	Managed Security Services â€" new economy relic or wave of the future?	2002
48	A comparison of Intrusion Detection systems	2001
49	Palladium, fraud, and surviving terrorism â€“ Compsec 2002	2002
50	Governments act to improve security	2003

51	Information Systems Risk Management: Key Concepts and Business Processes	2000
52	Digest of recent IT security press coverage	2002
53	Information Security in Multiprocessor Systems Based on the X86 Architecture	2000
54	The metamorphosis of malware writers	2006
55	International legal aspects of cryptography	2003
56	Protecting 21st Century Information â€" Itâ€™s Time for a Change	2001
57	Encountering encryption	2003
58	Security engineering and security RoI	2003
59	Roadmap to checking data migration	2003
60	White collar crime: a handmaiden of international tech terrorism	2002
61	The law, cybercrime, risk assessment and cyber protection	2003
62	Modeling network security	2006
63	A contest to evaluate IT security services management	2003
64	B2C Security â€" Be Just Secure Enough	2000
65	Careless about privacy	2003
66	The Evolution of Intrusion Detection Systems â€" The Next Step	2001
67	Public Key Infrastructure: Analysis of Existing and Needed Protocols and Object Formats for Key Recovery	2000
68	A framework for understanding and predicting insider attacks	2002
69	Information Security governance: COBIT or ISO 17799 or both?	2005
70	From secure wired networks to secure wireless networks - what are the extra risks?	2004
71	Americaâ€™s Internet Commerce and The Threat of Fraud	2000
72	Methods for preventing unauthorized software distribution	2003
73	The 10 deadly sins of information security management	2004
74	A Framework for the Implementation of Socio-ethical Controls in Information Security	2001
75	The economic approach of information security	2005
76	An enhancement of timestamp-based password authentication scheme	2002
77	A Framework for Understanding Vulnerabilities in Firewalls Using a Dataflow Model of Firewall Internals	2001
78	A Formalized Approach to the Effective Selection and Evaluation of Information Security Controls	2000

79	Applying information security governance	2003
80	Setting up an electronic evidence forensics laboratory	2003
81	University systems security logging: who is doing it and how far can they go?	2002
82	NIDS â€" Pattern Search vs. Protocol Decode	2001
83	Cyber-terrorism in context	2003
84	SSL Virtual Private Networks	2003
85	Hiding Digital Information Using a Novel System Scheme	2001
86	Managed Security Monitoring: Network Security for the 21st Century	2001
87	Next generation security for wireless: elliptic curve cryptography	2003
88	Implementing Public Key Infrastructures in a Dynamic Business Environment	2000
89	The role of criminal profiling in the computer forensics process	2003
90	Why we need a new definition of information security	2003
91	The concept of security and trust in electronic payments	2005
92	A comment on the Chen-Chung scheme for hierarchical access control	2003
93	Word Association Computer Passwords: The Effect of Formulation Techniques on Recall and Guessing Rates	2000
94	Cryptanalysis of an enhanced timestamp-based password authentication scheme	2003
95	Computer crimes: theorizing about the enemy within	2001
96	Implementing Information Security In The 21st Century â€" Do You Have the Balancing Factors?	2000
97	Improvement on Li et al.â€™s generalization of proxy signature schemes	2004
98	Non-PKI methods for public key distribution	2004
99	Information Security â€“ A Multidimensional Discipline	2001
100	Cryptanalyses of two key assignment schemes based on polynomial interpolations	2005
101	Cryptanalysis of two password-based authentication schemes using smart cards	2006
102	Authentication and Supervision: A Survey of User Attitudes	2000
103	Improving the ROI of the security management process	2004
104	An improvement of nonrepudiable threshold proxy signature scheme with known signers	2004
105	Trends In Cybercrime â€" An Overview Of Current Financial Crimes On The Internet	2001
106	New Names For Old â€" A Personal Surf Through Compsec 2000	2000

107	Investigative response: After the breach	2007
108	Decoding digital rights management	2007
109	Information Systems Audit Trails in Legal Proceedings as Evidence	2001
110	Dealing with contextual vulnerabilities in code: distinguishing between solutions and pseudosolutions	2003
111	A randomized RSA-based partially blind signature scheme for electronic cash	2005
112	On risk: perception and direction	2004
113	Spam, scams, chains, hoaxes and other junk mail	2002
114	The emergence of a comprehensive obligation towards computer security	2002
115	Love Conquers All?	2000
116	A traceable threshold signature scheme with multiple signing policies	2006
117	Information Security â€“ The Fourth Wave	2006
118	Usability and Security An Appraisal of Usability Issues in Information Security Methods	2001
119	Leading attackers through attack graphs with deceptions	2003
120	A Biometric Standard for Information Management and Security	2000
121	A user friendly remote authentication scheme with smart cards	2003
122	Privacy legislation: a comparison of the US and European approaches	2003
123	Computer forensics and electronic discovery: The new management challenge	2006
124	Security crisis management â€“ the basics	2002
125	Why access control is difficult	2002
126	A flexible date-attachment scheme on e-cash	2003
127	New hierarchical assignment without Public Key cryptography	2003
128	Re-engineering enterprise security	2006
129	The security challenges inherent in VoIP	2007
130	Much Ado About Nothing: Win32.Perrun	2002
131	Information Security Management: A Hierarchical Framework for Various Approaches	2000
132	Cyberterrorism?	2002
133	The 419 scam: information warfare on the spam front and a proposal for local filtering	2003
134	Improvement on the flexible tree-based key management framework	2005

135	Cryptography Regulations for E-commerce and Digital Rights Management	2001
136	Secure authentication scheme for session initiation protocol	2005
137	Information security governance: Due care	2006
138	An improvement of Hwangâ€“Leeâ€“Tang's simple remote user authentication scheme	2005
139	Information security obedience: a definition	2005
140	Information Lifecycle Security Risk Assessment: A tool for closing security gaps	2007
141	Efficient password authenticated key agreement using smart cards	2004
142	An empirical examination of the reverse engineering process for binary files	2006
143	The future of computer forensics: a needs analysis survey	2004
144	Search engines and privacy	2004
145	Ethical decision making: Improving the quality of acceptable use policies	2010
146	Fault trees for security system design and analysis	2003
147	Information security policy â€“ what do international information security standards say?	2002
148	A Practical Risk Analysis Approach: Managing BCM Risk	2000
149	Recent attacks on alleged SecurID and their practical implications	2005
150	A novel remote user authentication scheme using bilinear pairings	2006
151	Information security in networkable Windows-based operating system devices: Challenges and solutions	2007
152	Cryptanalyses and improvements of two cryptographic key assignment schemes for dynamic access control in a user hierarchy	2003
153	What InfoSec professionals should know about information warfare tactics by terrorists	2001
154	Security beliefs and barriers for novice Internet users	2008
155	Infection dynamics on the Internet	2005
156	Efficient proxy multisignature schemes based on the elliptic curve cryptosystem	2003
157	An empirical investigation of network attacks on computer systems	2004
158	Continuous auditing technologies and models: A discussion	2006
159	Secure information systems development - a survey and comparison	2005
160	Improvements on the WTLS protocol to avoid denial of service attacks	2005
161	Restraining and repairing file system damage through file integrity control	2004
162	From policies to culture	2004

163	Internet privacy law: a comparison between the United States and the European Union	2004
164	What InfoSec Professionals Should Know About Information Warfare Tactics by Terrorists	2002
165	Security considerations for incremental hash functions based on pair block chaining	2006
166	Information Security Management: An Approach to Combine Process Certification And Product Evaluation	2000
167	Matching key recovery mechanisms to business requirements	2005
168	Information Security â€" The Third Wave?	2000
169	Infection, imitation and a hierarchy of computer viruses	2006
170	Dealing with packet loss in the Interactive Chained Stream Authentication protocol	2005
171	Efficient identity-based RSA multisignatures	2008
172	Authentication and authorization infrastructures (AAIs): a comparative survey	2004
173	A novel approach for computer security education using Minix instructional operating system	2006
174	Hash channels	2005
175	Consensus ranking â€“ An ICT security awareness case study	2008
176	Efficient anonymous auction protocols with freewheeling bids	2003
177	A distributed systems approach to secure Internet mail	2005
178	Information security policyâ€™s impact on reporting security incidents	2005
179	A secure electronic voting protocol for general elections	2004
180	Efficient user identification scheme with key distribution preserving anonymity for distributed computer networks	2004
181	A new protocol to counter online dictionary attacks	2006
182	Analysis of vulnerabilities in Internet firewalls	2003
183	TBSE-an engineering approach to the design of accurate and reliable security systems	2004
184	Mitigation of network tampering using dynamic dispatch of mobile agents	2004
185	An assessment of website password practices	2007
186	Security surveys spring crop	2002
187	A quantitative method for ISO 17799 gap analysis	2006
188	Tele-Lab â€˜â€˜IT-Securityâ€™â€™ on CD: portable, reliable and safe IT security training	2004
189	Alternative Method for Unique RSA Primes Generation	2001
190	Hierarchical key assignment without public-key cryptography	2001

191	Aligning the information security policy with the strategic information systems plan	2006
192	Cryptanalysis of a Timestamp-Based Password Authentication Scheme	2002
193	A hybrid scheme for multicast authentication over lossy networks	2004
194	Access control in a hierarchy using one-way hash functions	2004
195	Authentication of users on mobile telephones e A survey of attitudes and practices	2005
196	The insider threat to information systems and the effectiveness of ISO17799	2005
197	Achieving Interoperability in a Multiple-Security-Policies Environment	2000
198	The availability of source code in relation to timely response to security vulnerabilities	2003
199	An Efficient and Practical Solution to Remote Authentication: Smart Card	2002
200	On the Security of Todayâ€™s Online Electronic Banking Systems	2002
201	Modelling and solving the intrusion detection problem in computer networks	2004
202	Access Control in Document-centric Workflow Systems â€“ An Agent-based Approach	2001
203	The Development of Access Control Policies for Information Technology Systems	2002
204	Building access control models with attribute exploration	2009
205	Real-time information integrity=system integrity+data integrity+continuous assurances	2005
206	Further analysis of password authentication schemes based on authentication tests	2004
207	A method for forensic analysis of control	2010
208	PING attack - How bad is it?	2006
209	Two proposed identity-based three-party authenticated key agreement protocols from pairings	2010
210	The design of a secure anonymous Internet voting system	2004
211	Robust remote authentication scheme with smart cards	2005
212	Capital market reaction to defective IT products: The case of computer viruses	2005
213	Semantic web-based social network access control	2011
214	Functional similarities between computer worms and biological pathogens	2007
215	Secure business application logic for e-commerce systems	2005
216	Information Security Governance: A model based on the Directâ€“Control Cycle	2006
217	Towards a location-based mandatory access control model	2006
218	An anonymous voting mechanism based on the key exchange protocol	2006

219	Value-focused assessment of ICT security awareness in an academic environment	2007
220	Feature deduction and ensemble design of intrusion detection systems	2005
221	Change trend of averaged Hurst parameter of traffic under DDOS flood attacks	2006
222	Towards information security behavioural compliance	2004
223	Security implications in RFID and authentication processing framework	2006
224	Critical study of neural networks in detecting intrusions	2008
225	Characterization of defense mechanisms against distributed denial of service attacks	2004
226	A novel mix-based location privacy mechanism in Mobile IPv6	2005
227	A Secure Identification and Key agreement protocol with user Anonymity (SIKA)	2006
228	Sinkhole intrusion in mobile ad hoc networks: The problem and some detection indicators	2005
229	Application of temporal and spatial role based access control in 802.11 wireless networks	2006
230	SSL/TLS session-aware user authentication revisited	2008
231	Attacks on the (enhanced) Yang-Shieh authentication	2003
232	A preliminary model of end user sophistication for insider threat prediction in IT systems	2005
233	Provably secure authenticated key exchange protocols for low power computing clients	2006
234	Real-time intrusion detection for high-speed networks	2005
235	Information systems security policies: a contextual perspective	2005
236	The Open Source approach â€“ opportunities and limitations with respect to security and privacy	2002
237	A qualitative study of usersâ€™ view on information security	2007
238	A resource-constrained group key agreement protocol for imbalanced wireless networks	2007
239	A framework for the governance of information security	2004
240	RBAC models â€“ concepts and trends	2003
241	Information security awareness in higher education: An exploratory study	2008
242	Query-directed passwords	2005
243	Radio frequency identification (RFID)	2006
244	Analysis of end user security behaviors	2005
245	Keyjacking: the surprising insecurity of client-side SSL	2005
246	Vulnerability forecasting - a conceptual model	2004

247	Certified electronic mail: Properties revisited	2010
248	New efficient user identification and key distribution scheme providing enhanced security	2004
249	User perceptions of security and usability of single-factor and two-factor authentication in automated telephone banking	2011
250	Teaching information systems security courses: A hands-on approach	2007
251	Use of K-Nearest Neighbor classifier for intrusion detection	2002
252	On Bricks and Walls: Why Building Secure Software is Hard	2002
253	A Simple Graphical Tool For Modelling Trust	2001
254	Security enhancement for the timestamp-based password authentication scheme using smart cards	2003
255	An adaptive method for anomaly detection in symmetric network traffic	2007
256	A behaviorist perspective on corporate harassment online: Validation of a theoretical model of psychological motives	2010
257	Authentication in a layered security approach for mobile ad hoc networks	2007
258	Comparative studies on authentication and key exchange methods for 802.11 wireless LAN	2007
259	SAD: web session anomaly detection based on parameter estimation	2004
260	A multinomial logistic regression modeling approach for anomaly intrusion detection	2005
261	Performance of the Java security manager	2005
262	Multi-dimensional credentialing using veiled certificates: Protecting privacy in the face of regulatory reporting requirements	2009
263	On the development of an internetwork-centric defense for scanning worms	2009
264	An efficient and fair buyerâ€“seller finger-printing scheme for large scale networks	2010
265	Adapting usage control as a deterrent to address the inadequacies of access controls	2009
266	Security issues in SCADA networks	2006
267	A new taxonomy of Web attacks suitable for efficient encoding	2003
268	Principles and requirements for a secure e-voting system	2002
269	Adaptable security mechanism for dynamic environments	2007
270	Enhanced three-party encrypted key exchange without server public keys	2004
271	Privacy-preserving programming using sython	2007
272	An oblique-matrix technique for data integrity assurance	2009
273	A secure extension of the Kwakâ€“Moon group signcryption scheme	2006
274	Hierarchical access control based on Chinese Remainder Theorem and symmetric algorithm	2002

275	Reducing threats from flawed security APIs: The banking PIN case	2009
276	Logical analysis of AUTHMAC_DH: a new protocol for authentication and key distribution	2004
277	Why users cannot use security	2005
278	An ID-based remote mutual authentication with key agreement scheme for mobile devices on elliptic curve cryptosystem	2009
279	Intrusion detection using text processing techniques with a kernel based similarity measure	2007
280	Layered security design for mobile ad hoc networks	2006
281	Efficient remote mutual authentication and key agreement	2006
282	Predation and the cost of replication: New approaches to malware prevention?	2006
283	Anti-keylogging measures for secure Internet login: An example of the law of unintended consequences	2007
284	An alternative architectural framework to the OSI security model	2004
285	Practical anonymous user authentication scheme with security proof	2008
286	RT-UNNID: A practical solution to real-time network-based intrusion detection using unsupervised neural networks	2006
287	ISRAM: information security risk analysis method	2005
288	Expected benefits of information security investments	2006
289	On the detection of pod slurping attacks	2010
290	Information Assurance for security protocols	2005
291	A survey and trends on Internet worms	2005
292	Comparing Java and .NET security: Lessons learned and missed	2006
293	A taxonomy of network and computer attacks	2005
294	A Novel Key Management Scheme Based on Discrete Logarithms and Polynomial Interpolations	2002
295	DNS-based email sender authentication mechanisms: A critical review	2009
296	Mining TCP/IP packets to detect stepping-stone intrusion	2007
297	A privacy-preserving clustering approach toward secure and effective data analysis for business collaboration	2007
298	Computer security impaired by legitimate users	2004
299	An active learning based TCM-KNN algorithm for supervised network intrusion detection	2007
300	Application-based anomaly intrusion detection with dynamic information flow analysis	2008
301	Self-efficacy in information security: Its influence on end usersâ€™ information security practice behavior	2009
302	Evaluating information security tradeoffs: Restricting access can interfere with user tasks	2007

303	Measuring, analyzing and predicting security vulnerabilities in software systems	2007
304	SDriver: Location-specific signatures prevent SQL injection attacks	2009
305	The use and usability of direction-based filtering in firewalls	2004
306	Securing communication using function extraction technology for malicious code behavior analysis	2009
307	Information security policy: An organizational-level process model	2009
308	Embedding biometric identifiers in 2D barcodes for improved security	2004
309	Anomaly-based network intrusion detection: Techniques, systems and challenges	2009
310	NetHost-Sensor: Investigating the capture of end-to-end encrypted intrusive data	2006
311	An approach to reliably identifying signs of DDOS flood attacks based on LRD traffic pattern recognition	2004
312	Reducing false positives in intrusion detection systems	2010
313	The impact that placing email addresses on the Internet has on the receipt of spam: An empirical analysis	2007
314	Profiling program behavior for anomaly intrusion detection based on the transition and frequency property of computer audit data	2006
315	A simple, configurable SMTP anti-spam filter: Greylists	2006
316	A practical key management scheme for access control in a user hierarchy	2002
317	Management of risk in the information age	2005
318	Holistic security requirement engineering for electronic commerce	2004
319	A prototype for assessing information security awareness	2006
320	Enterprise information security strategies	2008
321	An optimistic fair exchange protocol based on signature policies	2008
322	SEAS, a secure e-voting protocol: Design and implementation	2005
323	Security for a Multi-Agent System based on JADE	2007
324	Secure multiparty payment with an intermediary entity	2009
325	Can critical infrastructures rely on the Internet?	2005
326	The information security digital divide between information security managers and users	2009
327	Towards Web Service access control	2004
328	Trustedâ€¦orâ€¦trustworthy: the search for a new paradigm for computer and network security	2002
329	Probabilistic analysis of an algorithm to compute TCP packet round-trip time for intrusion detection	2007
330	Blind image steganalysis based on content independent statistical measures maximizing the specificity and sensitivity of the system	2009

331	Defending against spoofed DDoS attacks with path fingerprint	2005
332	Personalized cryptographic key generation based on FaceHashing	2004
333	How significant is human error as a cause of privacy breaches? An empirical study and a framework for error management	2009
334	Insider Threat Prediction Tool: Evaluating the probability of IT misuse	2002
335	Cryptanalysis of simple three-party key exchange protocol	2008
336	PAID: A Probabilistic Agent-Based Intrusion Detection system	2005
337	Implementing a passive network covert timing channel	2010
338	An improvement on efficient anonymous auction protocols	2005
339	Stability analysis of a SEIQV epidemic model for rapid spreading worms	2010
340	Information security management: An information security retrieval and awareness model for industry	2008
341	The effect of intrusion detection management methods on the return on investment	2004
342	Secure log management for privacy assurance in electronic communications	2008
343	Development of Information Security Baselines for Healthcare Information Systems in New Zealand	2002
344	A study on decision consolidation methods using analytic models for security systems	2007
345	A taxonomy and comparison of computer security incidents from the commercial and government sectors	2006
346	Detecting rogue access points using client-side bottleneck bandwidth analysis	2009
347	Empirical evaluation of SVM-based masquerade detection using UNIX commands	2005
348	New aspect-oriented constructs for security hardening concerns	2009
349	Information security culture: A management perspective	2010
350	Power system DNP3 data object security using data sets	2010
351	Improving information security awareness and behaviour through dialogue, participation and collective reflection. An intervention study	2010
352	Symbolic reachability analysis for parameterized administrative role-based access control	2011
353	A video game for cyber security training and awareness	2007
354	Specifying authentication using signal events in CSP	2009
355	Criteria to evaluate Automated Personal Identification Mechanisms	2008
356	Cyber security for home users: A new way of protection through awareness enforcement	2010
357	On the detection and identification of botnets	2010
358	A security privacy aware architecture and protocol for a single smart card used for multiple services	2010

359	Hybrid Key Escrow: A New Paradigm	2002
360	A schema for protecting the integrity of databases	2009
361	Peer-assisted carrying authentication (PACA)	2004
362	H2A: Hybrid Hash-chaining scheme for Adaptive multicast source authentication of media-streaming	2005
363	Automated containment of rootkits attacks	2008
364	Design of an enhancement for SSL/TLS protocols	2006
365	Acceptance of Subscriber Authentication Methods For Mobile Telephony Devices	2002
366	Phishing for user security awareness	2007
367	Extensions to the source path isolation engine for precise and efficient log-based IP traceback	2010
368	Multiple behavior information fusion based quantitative threat evaluation	2005
369	Retraining a keystroke dynamics-based authenticator with impostor patterns	2007
370	Opening Up The Enterprise	2000
371	Approach for selecting the most suitable Automated Personal Identification Mechanism (ASMSA)	2010
372	XML distributed security policy for clusters	2004
373	Extending the enforcement power of truncation monitors using static analysis	2011
374	A formal framework for real-time information flow analysis	2009
375	Keystroke dynamics-based authentication for mobile devices	2009
376	Improvement of keystroke data quality through artificial rhythms and cues	2008
377	User perceptions of security, convenience and usability for ebanking authentication tokens	2009
378	Building network attack graph for alert causal correlation	2008
379	A framework and taxonomy for comparison of electronic voting schemes	2006
380	Clustering subjects in a credential-based access control framework	2007
381	An ontology-based policy for deploying secure SIP-based VoIP services	2008
382	Predicting the intrusion intentions by observing system call sequences	2004
383	A survey of signature based methods for financial fraud detection	2009
384	Modeling and preventing TOCTTOU vulnerabilities in Unix-style file systems	2010
385	Why Johnny canâ€™t surf (safely)? Attacks and defenses for web users	2009
386	Efficient multi-server authentication scheme based on one-way hash function without verification table	2008

387	Anonymization models for directional location based service environments	2010
388	A model for deriving information security control attribute profiles	2003
389	A concise cost analysis of Internet malware	2009
390	Biometric attack vectors and defences	2007
391	Risk profiles and distributed risk assessment	2009
392	Securing The Electronic Market: The KEYSTONE Public Key Infrastructure Architecture	2000
393	Information security: The moving target	2009
394	Implementing enterprise security: a case study	2003
395	Scalable balanced batch rekeying for secure group communication	2006
396	A framework for behavior-based detection of user substitution in a mobile context	2007
397	Building lightweight intrusion detection system using wrapper-based feature selection mechanisms	2009
398	Improving user security behaviour	2003
399	Information security requirements â€“ Interpreting the legal aspects	2008
400	A personal mobile DRM manager for smartphones	2009
401	Detecting intrusion with rule-based integration of multiple models	2003
402	A real-time network intrusion detection system for large-scale attacks based on an incremental mining approach	2009
403	Issues and challenges in securing VoIP	2009
404	CIDS: An agent-based intrusion detection system	2005
405	On Incident Handling and Response: A state-of-the-art approach	2006
406	Assessing the security perceptions of personal Internet users	2007
407	Security middleware for enhancing interoperability of Public Key Infrastructure	2003
408	Advanced user authentication for mobile devices	2007
409	Providing secure execution environments with a last line of defense against Trojan circuit attacks	2009
410	A SIP-oriented SPIT Management Framework	2008
411	SMSSec: An end-to-end protocol for secure SMS	2008
412	Confidence in smart token proximity: Relay attacks revisited	2009
413	The challenges of understanding and using security: A survey of end-users	2006
414	Measuring IDS-estimated attack impacts for rational incident response: A decision theoretic approach	2009

415	Run-time label propagation for forensic audit data	2007
416	Recognition of electro-magnetic leakage information from computer radiation with SVM	2009
417	Access control for smarter healthcare using policy spaces	2010
418	Modeling and analyzing the spread of active worms based on P2P systems	2007
419	A survey of coordinated attacks and collaborative intrusion detection	2010
420	A framework and assessment instrument for information security culture	2010
421	Filtering XPath expressions for XML access control	2004
422	Is the mouse click mighty enough to bring society to its knees?	2003
423	A robust software watermarking for copyright protection	2009
424	Security and human computer interfaces	2003
425	A game-based intrusion detection mechanism to confront internal attackers	2010
426	A hybrid honeypot framework for improving intrusion detection systems in protecting organizational networks	2006
427	The Benefits of a Notification Process in Addressing the Worsening Computer Virus Problem: Results of a Survey and a Simulation Model	2001
428	Modeling the behavior of users who are confronted with security mechanisms	2011
429	Roving bugnet: Distributed surveillance threat and mitigation	2010
430	PKI-based trust management in inter-domain scenarios	2010
431	On the symbiosis of specification-based and anomaly-based detection	2010
432	Designing a cluster-based covert channel to evade disk investigation and forensics	2011
433	A qualitative analysis of software security patterns	2006
434	Security threats scenarios in trust and reputation models for distributed systems	2009
435	A taxonomy for information security technologies	2003
436	Survey of network security systems to counter SIP-based denial-of-service attacks	2010
437	A study of self-propagating mal-packets in sensor networks: Attacks and defenses	2011
438	Classifying data from protected statistical datasets	2010
439	Holistic security management framework applied in electronic commerce	2007
440	Steganographic Method for Secure Communications	2002
441	Human and organizational factors in computer and information security: Pathways to vulnerabilities	2009
442	Enforcing memory policy specifications in reconfigurable hardware	2008

443	Vulnerability Take Grant (VTG): An efficient approach to analyze network vulnerabilities	2007
444	Utilizing bloom filters for detecting flooding attacks against SIP based services	2009
445	Preventing massive automated access to web resources	2009
446	Applying digital rights management systems to privacy rights management	2002
447	Security analysis of GTRBAC and its variants using model checking	2011
448	Understanding usersâ€™ keystroke patterns for computer access security	2003
449	Giga Security	2002
450	An aspect-oriented approach for the systematic security hardening of code	2008
451	Evaluation of a low-rate DoS attack against application servers	2008
452	Cumulative notarization for long-term preservation of digital signatures	2004
453	Rico: a security proxy for mobile code	2004
454	Client-side cross-site scripting protection	2009
455	Containing large-scale worm spreading in the Internet by cooperative distribution of traffic filtering policies	2008
456	Fine-grained integration of access control policies	2011
457	A dynamic context-aware access control architecture for e-services	2006
458	An integral framework for information systems security management	2003
459	Utilizing fuzzy logic and trend analysis for effective intrusion detection	2003
460	Generalization of proxy signature-based on discrete logarithms	2003
461	Hybrid spam filtering for mobile communication	2010
462	A non-intrusive biometric authentication mechanism utilizing physiological characteristics of the human head	2007
463	SVision: A novel visual network-anomaly identification technique	2007
464	An analysis of the tools used for the generation and prevention of spam	2004
465	Runtime verification of cryptographic protocols	2010
466	Significantly improved performances of the cryptographically generated addresses thanks to ECC and GPGPU	2010
467	WhiteScript: Using social network analysis parameters to balance between browser usability and malware exposure	2011
468	Differentially secure multicasting and its implementation methods	2002
469	A feasible intrusion detector for recognizing IIS attacks based on neural networks	2008
470	Towards secure dynamic collaborations with group-based RBAC model	2009

471	Dynamic models for computer viruses	2008
472	An anomaly intrusion detection method by clustering normal user behavior	2003
473	Taming role mining complexity in RBAC	2010
474	Reconstruction of electronic signatures from eDocument printouts	2010
475	Pervasive authentication and authorization infrastructures for mobile users	2010
476	An incremental frequent structure mining framework for real-time alert correlation	2009
477	A framework for security assurance of access control enforcement code	2010
478	Fast detection and visualization of network attacks on parallel coordinates	2009
479	Collaborative privacy management	2010
480	Probabilistic model checking for the quantification of DoS security threats	2009
481	Digital signature of multicast streams secure against adaptive chosen message attack	2004
482	A generic mechanism for efficient authentication in B3G networks	2010
483	What the heck is this application doing? â€“ A security-by-contract architecture for pervasive services	2009
484	Legally â€œreasonableâ€ security requirements: A 10-year FTC retrospective	2011
485	Quantitative analysis of a certified e-mail protocol in mobile environments: A probabilistic model checking approach	2011
486	IP Traceback using header compression	2003
487	Standardizing vulnerability categories	2008
488	Modeling vulnerability discovery process in Apache and IIS HTTP servers	2011
489	Reliable and fully distributed trust model for mobile ad hoc networks	2009
490	A preliminary two-stage alarm correlation and filtering system using SOM neural network and K-means algorithm	2010
491	The Benefits of a Notification Process in Addressing the Worsening Computer Virus Problem: Results of a Survey and a Simulation Model	2002
492	Wavelet based Denial-of-Service detection	2006
493	The inference problem: Maintaining maximal availability in the presence of database updates	2010
494	Classification of web robots: An empirical study based on over one billion requests	2009
495	HMMPayl: An intrusion detection system based on Hidden Markov Models	2011
496	Understanding the limitations of S/MIME digital signatures for e-mails: A GUI based approach	2009
497	Real-time analysis of intrusion detection alerts via correlation	2006
498	RIP â€“ A robust IP access architecture	2009

499	Reverse OAuth: A solution to achieve delegated authorizations in single sign-on e-learning systems	2009
500	A secure peer-to-peer backup service keeping great autonomy while under the supervision of a provider	2010
501	Worm virulence estimation for the containment of local worm outbreak	2010
502	PENET: A practical method and tool for integrated modeling of security attacks and countermeasures	2009
503	Making security usable: Are things improving?	2007
504	Integrating Software Lifecycle Process Standards with Security Engineering	2002
505	An intruder model with message inspection for model checking security protocols	2010
506	A global security architecture for intrusion detection on computer networks	2008
507	WARP: A wormhole-avoidance routing protocol by anomaly detection in mobile ad hoc networks	2010
508	A probabilistic relational model for security risk analysis	2010
509	An operational model and language support for securing XML documents	2004
510	A framework of composable access control features: Preserving separation of access control concerns from models to code	2010
511	Managing key hierarchies for access control enforcement: Heuristic approaches	2010
512	Data remanence effects on memory-based entropy collection for RFID systems	2011
513	Dynamic subtree tracing and its application in pay-TV systems	2011
514	Generalizing cryptosystems based on the subset sum problem	2011
515	Secure localization with attack detection in wireless sensor networks.	2011
516	Is workplace surveillance legal in Canada?	2007
517	A note on the Ate pairing	2008
518	Requirements of federated trust management for service-oriented architectures	2007
519	At-private k-database information retrieval scheme	2001
520	A secret sharing scheme based on (t, n) threshold and adversary structure	2009
521	Counting equations in algebraic attacks on block ciphers	2010
522	Attacking a polynomial-based cryptosystem: Polly Cracker	2002
523	An attack on the isomorphisms of polynomials problem with one secret	2003
524	Biometric keys: suitable use cases and achievable information content	2009
525	Regular 2-ary right-to-left exponentiation algorithm with very efficient DPA and FA countermeasures	2010
526	Uncertain inference control in privacy protection	2009

527	All sail, no anchor II: Acceptable high-end PKI	2004
528	Computer security	2001
529	Building online trust through privacy practices	2007
530	Protocols useful on the Internet from distributed signature schemes	2004
531	Temporarily hidden bit commitment and lottery applications	2010
532	A general and efficient countermeasure to relation attacks in mix-based e-voting	2011
533	On the sequence of authorization policy transformations	2005
534	Efficient identity-based GQ multisignatures	2009
535	Breaking four mix-related schemes based on Universal Re-encryption	2007
536	Integrating software specifications into intrusion detection	2007
537	A new electronic check system with reusable refunds	2002
538	Hybrid authentication based on noisy channels	2003
539	A complete characterization of a family of key exchange protocols	2002
540	Efficient trace and revoke schemes	2010
541	Compliant cryptologic protocols	2002
542	Cryptanalysis of the new TTS scheme in CHES 2004	2006
543	Analysis of e-commerce protocols: Adapting a traditional technique	2003
544	Fair multi-party non-repudiation protocols	2003
545	Outbound authentication for programmable secure coprocessors	2004
546	Cryptoviral extortion using Microsoftâ€™s Crypto API. Can Crypto APIs help the enemy?	2006
547	Related-key rectangle attack on 36 rounds of the XTEA block cipher	2009
548	Key recovery for the commercial environment	2002
549	Analysing a stream authentication protocol using model checking	2004
550	A generalization of Paillierâ€™s public-key system with applications to electronic voting	2010
551	Theory and benefits of recursive certificate structures	2004
552	Bipartite modular multiplication with twice the bit-length of multipliers	2009
553	How to construct identity-based signatures without the key escrow problem	2010
554	Simplified security notions of direct anonymous attestation and a concrete scheme from pairings	2009

555	Zheng and Seberryâ€™s public key encryption scheme revisited	2003
556	Conference key agreement protocol with non-interactive fault-tolerance over broadcast network	2009
557	Security analysis of CRT-based cryptosystems	2006
558	Mobile Web services authentication using SAML and 3GPP generic bootstrapping architecture	2009
559	The efficiency of solving multiple discrete logarithm problems and the implications for the security of fixed elliptic curves	2004
560	A method for identifying Web applications	2009
561	Opacity generalized to transition systems	2008
562	A key pre-distribution scheme for wireless sensor networks: merging blocks in combinatorial design	2006
563	A coding approach to the multicast stream authentication problem	2008
564	Specifying and implementing privacy-preserving cryptographic protocols	2008
565	IDSIC: an intrusion detection system with identification capability	2008
566	A familiarity-based trust model for effective selection of sellers in multiagent e-commerce systems	2007
567	Improving cache attacks by considering cipher structure	2006
568	Low-randomness constant-round private XOR computations	2007
569	Intrusion and intrusion detection	2001
570	Delegation in role-based access control	2008
571	Generating visible RSA public keys for PKI	2004
572	PBAC: Provision-based access control model	2002
573	Efficient generation of secure elliptic curves	2007
574	Secure universal designated verifier signature without random oracles	2008
575	An algebraic approach to the verification of a class of Diffie-Hellman protocols	2007
576	Automated analysis of timed security: a case study on web privacy	2004
577	Flowchart description of security primitives for controlled physical unclonable functions	2010
578	Efficient online/offline identity-based signature for wireless sensor network	2010
579	A comprehensive simulation tool for the analysis of password policies	2009
580	Securing SOAP e-services	2002
581	A monitoring system for detecting repeated packets with applications to computer worms	2006
582	Conditional reactive simulatability	2008

583	Modeling contextual security policies	2008
584	PolicyUpdater: a system for dynamic access control	2006
585	Modification and optimization of a shuffling scheme: stronger security, formal analysis and higher efficiency	2011
586	How to obtain full privacy in auctions	2006
587	On hash functions using checksums	2010
588	On the security of the WinRAR encryption feature	2006
589	Networked cryptographic devices resilient to capture	2003
590	Privacy-preserving revocation checking	2009
591	Hierarchical time-based information release	2006
592	Achieving evenhandedness in certified email system for contract signing	2008
593	Two-party generation of DSA signatures	2004
594	Password hardening based on keystroke dynamics	2002
595	Escrow-free encryption supporting cryptographic workflow	2006
596	A probabilistic model for optimal insurance contracts against security risks and privacy violation in IT outsourcing environments	2007
597	Identity-based cryptography for grid security	2011
598	Verifiable shuffles: a formal model and a Paillier-based three-round construction with provable security	2006
599	Listega: list-based steganography methodology	2009
600	Execution transactions for defending against software failures: use and evaluation	2006
601	Secure outsourcing of sequence comparisons	2005
602	Content-based filtering of Web documents: the MaX system and the EUFORBIA project	2003
603	On the security of fair non-repudiation protocols	2005
604	The principle of guarantee availability for security protocol analysis	2010
605	A survey of certificateless encryption schemes and security models	2008
606	Identity-based key agreement protocols from pairings	2007
607	Keeping secrets in incomplete databases	2008
608	Using AVL trees for fault-tolerant group key management	2002
609	Intrusion detection for secure clustering and routing in Mobile Multi-hop Wireless Networks	2007
610	Limits of the BRSIM/UC soundness of Dolevâ€“Yao-style XOR	2008

611	Large-scale network intrusion detection based on distributed learning algorithm	2009
612	Authenticating mobile phone users using keystroke analysis	2007
613	A column dependency-based approach for static and dynamic recovery of databases from malicious transactions	2010
614	A hybrid scheme for securing fingerprint templates	2010
615	Estimating the maximum information leakage	2008
616	Adding support to XACML for multi-domain user to user dynamic delegation of authority	2009
617	Rigorous automated network security management	2005
618	Audit-based compliance control	2007
619	Fujisakiâ€“Okamoto hybrid encryption revisited	2005
620	A multi-layer framework for puzzle-based denial-of-service defense	2008
621	PKI design based on the use of on-line certification authorities	2004
622	Reactively secure signature schemes	2005
623	Bayesian rational exchange	2008
624	A new probabilistic rekeying method for secure multicast groups	2010
625	On the decidability of cryptographic protocols with open-ended data structures	2005
626	Principles of remote attestation	2011
627	Proposal on personal identifiers generated from the STR information of DNA	2002
628	Content-based image authentication: current status, issues, and challenges	2010
629	COVERAGE: detecting and reacting to worm epidemics using cooperation and validation	2007
630	Secure steganography based on embedding capacity	2009
631	Ensuring security in depth based on heterogeneous network security technologies	2009
632	The Elliptic Curve Digital Signature Algorithm (ECDSA)	2001
633	A lambic: a privacy-preserving recommender system for electronic commerce	2008
634	A graph-theoretical model of computer security	2004
635	Denial of service attacks and defenses in decentralized trust management	2009
636	Trust structures	2007
637	Analysis and application of Bio-Inspired Multi-Net SecurityModel	2010
638	Handling distributed authorization with delegation through answer set programming	2007

639	Offline firewall analysis	2006
640	Analyzing SLE 88 memory management security using Interacting State Machines	2005
641	Probabilistic timing covert channels: to close or not to close?	2011
642	Detection of file-based race conditions	2005
643	Secure group key establishment revisited	2007
644	SilentKnock: practical, provably undetectable authentication	2009
645	Key substitution attacks revisited: Taking into account malicious signers	2006
646	User interface design affects security: patterns in click-based graphical passwords	2009
647	Interactive two-channel message authentication based on Interactive-Collision Resistant hash functions	2009
648	A calculus for control flow analysis of security protocols	2004
649	A case study in hardware Trojan design and implementation	2011
650	Transformational typing and unification for automatically correcting insecure programs	2007
651	Enabling shared audit data	2005
652	A symbolic framework for multi-faceted security protocol analysis	2008
653	An incentive compatible reputation mechanism for ubiquitous computing environments	2007
654	A new hardware-assisted PIR with O(n) shuffle cost	2010
655	Discretionary capability confinement	2008
656	An analysis of accuracy experiments carried out over of a multi-faceted model of trust	2009
657	CRUST: cryptographic remote untrusted storage without public keys	2009
658	Game strategies in network security	2005
659	Formal validation of automated policy refinement in the management of network security systems	2010
660	A delegation model for extended RBAC	2010
661	Passive classification of wireless NICs during active scanning	2008
662	Computational probabilistic noninterference	2004
663	Elastic block ciphers: method, security and instantiations	2009
664	A formal graph based framework for supporting authorization delegations and conflict resolutions	2003
665	Management of access control policies for XML document sources	2003
666	Dynamic security labels and static information flow control	2007

667	Protecting data privacy through hard-to-reverse negative databases	2007
668	An overview of the verification of SET	2005
669	Database intrusion detection using sequence alignment	2010
670	Intrusion detection in voice over IP environments	2009
671	A framework for secure execution of software	2004
672	Controlled query evaluation for enforcing confidentiality in complete information systems	2004
673	Proposal on a multifactor biometric authentication method based on cryptosystem keys containing biometric signatures	2005
674	Using contourlet transform and cover selection for secure steganography	2010
675	Static use-based object confinement	2005
676	SAT-based model-checking for security protocols analysis	2008
677	Minimizing TTP's involvement in signature validation	2006
678	Negative representations of information	2009
679	A distributed digital rights management model for secure information-distribution systems	2004
680	OFMC: A symbolic model checker for security protocols	2005
681	Embedding renewable cryptographic keys into noisy data	2010
682	Flow-sensitive, context-sensitive, and object-sensitive information flow control based on program dependence graphs	2009
683	Symmetric authentication in a simulatable Dolevâ€“Yao-style cryptographic library	2005
684	Password-authenticated key exchange based on RSA	2010
685	Requirements engineering for trust management: model, methodology, and reasoning	2006
686	A system for securing push-based distribution of XML documents	2007
687	Certificate revocation system implementation based on the Merkle hash tree	2004
688	Understanding SPKI/SDSI using first-order logic	2006
689	Stack inspection and secure program transformations	2004
690	Enhancing grid security by fine-grained behavioral control and negotiation-based authorization	2009
691	Complete analysis of configuration rules to guarantee reliable network security policies	2008
692	Instruction-level security typing by abstract interpretation	2007
693	EXAM: a comprehensive environment for the analysis of access control policies	2010
694	Generalized XML security views	2009

695	PKI past, present and future	2006
696	Distributing security-mediated PKI	2006
697	A cognitive walkthrough of Autopsy Forensic Browser	2009
698	Using penetration testing feedback to cultivate an atmosphere of proactive security amongst end-users	2009
699	How well are information risks being communicated to your computer end-users?	2007
700	Understanding and transforming organizational security culture	2010
701	Impact of perceived technical protection on security behaviors	2009
702	Analysis of security-relevant semantics of BPEL in cross-domain defined business processes	2007
703	An empirical study of information security policy on information security elevation in Taiwan	2006
704	On the imbalance of the security problem space and its expected consequences	2007
705	Consumer motivations in taking action against spyware: an empirical investigation	2009
706	How perceptions of justice affect security attitudes: suggestions for practitioners and researchers	2009
707	Designing and aligning e-Science security culture with design	2010
708	Brand, knowledge, and false sense of security	2010
709	A framework for outsourcing IS/IT security services	2006
710	Justifying the need for a data security management plan for the UAE	2010
711	Trust, privacy, and security in electronic business: the case of the GCC countries	2006
712	A vocabulary test to assess information security awareness	2010
713	Human-related problems of information security in East African cross-cultural environments	2010
714	A review and future research directions of secure and trustworthy mobile agent-based e-marketplace systems	2010
715	Information security governance in Saudi organizations: an empirical study	2010
716	Expanding topological vulnerability analysis to intrusion detection through the incident response intelligence system	2010
717	Personalized cognitive passwords: an exploratory assessment	2011
718	A survey of intrusion detection and prevention systems	2010
719	A security standards' framework to facilitate best practices' awareness and conformity	2010
720	Security of personal data across national borders	2002
721	Evaluating the security controls of CAIS in developing countries	2007
722	Process-variance models in information security awareness research	2008

723	Information technology security management concerns in global financial services institutions Is national culture a differentiator?	2009
724	Management versus security specialists: an empirical study on security related perceptions	2008
725	A multidimensional approach to multilevel security	2008
726	Evaluating the security controls of CAIS in developing countries: an empirical investigation	2007
727	On security preparations against possible IS threats across industries	2006
728	Toward viable information security reporting systems	2007
729	Implementation and effectiveness of organizational information security measures	2008
730	A partial equilibrium view on security and privacy	2008
731	Electronic commerce: the information-security challenge	2000
732	Security-related concerns with geographic information systems and geographic mapping	2000
733	A test of interventions for security threats from social engineering	2008
734	Issues with information dissemination on global networks	2000
735	Formalizing information security requirements	2001
736	Security information management as an outsourced service	2006
737	Robust intrusion tolerance in information systems	2001
738	Information privacy compliance in the healthcare industry	2008
739	An integrated system theory of information security management	2003
740	Smart card technology for deploying a secure information management framework	2000
741	Biometrics in banking security: a case study	2008
742	Distributed component software security issues on deploying a secure electronic marketplace	2000
743	Information systems security in the Greek public sector	2001
744	A conceptual architecture for real-time intrusion monitoring	2000
745	Effects on employees' information security abilities by e-learning	2009
746	Awareness and challenges of Internet security	2000
747	A conceptual foundation for organizational information security awareness	2000
748	Improving protection mechanisms by understanding online risk	2007
749	An information privacy taxonomy for collaborative environments	2006
750	Critical analysis of different approaches to minimizing user-related faults in information systems security: implications for research and practice	2000

751	Incident response requirements for distributed security information management systems	2007
752	Shielding your company against information compromise	2000
753	Assessing and improving authentication confidence management	2010
754	Protecting personal privacy on the Internet	2002
755	Securing SCADA systems	2008
756	User-centred security applied to the development of a management information system	2007
757	Identifying security vulnerabilities through input flow tracing and analysis	2003
758	Vulnerability analysis and the practical implications of a server-challenge-based one-time password system	2010
759	Assessing image-based authentication techniques in a web-based environment	2010
760	Towards an insider threat prediction specification language	2006
761	Personalized biometric key using fingerprint biometrics	2007
762	Issue report on business adoption of Microsoft Passport	2006
763	System insecurity - firewalls	2002
764	Security guideline tool for home users based on international standards	2010
765	Controlling corporate e-mail, PC use and computer security	2001
766	An agent-based privacy-enhancing model	2008
767	Design of a multimedia traffic classifier for Snort	2007
768	Role of trust in e-business success	2004
769	Ontologies for information security management and governance	2008
770	Towards privacy in personal data management	2009
771	Inter-organizational intrusion detection using knowledge grid technology	2006
772	Reaching escape velocity. A practiced approach to information security management system implementation	2008
773	A strategic modeling technique for information security risk assessment	2007
774	Information security technologies as a commodity input	2005
775	Issues in cybersecurity: Understanding the potential risks associated with hackers/crackers	2002
776	Deployment of anti-virus software: A case study	2003
777	Attitudes of Australian information system managers against online attackers	2001
778	Maintaining privacy in anomaly-based intrusion detection systems	2005

779	Using continuous user authentication to detect masqueraders	2003
780	The utilization of trend analysis in the effective monitoring of information security. Part 1: the concept	2001
781	Internet anonymity practices in computer crime	2003
782	Dynamic content attacks on digital signatures	2005
783	A STOPE model for the investigation of compliance with ISO 17799-2005	2007
784	The importance of technology trust in Web services security	2002
785	System survivability: a critical security problem	2005
786	Evaluation and selection of an antivirus and content filtering software	2002
787	Domain names management: A strategy for electronic commerce security	2001
788	Mitigating the mobile agent malicious host problem by using communication patterns	2005
789	An improved two-tiered strategy to intrusion detection	2005
790	Intrusion detection: methods and systems. Part II	2003
791	On biometrics-based authentication and identification from a privacy-protection perspective Deriving privacy-enhancing requirements	2004
792	The utilization of trend analysis in the effective monitoring of information security. Part 2: the model	2002
793	A new model for monitoring intrusion based on Petri Nets	2001
794	Biometric authentication: Assuring access to information	2002
795	Security architectures for network clients	2003
796	Intelligent authentication, authorization, and administration (I3A)	2005
797	Towards secure sealing of privacy policies	2004
798	The information security management toolbox - taking the pain out of security management	2002
799	Security analyzers: Administrator assistants or hacker helpers?	2001
800	The application of information security policies in large UK-based organizations: an exploratory investigation	2003
801	A security risk management approach for e-commerce	2003
802	An exploration of wireless computing risks: Development of a risk taxonomy	2004
803	Information systems security from a knowledge management perspective	2005
804	A good-practice guidance on the use of PKI services in the public sector of the European Union member states	2005
805	Privacy and security concerns as major barriers for e-commerce: A survey study	2001
806	Private key generation from on-line handwritten signatures	2002

807	Towards maturity of information security maturity criteria: six lessons learned from software maturity criteria	2002
808	Quantifying the financial impact of IT security breaches	2003
809	Networks security measures using neuro-fuzzy agents	2003
810	Intrusion detection: The art and the practice. Part I	2003
811	E-commerce policies and customer privacy	2003
812	Cyberterrorism, computer crime, and reality	2004
813	Integrating security design into the software development process for e-commerce systems	2001
814	A new two-tiered strategy to intrusion detection	2004
815	A framework for analyzing e-commerce security	2002
816	E-enterprise security management life cycle	2005
817	Information systems security issues and decisions for small businesses: An empirical examination	2005
818	Usable set-up of runtime security policies	2007
819	A novel approach for regenerating a private key using password, fingerprint and smart card	2005
820	The biometric technologies business case: a systematic approach	2005
821	Novel biometric digital signatures for Internet-based applications	2001
822	An experimental comparison of secret-based user authentication technologies	2002
823	An automated framework for managing security vulnerabilities	2005
824	Incorporating WS-Security into a Web services-based portal	2004
825	Intelligent authentication, authorization, and administration (I3A)	2006
826	Information security: management's effect on culture and policy	2006
827	Embedding privacy in IT applications development	2004
828	Enhancing Web privacy and anonymity in the digital era	2004
829	Quantification, Optimization and Uncertainty Modeling in Information Security Risks: A Matrix-Based Approach	2002
830	Investigating the Impact of Publicly Announced Information Security Breaches on Three Performance Indicators of the Breached Firms	2009
831	Ensuring correctness, completeness, and Freshness for Outsourced Tree-Indexed Data	2008
832	Internet privacy: Interpreting key issues	2001
833	Examining digital piracy: Self-Control, Punishment, and Self-Efficacy	2009
834	On the Role of Human Mortality in Information System Security: From the Problems of Descriptivism to Non-Descriptive Foundations	2001

835	Social Issues in Electronic Commerce: Implications for Policy Makers	2001
836	Privacy Rights Management: Implementation Scenarios	2007
837	Information Technology as a Target and Shield in the Post 9/11 Environment	2005
838	The Detection of Data Errors in Computer Information Systems: Field Interviews with Municipal Bond Analysts	2000
839	Internet Privacy Policies: A Review and Survey of the Fortune 50	2005
840	Do Information Security Policies Reduce the Incidence of Security Breaches: An Exploratory Analysis	2005
841	Value-focused assessment of information system security in organizations	2006
842	Current directions in IS security research: towards socio-organizational perspectives	2001
843	On the security of open source software	2002
844	User behaviour towards protective information technologies: the role of national cultural differences	2009
845	Internet Usersâ€™ Information Privacy Concerns (IUIPC): The Construct, the Scale, and a Causal Model	2004
846	Choice and Chance: A Conceptual Model of Paths to Information Security Compromise	2009
847	An Empirical Analysis of Software Vendorsâ€™ Patch Release Behavior: Impact of Vulnerability Disclosure	2010
848	The Value of Intrusion Detection Systems in Information Technology Security Architecture	2005
849	The Economic Incentives for Sharing Security Information	2005
850	User Awareness of Security Countermeasures and Its Impact on Information Systems Misuse: A Deterrence Approach	2009
851	Managing Digital Piracy: Pricing and Protection	2004
852	An Extended Privacy Calculus Model for E-Commerce Transactions	2006
853	Let the Pirates Patch? An Economic Analysis of Software Security Patch Restrictions	2008
854	Managing Piracy: Pricing and Sampling Strategies for Digital Experience Goods in Vertically Segmented Markets	2005
855	A Value-at-Risk Approach to Information Security Investment	2008
856	The Security of Confidential Numerical Data in Databases	2002
857	The Random Oracle Methodology, Revisited	2004
858	Unconditional Security in Quantum Cryptography	2001
859	Beyond Separation of Duty: An Algebra for Specifying High-Level Security Policies	2008
860	Beyond Proof-of-Compliance: Security Analysis in Trust Management	2005
861	Efficient and Secure Authenticated Key Exchange Using Weak Passwords	2009
862	User-friendly and certificate-free grid security infrastructure	2011

863	Six Design Theories for IS Security Policies and Guidelines	2006
864	An Experimental Study on Ubiquitous Commerce Adoption: Impact of Personalization and Privacy Concerns	2008
865	Theoretical Explanations for Firmsâ€™ Information Privacy Behaviors	2005
866	Understanding Security Behaviors in Personal Computer Usage: A Threat Avoidance Perspective	2010
867	The Centrality of Awareness in the Formation of User Behavioral Intention toward Protective Information Technologies	2007
868	Toward Building Self-Sustaining Groups in PCR-based Tasks through Implicit Coordination: The Case of Heuristic Evaluation	2009
869	A Design Theory for Secure Information Systems Design Methods	2006
870	A Behavioral Analysis of Passphrase Design and Effectiveness	2009
871	Concern for Information Privacy and Online Consumer Purchasing	2006
872	A Semantic Approach to Secure Collaborative Inter-Organizational eBusiness Processes (SSCIOBP)	2008
873	Strategic, Tactical, & Operational Management Security Model	2009
874	Impact of Negative Message Framing on Security Adoption	2010
875	Understanding Situational Online Information Disclosure as a Privacy Calculus	2010
876	Social Network Analysis of a Criminal Hacker Community	2010
877	Computer Forensics: The Modern Crime Fighting Tool	2006
878	Protecting Personal Information Online: A survey of user privacy concerns and control techniques	2004
879	An Analysis of Security Threats of Electronic Election Systems	2006
880	A Block Cipher Based Upon Permutation, Substitution and Iteration	2007
881	Quest for Universal Identification - A Commentary	2009
882	Consumer's PCs: A Study of Hard Drive Forensics, Data Recovery, and Exploitation	2008
883	An Empirical Assessment of Factors Impeding Effective Password Management	2008
884	A Holistic Perspective on the Science of Computer Forensics	2005
885	Security Service Packages: Partitioning the Security Space	2006
886	Introducing the Information Technology Security Essential Body of Knowledge Framework	2009
887	Managing Privacy of User Generated Information in a Web 2.0 World	2010
888	Intrusion Detection Management System for eCommerce Security	2007
889	Customer Information: Protecting the Organization's Most Critical Asset from Misappropriation and Identity Theft	2006
890	Profiling User Behavior for Intrusion Detection Using Item Response Modeling	2007

891	A Cross-Cultural Analysis of Privacy Notices of the Global 2000	2007
892	Passwords: Do User Preferences and Website Protocols Differ From Theory?	2010
893	Information Technology Offshore Outsourcing Security Risks and Safeguards	2010
894	Growth Perspective of Information Security	2009
895	An Investigative Study: Health Care Workers as Security Threat Suppliers	2007
896	An Empirical Investigation of Factors Influencing Information Security Behavior	2008
897	Exoinformation Space Audits: An Information Richness View of Privacy and Security Obligations	2007
898	User Acceptance of Speech-Enabled Technologies for Configuration of Computer and Network Security	2010
899	Security Breaches, Privacy Intrusions, and Reporting of Computer Crimes	2005
900	Prime III: Defense-in-Depth Approach to Electronic Voting	2008
901	Making Sound Security Investment Decisions	2010
902	An Architecture for Automatic and Adaptive Defense	2007
903	A Research Framework for Information Systems Security	2006
904	An XML-Based Intelligent Agent Protocol Design Framework for Individualized Privacy Postures within Trusted Network Environments	2006
905	Misuse Cases for Identifying System Dependability Threats	2008
906	Examining the Impact of E-privacy Risk Concerns on Citizens' Intentions to Use E-government Services: An Oman Perspective	2009
907	Is Your Email Box Safe?	2010
908	Business and Security Executives Views of Information Security Investment Drivers: Results from a Delphi Study	2009
909	Determinants of Behavioral Responses to Online Privacy: The Effects of Concern, Risk Beliefs, Self-Efficacy, and Communication Sources on Self-Protection Strategies	2010
910	Employee Information Privacy Concerns with Employer Held Data: A Comparison of Two Prevalent Privacy Models	2010
911	The Role of Emotions in Shaping Consumers' Privacy Beliefs about Unfamiliar Online Vendors	2008
912	RFID Privacy Issues in Healthcare: Exploring the Roles of Technologies and Regulations	2010
913	Looking for Love in All the Wrong Places: A Security Case Study on Online Identity Theft	2007
914	Consumer Perception of Web Site Security Attributes	2010
915	Security and Privacy Governance: Criteria for Systems Design	2009
916	Do Concerns about Error in Data and Access to Data Affect Students' Feeling of Alienation?	2006
917	Dimensions of Network Security Planning For Web Services	2005

918	Information Assurance Technical Framework and End User Information Ownership: A Critical Analysis	2005
919	A Proposed Integrated Framework for Coordinating Computer Security Incident Response Team	2005
920	An Investigative Study: Consumers Password Choices on an E-Commerce Site	2005
921	Secure Web service-based resource sharing in ERP Networks	2005
922	Perceptions of Information Security in the Workplace: Linking Information Security Climate to Compliant Behavior	2005
923	Rating Certificate Authorities: A Market Approach to the Lemons Problem	2006
924	Case Study - The Case of a Computer Hack	2005
925	Ethics and Morality - a business opportunity for the Amoral?	2007
926	How Secure is Your Password? An Analysis of E-Commerce Passwords and their Crack Times	2006
927	The Ephemerizer: Making Data Disappear	2005
928	Systemic Risk redefining Digital Security	2005
929	Building User Authentication in An Inter-Organizational Information System	2006
930	RFID: a Systematic Analysis of Privacy Threats & A 7-point Plan to Address Them	2005
931	Managing Information Security: Demystifying the Audit Process for Security Officers	2006
932	Towards a Global Framework for Corporate and Enterprise Digital Policy Management	2006
933	Security Consistency in Information Ecosystems: Structuring the Risk Environment on the Internet	2006
934	Botnets: The Anatomy of a Case	2005
935	To opt-in, or to opt-out? That is the question. A Case Study	2006
936	An Evaluation of Size-based Traffic. Feature for Intrusion Detection	2007
937	Integrating Disaster Recovery Plan Activities into the System Development Life Cycle	2010
938	The Duality of Information Security Management: Fighting against Predictable and Unpredictable Threats	2008
939	Do anti-spam measures effectively cover the e-mail communication network? A formal approach	2007
940	Information Warfare: A Comparative Framework for Business Information Security	2005
941	A Model for Predicting Hacker Behavior	2007
942	Consideration of Risks and Internal Controls in Business Process Modelling	2009
943	SoapSY - Unifying Security Data from Various Heterogeneous Distributed Systems into a Single Database Architecture	2005
944	The Risk of Computerized Bureaucracy	2009
945	Organizational Learning for the Incident Management Process: Lessons from High Reliability Organizations	2008

946	The Effect of Spam and Privacy Concerns on E-mail Users' Behavior	2007
947	Workflows in Dynamic and Restricted Delegation	2009
948	Anchoring Information Security Governance Research: Sociological Groundings and Future Directions	2006
949	WIDS - A Wireless Intrusion Detection System for Detecting Man-in-the-middle Attacks	2005
950	The Ethics of IT Disaster Recovery Planning: Five Case Studies	2008
951	Methodology to Assess the Impact of Investments in Security Tools and Products	2005
952	An Exploratory Delphi Study among Small Business Executives on Adoption of Disaster Recovery Practices	2009
953	Just Trying to Be Friendly: A Case Study in Social Engineering	2008
954	Top Management Support, External and Internal Organizational Collaboration, and Organizational Flexibility in Preparation for Extreme Events	2009
955	The Effects of Security Education Training and Awareness Programs and Individual Characteristics on End User Security Tool Usage	2009
956	Threat Modeling the Enterprise	2009
957	Security Issues And Capabilities Of Mobile Brokerage Services And Infrastructures	2006
958	Mitigating Consumer Perceptions of Privacy and Security Risks with the Use of Residual RFID Technologies through Governmental Trust	2008
959	Addressing Internal Security Threats with Roaming User-Based Distributed Firewalls	2009
960	Challenges in Managing Information Security in Academic Institutions: Case of MDI in India	2007
961	Towards a Best Fit Between Organizational Security Countermeasures and Information Systems Misuse Behaviors	2007
962	Perceptual and Cultural Aspects of Risk Management Alignment: a case study	2008
963	Vulnerabilities and Patches of Open Source Software: An Empirical Study	2008
964	A Conceptual Model for Integrative Information Systems Security	2006
965	Incident Response Planning Using Collaboration Engineering Process Development and Validation	2008
966	Risk Management Standards - The Perception of ease of use	2010
967	Privacy-preserving discriminatory and nondiscriminatory pricing based electronic market clearing mechanisms	2010
968	Security, Privacy and Politics in India: A Historical Review	2010
969	Supporting Intrusion Detection Work Practice	2009
970	PriS-Tool: A Case Tool For Privacy-Oriented Requirements Engineering	2010
971	Proving Correctness of SCADA Security-Enhancement Models	2010
972	Security of Information Systems in Schools: An Evaluation using Audit and COBIT Interviews	2010
973	On Retaining Data Control to the Client in Infrastructure Clouds	2009

974	Aligning Security Awareness with Information Systems Security Management	2010
975	The Power of Discretion in IS Security	2010
976	A System Dynamics Model of Information Security Investments	2010
977	MILD DSS - Conceptual Architecture, Validation and Representation	2008
978	Automatically Compute Information Flow Quantity via Probabilistic Semantics	2009
979	On Security Metaphors and how they Shape the Emerging Practice of Secure Information Systems Development	2007
980	Mitigating Disaster: Improvising Information Technology in Response to Extreme Events	2009
981	A Framework to Facilitate Forensic Investigation of Falsely Advertised BGP Routes	2007
982	Modelling corporate wireless security and privacy	2005
983	Preventing application software piracy: An empirical investigation of technical copy protections	2007
984	Trustworthiness in electronic commerce: the role of privacy, security, and site attributes	2002
985	The role of external and internal influences on information systems security â€“ a neo-institutional perspective	2007
986	Internet privacy concerns and beliefs about government surveillance â€“ An empirical investigation	2008
987	Online trust: a stakeholder perspective, concepts, implications, and future directions	2002
988	Metrics for characterizing the form of security policies	2010
989	RFID, privacy and the perception of risk: A strategic framework	2007
990	Effective Information Security Requires a balance of social and technology factors	2010
991	Improving employees` compliance through information systems security training: An action research study	2010
992	User participation in Information Systems Security Risk Management	2010
993	Neutralization: New insights into the problem of employee information systems security policy violations	2010
994	Information Security Policy Compliance: An empirical study of rationality-based beliefs and information security awareness	2010
995	Avoidance of information technology threats: A theoretical perspective	2009
996	Fear appeals and information security behaviors: An empirical study	2010
997	The impact of malicious agents on the enterprise software industry	2010
998	Correlated failures, diversifications, and information security risk management	2011
999	Circuits of Power: A study of mandated compliance to an information systems security de jure standard in a government organization	2010
1000	Circuits of Power in Creating De Jure Standards: Shaping and international information systems security standard	2006

Apêndice C – Representação Árvore de *Keywords*

O próximo gráfico representa toda a estrutura da árvore de *keywords*. A estrutura é composta pelas *keywords* que estão dispostas pelas 85 classes de acordo com os 6 níveis de índice de especificidade.

Classe	Nível 1	Nível 2	Nível 3	Nível 4	Nível 5	Nível 6
1	Information Security managers Security professionals	Security agents				
		Security auditors (Sas)				
		Security Executives				
		Computer Investigative Specialists				
		Security manager				
2	Business Security	Small business security				
3	Security principles	Integrity	Information integrity			
			Storage integrity			
			File integrity			
			systemic integrity			
		Confidentiality	Inferential Disclosure			
			Information disclosure			
			Disclosure inference			
		Availability				
		Non repudiation Non-repudiation Nonrepudiation	Fair non-repudiation			
4	Threat	Technology threat				

		avoidance theory	
	Threats	unpredictable threats	
	Information Security threats	Threats taxonomy	STRIDE
	Security threats	Threat Analysis	
	Computer threats	Threat appraisal	
		Threat evaluation	
		Threat Modeling	
		DREAD	
		Threat prediction	Intrusion intention predicition
		Malicious peers	
		End user threats	
5	Vulnerability Vulnerabilities	Vulnerability assessment	Red teams
	Security vulnerability	Harmonised vulnerability categories	
	Security vulnerabilities	Common vulnerabilities and exposures (CVE) list	
	Information Security vulnerabilities	Vulnerability analysis	
		Security holes	
		Vulnerability database	
		Vulnerability discovery model (VDM)	
		Vulnerability forecasting	
		Vulnerability scanner (VS) vulnerability scanners Vulnerability scanners (VSs)	
		Information loss	
		Information leakage	

		Vulnerability Take-Grant model			
6	Anomaly	Anomaly detection			
		Anomaly visualization			
		Intrusive anomalies			
7	Attack Attacks Security attacks Computer attack	Intrusion	Intrusion attacks		
			Intrusion event		
			Network intrusions		
		Attack categories			
		Network attack Network attacks	network contamination		
			MAC address spoofing		
			Network covert channel Covert channels Covert communication	Covert timing channels	Passive covert timing channel
					Covert-timing channel free security
				Subliminal channel	
		Cache attack			
		Stepping-stone			
		Wormhole attacks			
		Attack description			
		Attack detection	visualization (PCAV)		
		Attack graph			
		Buffer overflow Buffer overflows			
		violation of safeguards			
		Cross-site scripting (XSS)			
		Attack			

		indicators	
		Multiphase attack analysis	
		Attack Modeling	
		Attack probabilities	
		Attack Response	
		Attack scenario Attack scenarios	Attack scenario extraction
		Attack target	
		Attack tree Attack trees	
		Fault attack	Fault tolerance Fault-tolerance
		Attack vector	
		Attacker model	
		Monotonic assumption	
		Minrank	
		Dictionary attack Dictionary attacks	Online dictionary attacks
		Interdependent strategies	
		Social media attacks	
		Forgery attack	
		Generic attacks	
		Internal attacker Internal attacks	
		Espionage	
		Spoofing	Tampering modes
		Web attacks	
		Code injection	
		VoIP attacks	
		Man-in-the-middle attack	
		Mass attacks	
		Relay attack	
		Relation attack	

		Side channel attack Side channel attacks			
		Timing attack			
		Algebraic attacks			
		targeted attacks			
		Identity Theft			
		Processor exhaustion			
		Resource exhaustion			
		Breach Security breach Security Breaches	data breaches		
		PING attack			
		Differential power attack			
		Denial of service DoS	Denial-of-Service (DoS) attack Denial-of-service attacks DoS attacks	Flooding attacks	Bloom filter
			Distributed Denial of Service (DDoS) Distributed Denial-of-Service DDoS DDoS/DoS	DDoS attack DDoS attacks Distributed denial-of-service attacks Distributed denial of service attacks	Distributed denial-of-service flood attacks
		Low-rate attack			
		Related-key rectangle attack			
		SQL Injection	SQLIA SQL injection attack		
		Internet attack visualization			
		Password			

		guessing attack			
		IP spoofing			
		Hacking	Hardware hacking		
			Shadowcrew		
			Microphone hijack		
		Simple power attack			
		API hijacking			
		Malware	Malicious code		
			Malicious IT	virtuous IT	
				severity	
			Trojans	Trojan circuit	
			Stealth malware		
			Bot Bots Botnet Botnets		
			Rootkits		
			Malware modeling		
			Malware protection		
			Computer worm Computer worms Worm	Probabilistic worm spread	
				Worm attack	
				Epidemic model	Vaccination
					Dynamic quarantine
				Local scanning worms	
				Global scanning worms	
				Worm containmen t	
				Worm defense	

				Worm detection
				worm propagation
				Worm virulence estimation
			Spyware	Roving bug
				Keylogger
			Phishing	
			Computer virus Computer viruses Virus Viruses	SIR
				Zombie Computers
				Cryptovirus
		Cracking	Computer Fraud and Abuse Act	
			Electronic Communications Privacy Act	
8	Cryptology	Cryptography	Incremental cryptography	
		Cryptographic	Chosen-cipher	Chosen ciphertext security
			Block ciphers Block cipher	XTEA
				Elastic block ciphers
				Variable-length block ciphers
			Two-channel cryptography	
			Constant-length ciphertext	
			blind decryption	
			Untraceable decryptor	
			Verifiable shuffles	
			Template privacy	

			Template protection			
			Template security			
			Constant-time Decryption			
			Identity based Crytography Identity-based cryptography	DKIM		
			Encryption	Secret Content Index File (SCIF)	Secret sharing	Shamir's secret-sharing scheme
						Adversary structure
				Seberry scheme		
				Zheng		
				Pairing-based cryptosystems		
				Hybrid encryption		
				Broadcast encryption		
				Universal re-encryption		
				Symmetric encryption		
				Homomorphic encryption		
				Dual-encryption		
				Element-wise encryption		
				Certificateless encryption		
				Encryption feature		
				Signcryption		
				Identity based	Hierarchical identity-Based	

				encryption	encryption
			Public verifiability		
			Cryptographic API		
			Cryptographic file systems		
			Cryptographic hash Hash	Hash algorithms	
				Adaptive hash-chaining	
				hash function Hash functions	Pair block chaining
					One-way Hash Function one-way hash functions
					Alleged SecurID hash function
					Iterated hash functions
				Merkle hash tree	
				Image hash	
				Collision resistance	
			Threshold cryptography		
			Cryptographic protocol Cryptographic protocols	Diffie-Hellman Diffiee Hellman	Decisional Diffie Hellman problem
					Gap Bilinear Diffie Hellman problem
				AES	
				RSA	SEM
					Multi-Prime RSA
					RSA cryptosystem
					RSA-CRT

				Okamoto-Uchiyama scheme
				EKE
				ElGamal
				Typed MSR
			Cryptography theory	
			Cryptosystem	Okamoto-Tanaka-Uchiyama
			Direct anonymous attestation	Bilinear maps
			Quantum cryptography	
			Elliptic curve cryptography ECC	Elliptic curve cryptosystem Elliptic curve cryptosystems
			SCADA cryptography	
			Certificateless cryptography	
			Simulatability	
			Approximate noninterference	
			Cryptographic key Cryptographic keys	Authenticated identity-based key agreement protocol
				Cryptographic Key Assignment
				Group key establishment
				Group key distribution protocols
				Group

				Rekey	
				Password-authenticated key	
				Key agreement Key-agreement	Group key agreement
					Conference key agreement
				key assignment	
				Session key	
				key concept	
				Key disabling	
				Key distribution	Distributed key distribution schemes
				Key entropy	
				Biometrics-cryptographic key	Biometric based key cryptography
				Key escrow	
				Key exchange	Secret key exchange
					Authenticated key exchange
					Key exchange graph
				key generation	
				Key hierarchies	
				Key management	Group key management
				Key pre-distribution	
				Batch rekeying	
				Key recovery	Software key recovery
					Commercial key recovery
					Key recovery attacks

				Key set protocol		
				Key sharing		
				Key substitution		
				Paillier public-key system		
				Public key Public-key	X.509	
					Public key cryptography Public-key cryptography	PKCS #1
					Public key encryption Public-key encryption	
					Public key generation	
					PKI Public key infrastructure Public-key infrastructure	SPKT
						SDSI
						Inter-domain PKI
						PKI requirements
			Certificate	Certificate authority		
				Central Authority		
				Public key certificates Public-key certificates		
				Certificate revocation	Certificate revocation lists	
					Certificate revocation trees	
				Recursive certificate		
				Certified execution		
				Attribute certificates		
				Veiled		

				certificate		
				Digital certificate Digital certificates		
			Signature	Id-based signature		
				Efficient signature schemes		
				One-time signature		
				Universal designated verifier signature	Random projection	
				Multisignature		
				Stream signature		
				Group signature		
				Block signature		
				Signature schemes	Reactive	Reactive simulatability /UC
					Tame transformation	
					Multivariate	
				Account signatures		
				Certificate-based signatures		
				Blind signature		
				Certificateless signatures		
				Distributed signatures		
				Intrusion signatures	Multiple signing policies	Multi-secret

				Digital signature Digital signatures	Fingerprint Fingerprinting Fingerprints	
					ECDSA	
					DSA	
					Digital signature amortization	
					Digital signature scheme	
				Electronic signature Electronic signatures		
				Identity-based signature Identity-based signatures		
				Self-certificated signatures		
				Visual signatures		
				Blinding signature		
				Partially blind signatures		
				Proxy signature	Proxy multi-signature scheme	Multi-proxy multi-signature scheme
						Multi-proxy signature scheme
					Proxy signatures scheme	Proxy threshold signature scheme
				Threshold signature	Threshold proxy signature	
		Cryptanalysis				
		Steganography	Watermarking	Digital		

				watermarking
			Linguistic steganography	
			Confused Document Encrypting Scheme (CDES)	
		Steganalysis	Image steganalysis	
			Embedding capacity	
9	Defences Defense Defense mechanism Defense mechanisms Control Controls Security mechanisms Countermeasure Countermeasures Security countermeasures Safeguard Safeguarding measure Security controls Protection Protections	Information control		
		Control flow		
		Control systems		
		End-User Controls		
		Inference control		
		Automated verification tools		
		Novel countermeasure		
		Preservation		
		Defence strategy		
		Innate defense		
		Client-side defense		
		Computer network defense		
		Usage Control		
		Behavioral control		
		Analysis of control		
		Statistical disclosure control		
		Internal controls		
		Deterrent Control		
		Collaborative		

		network defense	
		Automatic adaptive defense	
		Protection mechanisms	
		Copy protection	
		knowledge protection	
		Precautionary principle	
		Client-side protection	
		Intrusion prevention	
		Protection strategy	
10	Identity	Identifier	Statistically unique and cryptographically verifiable identifiers
		identity management	Identity management systems
		Identity verification	
11	Identification	RFID	Residual RFID
		Radio Frequency Identification	RFID security
		Radio Frequency Identification (RFID)	
		Identification capability	
		Automated Personal Identification	
		SAML	
		User identification	Secure login
			Login
		Web application	

		identification		
		Host identification		
		Personal identification	Biometric identification	
		Universal Identification		
		Keystroke dynamics	Keystroke dynamics-based	Retraining
		Keystroke Identification		
		DNA personal ID		
12	Authentication Authentication security Secure authentication	Token	USB tokens	
		user authentication		
		Authenticated channel		
		Authentication performance		
		Authentication protocols	Kerberos	
		Authentication scheme		
		Data source authentication		
		GAA		
		Knowledge-based authentication		
		Authentication tests	Physical unclonable function PUF	Controlled PUF
		Call center authentication		
		Content-based authentication		
		Composite authentication		
		APF		
		Covert authentication		
		Efficient authentication		
		HTTP digest		

		authentication		
		Individual authentication	Single sign-on	
			Biometric biometrics	Biometric key
				Biometric personal authenticati on
				Biometric verification
		Message authentication		
		Multi-pass authentication		
		Mutual authentication		
		Remote authentication		
		Smartcard Smart card Smart cards		
		Stream authentication	Stream authentication protocols	
		Symmetric authentication		
		Credentials	User credentials	
			Multi-dimensional credentialing	
			Credential validation service	
		Challenge questions	Client puzzles	
		Keystroke analysis		
		Password Passwords	Password authentication	
			Graphical passwords	
			PIN	PIN cracking
			password cracking	

			Password creation
			Password security
			Password Management
			Password Protocols
			Password recovery
			Password reset
			Password stress
			One-time passwords
			Strong Passwords
		Information-theoretically Secure authentication	
13	Access	Access security	Access Control security
		Access control Access controls	Time-Based access control
			Dynamic Access Control Problem
			Visual HIP
			Lattice-based access control models
			Mandatory access control models
			Optimistic Access Control
			Passphrases
			Access Control List
			Access control model
			Access control

			policies		
			Role-Based Access Control RBAC	GB-RBAC	Group-based RBAC
					GTRBAC
			OrBAC	OrBAC model	
			Access controller		
		Access granting			
		Access request evaluation time			
		Access to Data			
		Authorisation Authorization	Authorisation manager		
			Provisional action		
			authorization infrastructure		
			Delegated authorization		
			Delegation of authority		
			Authorization policy		
			Authorized view		
			Security view		
14	Computer network security Network security communication security	Firewall Firewalls	Application firewalls		
			Distributed firewalls		
			Internet firewalls		
			firewall testing		
			firewall vulnerability analysis		
		MNP			
		Secure multicast Multicast security			
		Wireless security WiFi security	Wireless sensor network security		

		Wired security			
		Secure clustering and routing			
		virtual private networks			
		B3G networks			
		Anomalous traffic			
		Perimeter security			
		B3G security			
		Host-based security with network components			
		SCADA network security			
		Network monitoring			
		IMS			
		Security Protocols			
		Protocol security	SSL	SSL/TLS	
				SSL/TLS session-aware	
				Enhanced SSL	
			Secure multiparty computation Secure multi-party computation		
			WTLS	WAP	WAP security
			Specification of security protocols		
			IP security protocols	ICMP	ICMP traceback
				IPsec	
			Auction protocol		
			ORCON policies		
			Communicatio n protocols	Mobile Authenticati	

				on Protocol
			OCSP	
			Secure two-party computation	Semi-trusted third party
			SCADA protocol security	Power system security SCADA
		Cheating File Database (CFD)		
		DNP3		
		Address obfuscating techniques		
		Anonymous communications		
		Secure group communication		
		VoIP security		
		Secure Electronic Transactions (SET)		
		Information flow	Information flow control	
		Information-flow security	Information flow anomaly detection	
			Real-time information flow	
			Dynamic information flow analysis	
15	Internet security	Internet crime	Cyberterrorism	
		untrusted servers		
		Internet threats		
		Internet login		
		Content-based filtering		
		Internet privacy	Internet users' information privacy	

			concerns		
			Internet Users' Information Privacy Concern (IUIPC)		
		Web security www security	Web content protection		
			Web service security		
		Internet worms	Active worms		
		Email security E-mail security	E-mail harvesters		
			E-mail hunter		
			SPF	Sender-ID	SIDF
			E-mail spam spam e-mail	spam	
				Anti-spam filter	Whitelist
					Greylist
				anti-spam measures	
				Anti-spoofing	
				Spam obfuscation	
				Spam over Internet Telephony (SPIT)	
				Spam SMS messages	
				spamming options	
			Electronic mail fraud		
			Mass e-mailers		
			S/MIME		
			Certified email Certified electronic mail	Certified e-mail probabilistic model checking	
		Electronic commerce security			

		Electronic commerce privacy			
		E-voting Electronic voting			
16	Security incidents	CERT CERT/CC			
17	Detection	Intrusion detection	Intrusion Detection and Protection		
			Online alert correlation		
			Intrusion Detection Management System		
			Distributed intrusion detection		
			Network intrusion detection	Snort	
				Real-time NIDS	
			TCM-KNN (Transductive Confidence Machines for K-Nearest Neighbors)		
			Specification-based approach		
			Anomaly intrusion detection		
			IDS Intrusion detection systems (IDS) Intrusion detection systems (IDSs) Intrusion detection system Intrusion	Alarms' distribution	
				ROC curves	
				IDS evaluation	
				Alarm correlation	Alert causal correlation
					K-means clustering
				Alert analysis	
				Correlation	

			detection systems	intractability	
				Correlation-based IDS	
				IDS systems and platforms	BAIT
		Misuse detection	Misuse intrusion detection		
		Honeypot Honeypots			
		Database intrusion			
		Misfeasor detection			
		Statistical detection			
		Masquerade detection	Masquerader detection		
		Signal detection			
		Detection mechanism	Sensor security		
		Stateful detection			
		Novelty detection	For novelty detection		
18	Information security requirements Security requirement Security requirements	Enterprise security requirements			
19	Security design	secure systems design			
		Protocol design			
		Security design principles			
		Security by design			
		Security modelling	Quantitative security modeling		
		Secure information systems development			

		Security by obscurity	
		Secure usability	Usable security
		HCI-S	
		Security GUI	
20	Security engineering	Security requirement engineering	
		Security/Software engineering	
		Security automata	
		Layered security approach	
21	Security architecture	Security models Security model	OSI Security model
			Formal security model
			Information security maturity model
			Hazard model
			Security consistency model
			Bell-Lapadula model
		Security Framework	
		Scalable security	
22	Secure architectures	Operating system security architecture	
23	Security hardware	Hardware security modules	
24	System security Security Systems	Security type system	
		Systems survivability	
		Distributed system security	
		Security	Security

		hardening	hardening plans	
			Security hardening patterns	
		Trusted computing	Trust and attestation	
		Grid security	Grid security infrastructure	
25	Data security	Data security at the point of use		
		Electronic data security		
		Physical data security		
		Data availability		
		Data confidentiality		
		Data destruction		
		Data integrity	Checksums	
		Data Exploitation		
		Data protect		
		Data sharing		
		Data Perturbation		
		Recovery	Data Recovery	
			On-line recovery	
		Multilevel data security		
		Content security		
		Text security	Authentic documents	CSP
26	Application security	Software security	Software protection	
			Software safety	
			Open source security	
			Security APIs	
			Security SDKs	
			Mobile code security	
			Patch release	

			time
			API flaws
			Virtual machine security secure virtual machine
			Java security
			.Net security
		Secure programming	Control-flow graph
			Static analysis
			Jif
			Bytecode verifier
			Code review
			Security patterns
			Code safety
		Language-based security Language based security	DSL
			Security-typed language
27	Database security	Controlled query evaluation	
28	Security planning		
29	Security evaluation		
	Security assessment		
	Security analysis		
30	safety analysis		
31	Security risk Information Security Risk	Risk analysis Information Security risk analysis	target and shield model
			Risk assessment Information Security risk assessment
			Information lifecycle security risk assessment

			Quantitative risk analysis	ALE	
			Paper-based risk analysis		
			Data lifecycle risk analysis		
		Risk evaluation			
		Risk exposure			
		Risk forecasting			
		Management of security risks Information systems risk management Security risk management Risk management	risk management standards		
			Classification	Classification methods	text categorization
				Taxonomy	
				Classification scheme	
			Models of risk management		
		Risk model			
		Risk modelling			
		Risk perception			
		Insider risks			
		Risk quantification			
		Disclosure risk			
		Business information risk			
32	Insider Inside security	Insider attacks			
		Insider misuse			
		Malicious insiders			
		Insider profiling			
		insider threat	Malicious act		
		Insider attack detection			
		Insider attack prediction			
33	Economics of IT security	IT investments			

	Economics of information security Economics of IS Economics of IS security	Downtime loss
		Security benefits
		Security costs and benefits
34	Information Security investments Information security investment Investments in Information Security security investment Security investments	ROSI
		Security RoI
		Benefits of security investments
		Security technology investment
35	IT security management Security management Information systems security management Information Security Management	Reporting
		Security management process
		Holistic security management
		Reliability management
36	Security policy Security policies Information Security Policy Information Security policies	Security policy quality
		Security policy adoption

	IS security policies	Digital Policy Management	
		Security policy implementation	
		Specification of security policies	
		Security policy management	Policy based management
		Information security policy components	
		Security policy composition	
		Formation security policy	
		Security and Privacy Policy Obligations	
		Encryption policy	
		Signature policy	
		Disclosure policy	
		Business policy	
		Policy anomalies	
		Policy design	
		Policy development	
		Policy enforcement	
		Policy refinement	
		Policy spaces	
		Policy scope	
		Policy uptake	
		Information Security metapolicies	
		Policy update	
		Employees' compliance with security policies	
37	Information Security Governance	Information systems security standardsNorm	security management code of practice

		s Standards	ISO/IEC standard	ISO17799 ISO 17799 ISO/IEC 17799 BS 7799
				ISO 27001 ISO/IEC 27001
				ISO/IEC 27002
			ISS de jure standards	
		Guideline Guidelines	Information Security guidelines	
		Direct-Control Cycle		
38	Information assurance Security assurance	Real-time assurances		
39	Information Security enforcement Security enforcement	Surveillance	Electronic Article Surveillance	
40	Agreements			
41	Security compliance Compliance Information	Information Security obedience		
		Compliance enforcement		
		Compliance management		
		Legislation	Regulations	Regulating service
				Health Insurance Portability and Accountability Act (HIPAA)
				Sarbanes-Oxley Act Sarbanes–

				Oxley	
				Health Information Technology for Economic and Clinical Health (HITECH)	
			Regulatory compliance		
			Legal compliance		
		Audit Auditing IS security auditing	Continuous auditing		
			IT audit IT auditing		
			Policy audit		
			Computer audit data		
			Audit objectives		
			Postaudit		
			Audit trails	Audit trail analysis	
		Digital rights	Digital rights management DRM	Enterprise DRM	
			Proprietary information protection	Copyright	Copyright protection
42	Information Security behavior Behavioral information security	Behavioral issues of information security			
		Security behaviour			
		protection-oriented behavior			
		Self-Protection Behavior			
		Social Aspects of Information	Social and Behavioral		

		Security	aspects of information security	
		Security attitudes		
		Information Security practices		
		Perceived security		
43	Information Security education Security education			
44	IS security training Security training			
45	Security awareness Awareness ICT Security awareness Information Security awareness	Training and awareness		
		Context awareness		
		E-privacy Awareness		
46	Certification Information security certification	process certification		
		Cross-certification		
		Certification authories		
		Blind certification		
47	Business Continuity	Business Continuity Planning		
		Disaster recovery Disaster response	Disaster Plans	IT Disaster Recovery Planning
		Contingency planning		
		Extreme event planning		
		High Reliability Organizations	Theory of high reliability	

			organizations	
		Distributed backup		
		Intrusion tolerance		
48	Critical infrastructure security	information infrastructure protection		
49	Information Security culture Security culture			
50	IS misuse Information systems misuse Computer misuse	Misuse models		
		Misuse Cases		
51	Computer abuse			
52	Computer crime	Piracy	Piracy deterrence	software piracy
			Digital piracy	
		Crime research		
		Criminals		
		Criminology theories		
		Crime deterrence	Punishment certainty	
			Punishment severity	
		Cyber crime		
		Cyber incidents		
		Corporate cyber harassment		
		Cybertrust		
		Professional cyber bullying		
		Fraud	Fraud detection	
			Social engineering	
			Fraud management	
		Forensics Computer forensics	Disk forensics	
			routing forensics	

		Information forensics	Forensic analysis	
			USB forensics	
			DIRT	
			Hard Drive Forensics	
			Digital forensics	
			Internet forensics	
			Cyber forensics	
			Software forensics	Trace-back mechanisms
			Traceability	User-controlled-traceability
				Packet tracing
			Evidence	Validity of evidence
				Evidence reasoning
				Evidential reasoning
			HTCIA	
			CTIN	
53	Information warfare	Offensive warfare		
		Defensive warfare		
		Weapons		
54	Security and Privacy	Security and privacy protection		
		concerns for information privacy Concern for Information Privacy (CFIP)	Employee Information Privacy Concern	
		information privacy privacy	Location privacy	
			E-privacy Risk Concerns	
			Corporate	

			privacy	
			Web privacy	
			Privacy-preserving revocation checking	
			Computer privacy	
			online privacy	Online Privacy Issues
			Privacy amplification	
			Privacy belief	
			privacy calculus	
			formal privacy	
			Privacy Concern privacy concerns	
			Privacy controls	
			Privacy enforcement	
			Privacy Enhancing Technologies (PETS) Privacy-enhancing technologies	
			Privacy infrastructure	
			Privacy law privacy laws Privacy regulations	
			Privacy Model Comparison	
			Privacy policy privacy policies	
			Privacy practices	
			Privacy protection	Private key protection

				E-privacy Protection	
			privacy requirements engineering	PriS	PriS-Tool
			privacy rights		
			Privacy Risk		
			Privacy-preserving clustering	Privacy-preserving clustering over	
			Privacy-preserving data mining		
			Privacy information hiding	Anonymity	Anonymity and privacy
					Anonymization
					Buyer anonymity
					Location k-anonymity
					Opacity
					Information hiding
					Non-deducibility
					User-controlled-anonymity
55	Security technologies Protective technology Security solutions	Security products	Protective information technologies		
			Notification process		
			anti-virus scanners		
			Anti-virus software antiviral software		
			File integrity analyzers		
			Anti-spyware		
		Security applications	NoScript		

		Secure hardware
		Secure coprocessors
56	Secure voting	
	Secret ballot	
57	End user security End-user security	
58	Safety property Safety problem	
59	Secure outsourcing	
60	Information Security baselines	
61	Safe contexts	
62	Security configuration	
63	Forward security	
64	Security binding	
65	Security-by-contracts	
66	Security for resource-constrained devices	
67	Quantitative security	
68	Inferential security	
69	Department of Homeland security	
70	Security proof	
71	Liveness property	
72	AACS	
73	Secure collaborations	
74	Information-theoretically Secure	
75	Security journals	

76	IS security research			
77	Information Security trends			
78	Security values			
79	Security knowledge			
80	Security Competencies			
81	Security perceptions			
82	Provable security			
83	Unconditional security			
84	Hackers Hacker	hacker groups		
		hacker profile		
		Hacker Tactics		
85	Incident Incidents	Incident Management process	Incident Handling	
			Incident Response	Incident response planning
				Intrusion response

Apêndice D – Lista de *Papers* vs. Classes

A tabela seguinte mostra para cada *paper* o total de classes diferentes que podem ser encontradas. É importante referir que este resultado provém da primeira tentativa de associar a um *paper* uma *tag*. Ainda é possível ver o nome do *journal* para o *paper* em análise e concluir a média de classes distintas que pode ser encontrada por cada um dos 14 *journals* explorados. Nesta tabela apenas são considerados os *papers* que na sua estrutura apresentam o campo "*keyword*".

Idpaper	Nome do Journal	Total Classes
1	Journal of Management Information Systems	3
2	Journal of Management Information Systems	1
3	Journal of Management Information Systems	2
4	Journal of Management Information Systems	2
5	Journal of Management Information Systems	3
6	Journal of Management Information Systems	1
7	Journal of Management Information Systems	2
8	Journal of Management Information Systems	1
9	Journal of Management Information Systems	3
10	Journal of Management Information Systems	1
11	Journal of Management Information Systems	1
12	Journal of Management Information Systems	1
13	Journal of Management Information Systems	3
14	Journal of Management Information Systems	4
15	Journal of Management Information Systems	2
16	Computers & Security	2
20	Computers & Security	3
22	Computers & Security	2
23	Computers & Security	4
30	Computers & Security	1
31	Computers & Security	3
38	Computers & Security	2
40	Computers & Security	3
44	Computers & Security	3
48	Computers & Security	3
51	Computers & Security	3
53	Computers & Security	2
68	Computers & Security	2
69	Computers & Security	3

70	Computers & Security	1
72	Computers & Security	1
73	Computers & Security	5
75	Computers & Security	3
76	Computers & Security	2
85	Computers & Security	2
91	Computers & Security	1
92	Computers & Security	2
93	Computers & Security	1
94	Computers & Security	2
95	Computers & Security	3
97	Computers & Security	1
98	Computers & Security	2
99	Computers & Security	1
100	Computers & Security	3
101	Computers & Security	3
102	Computers & Security	1
103	Computers & Security	1
104	Computers & Security	1
107	Computers & Security	3
108	Computers & Security	4
109	Computers & Security	2
111	Computers & Security	2
112	Computers & Security	4
116	Computers & Security	2
117	Computers & Security	5
121	Computers & Security	2
123	Computers & Security	1
126	Computers & Security	1
127	Computers & Security	1
128	Computers & Security	5
131	Computers & Security	4
132	Computers & Security	1
133	Computers & Security	3
134	Computers & Security	2
135	Computers & Security	4
136	Computers & Security	2
137	Computers & Security	1
138	Computers & Security	2
139	Computers & Security	1
140	Computers & Security	3
141	Computers & Security	4
142	Computers & Security	1
143	Computers & Security	1
144	Computers & Security	2

145	Computers & Security	2
147	Computers & Security	1
149	Computers & Security	2
150	Computers & Security	1
151	Computers & Security	1
154	Computers & Security	2
155	Computers & Security	2
156	Computers & Security	1
157	Computers & Security	3
158	Computers & Security	4
159	Computers & Security	2
160	Computers & Security	3
161	Computers & Security	3
162	Computers & Security	4
163	Computers & Security	2
165	Computers & Security	1
166	Computers & Security	3
167	Computers & Security	2
168	Computers & Security	1
169	Computers & Security	1
170	Computers & Security	1
171	Computers & Security	1
172	Computers & Security	4
174	Computers & Security	2
175	Computers & Security	1
176	Computers & Security	1
177	Computers & Security	2
178	Computers & Security	2
179	Computers & Security	2
180	Computers & Security	3
181	Computers & Security	3
182	Computers & Security	1
183	Computers & Security	4
184	Computers & Security	4
185	Computers & Security	2
187	Computers & Security	3
189	Computers & Security	1
190	Computers & Security	3
191	Computers & Security	3
192	Computers & Security	2
193	Computers & Security	2
194	Computers & Security	2
195	Computers & Security	1
196	Computers & Security	4
197	Computers & Security	1

199	Computers & Security	1
200	Computers & Security	2
201	Computers & Security	2
203	Computers & Security	2
204	Computers & Security	1
205	Computers & Security	4
206	Computers & Security	1
207	Computers & Security	2
208	Computers & Security	2
209	Computers & Security	2
210	Computers & Security	2
211	Computers & Security	3
212	Computers & Security	1
213	Computers & Security	1
214	Computers & Security	1
215	Computers & Security	2
216	Computers & Security	2
217	Computers & Security	2
218	Computers & Security	3
219	Computers & Security	3
220	Computers & Security	1
221	Computers & Security	2
222	Computers & Security	1
223	Computers & Security	3
224	Computers & Security	2
225	Computers & Security	3
226	Computers & Security	1
227	Computers & Security	5
228	Computers & Security	1
229	Computers & Security	1
230	Computers & Security	3
231	Computers & Security	2
232	Computers & Security	2
233	Computers & Security	1
234	Computers & Security	2
235	Computers & Security	2
236	Computers & Security	1
237	Computers & Security	2
238	Computers & Security	2
239	Computers & Security	2
241	Computers & Security	1
242	Computers & Security	1
243	Computers & Security	2
244	Computers & Security	2
245	Computers & Security	5

246	Computers & Security	1
247	Computers & Security	2
248	Computers & Security	4
249	Computers & Security	1
250	Computers & Security	1
251	Computers & Security	2
254	Computers & Security	2
255	Computers & Security	3
256	Computers & Security	2
257	Computers & Security	2
258	Computers & Security	3
259	Computers & Security	3
260	Computers & Security	2
261	Computers & Security	2
262	Computers & Security	5
263	Computers & Security	3
264	Computers & Security	3
265	Computers & Security	2
266	Computers & Security	4
267	Computers & Security	4
268	Computers & Security	2
269	Computers & Security	4
270	Computers & Security	3
271	Computers & Security	3
272	Computers & Security	2
273	Computers & Security	1
274	Computers & Security	2
275	Computers & Security	3
276	Computers & Security	2
278	Computers & Security	2
279	Computers & Security	2
280	Computers & Security	2
281	Computers & Security	3
282	Computers & Security	1
283	Computers & Security	2
284	Computers & Security	2
285	Computers & Security	4
286	Computers & Security	3
287	Computers & Security	1
288	Computers & Security	2
289	Computers & Security	2
290	Computers & Security	1
291	Computers & Security	2
292	Computers & Security	3
293	Computers & Security	3

294	Computers & Security	1
295	Computers & Security	3
296	Computers & Security	3
297	Computers & Security	2
299	Computers & Security	2
300	Computers & Security	2
301	Computers & Security	1
302	Computers & Security	2
303	Computers & Security	3
304	Computers & Security	3
305	Computers & Security	2
306	Computers & Security	2
307	Computers & Security	1
308	Computers & Security	4
309	Computers & Security	4
310	Computers & Security	3
311	Computers & Security	2
312	Computers & Security	1
313	Computers & Security	4
314	Computers & Security	3
315	Computers & Security	1
316	Computers & Security	2
317	Computers & Security	3
318	Computers & Security	3
319	Computers & Security	1
320	Computers & Security	3
321	Computers & Security	4
322	Computers & Security	2
323	Computers & Security	4
325	Computers & Security	1
326	Computers & Security	2
327	Computers & Security	1
329	Computers & Security	3
330	Computers & Security	2
331	Computers & Security	3
332	Computers & Security	1
333	Computers & Security	2
334	Computers & Security	4
335	Computers & Security	2
336	Computers & Security	3
337	Computers & Security	2
338	Computers & Security	3
339	Computers & Security	2
340	Computers & Security	5
341	Computers & Security	1

342	Computers & Security	3
343	Computers & Security	3
344	Computers & Security	1
345	Computers & Security	4
347	Computers & Security	2
348	Computers & Security	3
349	Computers & Security	1
350	Computers & Security	3
351	Computers & Security	1
352	Computers & Security	1
353	Computers & Security	3
354	Computers & Security	3
355	Computers & Security	4
356	Computers & Security	3
357	Computers & Security	2
358	Computers & Security	2
360	Computers & Security	2
361	Computers & Security	2
362	Computers & Security	4
363	Computers & Security	1
364	Computers & Security	1
366	Computers & Security	5
367	Computers & Security	2
368	Computers & Security	3
369	Computers & Security	3
371	Computers & Security	3
372	Computers & Security	1
373	Computers & Security	1
374	Computers & Security	3
375	Computers & Security	2
376	Computers & Security	2
377	Computers & Security	1
378	Computers & Security	4
379	Computers & Security	3
380	Computers & Security	2
381	Computers & Security	2
382	Computers & Security	2
383	Computers & Security	3
384	Computers & Security	2
385	Computers & Security	5
386	Computers & Security	2
387	Computers & Security	1
389	Computers & Security	2
390	Computers & Security	5
391	Computers & Security	2

393	Computers & Security	3
395	Computers & Security	2
396	Computers & Security	2
397	Computers & Security	2
399	Computers & Security	4
400	Computers & Security	2
401	Computers & Security	2
402	Computers & Security	3
403	Computers & Security	4
404	Computers & Security	2
405	Computers & Security	2
406	Computers & Security	3
407	Computers & Security	2
408	Computers & Security	1
409	Computers & Security	3
410	Computers & Security	2
411	Computers & Security	1
412	Computers & Security	3
414	Computers & Security	5
415	Computers & Security	2
416	Computers & Security	1
417	Computers & Security	2
418	Computers & Security	1
419	Computers & Security	3
420	Computers & Security	1
421	Computers & Security	1
422	Computers & Security	1
423	Computers & Security	2
424	Computers & Security	2
425	Computers & Security	2
426	Computers & Security	2
427	Computers & Security	2
428	Computers & Security	1
429	Computers & Security	3
430	Computers & Security	2
431	Computers & Security	3
432	Computers & Security	4
433	Computers & Security	2
434	Computers & Security	1
435	Computers & Security	9
436	Computers & Security	5
437	Computers & Security	3
438	Computers & Security	3
439	Computers & Security	4
440	Computers & Security	2

441	Computers & Security	1
442	Computers & Security	4
443	Computers & Security	5
444	Computers & Security	1
445	Computers & Security	2
447	Computers & Security	3
448	Computers & Security	2
450	Computers & Security	2
451	Computers & Security	2
452	Computers & Security	1
453	Computers & Security	2
454	Computers & Security	5
455	Computers & Security	1
456	Computers & Security	1
457	Computers & Security	1
458	Computers & Security	1
459	Computers & Security	2
460	Computers & Security	1
461	Computers & Security	1
462	Computers & Security	1
463	Computers & Security	3
464	Computers & Security	1
465	Computers & Security	2
466	Computers & Security	1
467	Computers & Security	2
469	Computers & Security	1
470	Computers & Security	2
471	Computers & Security	1
472	Computers & Security	2
473	Computers & Security	1
474	Computers & Security	2
475	Computers & Security	3
476	Computers & Security	3
477	Computers & Security	2
478	Computers & Security	2
479	Computers & Security	1
480	Computers & Security	2
481	Computers & Security	2
482	Computers & Security	3
483	Computers & Security	3
484	Computers & Security	1
485	Computers & Security	1
486	Computers & Security	1
487	Computers & Security	1
488	Computers & Security	2

489	Computers & Security	1
490	Computers & Security	1
491	Computers & Security	2
492	Computers & Security	2
493	Computers & Security	2
494	Computers & Security	1
495	Computers & Security	2
496	Computers & Security	3
497	Computers & Security	2
498	Computers & Security	4
499	Computers & Security	2
500	Computers & Security	4
501	Computers & Security	1
502	Computers & Security	1
504	Computers & Security	1
505	Computers & Security	2
506	Computers & Security	2
507	Computers & Security	1
508	Computers & Security	1
509	Computers & Security	2
510	Computers & Security	1
511	Computers & Security	4
512	International Journal of Information Security	2
513	International Journal of Information Security	1
514	International Journal of Information Security	1
516	International Journal of Information Security	2
517	International Journal of Information Security	1
519	International Journal of Information Security	2
520	International Journal of Information Security	1
521	International Journal of Information Security	2
522	International Journal of Information Security	1
523	International Journal of Information Security	1
524	International Journal of Information Security	2
525	International Journal of Information Security	3
526	International Journal of Information Security	2
527	International Journal of Information Security	2
528	International Journal of Information Security	4
529	International Journal of Information Security	1
530	International Journal of Information Security	1
531	International Journal of Information Security	1
532	International Journal of Information Security	3
533	International Journal of Information Security	1
534	International Journal of Information Security	2
535	International Journal of Information Security	2
536	International Journal of Information Security	2

538	International Journal of Information Security	3
539	International Journal of Information Security	2
540	International Journal of Information Security	2
542	International Journal of Information Security	2
543	International Journal of Information Security	2
544	International Journal of Information Security	1
545	International Journal of Information Security	2
546	International Journal of Information Security	2
547	International Journal of Information Security	2
548	International Journal of Information Security	1
549	International Journal of Information Security	1
550	International Journal of Information Security	2
551	International Journal of Information Security	1
552	International Journal of Information Security	2
553	International Journal of Information Security	1
554	International Journal of Information Security	4
555	International Journal of Information Security	1
556	International Journal of Information Security	2
557	International Journal of Information Security	2
558	International Journal of Information Security	3
560	International Journal of Information Security	3
561	International Journal of Information Security	1
562	International Journal of Information Security	1
563	International Journal of Information Security	1
564	International Journal of Information Security	5
565	International Journal of Information Security	4
567	International Journal of Information Security	2
569	International Journal of Information Security	4
570	International Journal of Information Security	1
571	International Journal of Information Security	1
572	International Journal of Information Security	2
573	International Journal of Information Security	1
574	International Journal of Information Security	2
575	International Journal of Information Security	1
576	International Journal of Information Security	1
577	International Journal of Information Security	2
578	International Journal of Information Security	3
579	International Journal of Information Security	1
580	International Journal of Information Security	2
581	International Journal of Information Security	4
583	International Journal of Information Security	3
584	International Journal of Information Security	2
585	International Journal of Information Security	1
586	International Journal of Information Security	1
587	International Journal of Information Security	3

588	International Journal of Information Security	2
589	International Journal of Information Security	3
590	International Journal of Information Security	2
591	International Journal of Information Security	3
592	International Journal of Information Security	1
593	International Journal of Information Security	1
594	International Journal of Information Security	2
595	International Journal of Information Security	1
596	International Journal of Information Security	1
597	International Journal of Information Security	2
598	International Journal of Information Security	3
599	International Journal of Information Security	1
600	International Journal of Information Security	1
601	International Journal of Information Security	3
602	International Journal of Information Security	2
603	International Journal of Information Security	3
605	International Journal of Information Security	1
606	International Journal of Information Security	1
607	International Journal of Information Security	2
609	International Journal of Information Security	3
610	International Journal of Information Security	1
611	International Journal of Information Security	1
612	International Journal of Information Security	2
613	International Journal of Information Security	3
614	International Journal of Information Security	2
615	International Journal of Information Security	2
616	International Journal of Information Security	2
617	International Journal of Information Security	2
618	International Journal of Information Security	3
619	International Journal of Information Security	2
620	International Journal of Information Security	2
621	International Journal of Information Security	3
622	International Journal of Information Security	1
623	International Journal of Information Security	1
624	International Journal of Information Security	2
625	International Journal of Information Security	1
626	International Journal of Information Security	4
627	International Journal of Information Security	2
628	International Journal of Information Security	2
630	International Journal of Information Security	1
631	International Journal of Information Security	2
632	International Journal of Information Security	1
633	International Journal of Information Security	2
634	International Journal of Information Security	2
635	International Journal of Information Security	2

637	International Journal of Information Security	1
638	International Journal of Information Security	1
639	International Journal of Information Security	2
640	International Journal of Information Security	1
641	International Journal of Information Security	2
643	International Journal of Information Security	3
644	International Journal of Information Security	3
645	International Journal of Information Security	1
646	International Journal of Information Security	2
647	International Journal of Information Security	2
648	International Journal of Information Security	3
649	International Journal of Information Security	1
650	International Journal of Information Security	3
651	International Journal of Information Security	2
652	International Journal of Information Security	2
654	International Journal of Information Security	1
655	International Journal of Information Security	2
657	International Journal of Information Security	1
658	International Journal of Information Security	1
659	International Journal of Information Security	2
660	International Journal of Information Security	1
661	International Journal of Information Security	2
662	International Journal of Information Security	2
663	International Journal of Information Security	2
664	International Journal of Information Security	1
665	International Journal of Information Security	1
667	International Journal of Information Security	1
668	International Journal of Information Security	1
669	International Journal of Information Security	1
670	International Journal of Information Security	1
671	International Journal of Information Security	2
672	International Journal of Information Security	2
673	International Journal of Information Security	3
674	International Journal of Information Security	1
676	International Journal of Information Security	1
677	International Journal of Information Security	3
678	International Journal of Information Security	1
679	International Journal of Information Security	4
680	International Journal of Information Security	1
681	International Journal of Information Security	2
682	International Journal of Information Security	2
683	International Journal of Information Security	1
684	International Journal of Information Security	2
685	International Journal of Information Security	1
686	International Journal of Information Security	2

687	International Journal of Information Security	1
688	International Journal of Information Security	1
689	International Journal of Information Security	2
690	International Journal of Information Security	3
691	International Journal of Information Security	3
692	International Journal of Information Security	2
693	International Journal of Information Security	1
694	International Journal of Information Security	1
695	International Journal of Information Security	2
696	International Journal of Information Security	2
697	Information Management and Computer Security	1
698	Information Management and Computer Security	1
699	Information Management and Computer Security	2
700	Information Management and Computer Security	2
701	Information Management and Computer Security	2
702	Information Management and Computer Security	1
703	Information Management and Computer Security	2
704	Information Management and Computer Security	1
705	Information Management and Computer Security	1
706	Information Management and Computer Security	2
707	Information Management and Computer Security	2
708	Information Management and Computer Security	1
709	Information Management and Computer Security	1
710	Information Management and Computer Security	3
711	Information Management and Computer Security	2
712	Information Management and Computer Security	1
713	Information Management and Computer Security	1
714	Information Management and Computer Security	2
715	Information Management and Computer Security	1
716	Information Management and Computer Security	2
717	Information Management and Computer Security	1
718	Information Management and Computer Security	2
719	Information Management and Computer Security	1
720	Information Management and Computer Security	2
721	Information Management and Computer Security	1
722	Information Management and Computer Security	1
723	Information Management and Computer Security	1
724	Information Management and Computer Security	2
725	Information Management and Computer Security	1
726	Information Management and Computer Security	1
727	Information Management and Computer Security	1
728	Information Management and Computer Security	1
729	Information Management and Computer Security	1
730	Information Management and Computer Security	3
732	Information Management and Computer Security	1

733	Information Management and Computer Security	2
734	Information Management and Computer Security	1
735	Information Management and Computer Security	2
736	Information Management and Computer Security	1
737	Information Management and Computer Security	1
738	Information Management and Computer Security	2
739	Information Management and Computer Security	3
740	Information Management and Computer Security	3
741	Information Management and Computer Security	2
742	Information Management and Computer Security	1
743	Information Management and Computer Security	1
744	Information Management and Computer Security	1
745	Information Management and Computer Security	1
746	Information Management and Computer Security	1
747	Information Management and Computer Security	1
748	Information Management and Computer Security	2
749	Information Management and Computer Security	2
750	Information Management and Computer Security	1
751	Information Management and Computer Security	1
753	Information Management and Computer Security	2
754	Information Management and Computer Security	2
755	Information Management and Computer Security	1
756	Information Management and Computer Security	1
757	Information Management and Computer Security	3
758	Information Management and Computer Security	2
759	Information Management and Computer Security	2
760	Information Management and Computer Security	1
761	Information Management and Computer Security	2
762	Information Management and Computer Security	2
763	Information Management and Computer Security	1
764	Information Management and Computer Security	2
766	Information Management and Computer Security	2
767	Information Management and Computer Security	2
768	Information Management and Computer Security	1
769	Information Management and Computer Security	1
770	Information Management and Computer Security	2
771	Information Management and Computer Security	2
772	Information Management and Computer Security	2
773	Information Management and Computer Security	3
774	Information Management and Computer Security	1
776	Information Management and Computer Security	1
777	Information Management and Computer Security	1
778	Information Management and Computer Security	3
779	Information Management and Computer Security	1
780	Information Management and Computer Security	1

781	Information Management and Computer Security	1
782	Information Management and Computer Security	2
785	Information Management and Computer Security	1
788	Information Management and Computer Security	2
789	Information Management and Computer Security	2
790	Information Management and Computer Security	2
791	Information Management and Computer Security	4
794	Information Management and Computer Security	1
796	Information Management and Computer Security	2
797	Information Management and Computer Security	2
799	Information Management and Computer Security	2
800	Information Management and Computer Security	2
801	Information Management and Computer Security	1
802	Information Management and Computer Security	2
804	Information Management and Computer Security	2
805	Information Management and Computer Security	1
806	Information Management and Computer Security	1
807	Information Management and Computer Security	1
809	Information Management and Computer Security	1
810	Information Management and Computer Security	3
811	Information Management and Computer Security	1
812	Information Management and Computer Security	2
814	Information Management and Computer Security	1
816	Information Management and Computer Security	1
817	Information Management and Computer Security	1
818	Information Management and Computer Security	2
819	Information Management and Computer Security	2
820	Information Management and Computer Security	2
821	Information Management and Computer Security	2
823	Information Management and Computer Security	3
824	Information Management and Computer Security	1
825	Information Management and Computer Security	2
826	Information Management and Computer Security	1
827	Information Management and Computer Security	2
828	Information Management and Computer Security	1
829	Information Resources Management Journal	2
830	Information Resources Management Journal	2
831	Information Resources Management Journal	2
833	Information Resources Management Journal	1
836	Information Resources Management Journal	2
837	Information Resources Management Journal	3
839	Information Resources Management Journal	2
840	Information Resources Management Journal	2
841	Information Systems Journal	1
842	Information Systems Journal	1

843	Information Systems Journal	2
844	Information Systems Journal	3
845	Information Systems Research	2
846	Information Systems Research	3
847	Information Systems Research	4
848	Information Systems Research	4
849	Information Systems Research	2
850	Information Systems Research	4
851	Information Systems Research	2
852	Information Systems Research	1
853	Information Systems Research	2
854	Information Systems Research	1
855	Information Systems Research	2
856	Information Systems Research	4
857	Journal Association for Computing Machinery	2
858	Journal Association for Computing Machinery	2
859	Journal Association for Computing Machinery	2
860	Journal Association for Computing Machinery	3
861	Journal Association for Computing Machinery	2
865	Journal of the Association for Information Systems	1
866	Journal of the Association for Information Systems	3
867	Journal of the Association for Information Systems	3
870	Journal of the Association for Information Systems	2
872	Journal of the Association for Information Systems	3
874	Journal of Computer Information Systems	1
875	Journal of Computer Information Systems	2
876	Journal of Computer Information Systems	2
877	Journal of Computer Information Systems	4
878	Journal of Computer Information Systems	1
881	Journal of Information Privacy and Security	3
882	Journal of Information Privacy and Security	2
883	Journal of Information Privacy and Security	2
884	Journal of Information Privacy and Security	2
886	Journal of Information Privacy and Security	4
887	Journal of Information Privacy and Security	1
888	Journal of Information Privacy and Security	1
889	Journal of Information Privacy and Security	2
890	Journal of Information Privacy and Security	2
892	Journal of Information Privacy and Security	3
893	Journal of Information Privacy and Security	2
894	Journal of Information Privacy and Security	3
895	Journal of Information Privacy and Security	2
896	Journal of Information Privacy and Security	2
897	Journal of Information Privacy and Security	1
901	Journal of Information Privacy and Security	1

902	Journal of Information Privacy and Security	2
903	Journal of Information Privacy and Security	1
905	Journal of Information Privacy and Security	3
906	Journal of Information Privacy and Security	2
907	Journal of Information Privacy and Security	2
908	Journal of Information Privacy and Security	2
909	Journal of Information Privacy and Security	2
910	Journal of Information Privacy and Security	2
911	Journal of Information Privacy and Security	1
912	Journal of Information Privacy and Security	3
913	Journal of Information Privacy and Security	5
914	Journal of Information Privacy and Security	1
915	Journal of Information Privacy and Security	1
916	Journal of Information Privacy and Security	2
918	Journal of Information Privacy and Security	3
920	Journal of Information Privacy and Security	1
921	Journal of Information Privacy and Security	1
923	Journal of Information System Security	2
926	Journal of Information System Security	2
927	Journal of Information System Security	2
929	Journal of Information System Security	1
930	Journal of Information System Security	2
931	Journal of Information System Security	5
932	Journal of Information System Security	2
933	Journal of Information System Security	3
934	Journal of Information System Security	2
936	Journal of Information System Security	3
937	Journal of Information System Security	2
938	Journal of Information System Security	3
939	Journal of Information System Security	1
941	Journal of Information System Security	1
942	Journal of Information System Security	2
943	Journal of Information System Security	2
945	Journal of Information System Security	2
946	Journal of Information System Security	4
947	Journal of Information System Security	1
949	Journal of Information System Security	3
950	Journal of Information System Security	1
952	Journal of Information System Security	1
953	Journal of Information System Security	4
954	Journal of Information System Security	1
956	Journal of Information System Security	1
957	Journal of Information System Security	1
958	Journal of Information System Security	3
959	Journal of Information System Security	1

960	Journal of Information System Security	6
961	Journal of Information System Security	4
962	Journal of Information System Security	2
963	Journal of Information System Security	1
964	Journal of Information System Security	2
965	Journal of Information System Security	2
966	Journal of Information System Security	2
967	Journal of Information System Security	2
968	Journal of Information System Security	1
969	Journal of Information System Security	2
970	Journal of Information System Security	1
971	Journal of Information System Security	2
972	Journal of Information System Security	3
973	Journal of Information System Security	2
974	Journal of Information System Security	3
975	Journal of Information System Security	1
976	Journal of Information System Security	3
979	Journal of Information System Security	2
980	Journal of Information System Security	1
981	Journal of Information System Security	2
982	Journal of Strategic Information Systems	2
983	Journal of Strategic Information Systems	2
984	Journal of Strategic Information Systems	1
985	Journal of Strategic Information Systems	1
986	Journal of Strategic Information Systems	2
987	Journal of Strategic Information Systems	1
988	Journal of Strategic Information Systems	1
989	Journal of Strategic Information Systems	3
991	Management Information Systems Quartely	2
992	Management Information Systems Quartely	2
993	Management Information Systems Quartely	2
994	Management Information Systems Quartely	5
995	Management Information Systems Quartely	3
996	Management Information Systems Quartely	3
998	Management Information Systems Quartely	2
999	Management Information Systems Quartely	1
1000	Management Information Systems Quartely	2

Apêndice E - Lista de *Stopwords*

O presente apêndice expõe de forma tabular toda a lista de *stopwords* que foram consideradas para a elaboração do projeto. A lista é composta por 906 palavras.

Stopword	Stopword	34	always	68	arising
1	a	35	am	69	around
2	able	36	among	70	as
3	about	37	amongst	71	a's
4	above	38	amoungst	72	aside
5	abst	39	amount	73	ask
6	accordance	40	an	74	asked
7	according	41	and/or	75	asking
8	accordingly	42	and-or	76	asks
9	across	43	announce	77	associated
10	actual	44	anon	78	at
11	actually	45	another	79	auth
12	added	46	any	80	available
13	adj	47	anybody	81	award
14	adopted	48	anyhow	82	away
15	affected	49	anymore	83	awfully
16	affecting	50	anyone	84	b
17	affects	51	anything	85	backed
18	after	52	anyway	86	backing
19	afterwards	53	anyways	87	backs
20	again	54	anywhere	88	be
21	against	55	apart	89	became
22	ah	56	apparently	90	because
23	ahead	57	appear	91	become
24	ain't	58	appreciate	92	becomes
25	all	59	appropriate	93	becoming
26	allow	60	approximately	94	been
27	allows	61	are	95	before
28	almost	62	area	96	beforehand
29	alone	63	areas	97	began
30	along	64	aren	98	begin
31	already	65	arent	99	beginning
32	also	66	aren't	100	beginnings
33	although	67	arise	101	begins

102	behind	145	concerning	188	downs
103	beings	146	consequently	189	downwards
104	believe	147	consider	190	dr
105	below	148	considered	191	du
106	beside	149	considering	192	due
107	besides	150	consisting	193	during
108	better	151	contain	194	e
109	beyond	152	containing	195	each
110	big	153	contains	196	early
111	bill	154	corresponding	197	ed
112	biol	155	could	198	edu
113	birthday	156	couldnt	199	eg
114	both	157	couldn't	200	eight
115	bottom	158	course	201	eighty
116	brief	159	cry	202	either
117	briefly	160	c's	203	eleven
118	but	161	currently	204	else
119	c	162	d	205	elsewhere
120	ca	163	date	206	embodiment
121	came	164	de	207	empty
122	can	165	definitely	208	ended
123	cannot	166	der	209	ending
124	cant	167	describe	210	ends
125	can't	168	described	211	enough
126	caption	169	desired	212	entirely
127	cases	170	despite	213	especially
128	cause	171	detail	214	et
129	causes	172	did	215	et-al
130	certain	173	didn't	216	etc
131	certainly	174	differ	217	even
132	changes	175	different	218	evenly
133	claim	176	differently	219	ever
134	clear	177	discussion	220	every
135	clearly	178	do	221	everybody
136	c'mon	179	does	222	everyone
137	co	180	doesnt	223	everything
138	com	181	doesn't	224	everywhere
139	come	182	doing	225	ex
140	comes	183	done	226	exactly
141	coming	184	don't	227	example
142	completely	185	down	228	except
143	comprises	186	downed	229	f
144	con	187	downing	230	faces

231	fact	274	given	317	hers
232	facts	275	gives	318	herse
233	far	276	giving	319	herself
234	felt	277	go	320	hes
235	few	278	goes	321	he's
236	fifteen	279	going	322	hi
237	fifth	280	gone	323	hid
238	fify	281	good	324	highest
239	fig	282	got	325	him
240	figs	283	gotten	326	himse
241	fill	284	great	327	himself
242	find	285	greater	328	his
243	finds	286	greatest	329	hither
244	fire	287	greetings	330	honor
245	first	288	grouped	331	hopefully
246	five	289	grouping	332	how
247	fix	290	groups	333	howbeit
248	followed	291	h	334	however
249	following	292	had	335	hundred
250	follows	293	hadn't	336	i
251	former	294	happens	337	i'd
252	formerly	295	hardly	338	ie
253	forth	296	has	339	if
254	forty	297	hasnt	340	ignored
255	forward	298	hasn't	341	i'll
256	found	299	have	342	im
257	four	300	haven't	343	i'm
258	from	301	having	344	immediate
259	front	302	he	345	immediately
260	full	303	hed	346	importance
261	fully	304	he'd	347	important
262	further	305	he'll	348	in
263	furthered	306	hello	349	inasmuch
264	furthering	307	help	350	inc
265	furthermore	308	hence	351	indeed
266	furthers	309	her	352	indicate
267	g	310	here	353	indicated
268	gave	311	hereafter	354	indicates
269	generally	312	hereby	355	inner
270	get	313	herein	356	insofar
271	gets	314	heres	357	instead
272	getting	315	here's	358	interest
273	give	316	hereupon	359	interested

360	interesting	403	likely	446	must
361	interests	404	line	447	my
362	into	405	little	448	myse
363	invention	406	'll	449	myself
364	inward	407	long	450	n
365	isn't	408	longer	451	na
366	itd	409	longest	452	name
367	it'd	410	look	453	namel
368	items	411	looking	454	namely
369	it'll	412	looks	455	nay
370	its	413	ltd	456	nd
371	it's	414	m	457	near
372	itse	415	made	458	nearly
373	itself	416	mainly	459	necessarily
374	i've	417	make	460	necessary
375	j	418	makes	461	need
376	just	419	man	462	needed
377	k	420	many	463	needing
378	keep	421	may	464	needs
379	keeps	422	maybe	465	neither
380	kept	423	me	466	never
381	kg	424	mean	467	nevertheless
382	kind	425	meantime	468	newer
383	km	426	meanwhile	469	newest
384	knew	427	meet	470	next
385	know	428	meets	471	nine
386	known	429	men	472	ninety
387	knows	430	merely	473	no
388	l	431	mg	474	nobody
389	largely	432	might	475	non
390	last	433	mill	476	none
391	lately	434	million	477	nonetheless
392	later	435	mine	478	noone
393	latest	436	ml	479	nor
394	latter	437	more	480	normally
395	latterly	438	moreover	481	nos
396	less	439	most	482	not
397	lest	440	mostly	483	noted
398	let	441	move	484	nothing
399	lets	442	mr	485	novel
400	let's	443	mrs	486	now
401	like	444	much	487	nowhere
402	liked	445	mug	488	numbers

489	o	532	particular	575	quite
490	obtain	533	particularly	576	qv
491	obtained	534	parting	577	r
492	obviously	535	parts	578	ran
493	often	536	past	579	rather
494	oh	537	per	580	rd
495	ok	538	perhaps	581	re
496	okay	539	place	582	readily
497	old	540	placed	583	really
498	older	541	places	584	reasonably
499	oldest	542	please	585	recent
500	omitted	543	plus	586	recently
501	on	544	pointed	587	ref
502	once	545	pointing	588	refs
503	one	546	points	589	regarding
504	ones	547	poorly	590	regardless
505	one's	548	possible	591	regards
506	only	549	possibly	592	related
507	onto	550	potentially	593	relatively
508	opened	551	pp	594	reprinted
509	opening	552	predominantly	595	respectively
510	opens	553	preferably	596	resulted
511	or	554	preferred	597	resulting
512	ord	555	present	598	results
513	ordering	556	presented	599	room
514	orders	557	presenting	600	rooms
515	other	558	presents	601	run
516	others	559	presumably	602	s
517	otherwise	560	previously	603	said
518	ought	561	primarily	604	same
519	our	562	probably	605	saw
520	ours	563	problems	606	say
521	ourselves	564	promptly	607	saying
522	out	565	proud	608	says
523	outside	566	provide	609	sec
524	overall	567	provided	610	second
525	owing	568	provides	611	secondly
526	own	569	pt	612	seconds
527	p	570	put	613	section
528	page	571	puts	614	see
529	pages	572	q	615	seeing
530	part	573	que	616	seem
531	parted	574	quickly	617	seemed

618	seeming	661	someone	704	the
619	seems	662	somethan	705	their
620	seen	663	something	706	theirs
621	sees	664	sometime	707	them
622	selves	665	sometimes	708	themselves
623	sensible	666	somewhat	709	then
624	sent	667	somewhere	710	thence
625	serious	668	soon	711	there
626	seriously	669	sorry	712	thereafter
627	seven	670	specifically	713	thereby
628	several	671	specified	714	thered
629	shall	672	specify	715	there'd
630	she	673	specifying	716	therefore
631	shed	674	spp	717	therefrom
632	she'd	675	still	718	therein
633	she'll	676	stop	719	there'll
634	shes	677	strongly	720	thereof
635	she's	678	sub	721	therere
636	should	679	substantially	722	there're
637	shouldn't	680	successfully	723	theres
638	show	681	such	724	there's
639	showed	682	sufficiently	725	thereto
640	showing	683	suggest	726	thereupon
641	shown	684	suitable	727	there've
642	showns	685	sup	728	these
643	shows	686	sure	729	they
644	sides	687	t	730	theyd
645	significant	688	taken	731	they'd
646	significantly	689	takes	732	they'll
647	similar	690	taking	733	theyre
648	similarly	691	tell	734	they're
649	since	692	ten	735	they've
650	sincere	693	tends	736	thick
651	six	694	th	737	thin
652	sixty	695	than	738	thing
653	slightly	696	thank	739	things
654	smaller	697	thanks	740	think
655	smallest	698	thanx	741	thinks
656	so	699	that	742	this
657	so-called	700	that'll	743	thorough
658	some	701	thats	744	thoroughly
659	somebody	702	that's	745	those
660	somehow	703	that've	746	thou

747	though	790	until	833	weren't
748	thoughh	791	unto	834	we've
749	thought	792	up	835	what
750	thoughts	793	upon	836	whatever
751	thousand	794	ups	837	what'll
752	throug	795	upward	838	whats
753	through	796	us	839	what's
754	throughout	797	used	840	what've
755	thru	798	useful	841	when
756	thus	799	usefully	842	whence
757	til	800	usefulness	843	whenever
758	tip	801	uses	844	where
759	to	802	using	845	whereafter
760	today	803	usually	846	whereas
761	together	804	uucp	847	whereby
762	too	805	v	848	wherein
763	took	806	various	849	wheres
764	top	807	've	850	where's
765	toward	808	versus	851	whereupon
766	towards	809	very	852	wherever
767	tried	810	via	853	whether
768	tries	811	viz	854	which
769	truly	812	vol	855	while
770	try	813	vols	856	whim
771	trying	814	vs	857	whither
772	ts	815	w	858	who
773	t's	816	want	859	whod
774	turn	817	wanted	860	who'd
775	turned	818	wanting	861	whoever
776	turning	819	wants	862	whole
777	turns	820	was	863	who'll
778	twelve	821	wasn't	864	whom
779	twenty	822	ways	865	whomever
780	twice	823	we	866	whos
781	two	824	wed	867	who's
782	u	825	we'd	868	whose
783	un	826	welcome	869	why
784	under	827	well	870	widely
785	undergoing	828	we'll	871	will
786	unfortunately	829	wells	872	willing
787	unless	830	went	873	wish
788	unlike	831	were	874	within
789	unlikely	832	we're	875	without

876	wonder	887	yes	898	you're
877	won't	888	yet	899	yours
878	words	889	you	900	yourself
879	working	890	youd	901	yourselves
880	works	891	you'd	902	you've
881	would	892	you'll	903	z
882	wouldn't	893	young	904	zero
883	x	894	younger	905	al
884	y	895	youngest	906	ate
885	year	896	your		
886	years	897	youre		

Apêndice F – *Tag* - Exploração da Fórmula

A tabela seguinte mostra para cada paper qual a tag que lhe foi atribuída segundo a fórmula do "calculovalor".

Idpaper	Keyword Atribuída
1	knowledge protection
2	online privacy
3	electronic mail fraud
4	security attacks
5	data sharing
6	privacy
7	data breaches
8	privacy
9	intrusion response
10	economics of IS security
11	security investments
12	mass attacks
13	crime deterrence
14	ROC curves
15	evidential reasoning
16	BS 7799
20	password
22	Intrusion prevention
23	Encryption
30	Intrusion detection
31	Data security
38	biometrics
40	Login
44	Compliance management
48	Intrusion detection systems
51	key concept
53	Cryptography
68	insider attacks
69	ISO 17799
70	Wireless security
72	cryptography
73	Information security risk analysis
75	ALE
76	Password authentication
85	Secret Content Index File (SCIF)

91	PKI
92	Access control
93	password security
94	password authentication
95	malicious act
97	Proxy signatures scheme
98	X.509
99	certification
100	Key management
101	Key-agreement
102	biometrics
103	Security management
104	Threshold proxy signature
107	Breach
108	DRM
109	audit trails
111	Partially blind signatures
112	Risk assessment
116	Multi-secret
117	Sarbanes-Oxley
121	cryptography
123	Computer forensics
126	Blinding signature
127	key assignment
128	Vulnerability
131	standards
132	cyberterrorism
133	spam
134	Key management
135	Security applications
136	Key agreement
137	Information security governance
138	Hash function
139	Information security obedience
140	Data lifecycle risk analysis
141	Session key
142	Software protection
143	Digital forensics
144	Privacy
145	Security policy quality
147	information security policy
149	Alleged SecurID hash function
150	Smart card
151	Management of security risks
154	Internet threats

155	Virus
156	Elliptic curve cryptosystem
157	Network intrusions
158	Continuous auditing
159	Confidentiality
160	WTLS
160	WAP
161	File integrity
162	Security behaviour
163	Information privacy
165	Hash function
165	Pair block chaining
166	standards
167	Key recovery
168	Certification
169	Computer viruses
170	Individual authentication
171	Digital signature
172	Public key certificates
174	Subliminal channel
175	Information security awareness
176	anonymity
177	Privacy
178	Computer abuse
179	Blind signature
180	Key distribution
181	Hash functions
182	Internet firewalls
183	Risk modelling
184	File integrity analyzers
185	Passwords
187	BS 7799
189	key generation
190	one-way hash functions
191	Security breaches
192	Password authentication
193	Group key distribution protocols
194	One-way Hash Function
195	Biometric
196	ISO17799
197	information security metapolicies
199	password
200	WAP security
201	Audit trail analysis
203	Access control

204	Mandatory access control models
205	Information integrity
206	Password authentication
207	Digital forensics
208	ICMP
209	Key agreement
210	Secret sharing
211	Login
212	Viruses
213	Access control
214	Computer worm
214	Worm detection
215	SSL
216	Direct-Control Cycle
217	Bell-Lapadula model
218	Anonymity
219	ICT security awareness
220	Intrusion detection
221	Distributed denial-of-service flood attacks
222	IT auditing
223	RFID
224	Intrusion detection systems
225	Distributed denial of service attacks
226	Location privacy
227	Key agreement
228	Intrusion detection
229	RBAC
230	SSL/TLS session-aware
231	Password authentication
232	Insider misuse
233	Authenticated key exchange
234	Intrusion detection
235	Security policy implementation
236	Privacy
237	Information security management
238	Group key agreement
239	Business information risk
241	Awareness
242	User authentication
243	RFID security
244	Passwords
245	PKI
246	Vulnerability scanner (VS)
247	Certified electronic mail
248	Key distribution

249	Authentication
250	Information assurance
251	text categorization
254	cryptography
255	Anomaly detection
256	Corporate cyber harassment
257	Layered security approach
258	Key exchange
259	Intrusion detection
260	Classification
261	Access controller
262	Public key encryption
263	Worm containment
264	Fingerprinting
265	Optimistic Access Control
266	SCADA network security
267	Taxonomy
268	Electronic voting
269	Cryptographic protocol
270	EKE
271	Privacy
272	Storage integrity
273	Public-key cryptography
274	Cryptographic Key Assignment
275	PIN cracking
276	Key distribution
278	Key agreement
279	Intrusion detection
280	Security protocols
281	Session key
282	Virus
283	Keylogger
284	OSI security model
285	Key distribution
286	Intrusion detection systems
287	Quantitative risk analysis
288	Attack probabilities
289	USB forensics
290	Cryptographic protocol
291	Internet worms
292	Virtual machine security
293	Taxonomy
294	Cryptographic key assignment
295	Spam
296	Intrusion detection

297	Random projection
299	TCM-KNN (Transductive Confidence Machines for K-Nearest Neighbors)
300	Information flow anomaly detection
301	Information security practices
302	SEM
303	Quantitative security modeling
304	SQL injection attack
305	Anti-spoofing
306	Malicious code
307	Policy development
308	Digital watermarking
309	IDS systems and platforms
310	Encryption
311	Anomaly intrusion detection
311	Distributed denial-of-service attacks
312	Intrusion detection systems
313	Spam
314	Intrusion detection
315	Spam
316	key management
317	Risk analysis
318	Security requirement engineering
319	Information security awareness
320	Security costs and benefits
321	PKI
322	Electronic voting
323	Encryption
325	Worm attack
326	Risk perception
327	Authorisation manager
329	Intrusion detection
330	Image steganalysis
331	DDoS
332	Biometric based key cryptography
333	Privacy
334	misfeasor detection
335	Password-authenticated key
336	Intrusion detection
337	Passive covert timing channel
338	Auction protocol
338	Anonymity
339	Dynamic quarantine
340	Information security awareness
341	Intrusion detection systems

342	Digital signatures
343	information privacy
344	Security controls
345	Taxonomy
347	Intrusion detection
348	Software security
349	Information security culture
350	Access control security
351	Awareness
352	Role-based access control
353	Training and awareness
354	CSP
355	Biometric verification
356	Regulating service
357	Botnet
358	Smart cards
360	Watermarking
361	Password recovery
362	Adaptive hash-chaining
363	Rootkits
364	SSL/TLS
366	Phishing
367	Packet tracing
368	Evidence reasoning
369	Keystroke dynamics-based
371	Biometric verification
372	Security policy
373	Software safety
374	Covert-timing channel free security
375	Biometrics
376	Password authentication
377	Authentication
378	Alert causal correlation
379	Electronic voting
380	Access control
381	Attack description
382	Intrusion intention prediction
383	Fraud detection
383	Account signatures
384	Vulnerabilities
385	Phishing
386	One-way hash function
387	Anonymization
389	Risk assessment
390	Biometric

391	Risk management
393	Security breaches
395	Batch rekeying
395	Group key management
396	Masquerader detection
397	Intrusion detection system
399	ISO/IEC 17799
400	Digital Rights Management
401	intrusion detection systems
402	Real-time NIDS
403	VoIP security
404	Security agents
405	Incident Handling
406	Security attitudes
407	PKI
408	Keystroke analysis
408	Biometrics
409	Trojan circuit
410	Spam over Internet Telephony (SPIT)
411	Cryptography
412	Relay attack
414	Incident response
415	Computer forensics
416	Information leakage
417	Access control
418	Active worms
419	IDS systems and platforms
420	Information security culture
421	Access control
422	Defensive warfare
423	Copyright protection
424	Firewalls
425	Intrusion detection
426	Snort
427	computer virus
428	Perceived security
429	Roving bug
430	Inter-domain PKI
431	Specification-based approach
432	Information hiding
433	Security patterns
434	Threats taxonomy
435	digital signatures
436	DDoS
437	Sensor security

438	Classification methods
439	Holistic security management
440	steganography
441	Red teams
442	Covert channels
443	Access control
444	Bloom filter
445	Web content protection
447	GTRBAC
448	Biometrics
450	Security hardening patterns
451	Denial of service
452	Certification
453	Mobile code security
454	Cross-site scripting (XSS)
455	Internet worms
456	Access control
457	Access control
458	security management
459	Intrusion detection
460	Digital signature
461	Spam SMS messages
462	Biometrics
463	Anomaly visualization
464	Spam
465	SSL
466	ECDSA
467	NoScript
469	Intrusion detection
470	Group-based RBAC
471	SIR
472	Intrusion Detection
473	Role-based access control
474	Electronic signatures
475	Authorization
476	Online alert correlation
477	Access control
478	DDoS attacks
479	Privacy infrastructure
480	Denial of service
481	One-time signature
482	B3G networks
483	Policy enforcement
484	Legal compliance
485	Certified e-mail probabilistic model checking

486	DoS
487	Vulnerability scanners (VSs)
488	Vulnerability discovery model (VDM)
489	Threshold cryptography
490	K-means clustering
491	computer virus
492	DDoS/DoS
493	Disclosure inference
494	Web security
495	Network intrusion detection
496	Digital signatures
497	Alert analysis
498	Attack detection
499	Delegated authorization
500	Distributed backup
501	Global scanning worms
502	Attack Modeling
504	Security Engineering
505	Security protocols
506	IDS
507	Wormhole attacks
508	Risk assessment
509	Element-wise encryption
510	Access controls
511	Key hierarchies
512	RFID
513	Fingerprinting
514	Okamoto-Tanaka-Uchiyama
516	Privacy
517	Pairing-based cryptosystems
519	Privacy
520	Secret sharing
521	AES
522	Public key cryptography
523	Public key cryptography
524	Biometric key
525	RSA-CRT
526	Privacy protection
527	PKI
528	Vulnerability
529	Privacy policy
530	Distributed signatures
531	Public verifiability
532	Relation attack
533	Authorization policy

534	Public-key cryptography
535	Universal re-encryption
536	Intrusion detection
538	Key sharing
539	Key exchange graph
540	Broadcast encryption
540	Copyright protection
542	Public key
543	Cryptographic protocol
544	Non-repudiation
545	Secure coprocessors
546	Cryptovirus
547	Block cipher
547	XTEA
548	Commercial key recovery
549	Stream authentication protocols
550	Public-key encryption
551	Recursive certificate
552	RSA
553	Identity-based signatures
554	User-controlled-anonymity
555	Public key encryption
556	Conference key agreement
557	RSA
558	Mobile Authentication Protocol
560	Fingerprinting
561	Opacity
562	Key pre-distribution
563	Stream authentication
564	Specification of Security Protocols
565	Fingerprints
567	Block cipher
569	Intrusion signatures
570	Authorization
570	Role-based access control
571	RSA
572	Access control model
573	Public key cryptography
574	Universal designated verifier signature
575	Diffie-Hellman
576	Web privacy
577	Physical unclonable function
578	Wireless sensor network security
578	Id-based signature
579	Password

580	Digital certificates
581	Computer worms
583	OrBAC
584	Access control
585	ElGamal
586	Homomorphic encryption
587	Iterated hash functions
588	Encryption feature
589	Cryptographic protocols
590	Anonymity and privacy
591	Hierarchical identity-Based encryption
592	Certified email
593	Digital signature
594	Key generation
595	Identity based encryption
596	Privacy protection
597	Public key infrastructure
598	Paillier public-key system
599	Linguistic steganography
600	Buffer overflows
601	Privacy
601	Secure multiparty computation
602	User credentials
603	Fair non-repudiation
605	Public-key cryptography
606	Identity based Crytography
607	Controlled query evaluation
609	Intrusion detection system
610	Reactive simulatability/UC
611	Intrusion detection system
612	Biometrics
613	On-line recovery
614	Fingerprint
615	Covert channels
616	RBAC
617	IP security protocols
618	Access control
619	Public-key cryptography
620	Client puzzles
621	PKI
622	Reactive
623	Cryptographic protocols
624	Key management
625	Cryptographic protocols
626	Trust and attestation

627	Biometric personal authentication
628	Image hash
630	Steganalysis
630	Embedding capacity
631	Network security
632	Signature schemes
633	Privacy protection
634	Social engineering
635	Denial of service
637	MNP
638	Access control
639	Network Security
640	Smart cards
641	Covert timing channels
643	Group key establishment
644	Cryptography
645	Digital signature
646	Usable security
647	Two-channel cryptography
648	Cryptographic protocols
649	Hardware hacking
650	Information flow control
651	Audit
652	Security protocols
654	Privacy
655	Language-based security
657	Cryptographic file systems
658	Network security
659	Policy refinement
660	OrBAC model
661	Wireless security
662	Simulatability
663	Elastic block ciphers
664	Access control
665	Access control
667	Privacy
668	Secure Electronic Transactions (SET)
669	Database intrusion
670	Correlation-based IDS
671	Software protection
672	Inference control
673	Public-key cryptography
674	Steganography
676	Security protocols
677	Digital signature

678	Privacy information hiding
679	Digital rights management
680	Security protocols
681	Biometrics
682	Software security
683	Symmetric authentication
684	Key exchange
685	Security Engineering
686	Security policies
687	Public-key infrastructure
688	SPKT
689	Language-based security
689	Static analysis
690	Grid security
691	Intrusion Detection systems
692	Information flow
693	Access control
694	Security view
695	PKI
696	SEM
697	Crime research
698	Data security
699	Risk analysis
700	Risk management
701	Risk analysis
702	Data security
703	Security products
704	Data security
705	Data security
706	Surveillance
707	Data integrity
708	Anti-virus software
709	Data security
710	Data integrity
711	Privacy
712	Data integrity
713	Data integrity
714	Data integrity
715	Network security
716	Intrusion detection systems
717	Passwords
718	Intrusion detection systems
719	Data integrity
720	Privacy
721	Data security

722	Data security
723	Data security
724	Risk analysis
725	Data security
726	Data security
727	Data security
728	Data security
729	Data security
730	Business policy
732	Security management
733	Risk analysis
734	Security management
735	Risk assessment
736	Data security
737	Fault tolerance
738	Privacy
739	Control systems
740	Digital certificates
741	Biometrics
742	Network security
743	Security management
744	Access control
745	Data security
746	Network security
747	Network security
748	Risk management
749	Privacy
750	Security management
751	Data security
753	Message authentication
754	Computer privacy
755	Data security
756	Data security
757	Security products
758	Message authentication
759	Message authentication
760	Data security
761	Biometrics
762	Message authentication
763	Encryption
764	Risk management
766	Privacy
767	Signal detection
768	Data security
769	Data security

770	Privacy
771	Data security
772	Standards
773	Risk assessment
774	Data security
776	Viruses
777	Hacking
778	Encryption
779	Data security
780	Hacking
781	Criminals
782	Digital signatures
785	Data security
788	Cryptography theory
789	Data security
790	Information disclosure
791	Privacy
794	Biometrics
796	Digital signatures
797	Privacy
799	Security products
800	Policy audit
801	Risk analysis
802	Risk management
804	privacy
805	Privacy
806	Biometrics
807	Standards
809	Hacking
810	Computer viruses
811	Computer privacy
812	Computer crime
814	Data security
816	Data security
817	Data security
818	Business policy
819	Encryption
820	Data security
821	Biometrics
823	Risk management
824	Data security
825	Security products
826	Information control
827	Privacy
828	Privacy

829	Risk exposure
830	Availability
831	untrusted servers
833	digital piracy
836	data protect
836	privacy regulations
837	computer privacy
839	electronic commerce privacy
840	formation security policy
841	security values
842	IS security research
843	code review
844	protective information technologies
845	information privacy
845	Internet users' information privacy concerns
846	information systems risk management
847	patch release time
848	ROC curves
849	security technology investment
850	IS misuse
851	copyright
852	privacy calculus
853	software piracy
854	piracy
855	information assurance
856	Confidentiality
857	Correlation intractability
858	Quantum cryptography
859	Access control
860	Access control
861	cryptography
865	privacy
866	severity
867	spyware
870	passphrases
872	Role-Based Access Control
874	detection
875	Privacy belief
876	Shadowcrew
877	BAIT
878	online privacy
881	Universal Identification
882	Data Recovery
883	Password Management
884	Computer forensics

886	Security Framework
887	Online Privacy Issues
888	Intrusion Detection
889	Identity Theft
890	Intrusion Detection and Protection
892	Password Protocols
893	Information Security Risk
894	Information Security Practices
895	Passwords
896	Social and Behavioral Aspects of Information Security
897	Security and Privacy Policy Obligations
901	Investments in Information Security
902	worm propagation
903	privacy
905	Misuse Cases
906	E-privacy Risk Concerns
907	Privacy
908	Security Executives
909	Online Privacy
910	Employee Information Privacy Concern
911	Privacy Belief
912	Health Insurance Portability and Accountability Act (HIPAA)
913	Phishing
914	enhanced SSL
915	Privacy
916	Privacy
918	defence strategy
920	Passwords
921	Web Service Security
923	Public Key Infrastructure
926	Passwords
927	blind decryption
929	user authentication
930	RFID
931	Auditing
932	Enterprise DRM
933	spam
934	Botnets
936	Intrusion detection
937	Disaster recovery
938	risk analysis
939	spam
941	hacker profile
942	audit objectives
943	Intrusion Detection Systems

945	Incident Management process
946	spam e-mail
947	grid security
949	intrusion detection systems
950	IT Disaster Recovery Planning
952	business continuity planning
953	Social Engineering
954	extreme event planning
956	Threat Modeling
956	STRIDE
957	Security Requirements
958	Privacy Risk
959	Distributed firewalls
960	Computer Viruses
961	information systems misuse
962	Social Aspects of Information Security
963	vulnerabilities
964	security model
965	incident response planning
966	risk assessment
967	Secure multi-party computation
968	privacy
969	intrusion detection
970	PriS-Tool
971	SCADA protocol security
972	information security maturity model
973	secure virtual machine
974	ISO/IEC 27001
975	systemic integrity
976	security modelling
979	hacker profile
980	disaster response
981	ICMP traceback
982	Corporate privacy
983	Software piracy
984	Privacy
985	Sarbanes-Oxley Act
986	Privacy
987	Privacy
988	Security policies
989	RFID
991	employees' compliance with security policies
992	Sarbanes-Oxley Act
993	IS security policies
994	behavioral issues of information security

995	virtuous IT
996	threat appraisal
998	downtime loss
999	ISS de jure standards
1000	security management code of practice

Apêndice G – *Tag* Final

A próxima tabela identifica a tag final por cada paper. Poderão ser visto casos em que para um paper foram identificadas mais do que uma *keyword*.

Id Paper	Keyword
1	Safe contexts
1	Security professionals
2	online privacy
3	Phishing
4	security attacks
5	data sharing
6	privacy
7	data breaches
8	privacy
9	Intrusion detection
9	Intrusion prevention
10	economics of IS security
11	security investments
12	targeted attacks
13	crime deterrence
14	Postaudit
15	evidential reasoning
16	Information security certification
17	Guidelines
17	Security Breaches
17	violation of safeguards
18	Information Security Management
19	Information Security Governance
20	password
21	Intrusion detection
22	Intrusion prevention
23	Encryption
24	Cyber crime
24	privacy
25	Communication security
26	Virus
27	Computer forensics
27	Forensics
28	Information Security Managers
29	Web Security

30	Intrusion detection
31	Business Security
32	Software safety
33	Legislation
34	Risk analysis
35	Denial of service attacks
36	Policy update
37	Security mechanisms
38	Authentication
39	Spoofing
40	Authentication
41	Key recovery
42	Intrusion detection systems
43	IS Security Research
44	Compliance management
45	Firewall
45	Firewalls
46	Access security
46	Audit
47	Secure outsourcing
48	Intrusion detection systems
49	Fraud
49	Frauds
50	Protection
51	Risk management
52	IS Security Research
53	Security Systems
54	Malware
55	Cryptography
56	Protection
57	Encryption
58	Security RoI
59	Data Integrity
60	Criminals
60	Frauds
61	Cybercrime
62	Network security
63	Intrusion
63	Intrusion detection
64	Secure Electronic Transactions
65	privacy
66	Intrusion
66	Intrusion detection
66	Security products
67	PKI

67	Public key infrastructure
68	insider attacks
69	ISO 17799
70	Wireless security
71	Fraud
72	cryptography
73	Information Security Management
74	Controls
75	ALE
76	Password authentication
77	Firewall
77	Firewalls
78	SET
79	Internal controls
79	Legislation
79	Reporting
80	Evidence
81	Privacy law
82	Firewall
82	Intrusion
82	Intrusion detection
82	Intrusions
82	VS
83	Cyber-terrorism
84	IPsec
85	Confused Document Encrypting Scheme (CDES)
86	Network monitoring
86	Internet security
87	Elliptic curve cryptography
88	Public key
88	Public key infrastructures
89	Evidence
90	Security professionals
91	PKI
92	Access control
93	password security
94	password authentication
95	malicious act
96	Information Security Policy
97	Proxy signatures scheme
98	X.509
99	certification
100	Key management
101	Password
102	Authentication

103	Security management
104	Threshold proxy signature
105	Fraud
105	Criminal
106	Information Assurance
107	Breach
108	Piracy
109	audit trails
110	Worm
111	RSA
112	Risk assessment
113	spam
114	Standards
115	Security journal
116	Multiple signing policies
117	Sarbanes-Oxley
118	Taxonomy
119	Red teams
120	Biometric
121	Remote authentication
122	Legislation
122	privacy
123	Computer forensics
124	Firewall
125	Access control
126	Blinding signature
127	Cryptography
128	Vulnerability
129	Attack
129	Control
130	Virus
131	Guidelines
132	cyberterrorism
133	spam
134	Key management
135	Cryptography
136	Secure authentication
137	Information security governance
138	Smart card
139	Information security obedience
140	Information lifecycle security risk assessment
141	Password
142	Software protection
143	Digital forensics
144	Privacy

145	Security policy quality
146	Public-key
147	information security policy
148	Risk analysis
149	Alleged SecurID hash function
150	Smart card
151	Management of security risks
152	Cryptographic Key Assignment
152	key assignment
153	Information warfare
154	Internet threats
155	Virus
156	Elliptic curve cryptosystem
157	Network intrusions
158	Continuous auditing
159	Secure information systems development
160	WTLS
161	File integrity
162	Security behaviour
163	Privacy law
164	Attack
164	Security professionals
165	Pair block chaining
166	process certification
167	Key recovery
168	Certification
169	Computer viruses
170	Individual authentication
171	RSA
172	Authentication
173	Security mechanism
174	Digital signature
175	Information security awareness
176	anonymity
177	Privacy
178	Incidents
179	Electronic voting
180	Key distribution
181	Online dictionary attacks
182	Internet firewalls
183	Security engineering
184	Tampering modes
185	Passwords
186	Evidence
187	BS 7799

188	Access security
189	RSA
190	key assignment
191	Information Security Policy
192	Password authentication
193	Authentication protocols
194	Access control
195	Biometric
196	insider threat
197	information security metapolicies
198	Availability
199	password
200	WAP security
201	Intrusion detection
202	Access control
203	Access control
203	Access control policies
204	Lattice-based access control models
205	Internal controls
205	Risk management
206	Password authentication
207	Analysis of control
208	PING attack
209	Key agreement
210	Secret sharing
211	Remote authentication
211	Smart cards
212	Viruses
213	Access control
214	Worm detection
215	SSL
216	Direct-Control Cycle
217	Access control
218	Key exchange
219	ICT security awareness
220	Intrusion detection
221	Anomaly detection
222	IT auditing
223	RFID
224	Intrusion detection systems
225	Distributed denial of service attacks
226	Location privacy
227	Key agreement
227	Anonymity
228	Intrusion detection

229	RBAC
230	SSL/TLS session-aware
231	Password authentication
232	insider threat
233	Authenticated key exchange
234	Intrusion detection
235	Security policy implementation
236	Privacy
237	Awareness
238	Group key agreement
239	Information Security Governance
240	RBAC
241	Awareness
242	User authentication
242	Passwords
243	RFID security
244	Passwords
245	SSL
246	Vulnerability forecasting
247	Certified electronic mail
248	Key distribution
249	Authentication
250	Information assurance
251	Intrusion detection
252	Software security
253	Trusted computing
254	Password
255	Anomaly detection
256	Corporate cyber harassment
257	Layered security approach
258	Key exchange
259	Anomaly detection
260	Intrusion detection
261	Security manager
262	Multi-dimensional credentialing
263	Worm containment
264	Digital watermarking
265	Usage Control
266	Intrusion detection systems
267	Taxonomy
268	Electronic voting
269	Cryptographic protocol
270	Key exchange
271	Privacy
272	Data integrity

273	Signcryption
274	Access control
275	Security APIs
276	Key distribution
277	Threats
278	Key agreement
279	Intrusion detection
280	Security protocols
281	Authentication
282	Virus
283	Internet login
284	OSI security model
285	Authentication
286	Intrusion detection systems
287	Quantitative risk analysis
288	Attack probabilities
289	USB forensics
290	Cryptographic protocol
291	Internet worms
292	.Net security
293	Taxonomy
294	Cryptographic key assignment
295	SIDF
296	Stepping-stone
297	Privacy-preserving clustering
298	SET
299	TCM-KNN (Transductive Confidence Machines for K-Nearest Neighbors)
300	Dynamic information flow analysis
301	Information security practices
302	SEM
303	Vulnerabilities
304	SQL injection attack
305	Firewalls
306	Malicious code
307	Policy development
308	Digital watermarking
309	IDS systems and platforms
310	Encryption
311	Statistical detection
312	Intrusion detection systems
313	Spam
314	Computer audit data
315	Anti-spam filter
316	key management

317	Risk analysis
318	Security requirement engineering
319	Information security awareness
320	Security costs and benefits
321	Electronic signature
322	Electronic voting
323	Encryption
324	Electronic Commerce Security
325	Worm attack
326	Information Security managers
327	Access control
328	Encryption
329	Intrusion detection
330	Image steganalysis
331	DDoS
332	Biometric based key cryptography
333	Privacy
334	Threat prediction
335	Cryptanalysis
336	Intrusion detection
337	Passive covert timing channel
338	Authentication tests
339	Dynamic quarantine
340	Information Security Management
341	Intrusion detection systems
342	Integrity
343	Information Security baselines
344	Security controls
345	Taxonomy
346	Access
346	Detection
347	Masquerade detection
348	Security hardening
349	Information security culture
350	DNP3
351	Awareness
352	Role-based access control
353	Training and awareness
354	CSP
355	Automated Personal Identification
356	Information Security awareness
357	Botnet
358	privacy
359	Key escrow
359	Key recovery

360	Watermarking
361	Password recovery
362	Data source authentication
362	Non-repudiation
363	Rootkits
364	SSL/TLS
365	Fingerprint
366	Phishing
367	Packet tracing
368	Threat evaluation
369	Retraining
370	Security architecture
371	Automated Personal Identification
372	Security policy
373	Software safety
374	Real-time information flow
375	Keystroke dynamics
375	user authentication
376	Keystroke dynamics
377	Authentication
378	Alert causal correlation
379	Electronic voting
380	Access control
381	Attack description
382	Intrusion intention prediction
383	Fraud detection
384	Fraud Detection
385	Phishing
386	One-way hash function
387	Anonymization
388	SET
389	Risk assessment
390	Biometric
391	Risk management
392	PKI
392	Public key infrastructure
393	Security breaches
394	Information Security Management
395	Batch rekeying
395	Secure group communication
396	Masquerader detection
397	Intrusion detection system
398	Security behaviour
399	Information security requirements
400	Digital Rights Management

401	intrusion detection systems
402	Real-time NIDS
403	VoIP security
404	Intrusion detection
405	Incident Handling
406	Security perceptions
407	PKI
408	user authentication
409	Trojan circuit
410	Spam over Internet Telephony (SPIT)
410	spam
411	Cryptography
412	Token
413	Standard
414	Incident response
415	Audit
416	Information leakage
417	Access control
417	Policy spaces
418	Active worms
419	IDS systems and platforms
420	Information security culture
421	Access control
422	Defensive warfare
423	Copyright protection
424	Firewalls
425	Intrusion detection
425	Detection mechanism
426	Honeypots
427	computer virus
427	Notification process
428	Perceived security
429	Surveillance
430	Cross-certification
431	Specification-based approach
432	Information hiding
433	Security patterns
434	Threats taxonomy
435	digital signatures
436	DDoS
437	Sensor security
438	Statistical disclosure control
439	Holistic security management
440	Cryptanalysis
441	Red teams

442	Covert channels
443	Vulnerability
444	Bloom filter
445	Web content protection
446	Digital rights management
446	DRM
446	privacy rights
447	GTRBAC
448	Keystroke Identification
449	Detection
449	Threats
450	Security hardening patterns
451	Denial of service
452	Certification
453	Mobile code security
454	Cross-site scripting (XSS)
455	Internet worms
456	Access control
457	Access control
458	security management
459	Intrusion detection
460	Proxy signature
461	Content-based filtering
462	Biometrics
463	Intrusion detection
464	Spam
465	SSL
466	ECC
467	Malware
468	Key distribution
469	Intrusion detection
470	Group-based RBAC
471	SIR
472	Intrusion Detection
473	Role-based access control
474	Electronic signatures
475	Authorization
476	Online alert correlation
477	Access control
478	DDoS attacks
479	Privacy infrastructure
479	Privacy-enhancing technologies
480	Attacker model
481	Block signature
481	Stream signature

482	B3G networks
483	Policy enforcement
484	Legal compliance
485	Certified e-mail
486	DoS
487	Vulnerability scanners (VSs)
488	Vulnerability discovery model (VDM)
489	Threshold cryptography
490	K-means clustering
491	computer virus
491	Notification process
492	DDoS/DoS
493	Availability
494	Web security
495	Network intrusion detection
496	Digital signatures
497	Alert analysis
498	Attack detection
499	Single sign-on
500	Distributed backup
500	Access control
500	Malicious peers
501	Worm virulence estimation
502	Attack Modeling
502	Attack trees
503	Evidence
503	Usable security
504	Security Engineering
505	Security protocols
506	IDS
507	Wormhole attacks
508	Risk assessment
509	Element-wise encryption
509	Digital signature
510	Access controls
511	Key hierarchies
512	Fingerprint
513	Fingerprinting
514	Okamoto-Tanaka-Uchiyama
515	Attack detection
516	Surveillance
517	Pairing-based cryptosystems
518	SET
519	Privacy
520	Adversary structure

521	Block ciphers
522	Public key cryptography
523	Public key cryptography
524	Biometric key
525	Differential power attack
525	Fault attack
526	Privacy protection
527	PKI
528	Vulnerability
529	Privacy practices
530	Distributed key distribution schemes
531	Public verifiability
532	E-voting
533	Authorization policy
534	Identity-based signature
534	Multisignature
535	Universal re-encryption
536	Intrusion detection
537	Signature
538	Hash functions
539	Information-theoretically Secure
540	Copyright protection
541	Cryptosystems
542	Minrank
543	Cryptographic protocol
544	Non-repudiation
545	Secure coprocessors
546	Cryptovirus
547	Related-key rectangle attack
547	XTEA
548	Commercial key recovery
549	Stream authentication protocols
550	Electronic voting
551	Recursive certificate
552	RSA
553	Identity-based signatures
553	Key escrow
554	Direct anonymous attestation
555	Public key encryption
556	Conference key agreement
557	RSA
558	SAML
559	Cryptographic
560	Web application identification
561	Opacity

562	Key pre-distribution
563	Stream authentication
564	privacy
565	Identification capability
566	Electronic commerce security
567	Cache attack
568	Protocol design
569	Intrusion detection
570	Role-based access control
571	RSA
572	Access control model
573	Public key cryptography
574	Universal designated verifier signature
575	Diffie-Hellman
576	Web privacy
577	Certified execution
577	Controlled PUF
578	Id-based signature
579	Password
580	Access control
581	Computer worms
582	Cryptographic protocols
582	Simulatability
583	OrBAC
584	Access control
585	ElGamal
586	Homomorphic encryption
587	Checksums
588	Encryption feature
589	Key disabling
590	Privacy-preserving revocation checking
591	Hierarchical identity-Based encryption
592	Certified email
593	Digital signature
594	Key generation
595	Identity based encryption
596	Privacy protection
597	Identity-based cryptography
598	Verifiable shuffles
599	Linguistic steganography
600	Buffer overflows
601	Secure outsourcing
602	Access control
603	Fair non-repudiation
604	Availability

604	SET
605	Certificateless encryption
606	Identity based Crytography
607	Controlled query evaluation
608	Group key management
609	Secure clustering and routing
610	Reactive simulatability/UC
611	Intrusion detection system
612	Keystroke analysis
613	On-line recovery
614	Fingerprint
615	Covert channels
616	Delegation of authority
617	IP security protocols
618	Access control
618	Audit
619	Public-key cryptography
620	Client puzzles
621	PKI
622	Reactive
623	Cryptographic protocols
624	Key management
625	Cryptographic protocols
626	Trust and attestation
627	Biometric personal authentication
628	Image hash
629	Worm attack
629	Worm detection
630	Steganography
630	Embedding capacity
631	Network security
631	Security technologies
632	ECDSA
633	Semi-trusted third party
634	Social engineering
635	Denial of service
636	Credential
637	MNP
638	Authorization
639	Network Security
640	Smart cards
641	Covert timing channels
642	Integrity
643	Group key establishment
644	Cryptography

645	Key substitution
646	Graphical passwords
647	Hash functions
648	Security Protocols
649	Hardware hacking
650	Information flow control
651	Audit
652	Security protocols
653	Trusted Computing
654	Privacy
655	Access control
656	Internet security
657	Cryptographic file systems
658	Network security
659	Policy refinement
660	Role-Based Access Control
661	Wireless security
662	Simulatability
663	Elastic block ciphers
664	Authorization
665	Access control
666	Information flow control
667	Privacy
668	Secure Electronic Transactions (SET)
669	Database intrusion
670	Correlation-based IDS
671	Software protection
672	Confidentiality
673	Public-key cryptography
674	Steganography
675	Static analysis
676	Security protocols
677	Digital signature
678	Privacy information hiding
679	Digital rights management
680	Security protocols
681	Cryptographic keys
682	Information flow control
683	Symmetric authentication
684	Key exchange
684	RSA
685	Security Engineering
686	Security policies
687	Certificate revocation
687	Merkle hash tree

688	SDSI
689	Static analysis
690	Grid security
690	Authorization
690	Behavioral control
691	Network security
692	Information flow
693	Access control
694	Security view
695	PKI
696	Trusted computing
697	Crime research
698	Data security
699	Risk analysis
700	Risk management
701	Risk analysis
702	Data security
703	Security products
704	Data security
705	Data security
706	Surveillance
707	Data integrity
708	Anti-virus software
709	Data security
710	Data integrity
711	Privacy
712	Data integrity
713	Data integrity
714	Data integrity
715	Network security
716	Intrusion detection systems
717	Passwords
718	Intrusion detection systems
719	Data integrity
720	Privacy
721	Data security
722	Data security
723	Data security
724	Risk analysis
725	Data security
726	Data security
727	Data security
728	Data security
729	Data security
730	privacy

731	Perceived security
732	Security management
733	Risk analysis
734	Security management
735	Risk assessment
736	Data security
737	Fault tolerance
738	Privacy
738	Legislation
739	Risk management
740	Smart cards
741	Biometrics
742	Network security
743	Security management
744	Access control
745	Data security
746	Network security
747	Network security
748	Privacy
749	Privacy
750	Security management
751	Data security
752	Espionage
752	Weapon
752	SET
753	Message authentication
754	Computer privacy
755	Data security
756	Data security
757	Security products
758	Data security
759	Message authentication
760	Data security
761	Biometrics
762	Message authentication
763	Encryption
764	Risk management
765	Control
766	Privacy
767	Signal detection
768	Data security
769	Data security
770	Privacy
771	Data security
772	Standards

773	Risk assessment
774	Data security
775	Attacks
776	Viruses
777	Hacking
778	privacy
779	Data security
780	Hacking
781	Criminals
782	Digital signatures
783	ISO 17799
784	Web service security
785	Data security
786	Computer viruses
786	Virus
786	Viruses
787	Electronic commerce security
788	Cryptography theory
789	Data security
790	Information disclosure
791	Privacy
791	Identification
792	Hacking
793	Audit
793	Intrusion
793	Intrusion detection
794	Biometrics
795	Integrity
796	Digital signatures
797	Privacy
798	Information Security Management
798	Security management
799	Security products
800	Policy audit
801	Risk analysis
802	Risk management
803	Information systems security management
803	Security knowledge
803	Security management
803	Systems security
804	Privacy
805	Privacy
806	Biometrics
807	Standards
808	Denial of service attacks

809	Hacking
810	Computer viruses
811	Computer privacy
812	Computer crime
813	Countermeasures
813	Identification
813	Security assurance
813	Security design
813	Security engineering
813	Systems security
814	Data security
815	Threats
816	Data security
817	Data security
818	Business policy
819	Encryption
820	Identification
821	Biometrics
822	PIN
823	Risk management
824	Data security
825	Security products
826	Information control
827	Privacy
828	Privacy
829	Risk exposure
830	Confidentiality
830	Integrity
831	untrusted servers
832	privacy concerns
833	digital piracy
834	Security Breaches
835	Social Aspects of Information Security
836	data protect
836	privacy rights
837	target and shield model
837	Theory of high reliability organizations
838	Evidence
839	Internet privacy
840	formation security policy
841	security values
842	IS security research
843	code review
844	protective information technologies
845	information privacy

846	information systems risk management
847	patch release time
848	ROC curves
849	Security Breaches
850	Security countermeasures
851	Digital piracy
852	privacy calculus
853	software piracy
854	piracy
855	security investment
856	Data Perturbation
857	Correlation intractability
858	Quantum cryptography
859	Policy design
860	Access control
861	Passwords
862	Public key infrastructures
863	Guidelines
864	privacy concerns
865	privacy
865	information privacy
866	Threat
867	Awareness
868	Awareness
869	System security
870	passphrases
871	privacy concerns
871	Privacy Risk
872	Security knowledge
873	Security model
874	detection
875	Information disclosure
876	Hacker
877	Computer forensics
878	privacy concerns
879	Security threats
879	Threats
880	Block cipher
881	Universal Identification
882	Data Recovery
882	Hard Drive Forensics
883	Password Management
884	Computer forensics
885	Access control
885	Smart card

885	SET
886	Security training
887	Online Privacy Issues
888	Intrusion Detection Management System
889	Identity Theft
890	Intrusion Detection and Protection
891	online privacy
891	privacy policies
891	Privacy policy
892	Strong Passwords
893	Information Security Risk
893	Safeguard
894	Information Security policies
895	Passwords
896	Social and Behavioral Aspects of Information Security
897	Security and Privacy Policy Obligations
898	Network security
898	Security configuration
898	Security Systems
899	Security Breaches
899	Criminal
899	Intrusions
900	Electronic voting
901	Investments in Information Security
902	worm propagation
903	privacy
904	privacy
905	Misuse Cases
906	E-privacy Risk Concerns
907	Privacy
908	Security Executives
908	Information security investment
909	Online Privacy
910	Employee Information Privacy Concern
911	Privacy Belief
912	Privacy Enhancing Technologies (PETS)
913	Identity Theft
914	enhanced SSL
915	Privacy
916	Privacy
917	Standards
918	Information assurance
919	Incident Response
919	Virus
920	Passwords

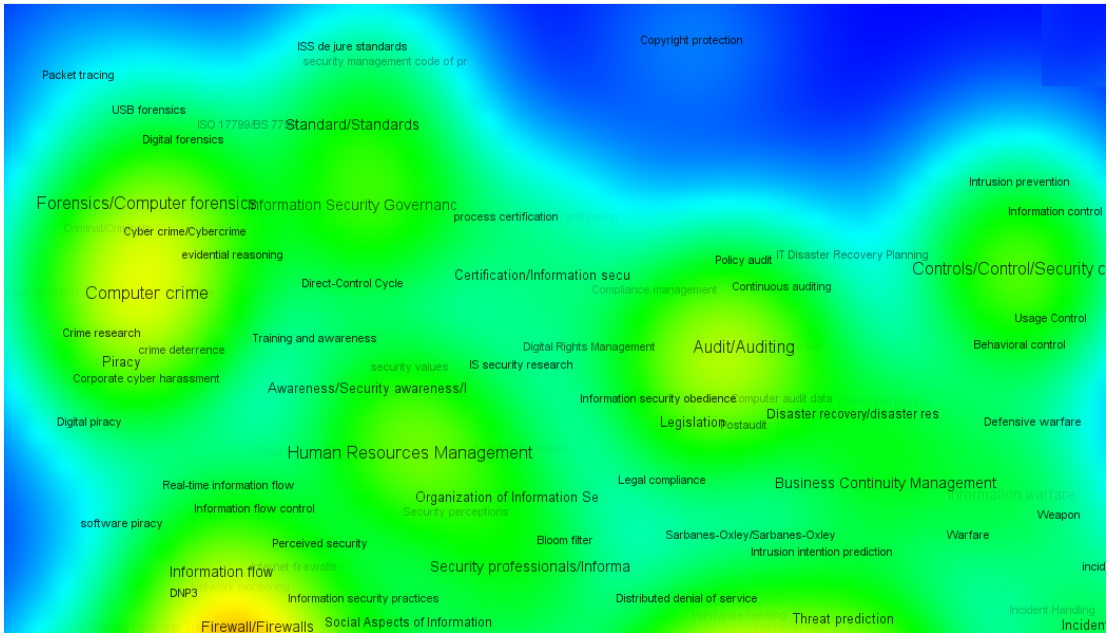
921	Web Service Security
922	Viruses
923	Public Key Infrastructure
924	Attack Response
925	Security principles
926	Passwords
927	Key management
928	Business information risk
929	user authentication
930	RFID
930	privacy
931	Auditing
932	Digital Policy Management
933	Security consistency model
934	Botnets
935	privacy
936	Intrusion detection
937	Disaster recovery
938	unpredictable threats
939	anti-spam measures
940	Information warfare
940	Systems security
941	hacker profile
942	Internal controls
943	Intrusion Detection Systems
944	Security risk
945	Incident Management process
946	spam e-mail
947	grid security
948	Information Security Governance
948	Protection
949	Intrusion detection systems
950	IT Disaster Recovery Planning
951	Availability
951	Risk analysis
952	Disaster recovery
953	Social Engineering
954	extreme event planning
955	Awareness
955	Security risks
956	Threat Modeling
957	Security Requirements
958	Privacy Risk
958	Residual RFID
959	Distributed firewalls

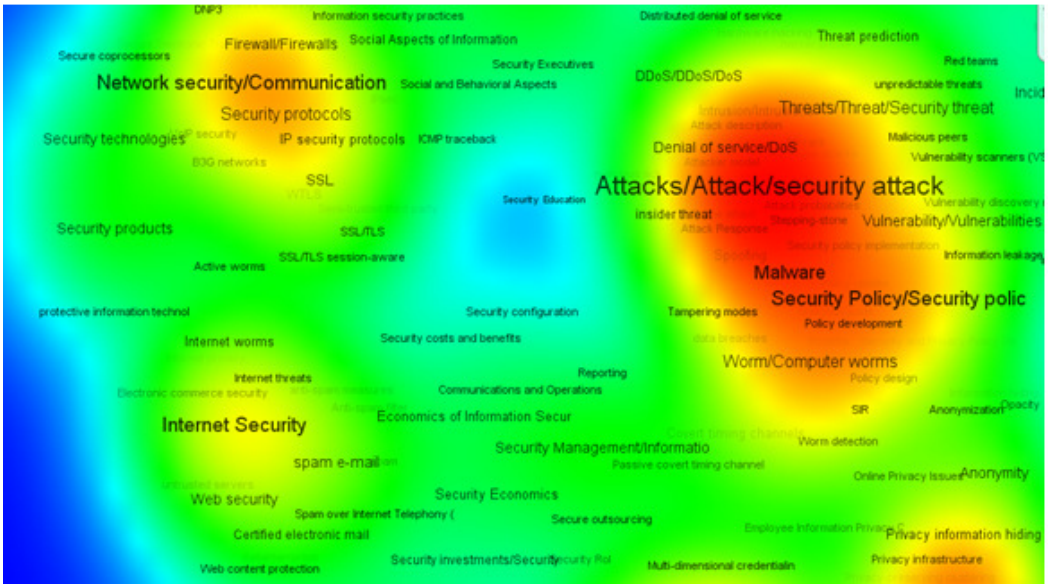
960	Security management
961	information systems misuse
961	Security countermeasures
962	Risk management
963	vulnerabilities
964	Security management
965	incident response planning
966	risk management standards
967	Secure multi-party computation
968	privacy
969	intrusion detection
970	PriS-Tool
971	SCADA protocol security
972	Audit
973	secure virtual machine
974	Information systems security management
974	Security awareness
975	systemic integrity
976	Security management
977	severity
978	Information flow
979	hacker profile
980	disaster response
981	ICMP traceback
982	privacy policies
983	Software piracy
984	Privacy
985	Sarbanes-Oxley Act
986	Privacy
986	Surveillance
987	Privacy
988	Security policies
989	RFID
989	privacy
990	Guidelines
990	Security Executives
991	IS security training
992	Security risk management
993	IS security policies
994	Information Security awareness
994	Compliance
994	Information Security Policy
995	virtuous IT
996	Information assurance
997	Economics of Information Security

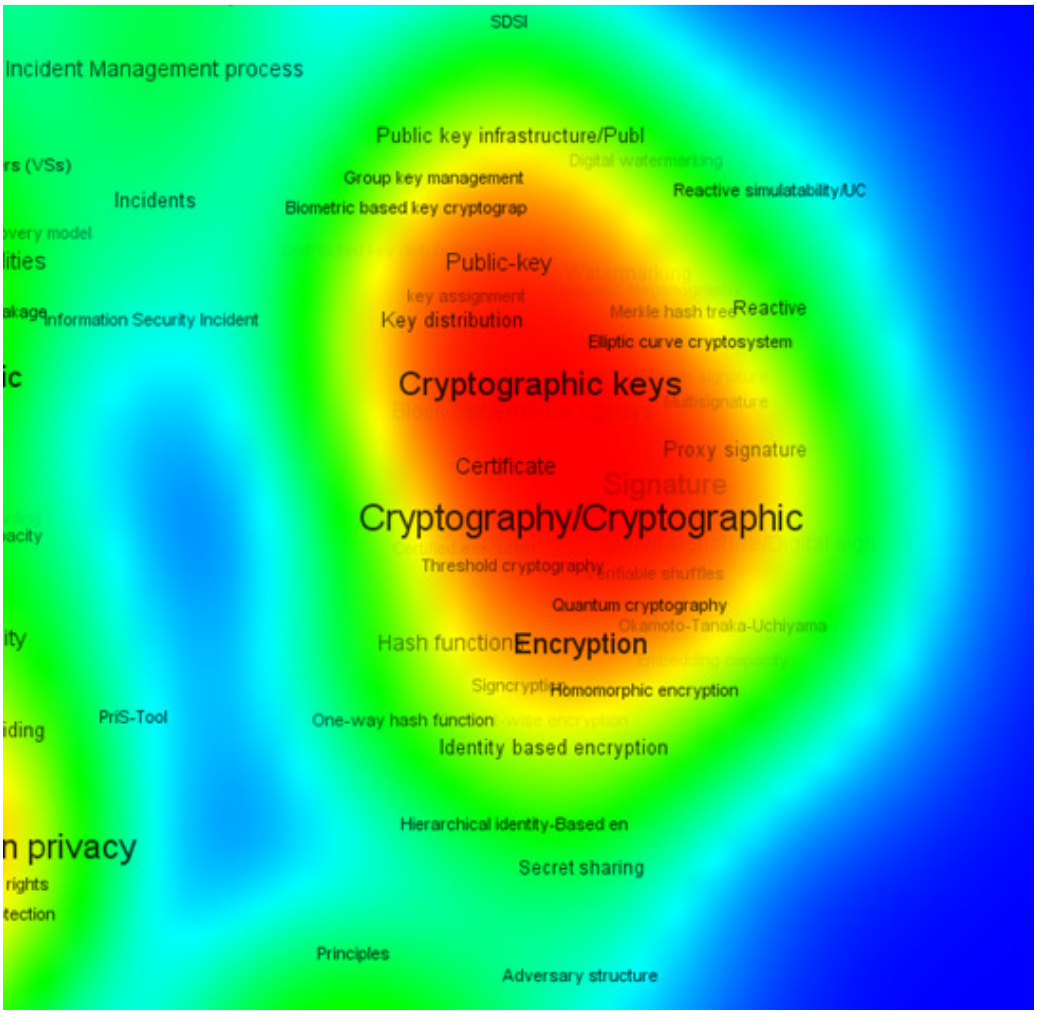
998	Risk management
999	ISS de jure standards
1000	security management code of practice

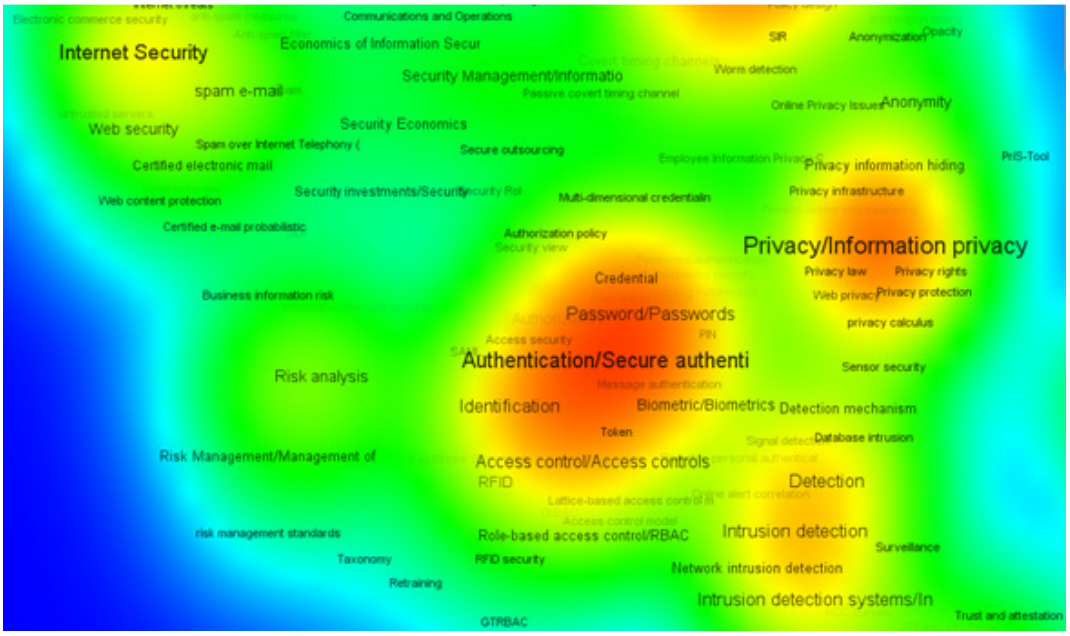
Apêndice H - Apresentação Visual do Esquema

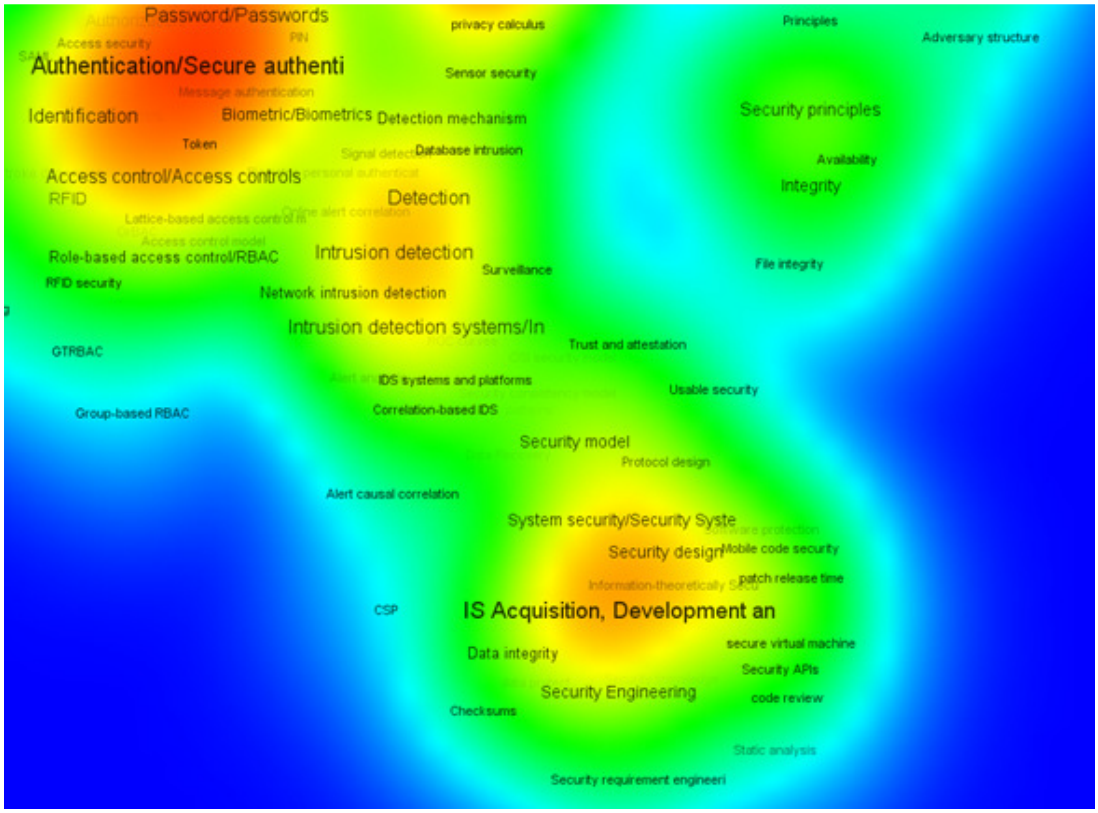
As figuras que se seguem representam o esquema de classificação para a literatura de SSI. É um esquema visual que ilustra claramente a dimensão que uma área de investigação de SSI poderá abranger. É de relembrar que este esquema apenas considera as keyword identificadas como tags nos papers. Dada a dimensão do esquema não foi possível representá-lo numa única figura e por isso está repartido por várias figuras.




Information security practices
Distributed denial of service
Firewall/Firewalls
Social Aspects of Information
Threat prediction
Secure coprocessors
Security Executives
Red teams
Network security/Communication
Social and Behavioral Aspects
DDoS/DDoS/DoS
unpredictable threats
Security protocols
Threats/Threat/Security threat
Security technologies
IP security protocols
ICMP traceback
Malicious peers
Denial of service/DoS
SSL
Attacks/Attack/security attack
Security Education
insider threat
Security products
SSL/TLS
Vulnerability/Vulnerabilities
SSL/TLS session-aware
Active worms
Malware
Security Policy/Security polic
protective information technol
Security configuration
Tampering modes
Policy development
Internet worms
Security costs and benefits
Worm/Computer worms
Internet threats
Reporting
Communications and Operations
Internet Security
Economics of Information Secur
SIR
Anonymization
Security Management/Informatio
Worm detection
spam e-mail
Online Privacy Issues
Anonymity
Web security
Security Economics
Spam over Internet Telephony (
Secure outsourcing
Certified electronic mail
Privacy information hiding
Web content protection
Security investments/Security
Multi-dimensional credentialin
Privacy infrastructure


SDSI
Incident Management process
Public key infrastructure/Publ
Digital watermarking
Group key management
Reactive simulatability/UC
Incidents
Biometric based key cryptograp
Public-key
key assignment
Information Security Incident
Key distribution
Reactive
Elliptic curve cryptosystem
Cryptographic keys
Proxy signature
Certificate
Signature
Cryptography/Cryptographic
Threshold cryptography
Quantum cryptography
Okamoto-Tanaka-Uchiyama
Hash function
Encryption
Homomorphic encryption
PriS-Tool
One-way hash function
Identity based encryption
Hierarchical identity-Based en
Secret sharing
Principles
Adversary structure


Internet Security
Economics of Information Secur
Communications and Operations
Security Management/Informatio
spam e-mail
Web security
Security Economics
Spam over Internet Telephony (
Certified electronic mail
Web content protection
Certified e-mail probabilistic
Security investments/Security
Secure outsourcing
Multi-dimensional credentialin
Authorization policy
Business information risk
Credential
Password/Passwords
Authentication/Secure authenti
Risk analysis
Identification
Biometric/Biometrics
Detection mechanism
Token
Database intrusion
Risk Management/Management of
Access control/Access controls
RFID
Detection
risk management standards
Role-based access control/RBAC
Intrusion detection
Surveillance
Taxonomy
RFID security
Network intrusion detection
Retraining
Intrusion detection systems/In
GTRBAC
Trust and attestation
Worm detection
Online Privacy Issues
Anonymity
Privacy information hiding
Privacy infrastructure
Privacy/Information privacy
Privacy law
Privacy rights
Privacy protection
privacy calculus
Sensor security
SIR
Anonymization
PriS-Tool


Password/Passwords
privacy calculus
Authentication/Secure authenti
Sensor security
Identification
Biometric/Biometrics
Detection mechanism
Token
Database intrusion
Access control/Access controls
RFID
Detection
Role-based access control/RBAC
Intrusion detection
Surveillance
RFID security
Network intrusion detection
Intrusion detection systems/In
GTRBAC
Trust and attestation
IDS systems and platforms
Usable security
Group-based RBAC
Correlation-based IDS
Security model
Protocol design
Alert causal correlation
System security/Security Syste
Security design
Mobile code security
patch release time
CSP
IS Acquisition, Development an
Data integrity
secure virtual machine
Security APIs
Security Engineering
code review
Checksums
Static analysis
Security requirement engineeri
Principles
Adversary structure
Security principles
Availability
Integrity
File integrity

Referências

(ISC)². (1996). (ISC)² Inspiring a Safe and Secure Cyber World Retrieved 7 Junho de 2012, from https://www.isc2.org/

(ISO), I. O. f. S. (2008). The ISO 27000 Directory Retrieved 1 Abril de 2013, from http://www.27000.org/iso-27002.htm

ACM. (1998). ACM Computing Classification System ToC Retrieved 12 Março de 2011

Alter, S. (1996). *Information Systems: A Management Perspective*: Benjamin/Cummings Publishing Company.

Amaral, L. A. M. d. (1994). *Praxis - Um Referencial para o Planeamento de Sistemas de Informação.* Tese de Doutoramento, Universidade do Minho, Guimarães.

Anderson, R. (1993). Why Cryptosystems Fail. *Communication of the ACM, 37*(11), 32-44.

Antman, M. E. M., Lau, M. J., Kupelnick, B., Mosteller, P. F., e Chalmers, M. T. C. (1992). A comparison of results of meta-analyses of randomized control trials and recommendations of clinical experts: Treatments for myocardial infarction. *Journal of the American Medical Association, 268*(2), 240-248. doi: 10.1001/jama.1992.03490020088036

Armstrong, H. (2000). *Managing information security in healthcare – an action research experience.* Paper presented at the Proceedings of the Sixteen Annual Working Conference on Information Security.

Artelsmair, C., Essmayr, W., Lang, P., Wagner, R., e Weippl, E. (2002). CoSMo: An Approach Towards Conceptual Security Modeling *Springer*, 557-566.

Barki, H., Rivard, S., e Talbot, J. (1988). An Information Systems Keyword Classification Scheme. *MIS Quarterly, 12*, 299-322.

Barki, H., Rivard, S., e Talbot, J. (1993). A keyword classification scheme for IS research literature: An update. *MIS Quarterly, 17*(2), 209-226.

Barnes, B. H. (1998). Computer security research: a British perspective. *IEEE Software, 15*(5), 30-33. doi: 10.1109/52.714619

Barros, A. J. P. d., e Lehfeld, N. A. d. S. (1986). *Fundamentos de Metodologia*. São Paulo: McGraw-Hill.

Baskerville, R. (1988). *Designing information systems security*.

Baskerville, R. (1989). Logical Controls Specification: an Approach to Information Systems Security. *Systems Development for Human Progress*, 241-255.

Baskerville, R. (1991). Risk analysis: an interpretive feasibility tool in justifying information systems security. *European Journal of Information Systems, 1*(2), 121-130. doi: 10.1057/ejis.1991.20

Baskerville, R. (1992). The Developmental Duality of Information Systems Security *Journal of Management Systems, 4*(1), 1-12.

Baskerville, R. (1993). Information systems security design methods: implications for information systems development. *ACM Computing Surveys (CSUR), 25*(4), 375-414.

Benbasat, I., e Zmud, R. W. (1999). Empirical research in information systems: the practice of relevance. *MIS Quarterly, 23*(1), 3-16. doi: 10.2307/249403

Boland, R. J., e Hirschheim, R. (1987). *Critical Issues in Information Systems Research* (First Ed. ed.). Universidade da Califórnia: John Wiley & Sons.

Booysen, H. A. S., e Eloff, J. H. P. (1995). *A Methodology for the development of secure Application Systems.* Paper presented at the Proceeding of the 11th IFIP TC11 International Conference on Information Security.

Brascher, M., e Café, L. (2008). *Organização da Informação ou Organização do Conhecimento?* Paper presented at the Encontro Nacional de Pesquisa em Ciência da Informação, S.Paulo.

Bretschneider, S., e Wittmer, D. (1993). Organizational Adoption of Microcomputer Technology: The Role of Sector *Journal: Information Systems Research, 4*(1), 88-108. doi: 10.1287/isre.4.1.88

Browne, P. S. (1979). *Security, checklist for computer center self-audits*. Washington, D.C

Buckingham, R. A., Hirschheim, R. A., Land, F. F., e Tully, C. J. (1987). *Information Systems Curriculum: a Basis for Course Design*. Cambridge University Press, 1987.

Burrell, G., e Morgan, G. (1979). *Sociological Paradigms and Organizational Analysis: Elements of the Sociology of Corporate Life* (First Edition ed.). London: Heinemann Education Books.

Campos, M. L. d. A., e Gomes, H. E. (2007). Taxonomia e Classificação: a categorização como princípio *VIII ENANCIB – Encontro Nacional de Pesquisa em Ciência da Informação*.

Carvalho, J. Á. (2000). *Information System? Which One Do You Mean?* Paper presented at the Information Systems Concepts: An Integrated Discipline Emerging, Proceedings of the ISCO 4 Conference, Leiden, Holanda.

Carvalho, J. Á. (2001). *Tecnologias da Informação nas Organizações.* Universidade do Minho, Guimarães.

Carvalho, J. Á. (2012). *Informação, Sebenta Fundamentos dos Sistemas de Informação.* Universidade do Minho, Guimarães.

Carvalho, J. Á., e Amaral, L. (1993). Matriz de Actividades: Um enquadramento Conceptual para as Actividades de Planeamento e Desenvolvimento de Sistemas de Informação. *Sistemas de Informação, 1*, 37-48.

Chaula, J. A., Yngström, L., e Kowalski, S. (2005). *A Framework for Evaluation of Information Systems Security*. Paper presented at the Peer-reviewed Proceedings of the ISSA 2005 New Knowledge Today Conference, South Africa.

Chua, W. F. (1986). Radical Developments in Accounting Thought. 601-632.

Claver, E., Gonzalez, R., e Llopis, J. (1999). An analysis of research in information systems (1981-1997).

Coulter, N., French, J., Glinert, E., Horton, T., Mead, N., Rada, R., . . . Wierzbicki, C. (1998). Computing Classification System 1998: Current Status and Future Maintenance Report of the CCS Update Committee. In C. Reviews (Ed.).

Cruz, A. J. R. d. (2007). *Data Mining via Redes Neuronais Artificiais e Máquinas de Vetores de Suporte.* Universidade do Minho, Braga.

Dahlberg, I. (2006). Teoria do conceito. Ciência da Informação. *Knowledge Organization, 7*(2), 101 - 107.

Damas, L. (2005). *SQL - Structured Query Language* (6ª Edição Actualizada e Aumentada ed.).

Denzin, N. K., e Lincoln, Y. S. (2005). *The Sage handbook of qualitative research*: Sage Publications, Inc.

Dhillon, G. (1997). *Managing Information System Security*. London: MacMillan.

Dhillon, G., e Backhouse, J. (1994). *Responsibility analysis: a basis for understanding complex managerial situation*. Paper presented at the International System Dynamics Conference University of Stirling, Scotland.

Dhillon, G., e Backhouse, J. (1996a). Risks in the Use of Information Technology Within Organizations *International Journal of Information Management, 16*(1), 65-74.

Dhillon, G., e Backhouse, J. (1996b). Structures of responsibility and security of information systems. *European Journal of Information Systems, 5*(1), 2-9. doi: 10.1057/ejis.1996.7

Dhillon, G., e Backhouse, J. (2000a). Information System Security Management in the New Millennium. *Communication of the ACM, 43*(7), 125-128.

Dhillon, G., e Backhouse, J. (2000b). Technical opinion: Information system security management in the new millennium. *Communications of the ACM, 43*(7), 125-128. doi: 10.1145/341852.341877

Dhillon, G., e Backhouse, J. (2001). Current directions in IS security research: towards socio-organizational perspectives. *Information Systems Journal (2001), 11*, 127-153.

Dilly, R. (1999). Data Mining: an introduction. Queens University of Belfast: Belfast: Parallel Computer Centre.

Ein-Dor, P., e Segev, E. (1993). A Classification of Infonnation Systems: Analysis and Interpretation. *Information Systems Research, 4*(2), 166-204.

Elberrichi, Z., e Aljohar. (2007). "N-grams in Texts Categorization.". *Scientific Journal of King Faisal University (Basic and Applied Sciences), 8*(2).

Ellmer, E., Pernul, G., e Kappel, G. (1995). *Object-Oriented Modeling of Security Semantics.* Paper presented at the Proceedings of the 11th Annual Computer Security and Applications Conference,, New Orleans, Louisiana, USA.

Falkenberg, E. D., Hesse, W., Lindgreen, P., Nilsson, B. E., Oei, J. L. H., Rolland, C., . . . Voss, K. (1998). Frisco: A Framework of Information Systems Concepts. In I. F. f. I. Processing (Ed.).

Fan, W., e Wallace, L. (2005). Tapping into the Power of Text Mining, Communications of ACM. Department of Accounting and Information Systems - Virginia Polytechnic Institute and State University, et al.:

.

Feldman, R., e Sanger, J. (2007a). *The Text Mining Handbook*. New York: United States of America.

Feldman, R., e Sanger, J. (2007b). *The Text Mining Handbook: Advanced Approaches in Analyzing Unstructured Data*. Cambridge: Cambridge University Press

Fernández-Medina, E., e Piattini, M. (2003). Designing Secure Databases for OLS. *Springer*, 886-895.

Fernández-Medina, E., Trujillo, J., Villarroel, R., e Piattini, M. (2004). Extending UML for Designing Secure Data Warehouses. *Springer*, 217-230.

Fernandez, E. B. (2004). A methodology for secure software design. *Springer*.

Frisinger, A. (2001). *Improving the protection of assets in open distributed systems by use of X-ifying risk analysis.* Paper presented at the Proceedings of the IFIP TC11 Sixteenth International Conference on Information Security, Paris, France.

Galliers, R. D. (1992a). Choosing information systems research approach. *Information Systems Research*, 144-162.

Galliers, R. D. (1992b). *Information Systems Research: Issues, Methods and Practical Guidelines* (First Edition ed.): Alfred Waller, 1994.

Gehrke, M., Pfitzmann, A., e Rannenberg, K. (1992). *Information Technology Security Evaluation Criteria (ITSEC) – a Contribution to Vulnerability?* Paper presented at the PROCESSING of the IFIP 12th World Computer Congress, Madrid, Spain,.

Georg, G., Ray, I., e France, R. (2001). Using Aspects to Design a Secure System. *IEEE*.

Gouveia, L. B., e Ranito, J. (1997). Sistemas de Informação de Apoio Á Gestão *Inovação e Governação nas Autarquias*: Sociedade Portuguesa de Inovação.

Guarro, S. B. (1987). Principles and procedures of the LRAM approach to information systems risk analysis and management. *Computer and Security, 6*(6), 493-504.

Habermas, J. (1984). *The Theory of Communicative Action - Reason and the Rationalisation of Society* (Vol. 1). Boston, MA, USA.

Habermas, J. (1987). *The Theory of Communicative Action–The Critique of Functionalist Reason* (Vol. 2). Boston, MA, USA.

Halliday, S., Badenhorst, K., e Solms, R. v. (1996). A business approach to effective information technology risk analysis and management. *Information Management & Computer Security, 4*(1), 19-31.

Han, J., Kamber, M., e Pei, J. (2011). *Data Mining-Concepts and Techniques* (Third Edition ed.): Morgan Kaufmann.

Harris, S. (2012). *CISSP All-in-One Exam Guide* (6th Ed. ed.): McGraw-Hill Osborne Media.

Hearst, M. A. (1999, June 20-26). *Untangling text data mining.* Paper presented at the In Proceedings of ACL'99: the 37th Annual Meeting of the Association for Computational Linguistics, University of Maryland.

Herath, T., e Rao, H. R. (2009). Encouraging information security behaviors in organizations: Role of penalties, pressures and perceived effectiveness. *Decision Support Systems, 47*, 154-165. doi: 10.1016

Herrmann, G., e Pernul, G. (1999). Viewing business-process security from different perspectives *Journal International Journal of Electronic Commerce, 3*(3), 89-103.

Hevner, A. R., March, S. T., Park, J., e Ram, S. (2004). Design Science in Information Systems Research *MIS Quarterly, 28*(1), 75-105.

Hitchings, J. (1995). *Achieving an integrated design: the way forward for information security.* Paper presented at the Proceedings of the IFIP TC11 11th international conference on information security.

Hjerland, B. (2003). Fundamentals of knowledge organization. *Knowledge Organization, 30,* 87-111.

Hodge, G. (2000). *Systems of Knowledge Organization for Digital Libraries: Beyond Traditional Authority Files*: 2000.

Hutchinson, W., e Warren, M. (2000). *Using the viable systems model to develop an understanding information system security threats to an organisation.* Paper presented at the Proceedings of the 1st Australian Information Security Management Workshop.

Iivari, J., e Hirschheim, R. (1996). Analyzing Information Systems Development: A Comparison and analysis of eight is development. *Information Systems, 21*(7), 551-575.

Iivari, J., e Kerola, P. (1983). A Sociocybernetic framework for the feature analysis of information systems design methodologies. *Information Systems Design Methodologies*, 87-139.

ISACA. (2006). ISACA - Trust, and value from, information system Retrieved 23 Abril 2013, from http://www.isaca-lisbon.org/cism.html

ISO/IEC17799, B. (2000). Code of Practice for Information Security Management. Department of Trade and Industry.

Jacobs, P. S. (1992). Joining Statistic with NLP for text Categorization. *ACM Computing Surveys (CSUR)*, 178-185.

James, H. L. (1996). *Managing information systems security: a soft approach.* Paper presented at the Proceedings of the Information Systems Conference of New Zealand.

Jarvinen, P. H. (2000). *Research Questions Guiding Selection of an Appropriate Research Method.* Paper presented at the Proceedings of ECIS2000, Wien: Vienna University of Economics and Business Adminstration.

Jean, W., e T., W. R. (2002). Analyzing the Past to Prepare for the Future: Writing a Literature Review. *MIS Quarterly, 26*(2), XIII-XXIII.

Jordan, E. (1993). Executive information systems for the chief information officer. *International Journal of Information Management, 13*(4), 249-259.

Jurjens, J. (2002). UMLsec-Extending UML for Secure Systems Development. *Springer*, 412-425.

Karyda, M., Kokolakis, S., e Kiountouzis, E. (2001). *Redefining Information Systems Security: Viable Information Systems.* Paper presented at the In Proceedings of the IFIP TC11 16th International Conference on Information Security Paris, France.

Khwaja, A. A., e Urban, J. E. (2002a). A synthesis of evaluation criteria for software specifications and specification techniques. *International Journal of Software Engineering and Knowledge Engineering, 12*(5), 581-599. doi: 10.1142/s0218194002001062

Khwaja, A. A., e Urban, J. E. (2002b). A synthesis of evaluation criteria for software specifications and specification techniques. *International Journal of Software Engineering and Knowledge Engineering, 12*(5), 581-599.

Kim, D., e Solomon, M. G. (2010). *Fundamentals of Information Systems Security*: Jones and Bartlett Publishers, Inc.

Kothari, C. (2004). *Research methodology: methods and techniques* (2. Ed. ed.): New Age International.

Krauss, L. (1972). *SAFE: Security Audit and Field Evaluation for Computer Facilities and Information Systems*. New York, United States.

Liebenau, J., e Backhouse, J. (1990). *Understanding information: an introduction*. London, UK: Macmillan Publishers Limited.

Linoff, G. S., e Berry, M. J. A. (1997). *Data Mining Techiques. For Marketing, Sales, and Customer Support* (Vol. Third Edition): John Wiley & Sons, Inc.

Liu, L., Yu, E., e Mylopoulos, J. (2003). Security and privacy requirements analysis within a social setting. *Journal of Lightwave Technology, 3*, 151-161. doi: 10.1109/ICRE.2003.1232746

Livari, J. (1992). The organizational fit of information systems. *Information Systems Journal, 2*(1), 3-29. doi: 10.1111/j.1365-2575.1992.tb00064.x

Ives, B., Hamilton, S., e G, B., Davis (1980). A Framework For Research in Computer-Based Management Information Systems. *Management Science, 26*(9), 910-934.

Magalhães, R. (1997). Sistemas de Informação: Definição, Origens e Perspectivas para Portugal *Sistemas de Informação* (Vol. 6, pp. 53-56).

March, S. T., e Smith, G. F. (1995). Design and natural science research on information technology. *Decision Support Systems, 15*(4), 251-266.

Marks, D. G., Sell, P. J., e Thuraisinghm, B. M. (1996). MOMT: A multilevel object modeling technique for designing secure database applications.

McCumber, J. (1991). *Information systems security: A comprehensive model.* Paper presented at the Proceedings of the 14th National Computer Security Conference.

McDermott, J., e Fox, C. (1999). *Using Abuse Case Models for Security Requirements.* Paper presented at the Proceedings of the 15th Annual Computer Security Applications Conference (ACSAC).

McLean, E. R. (1982). Information Systems and its Underlying Disciplines: A Summary of the Papers. *DATA BASE Fall, 14*(1), 3-6. doi: 10.1145/1017702.1017704

Mentzas, G. (1994). A Functional Taxonomy of Computer Based Information Systems. *International Journal of Information Management, 14*(6), 397-410.

Monteiro, L. d. O., Gomes, I. R., e Oliveira, T. (2006). *Etapas do Processo de Mineração de Textos – uma abordagem aplicada a textos em Português do Brasil.* Paper presented at the Anais do XXVI Congresso da SBC - I Workshop de Computação e Aplicações, Campo Grande.

Morgan, G. (2006). *Images of Organization* (Updated ed.): Sage Publications, Inc.

Moulton, R. T., e Moulton, M. E. (1996). Electronic Communications Risk Management: A Checklist for Business Managers. *Computer and Security, 15*(5), 377-386.

Murine, G., e Carpenter, J. C. (1984). *Measuring computer system security using software security metrics* Paper presented at the Proceedings of the 2nd IFIP international conference on Computer security: a global challenge, Toronto, Ontario, Canada

Myers, M. D. (1997a). Qualitative Research in Information Systems *ACM Computing Surveys (CSUR)*. MISQ Discovery.

Myers, M. D. (1997b). Qualitative Research in Information Systems. [Article]. *MIS Quarterly, 21*(2), 241-242.

O'Brien, J. A. (2000). *Introduction to Information Systems: Essentials for the Internetworked Enterprise*: Irwin/McGraw-Hill.

O´Brien, J. A., e Marakas, G. M. (2009). *Introduction to Information Systems* (15th Edition ed.): McGraw-Hill/irwin.

Parker, D. (1998). *Fighting computer crime: a new framework for protecting information*. New York, USA: John Wiley & Sons.

Passarin, D. (2005). Text Mining no Aperfeiçoamento de Consultas e Definição de Contextos de uma Central de Notícias Baseada em RSS. Palmas: Centro Universitário Luterano de Palmas.

Peixoto, D., Batista, G., e Capela, M. J. (2003-2004). *Text Categorization - Categorização de Textos - Processos, Métodos e Aplicações*. Ciências Documentais.

Pernul, G. (1992). Security Constraint Processing During Multilevel Secure Database Design. *IEEE*, 75-84.

Pernul, G., e Quirchmayr, G. (1994). Organizing MLS Databases from a Data Modelling Point of View *IEEE*, 84-105.

Pernul, G., Tjoa, A. M., e Winiwarter, W. (1998). Modelling data secrecy and integrity. *Data & Knowledge Engineering, 26*, 291-308.

Poore, R. S. (1999). Generally Accepted System Security Principles *Information Systems Security, 8*(3), 27.

Priebe, T., e Pernul, G. (2001). A Pragmatic Approach to Conceptual Modeling of OLAP Security. *Springer*, 311-324.

Reddy, G. S., Srinivasu, R., Rikkula, S. R., e Rao, V. S. (2009). Management information system to help managers for providing decision making in an organization. *International Journal of Reviews in Computing*, 1-6.

Röhm, A. W., e Pernul, G. (2000). COPS: A Model and Infrastructure for Secure and Fair Electronic Markets. *Decision Support Systems, 29*(4), 434-455.

Rue, L. W., e Holland, P. G. (1986). *Strategic management : concepts and experiences*.

Sá-Soares, J. F. d. (2005). *Interpretação da Segurança de Sistemas de Informação Segundo a Teoria da Acção.* Universidade do Minho, Guimarães.

Saltmarsh, T., e Browne, P. (1983). Data processing – risk assessment. In J. W. a. S. Ltd (Ed.), (Vol. 2, pp. 93-116). New York: Advances in Computer Security Management.

Sanders, P., Furrell, S., e Warren, M. (1996). Baseline Security Guidelines for Health Care Management. *In the SEISMED Consortium (eds), Data Security for Health Care, 31*(Management Guidelines, Baseline Security Guidelines for Health Care Management), 82-107.

Schein, E. H. (1972). *Organizational psychology* (2 ed.). Universidade de Michigan: Prentice-Hall.

Sebastiani, F. (2002). Machine learning in automated text categorization. *ACM Computing Surveys, 34*(1), 1-47.

Simon, H. A. (1996). *The Sciences of the Artificial* (3 Edition ed.): MIT Press.

Siponen, M. (2002). *Designing secure information systems and software.* University of Oulu, Oulu.

Siponen, M. (2005a). Analysis of modern IS security development approaches: towards the next generation of social and adaptable ISS methods. *Information and Organization, 15*, 339-375.

Siponen, M. (2005b). An analysis of the traditional IS security approaches: implications for research and practice. *European Journal of Information Systems, 14*(3), 303-315. doi: 10.1057/palgrave.ejis.3000537

Siponen, M., e Oinas-kukkonen, H. (2007). A Review of Information Security Issues and Respective Contributions. *Data Base For Advances In Information Systems, 38*(1), 60-80.

SSE-CMM. (2003). Systems security engineering capability maturity model. *Version 3.0*.

Stacey, T. R. (1996). The information security program maturity grid. *Information Systems Security, 5*(2), 22-33.

Straub, D. W., e Welke, R. J. (1998). Coping with systems risk: security planning models for management decision making. *MIS Quarterly, 22*(4), 441-464.

Strens, R., e Dobson, J. (1993). How Responsibility Modelling Leads to Security Requirements *ACM*, 143-149.

Tan, A.-H. (1999). *Text Mining: The state of the art and the challenges.* Paper presented at the In Proceedings of the PAKDD 1999 Workshop on Knowledge Disocovery from Advanced Databases Singapore.

Thomas, R. K., e Sandhu, R. S. (1994). Conceptual Foundations for a Model of Task-based Authorizations. *IEEE*, 66-79.

Tipton, H. F. (2009). *Official (ISC) 2 guide to the CISSP CBK*: Auerbach Publications.

Varajão, J. E. Q. (1998). *A Arquitectura da Gestão de Sistemas de Informação*: FCA.

Villarroel, R., Fernández-Medina, E., e Piattini, M. (2005). Secure information systems development – a survey and comparison. *Computers & Security, 24*, 308-321.

Vivas, J. L., Montenegro, J. A., e López, J. (Cartographer). (2003). Towards a Business Process-Driven Framework for Security Engineering with the UML.

Wadsworth, B. J., e Rovai, E. (1992). *Inteligência e Afetividade da Criança na teoria de Piaget*. São Paulo.

Waegel, D. (2006). The Development of Text-Mining Tools and Algorithms. CiteSeer.

Webster, J., e Watson, R. T. (2002). Analyzing the Past to Prepare for the Future: Writing a Literature Review. *Management Information Systems Quarterly, 26*(2), XIII-XXIII.

Westin, S., Roy, M., e Kim, C. K. (1994). Cross-fertilization of knowledge: The case of MIS and its reference disciplines. *Information Resources Management Journal, 7*(2), 24-34.

Whitman, M. E., e Mattord, H. J. (2011). *Principles of Information security* (Four Edition ed.): Course Technology.

Winograd, T. (1996). *Bringing design to software* (1 Edition ed.): ACM Press.

Wood, C. C., W.Banks, W., B.Guarro, S., Garcia, A. A., Hampel, V. E., e Sartorio, H. P. (1987). *Computer security: a comprehensive controls checklist*. New York, United State.

Woods, E. (2004). The corporate taxonomy: creating a new order. *KMWorld, 13*(7).

Yang, H.-C., e Lee, C.-H. (2005). A text mining approach for automatic construction of hypertexts. *Expert Systems with Applications, 29*(4), 723-734.

Zorrinho, C. (1991). *Gestão da Informação*: Editorial Presença.

Zorrinho, C. (1995). *Gestão da Informação: Condição para Vencer*: IAPMEI.