



The use of **Taxonomies** as a way to achieve **Interoperability** and improved **Resource Discovery** in DSpace-based Repositories

Miguel Ferreira
mferreira@dsi.uminho.pt

Ana Alice Baptista
analice@dsi.uminho.pt





Conteúdos

- Repositórios Institucionais @ UMinho
- Contextualização do problema
 - Porquê metainformação?
 - Porquê vocabulários controlados?
- Tipos de vocabulários controlados
 - Taxionomias, Thesauri e Ontologias
- Add-on de vocabulários controlados
- Conclusões
- Trabalho futuro



Repositórios Institucionais

- O que são?
 - **Sistemas** de informação que servem para **armazenar, preservar e disseminar** a produção intelectual de uma organização
- RepositoriUM
 - Repositório Institucional da UM
 - <http://repositorium.sdum.uminho.pt>
 - Artigos, teses, dissertações, apresentações, literatura cinzenta, etc
- Papadocs
 - Repositório dos alunos
 - <http://papadocs.dsi.uminho.pt>
 - **Trabalhos** práticos, relatórios, material didáctico, etc.
 - Lic. Informática Gestão, Mestrado em Sistemas de Informação, Eng. Civil, Arquitectura, outras?
 - **Plataforma de testes** para os add-ons desenvolvidos
- DSpace
 - Desenvolvido em parceria pelo **MIT** e a **HP**
 - Utilizado em todo o **mundo**
 - Suporta todo o tipo de **formatos** digitais (e.g. PDF, Áudio, Vídeo, Imagem, Texto, etc.)
 - **Open-source**, Java, PostgreSQL, Tomcat



Contributo da Universidade do Minho

- Tradução do DSpace para língua **portuguesa**
 - Serviços de Documentação da U. Minho
 - Utilizado em várias universidades em Portugal e Brasil
 - Universidade Fernando Pessoa
 - Universidade de São Paulo (USP)
 - Universidade Estadual de São Paulo
 - Tribunal Justiça do Brasil
 - Universidade Federal de minhas Gerais
 - Universidade Federal do Paraná
 - ...
- Desenvolvimento de **extensões** à plataforma sob a forma de **add-ons**
 - Comentários
 - Recomendação
 - Teia de comunicação
 - **Vocabulários controlados**



Contributo da Universidade do Minho

[Papadocs >](#)
[Mestrado em Sistemas de Informação >](#)
[Práticas de Tecnologias de Informação II >](#)

Título: Detecção de Intrusões / Data Mining

Comentários:

- [-] **exemplo de comentário (2)**
 - [-] **comentar comentários em grafo (1)**
 - [+] **comentário ao comentário agregador**
[Responder] [Marcar] [Teia de Comunicação]
Comentador: Ana A. Baptista
Data: 23-Jun-2004 09:43:18
Exemplo de comentário ao comentário agregador
- [-] **exemplo 1 (4)**
 - [+] **comentário ao comentário**
[Responder] [Marcar] [Teia de Comunicação]
Comentador: Ana A. Baptista
Data: 25-Mai-2004 15:49:30
exemplo 2
 - [+] **comentar comentários em grafo (1)**
 - [+] **teste**

5 pessoas relacionadas

- [Paulo Cortez](#)
- [Teresa Nunes](#)
- [Anabela Moreira Bernardino](#)
- [...](#)
- [...](#)

Relacionados

- [...](#)
- [...](#)
- [...](#)
- [...](#)
- [...](#)
- [...](#)
- [...](#)
- [...](#)
- [...](#)
- [...](#)

Apare

Fic

Rela

Análise Preliminar.zip

Anexo ao Relatório

1239Kb

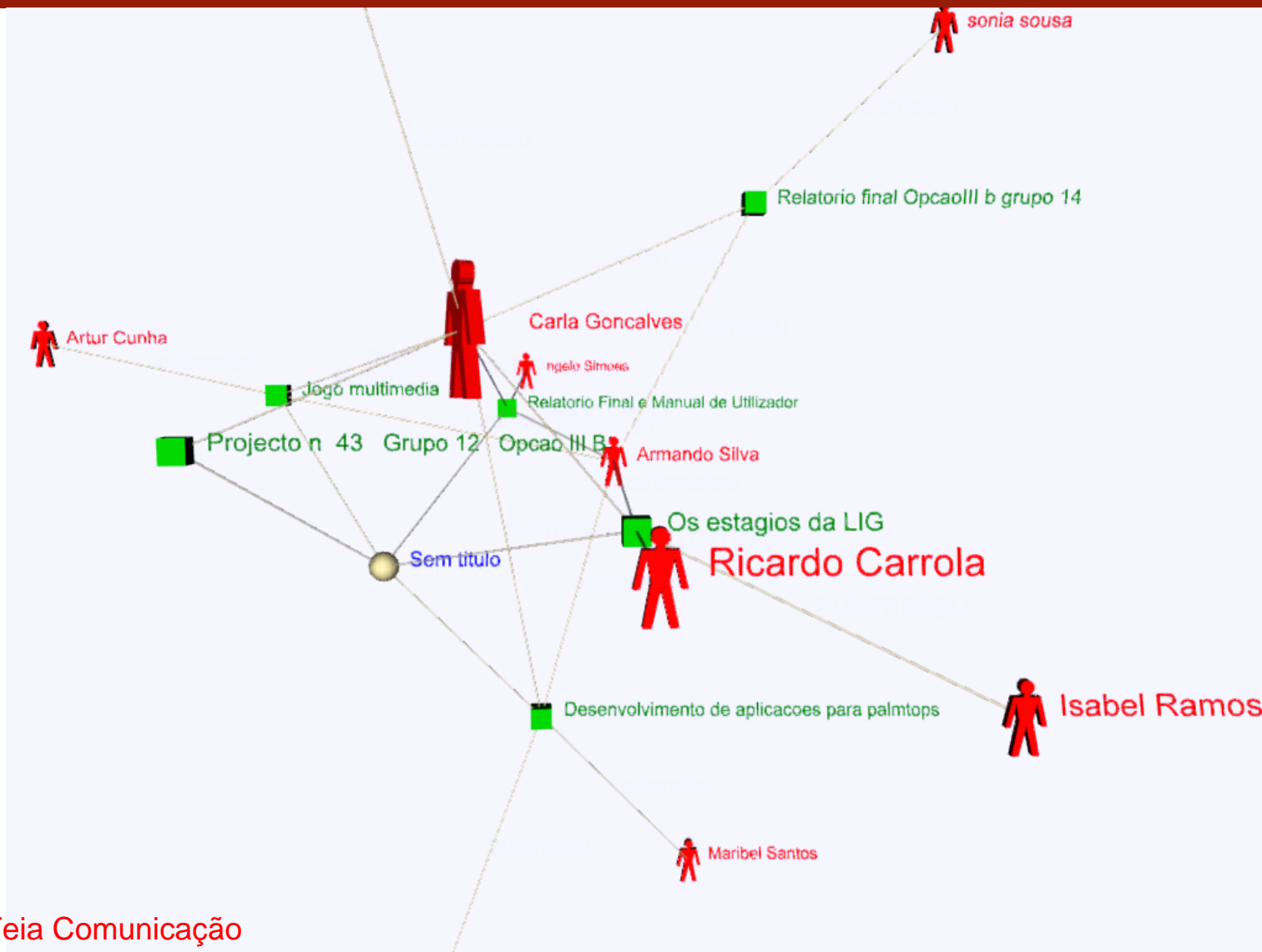
Add-on Comentários

[Mostrar registo em formato completo](#)

Add-on Recomendação



Contributo da Universidade do Minho





Contextualização do problema

- Porquê metainformação?
 - **Descrever** o objecto custodiado
 - Facilitar ao **acesso** aos objectos preservados
 - e.g. título, descrição, autores, **palavras-chave**
- Como obter metainformação?
 - Automaticamente
 - Pouco preciso
 - Manualmente
 - Catalogadores **profissionais** (arquivistas, bibliotecários, ...)
 - Catalogadores **amadores** (auto-arquivo)
- Problemas da catalogação manual
 - Demasiado **onerosa**
 - **Baixa concordância** entre palavras-chave atribuídas por catalogadores distintos
 - Problemas acrescidos no **acesso** à informação



Vocabulários Controlados



fish



reptiles



birds



mammals



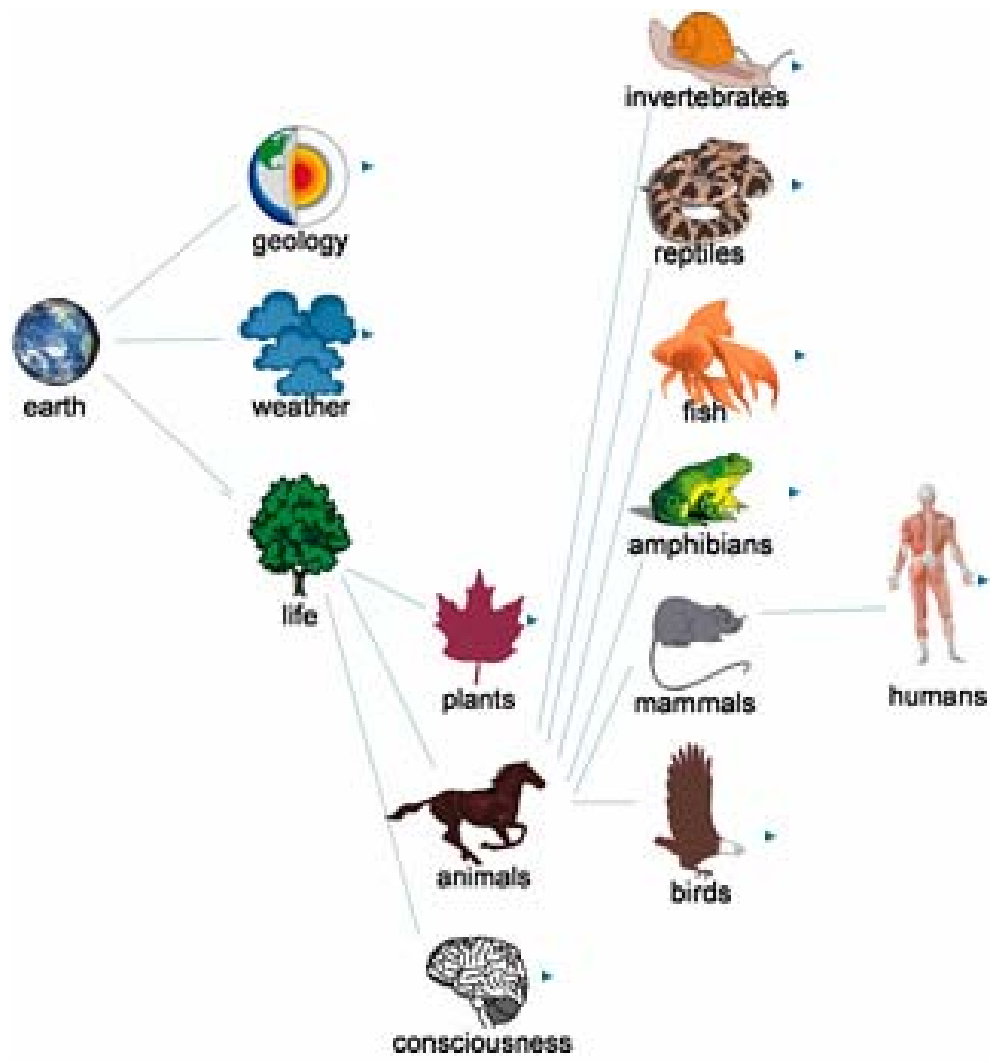
invertebrates



humans

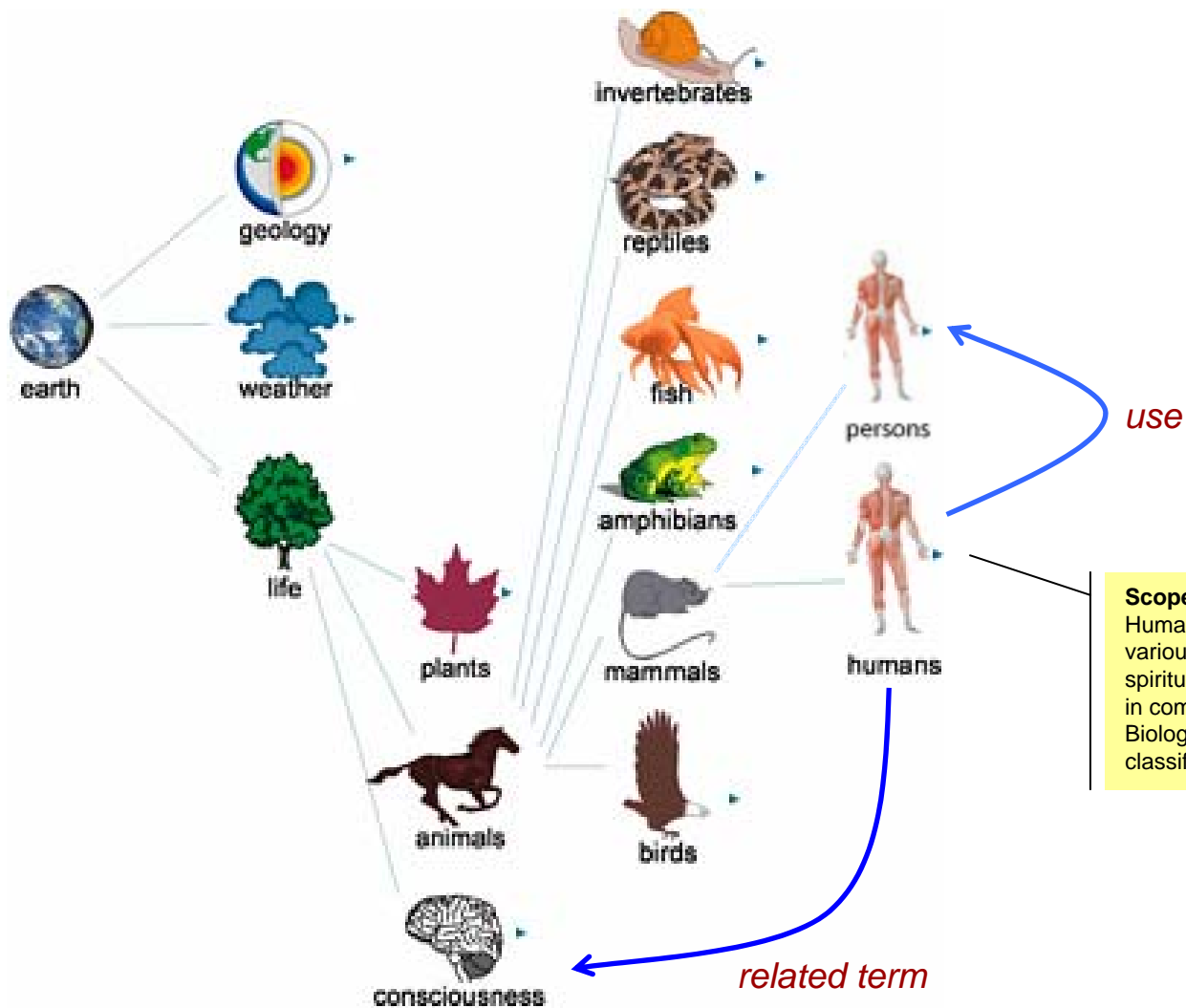


Taxionomias





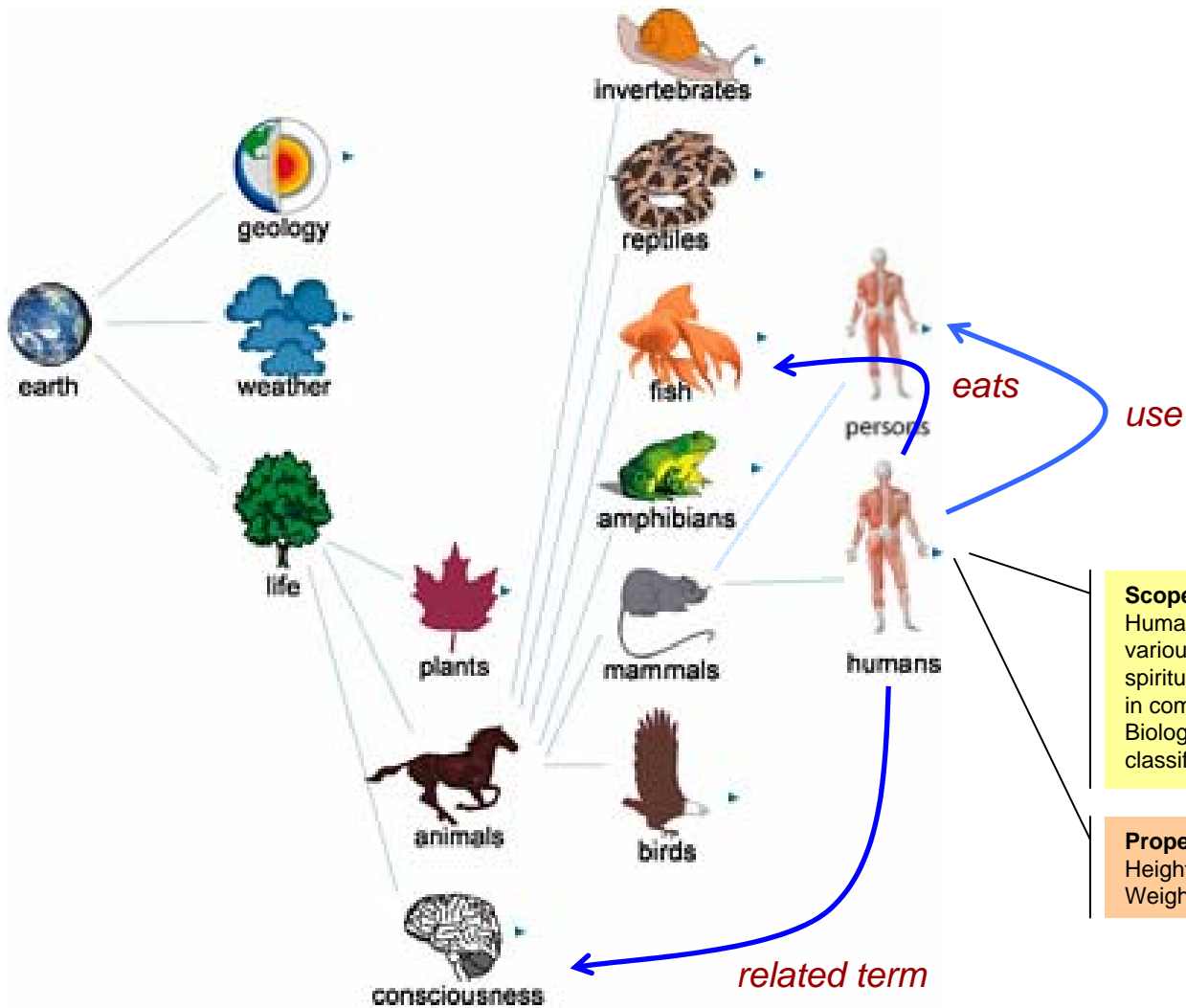
Thesaurus



Scope note
Human beings are defined variously in biological, spiritual, and cultural terms, or in combinations thereof. Biologically, they are classified as Homo sapiens.



Ontologias



Scope note
Human beings are defined variously in biological, spiritual, and cultural terms, or in combinations thereof. Biologically, they are classified as Homo sapiens.

Properties:
Height
Weight

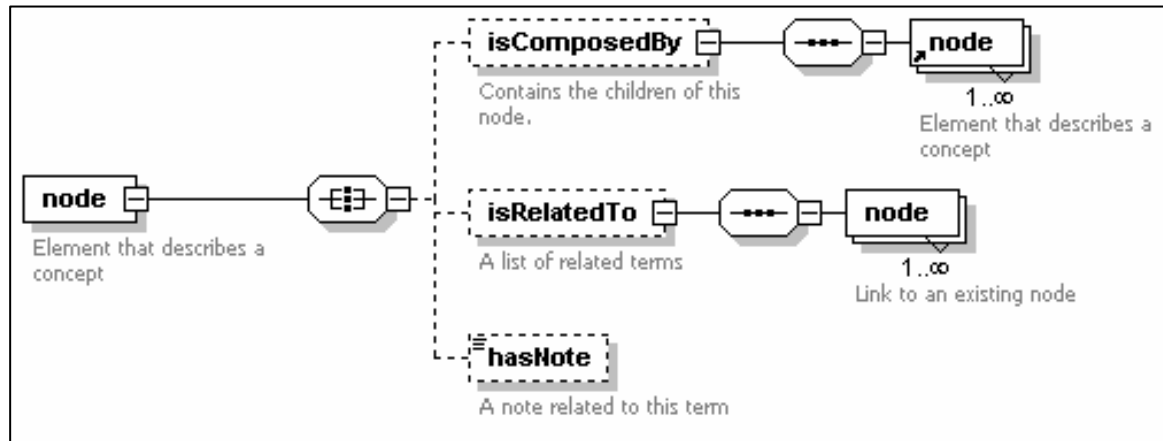


Add-on de Vocabulários Controlados

- Vocabulário controlado baseado numa **taxionomia**
- Reutilização de taxionomias já existentes
 - **ACM** Computing Classification System
 - **Engineering Index** Thesaurus
- Taxionomias descritas em **XML**
- Conversão para **HTML** recorrendo a **XSLT**

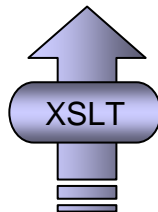


Add-on de Vocabulários Controlados



- + ACMCCS98
 - + General Literature
 - + Hardware
 - Computer Systems Organization
 - + GENERAL
 - PROCESSOR ARCHITECTURES
 - General
 - + Single Data Stream Architectures
 - + Multiple Data Stream Architectures (Multiprocessors)
 - + Other Architecture Styles
 - + Parallel Architectures
 - + Miscellaneous

HTML



<XML>



Catálogo e recuperação de informação

- e.g. Uma sebenta sobre programação em Java
 - ACMCCS98/Software/PROGRAMMING TECHNIQUES/Object-oriented Programming
 - ACMCCS98/General Literature

Seleccione os conceitos a pesquisar

Filtro:

- ACMCCS98
 - General Literature
 - GENERAL
 - INTRODUCTORY AND SURVEY
 - REFERENCE (e.g., dictionaries, encyclopedias, glossaries)
 - MISCELLANEOUS
 - Hardware
 - Computer Systems Organization
 - Software
 - Data
 - Theory of Computation
 - Mathematics of Computing
 - Information Systems
 - Computing Methodologies
 - Computer Applications
 - Computing Milieux



Conclusões

- **Simple**s de utilizar e instalar
- Simplifica a **descrição** e a **pesquisa** de informação no repositório
- Permite **generalização** de conceitos
- **Schema simple**s ao contrário da maioria das normas
- Versão oficial do **ACM CCS** em XML disponível na página da ACM



Trabalho futuro

- **Migração de tecnologia** do Papadocs para o RepositoriUM
- Integração de add-ons no **código oficial do DSpace**
- **Modelos** de representação mais **complexos**
 - Thesauri, ontologias
 - Descrição do ACMCCS em OWL
- **Classificação automática** de documentos ou apenas sugestão de termos de classificação
- Possível necessidade de efectuar a "**junção**" de taxionomias cujos conceitos se sobrepõem



Questões?

<http://dspace-dev.dsi.uminho.pt>

Miguel Ferreira
mferreira@dsi.uminho.pt

Ana Alice Baptista
analice@dsi.uminho.pt