

## OS MOTORES DE BUSCA NUMA PERSPECTIVA COGNITIVA

---

Tiago Tavares

Instituto de Educação e Psicologia da Universidade do Minho

mrtavares@gmail.com

Vitor Diegues

Instituto de Educação e Psicologia da Universidade do Minho

v.diegues70@gmail.com

José Ferreira

Instituto de Educação e Psicologia da Universidade do Minho

jcferreira12@gmail.com

Luísa Domingues

Instituto de Educação e Psicologia da Universidade do Minho

luisa70@gmail.com

Jorge Costa

Instituto de Educação e Psicologia da Universidade do Minho

jorcos@gmail.com

Lia Raquel Oliveira

Universidade do Minho

lia@iep.uminho.pt

### Resumo

Nos últimos anos têm-se assistido a uma verdadeira corrida no mundo dos motores de busca com o Google a dominar o mercado e outros a tentarem acompanhá-lo. No entanto, já desde os anos 40 que se tem abordado e investigado esta questão. Os motores de busca indexam a informação disponível na Internet para a tornar mais acessível. Ao darem-nos acesso à informação, estão a estender algum do seu poder ao utilizador. Organizam a informação digital da mais relevante para a menos relevante, utilizando para isso diversos critérios de seriação. É, então, necessário desenvolver estratégias de pesquisa para acedermos melhor à informação e podermos criar conhecimento. Se compararmos o funcionamento dos motores de busca com o da nossa própria cognição, verificamos que existem bastantes semelhanças. Esse paralelismo cognitivo faz-nos reflectir sobre o futuro dos motores de busca. Será que conseguirão, eventualmente, pensar?

Palavras-chave: motores de busca, aprendizagem na web, aprendizagem em rede, cognição.

### Abstract

In the last years one has seen a remarkable race in the world of search engines with Google controlling the market and others just trying to keep up. Nonetheless, the attempt at improving these systems dates back to the 40s. Search engines index the information that is available on the Internet to make it more accessible. By giving us access to the information, they are extending some of its power to the user. They organize the digital information from the most relevant to the least relevant using, for that purpose, seriation criteria. It is, therefore, necessary to develop searching strategies for us to create knowledge. If we compare search engines with our own cognition, we can see that there are similarities. This cognitive parallelism makes us reflect about the future of search engines. Will they, eventually, be able to think?

Keywords: search engines, web-learning, network learning, cognition

## Motores de Busca

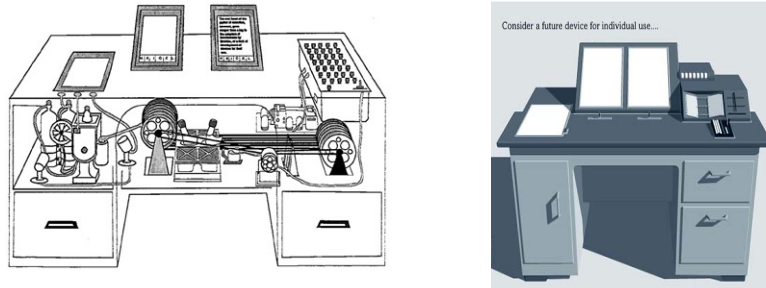
### Evolução

Vannevar Bush propôs, em meados dos anos 40, no artigo intitulado “As We May Think”, um mecanismo capaz de articular a informação em função das necessidades do utilizador, actuando como uma “extensão de memória”. Este dispositivo electromecânico foi designado por Memex (“*memory expander*”) e destinava-se a armazenar documentos pesquisáveis por índice.



Fig. 3 – Artigo de Vannevar Bush – “As we may think”

O Memex deveria permitir ao utilizador definir associações entre informação, ligando um texto a outro, reflectindo a capacidade humana de associar informação. Esta ideia visionária ilustrou a necessidade de criação de um mecanismo que permitisse armazenar o conhecimento acumulado. Mais, era necessário que esse mecanismo permitisse uma fácil consulta dos documentos e, simultaneamente, fosse capaz de articular a informação armazenada em função das necessidades do utilizador. Este mecanismo deveria funcionar como uma extensão de memória para o ser humano, uma vez que permitiria o armazenamento e processamento de grandes quantidades de informação.



**Fig. 4 – O Memex idealizado por Vannevar Bush.**

Estas ideias foram retomadas por Douglas Engelbart, no início dos anos 60, com o programa “Augment”. Em 1965 Theodore Nelson cria um projecto que se destinava a ser o repositório de tudo o que a humanidade tinha escrito. Este projecto, conhecido por Xanadu nunca chegou a ser totalmente concretizado. A sua grande virtude residiu no reforço da importância de ser criado um sistema capaz de armazenar e processar grandes quantidades de informação, ao mesmo tempo que criava a designação de hipertexto, cuja principal funcionalidade residia na capacidade não linear de acesso à informação, em função das motivações do utilizador.

O primeiro motor de busca surgiu em 1990 - Archie - criado por Alan Emtage, estudante universitário. “Este motor de busca ‘rapava’ os arquivos dentro da Internet (daí o seu nome “Archie”) e construía um índice de cada ficheiro que encontrava” (Battelle, 2006).

Em 1993, estudantes da Universidade de Nevada criaram o Verónica. Tratava-se de uma versão do Archie com uma melhoria no Gopher o que o tornava bastante mais apto para trabalhar com o protocolo FTP. Esta melhoria técnica constituiu um passo em frente na precisão da busca, aproximando-a dos actuais padrões. No entanto, quer ao Archie quer ao Verónica faltavam ainda as capacidades semânticas, uma vez que não indexavam todo o texto do documento, mas apenas o título. Na prática, isto significava que “o utilizador tinha de saber, ou inferir, o título do documento que procurava” (Battelle, 2006).

Com o WWW o Archie e o Verónica tornaram-se ultrapassados. O número de páginas Web crescia exponencialmente o que levou ao aparecimento do primeiro motor de busca para a Web – o Wanderer.com - criado por Matthew Gray. Este motor de busca foi rapidamente ultrapassado por outros mais potentes que iam aparecendo. Em 1994 surge o WebCrawler.com, desenvolvido por Brian Pinkerton. Este motor de busca foi mais tarde adquirido pela AOL (America On Line) tendo-se tornado o primeiro motor de busca verdadeiramente democrático, dada a sua interface de navegação simples e a capacidade de busca de texto inteiro.

Em Maio de 1994, Michael Mauldin cria o Lycos, motor de busca inovador que usava algoritmos matemáticos mais sofisticados para determinar o significado das páginas, tendo-se tornado o primeiro a utilizar os links (ligações) como base de relevância. Em Outubro de 1995, nasce o Excite, um motor de busca pioneiro em procedimentos considerados, actualmente, imprescindíveis na Web.

Quando a “Internet era jovem e ‘Google’ era ainda um erro comum de ortografia para muita gente” (Battelle, 2006), Louis Monier cria o AltaVista. A capacidade deste motor de busca, associado ao desenvolvimento da tecnologia dos processadores, permitiram ao AltaVista criar um autêntico índice da (ainda) jovem Web, até então nunca conseguido. Trata-se, efectivamente, do primeiro motor de busca propriamente dito. Em 1995, Jerry Yang e David Filo lançam o Yahoo. A Yahoo detinha uma parte única do mercado, pois criara um directório que permitia a especialistas e aos utilizadores comuns o acesso à Web. Em 1998, Larry Page e Sergey Brin lançam o Google. Este motor de busca incorporou um sistema inovador de escalonamento de páginas designado por PageRank que permitiu resolver o problema dos links nas páginas, ou seja, na prática conseguia efectuar uma triagem das páginas, escalonando-as por sequência. O sucesso do Google torna-se inegável.

### **Conceito**

Um motor de busca é uma aplicação informática que encontra informações contidas nos sites. Todos desempenham três funções chave: procurar informações por palavras; constituir um índice e indexá-las ao local onde foram encontradas; permitir ao cibernauta conjugar várias palavras que estejam contidas nesse índice, de forma a criar novas combinações. Para além de procurarem nos cabeçalhos, URL, nos títulos e nos textos, utilizam os *metadados* dos *sites* para ‘perceberem’ se essas páginas interessam ou não para a busca que estamos a efectuar. Os metadados (*meta tags*) são etiquetas de informação que um dono de um *site* define como sendo um resumo do seu *site* e que estabelece como é que este deve ser indexado.

Para Peixoto (2008),

os motores de busca utilizam software conhecido como 'aranhas' ou 'robots' que percorrem 'toda' a Internet em busca da informação (...) que se pretende (...) sempre que se introduz uma palavra ou um conjunto de palavras (...) as bases de dados são percorridas em busca de documentos ou sites que lhe correspondam (...) num processo designado por Web Crawling.

Este implica que as aranhas consultem muitas páginas até encontrarem o pretendido. A pesquisa começa, logicamente, pelos servidores que têm maior tráfego e pelas páginas mais procuradas.

Devido às alterações que ocorrem nas páginas, o trabalho das aranhas nunca está completo. Há sempre novas palavras para indexar a novas localizações. No entanto, a nossa busca tem um resultado, que corresponde ao momento em que a aranha encontra uma série de páginas que respondem ao critério da busca. Nessa altura, os dados são recolhidos para o index dos motores de busca, que cria uma base de dados com essa informação. A forma como a informação é indexada depende de cada motor de busca, podendo ser feita por palavras, títulos, URL's ou por directorias. Segundo Lancaster (1993), “o propósito principal da elaboração de índices e resumos é construir representações de documentos publicados numa forma que se preste à sua inclusão em algum tipo de base de dados”.

Um motor de busca é composto por cinco componentes principais: o *crawler*, o *repositório*, o *indexador*, o *ordenador* e o *apresentador*. O *crawler* descobre e recolhe automaticamente conteúdos da Web, seguindo os links contidos nas páginas. Apenas os conteúdos que o crawler seja capaz de encontrar e recolher, poderão vir a constar em resultados de pesquisas no motor de busca. O *repositório* armazena as páginas recolhidas de modo a que possam ser indexadas e mostradas em cache. O *indexador* extrai as palavras dos conteúdos Web e cria um índice invertido. Caso não seja possível extrair correctamente as palavras de uma página, esta dificilmente será retornada como resultado de pesquisas. O *ordenador* ordena as páginas que contenham os termos pesquisados por um utilizador de modo a que as mais relevantes sejam apresentadas nos primeiros lugares. As páginas que não tenham sido escritas considerando os requisitos dos motores de busca são relegadas para posições mais baixas em relação a páginas optimizadas para motores de busca. O *apresentador* gere a interface de utilização do motor de busca. Ele recebe os termos pesquisados pelos utilizadores, acede à informação dos índices e apresenta os resultados da pesquisa na forma de links para as páginas.

### **Em busca da informação**

Para identificar a informação relevante, necessária e correcta, cada utilizador necessita de localizar e avaliar a fonte da informação. Os motores de busca são uma ferramenta fundamental para esta tarefa.

Segundo a enciclopédia online Wikipédia (2008a), os motores de busca trabalham armazenando informações sobre um grande número de páginas, que obtêm do WWW. Como já foi referido anteriormente, Peixoto (2008) identifica dois tipos de motores de busca: os que indexam, por título e URL, toda a informação onde encontram a palavra ou o conjunto de

palavras a pesquisar, conhecidos como "webcrawlers", sendo o mais conhecido o Google e aqueles que funcionam com base em directorias, sendo o mais conhecido o Yahoo.

Nos motores de busca indexadores, a actividade de pesquisa de informação é realizada por um *Web Crawler*<sup>1</sup> - um Web browser automatizado que segue cada link que vê. As exclusões podem ser feitas pelo uso do robots.txt<sup>2</sup>. O conteúdo de cada página então é analisado para determinar como deverá ser indexado (por exemplo, as palavras são extraídas de títulos, cabeçalhos ou campos especiais chamados meta tags). Os dados sobre as páginas são armazenados num banco de dados indexado para uso nas pesquisas futuras. Alguns sistemas, como o do Google, armazenam toda ou parte da página de origem, a cache, assim como informações sobre as páginas e alguns armazenam cada palavra de cada página encontrada, como o AltaVista. Esta página em cache guarda sempre o próprio texto da busca. Os directórios, como o Yahoo, são bases de dados que utilizam uma estrutura hierárquica.

Se o utilizador estiver interessado em encontrar informação sobre a aurora boreal, poderá seguir o caminho através de categorias de assuntos: ciência, astronomia, auroras boreais. A partir daqui pode ligar-se a vários sites Web que mostram imagens da aurora boreal a partir da terra e do espaço, explicam o fenómeno e indicam onde o mesmo pode se encontrado (Jonassen, 2000, p.201).

A facilidade de encontrar informação é maior num directório do que num indexador, já que a informação aí incluída é analisada, agrupada e categorizada por intervenção humana.

## Motores de busca – uma ferramenta de aprendizagem

### Curiosidade VS Fidelidade

Search engines have become one of the most important Web services because of the rapid growth in the amount of information available on the Internet. Increasingly advanced features are being developed to improve the standard of service provided to search engine users. [...] By knowing what kind of features motivate usage, search engine designers can focus on and improve the features that stimulate and strengthen a user's motivation (Wu, Chuang, & Chen, 2008).

Wu, Chuang e Chen (2008, p.1830) referem que o recurso a um determinado motor de busca e não outro pode ser explicado por duas razões: "try-out motivation" ou "keep-using

---

<sup>1</sup> A **web crawler** (also known as a **web spider**, **web robot**, or - specially in the FOAF community—**web scutter**) is a program or automated script that browses the World Wide Web in a methodical, automated manner. Other less frequently used names for web crawlers are **ants**, **automatic indexers**, **bots**, and **worms** (Wikipedia, 2008b)

<sup>2</sup> **Robot** é um programa de computador que percorre automaticamente as páginas da Internet em busca de documentos, a fim de indexá-los, validá-los ou monitorar alterações de conteúdo. Para controlar as actividades desses *robots* durante suas buscas, opcionalmente, webmasters podem criar um arquivo chamado **robots.txt** no directório raiz de um determinado endereço web. **Robots.txt** é um arquivo no formato texto (.txt) que funciona como "filtro" para os Crawlers e robots dos motores de busca da Internet, permitindo ou bloqueando o acesso a partes ou à totalidade de um determinado site. (Wikipedia, 2008c)

motivation”. O primeiro tipo de motivação ocorrerá na fase de teste das suas funcionalidades. Esta fase pode ocorrer se formos programadores ou se, como utilizadores comuns, o estivermos a explorar pela primeira vez. O segundo tipo de motivação ocorrerá se houver algum tipo de fidelização. Esta fidelização ocorre quando o utilizador recorre ao motor de busca para adquirir informação. Há uma motivação contínua que só deixará de existir caso o motor de busca seja alterado e deixe de ir ao encontro das necessidades do utilizador.

## Web-Aprendizagem

Using the Internet as a vast online library that it is, requires multiple skills. Effective information gleaned from the Internet combines expertise in searching for information, evaluating the worth of that information, and then organizing the information to make it more readily usable. (Jonassen, Howland, Marra, & Crismond, 2008, p.15)

Uma pesquisa eficaz é precedida por uma necessidade ou intencionalidade. Pesquisamos porque precisamos de informação. Podemos considerar que “navegar na Internet” é uma actividade sem rumo, à deriva e que “pesquisar na Internet” é uma actividade consciente, com um propósito, um objectivo. Se entendermos a pesquisa na Internet como uma actividade cognitiva, é possível decompor a actividade em vários elementos para conseguir compreender e melhorar a pesquisa. Wu, Chuang e Chen (2008, p.1831) dividem esse processo da seguinte forma:

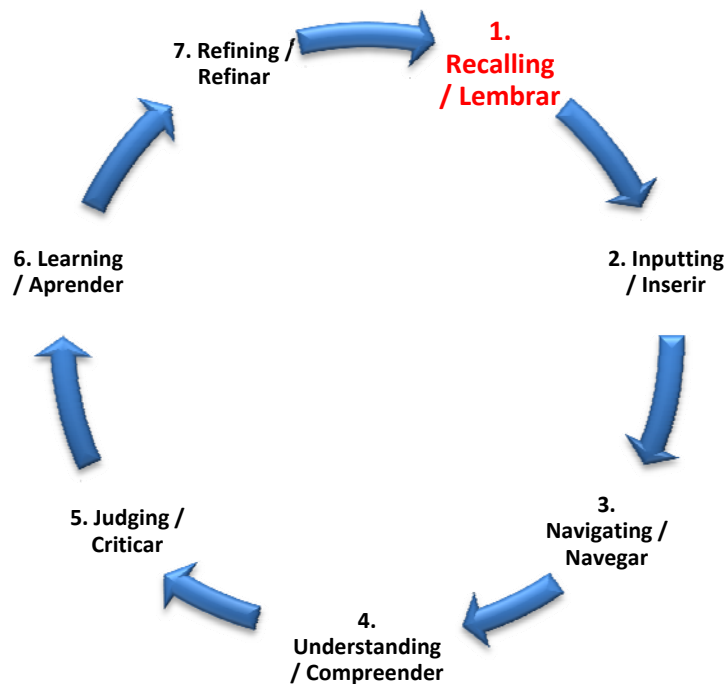


Fig. 5 – Ciclo de Pesquisa de Wu, Chuang e Chen (2008)

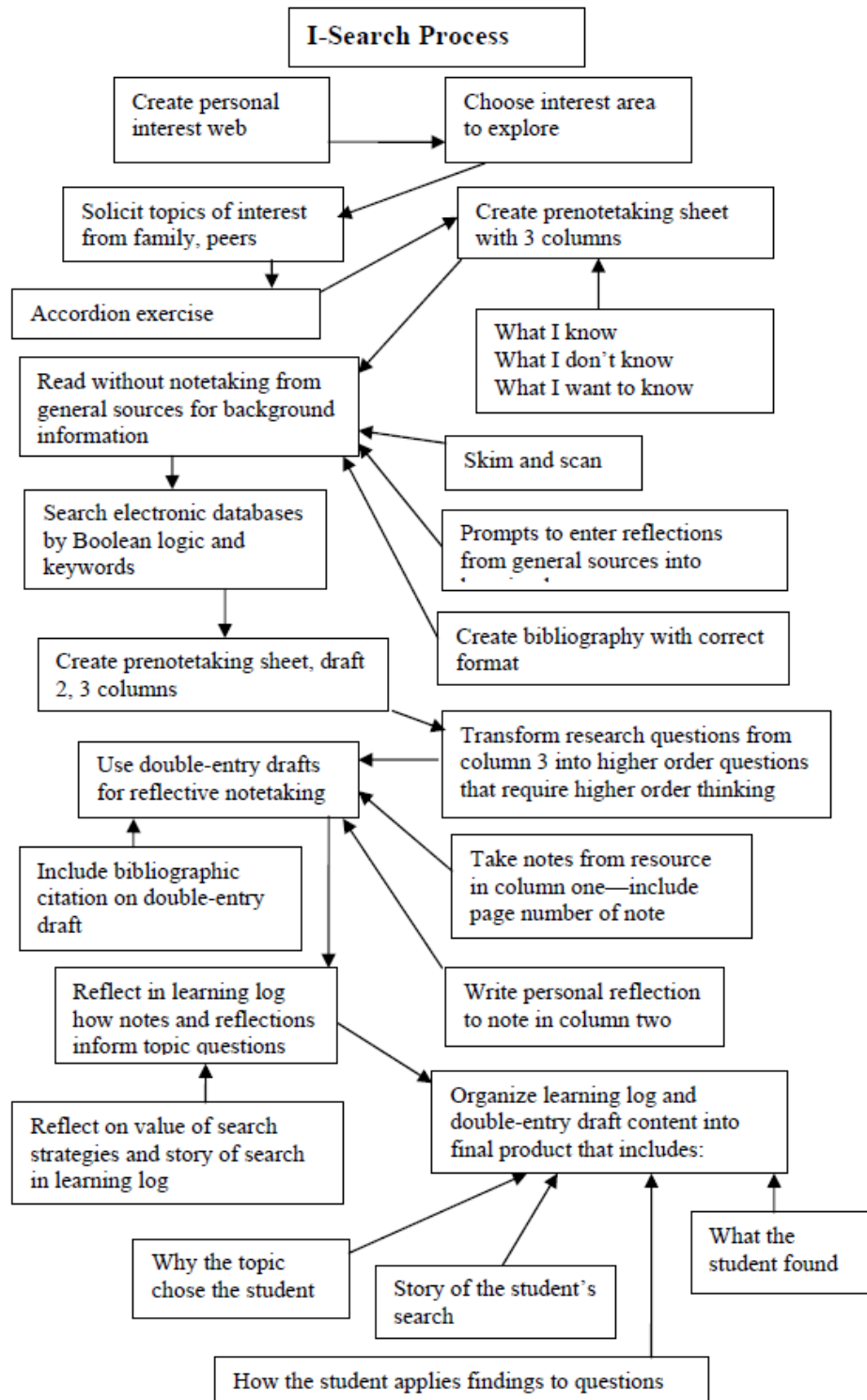
Já Jonassen e Colaric (2000, p. 202) descrevem-no de outra forma:



Fig. 6 – Ciclo de pesquisa de Jonassen e Colaric (2000).

Tallman e Joyce (2005, p.2) apresentam ainda outro processo denominado de I-Search onde desenham uma orientação mais detalhada sobre o processo de procura de informação:





Copyright J. Tallman and M. Joyce, 2005

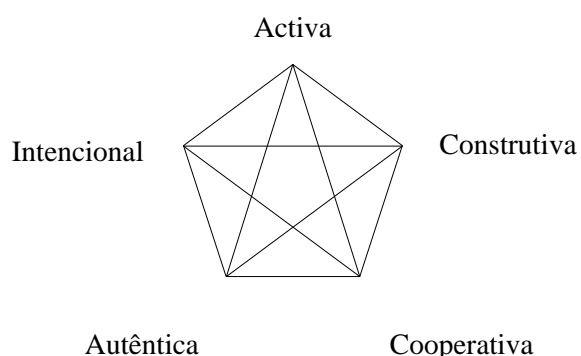
Fig. 7 – I-Search Process (Tallman e Joyce, 2005).

Os três ciclos de pesquisa referidos são bastante semelhantes e servem como orientação primária para a criação de uma aprendizagem significativa, onde se use menos a simples memorização e mais o relacionamento entre o que está já aprendido e o que se está a conhecer.

Ao procurar informação num dos muitos motores de busca disponíveis actualmente (Google, Yahoo, MSN Live, AOL, Ask...) o utilizador está a agir sobre a informação. Tem de usar aquilo que já conhece para iniciar a pesquisa e tentar descobrir o que ainda não conhece. Dos muitos resultados que surgem há que filtrar os que são importantes e organizar a informação de uma forma lógica, coerente e útil. Caso não encontre o que procura, é necessário voltar a usar os pré-conhecimentos para reorientar a pesquisa com outras palavras-chave.

Uma componente comum aos três ciclos de pesquisa descritos é a crítica. Jonassen & Colaric (2000, p. 210) referem que esta é a componente mais invocada e a criativa a menos, que só se encontra presente na escolha de uma estratégia para pesquisar. Tal permite-nos retirar algumas conclusões sobre o real valor da aprendizagem através das pesquisa na Internet. A componente crítica é usada durante quase todo o processo, pois o utilizador tem de verificar a relevância do que encontra para o seu objectivo global. No entanto, a aprendizagem só ocorre quando o utilizador relaciona as suas descobertas com aquilo que já sabia. Dessa forma, é-nos possível dizer que esta é apenas uma fase da aprendizagem no sentido de que ela seja verdadeiramente significativa.

Jonassen, Howland, Marra, & Crismond (2008, p. 3) referem que a aprendizagem significativa possui cinco grandes características: é activa, intencional, construtiva, autêntica e cooperativa. Se compararmos as características da aprendizagem significativa com as que uma pesquisa significativa deve ter, podemos concluir que existe paralelismo.



**Fig. 8 – Jonassen & Colaric (2008, p.3)**

Uma pesquisa é intencional na medida em que temos um objectivo de descobrir e de conhecer. Há uma curiosidade ou necessidade precedente.

Uma pesquisa é activa na medida em que temos de agir sobre algo, quer se trate de uma biblioteca, um livro ou um motor de busca. É através da interacção com o que nos rodeia que somos activos, tanto como observadores como manipuladores.

A característica construtivista está presente na interação cognitiva entre os pré-conhecimentos e os novos. A criação de um conhecimento significativo ocorre quando o utilizador tenta construir um novo conhecimento através da desconstrução do que é novo e da integração no antigo.

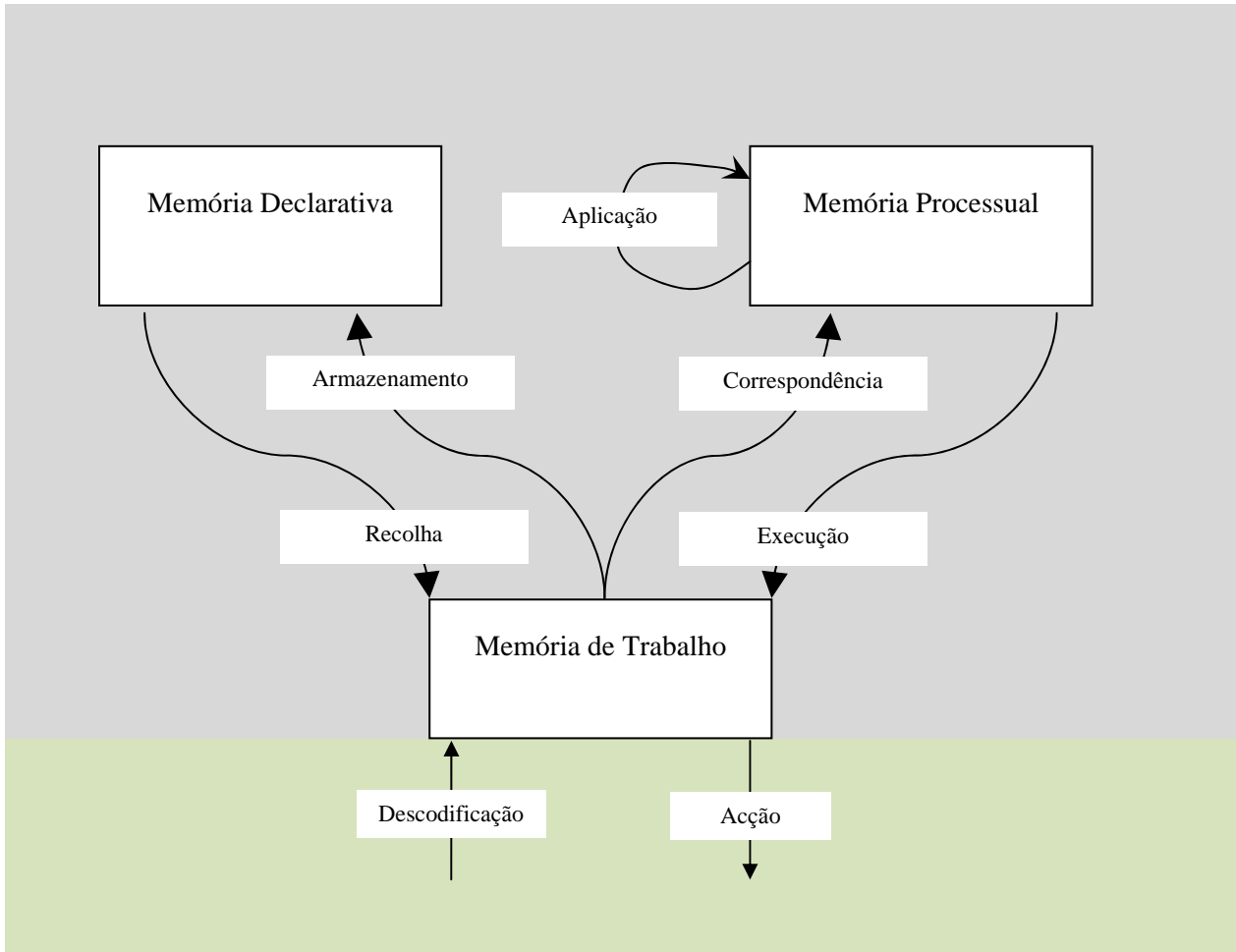
Pesquisar na Internet é uma actividade normalmente perspectivada como sendo individual e solitária. Se reflectirmos sobre o processo de seriação dos sites, podemos chegar a uma outra conclusão. Um dos elementos para a classificação da importância de um site é o número de visitantes que tem. Dessa forma, pode-se assistir a uma afinação colaborativa da qualidade dos sites, mesmo não havendo contacto directo entre os participantes.

A autenticidade também se encontra presente nos motores de busca. A maior parte da informação que está disponível na Internet não foi desenvolvida artificialmente para um propósito educativo. São pessoas / empresas reais que por necessidade ou vontade criam conteúdos contextualizados na Internet. Essa contextualização dos conteúdos autentifica e valida a pesquisa.

É importante seguir determinados passos quando se quer encontrar qualquer tipo de informação e é também importante que a nossa pesquisa seja significativa. Os motores de busca devem, no entanto, ser considerados como apenas um meio na construção de conhecimento e não um fim. São apenas uma das fases. É uma fase que é pautada predominantemente pela componente crítica e não criativa. A fase de criação é posterior à fase de pesquisa.

### **Metacognição**

Nos últimos anos, têm-se assistido a um interesse acrescido nos processos cognitivos que estão por trás da aprendizagem. Um dos exemplos disso mesmo é o enfoque que o portefólio tem tido como ferramenta reflexiva e representativa da evolução. Acreditamos que os motores de busca podem ser encarados numa perspectiva metacognitiva, na medida em que interagem de uma forma cíclica com a nossa cognição e também porque o seu próprio funcionamento assemelha-se ao da nossa cognição. Marzano (1998, p. 245) descreve o nosso sistema cognitivo da seguinte forma:



**Fig. 9 – Sistema cognitivo humano (Marzano 1998, p. 245) .**

Num contexto metacognitivo, a memória declarativa contém informação sobre a natureza dos objectivos, planos, linhas temporais, recursos e as suas interacções. Por exemplo, se pedirmos que se pesquise informação sobre carros, o utilizador sabe o que são carros, sabe que tem de fazer um plano de pesquisa, organizar recursos e estabelecer metas. A memória processual coloca todas as etapas em prática. Com a memória processual, o utilizador cria o plano de pesquisa, escolhe quais as melhores fontes de informação e estabelece metas. (Marzano, 1998, p. 253)

No funcionamento dos motores de busca, pode encontrar-se algo semelhante. Belew (2000, p. 10) simplifica o funcionamento de um sistema de busca através da seguinte figura:

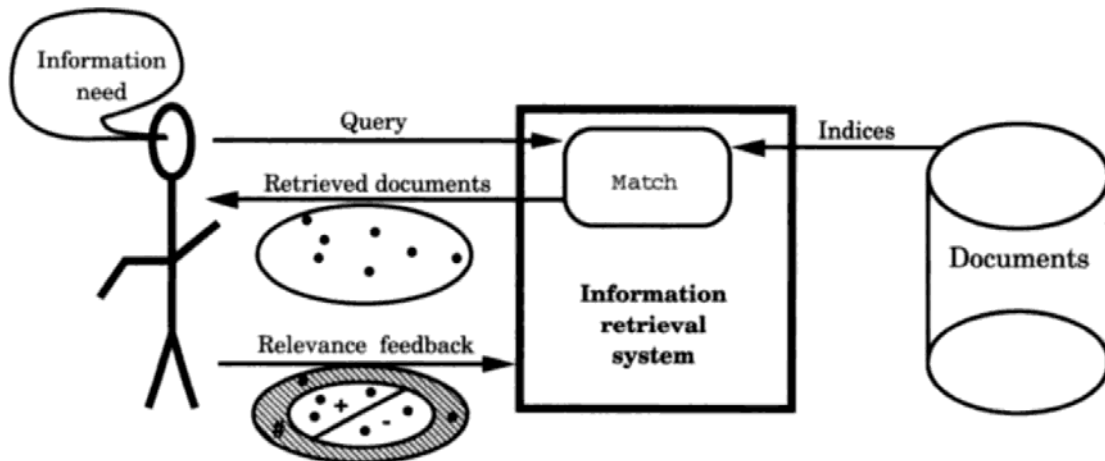


Fig. 10 – Ferramentas e processo de um sistema de busca (Belew, 2000, p.10).

Como se pode notar, há semelhanças no funcionamento dos dois sistemas. Se sobrepuermos as estruturas gráficas podemos ficar com uma ideia mais clara em como há processos equivalentes.

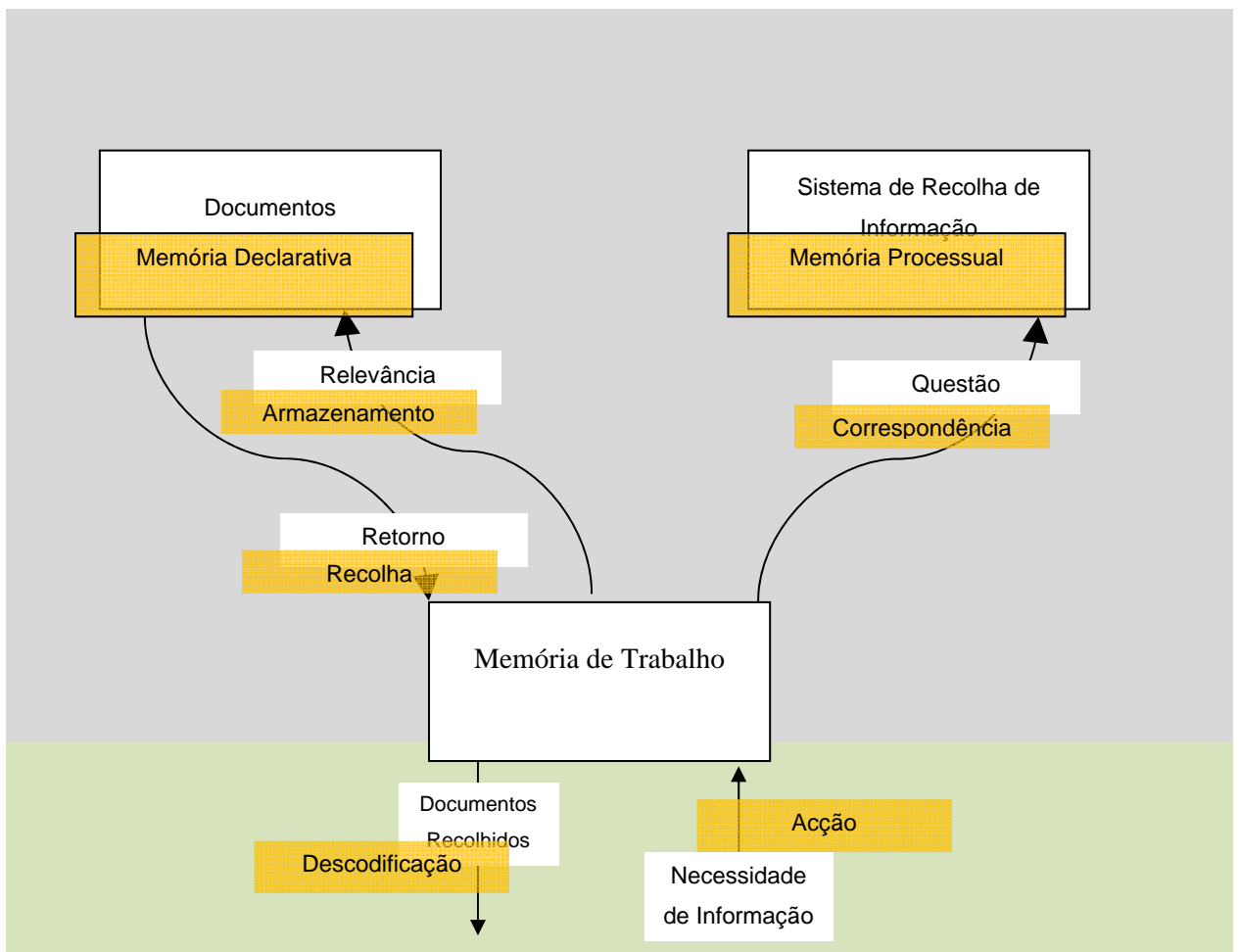


Fig. 11 – Sobreposição do funcionamento de um motor de busca e do funcionamento da nossa cognição.

Iniciamos a pesquisa porque temos uma necessidade de informação e colocamos uma questão através de conceitos, sintetizados em palavras-chave, ao motor de busca. Através de um sistema de recolha de informação, o motor de busca estabelece uma correspondência aos documentos. Ao dar retorno, o motor de busca está igualmente a receber informação sobre a relevância dos documentos que está a fornecer.

Será então que os motores de busca funcionam desta maneira por razões puramente programáticas ou será que os programadores se basearam no funcionamento da nossa própria cognição? Se os motores de busca foram feitos à nossa imagem então será que eventualmente vão conseguir perceber-nos assim como nós percebemos quem fala a nossa linguagem? E sentir?

## Referências

- Battelle, J. (2006). *The Search*. Cruz Quebrada: Casa das Letras.
- Belew, R. K. (2000). *Finding Out About: A Cognitive Perspective on Search Engine Technology and the WWW*. Cambridge: Cambridge University Press.
- Jonassen, D. H., & Colaric, S. (2000). Ferramentas de pesquisa intencional de informação enquanto ferramentas cognitivas. In D. H. Jonassen (Ed.). *Computadores como Ferramentas cognitivas*. Porto: Porto Editora. Pp. 195-214.
- Jonassen, D., Howland, J., Marra, R. M., & Crismond, D. (2008). Investigating with Technologies. In D. Jonassen, J. Howland, R. M. Marra, & D. Crismond (Eds.). *Meaningful Learning with Technology*. New Jersey: Pearson Education, Inc. Pp. 13-39.
- Wikipédia (2008a). Motores de busca. Acedido em 4 de Novembro de 2008 em [http://pt.wikipedia.org/wiki/Motor\\_de\\_busca](http://pt.wikipedia.org/wiki/Motor_de_busca).
- Wikipédia (2008b) Webcrawler. Acedido em 4 de Novembro de 2008 em <http://pt.wikipedia.org/wiki/webcrawler>.
- Wikipédia (2008c) Robots. Acedido em 4 de Novembro de 2008 em <http://pt.wikipedia.org/wiki/robots>.
- Peixoto, P. (2008). Motores de busca. Acedido em 3 de Novembro de 2008, de [http://www4.fe.uc.pt/fontes/pesquisa\\_na\\_internet/motores\\_busca/motores\\_de\\_busca.htm](http://www4.fe.uc.pt/fontes/pesquisa_na_internet/motores_busca/motores_de_busca.htm).
- Tallman, I. J., & Joyce, Z. M. (2005). *The I-Search: A Powerful Collaborative Planning Tool*. American Association of School Librarians. Pittsburgh, Pennsylvania.
- Wu, L.-L., Chuang, Y.-L., & Chen, P.-Y. (2008). Motivation for Using Search Engines: A Two-Factor Model. *Journal of the American Society for Information Science and Technology* , Pp. 1829-1840.

Nota: Este trabalho integra-se no Projecto Colectivo DesignDem2 - Design de Dispositivos de Educação Mediatizada: processos, ambientes e objectos de aprendizagem (PC11-LIII-2008), CIEd.