

Pesquisa e Publicação de Informação

Introdução à Pesquisa de Informação

Nuno D. Mendes

Licenciatura em Sistemas e Tecnologias de Informação

30 Mar 2012
ISEGI – UNL

Pesquisa de Informação

- ▶ Trata da Representação, Armazenamento, Organização e Acesso a Informação
- ▶ Procura satisfazer Necessidades de Informação do Utilizador

Necessidade de Informação do Utilizador

Obter todos os documentos sobre Universidades com ofertas de cursos superiores na área da gestão da informação localizadas perto de uma cidade com mais de 20,000 habitantes e que apresente a morada, telefone e nota mínima de entrada no curso no último concurso de acesso ao Ensino Superior

Necessidade de Informação do Utilizador

Obter todos os documentos sobre Universidades com ofertas de cursos superiores na área da gestão da informação localizadas perto de uma cidade com mais de 20,000 habitantes e que apresente a morada, telefone e nota mínima de entrada no curso no último concurso de acesso ao Ensino Superior

Pedido de Informação – Query

Normalmente convertida em **palavras-chave** ou **termos de pesquisa**:
universidade + licenciatura + gestão da informação + morada + telefone + nota mínima de acesso

Recuperação de Dados

- ▶ e.g., Que documentos contêm um dado conjunto de palavras-chave.
- ▶ Semântica bem definida (e.g. ocorrência de palavras, queries SQL em bases de dados relacionais, etc)
- ▶ A recuperação de um só objecto que não cumpra os critérios de pesquisa = **Falha!**

Recuperação de Dados

- ▶ e.g., Que documentos contêm um dado conjunto de palavras-chave.
- ▶ Semântica bem definida (e.g. ocorrência de palavras, queries SQL em bases de dados relacionais, etc)
- ▶ A recuperação de um só objecto que não cumpra os critérios de pesquisa = **Falha!**

Pesquisa de Informação

- ▶ e.g., Obter informação sobre um assunto ou um tópico
- ▶ Semântica mal-definida, pesquisas em língua natural (frequentemente ambígua)
- ▶ Pequenos erros são tolerados

Pesquisa de Informação vs. Recuperação de Dados

Recuperação de Dados

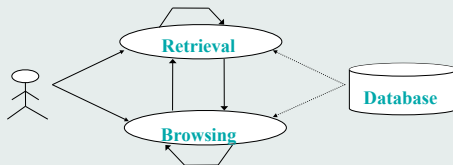
- ▶ e.g., Que documentos contêm um dado conjunto de palavras-chave.
- ▶ Semântica bem definida (e.g. ocorrência de palavras, queries SQL em bases de dados relacionais, etc)
- ▶ A recuperação de um só objecto que não cumpra os critérios de pesquisa = **Falha!**

Pesquisa de Informação

- ▶ e.g., Obter informação sobre um assunto ou um tópico
- ▶ Semântica mal-definida, pesquisas em língua natural (frequentemente ambígua)
- ▶ Pequenos erros são tolerados

Sistema de Pesquisa de Informação

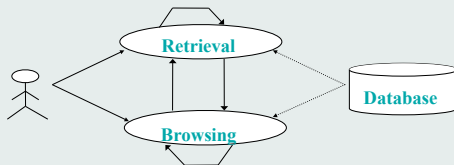
- ▶ Procura interpretar o conteúdo dos objectos informacionais
- ▶ Constrói uma ordenação de resultados que reflecte a sua relevância para a pesquisa efectuada
- ▶ o Sistema é tanto melhor quanto mais eficaz fôr em captar a noção de relevância



Pesquisa/Retrieval

- ▶ O utilizador procura informação (ou dados) com um certo propósito definido
- ▶ Converte a sua necessidade de informação numa query

Tarefas do Utilizador

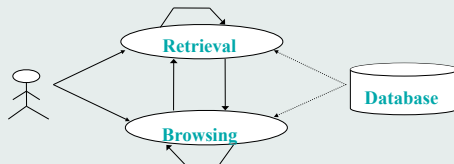


Pesquisa/Retrieval

- ▶ O utilizador procura informação (ou dados) com um certo propósito definido
- ▶ Converte a sua necessidade de informação numa query

Navegação/Browsing

- ▶ O utilizador procura informação genérica sobre um tópico, mas
- ▶ Pode querer navegar entre tópicos relacionados
- ▶ A necessidade de informação não está bem definida
- ▶ A relação com o sistema é interactiva e envolve tarefas de pesquisa e navegação intercaladas

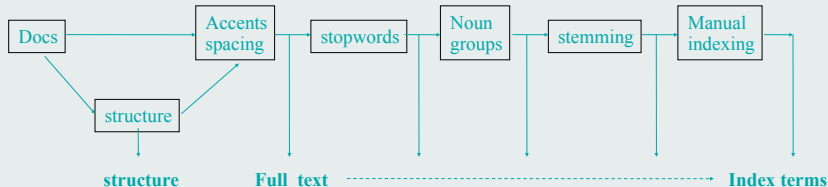


Filtragem

Quando a pesquisa de informação é feita automática e sistematicamente pelo sistema de pesquisa de informação a partir de um repositório dinâmico de objectos informacionais é designada de *filtragem*.

(e.g. envio periódico de notícias relevantes para uma query definida pelo utilizador – alertas Google News)

Visão Lógica dos Documentos



- ▶ Podem considerar o texto integral (+ preciso?)
- ▶ Podem considerar apenas termos de indexação definidos por um humano (+ conciso, - preciso)
- ▶ Podem adoptar representações intermédias
 - ▶ Realizar operações sobre o texto para eliminar palavras muito comuns (stopwords) como artigos ou preposições
 - ▶ Considerar apenas a raiz das palavras (ignorar género e número ou marcas de modo e tempo)
 - ▶ Considerar apenas substantivos (ignorar verbos e adjektivos)
- ▶ Podem tirar partido da estrutura do documento (capítulos, secções, etc)

Até ao advento da WWW

- ▶ Classificação e Categorização de documentos
- ▶ Suporte de vários sistemas e linguagens
- ▶ Interfaces com o Utilizador e Visualização da Informação

Até ao advento da WWW

- ▶ Classificação e Categorização de documentos
- ▶ Suporte de vários sistemas e linguagens
- ▶ Interfaces com o Utilizador e Visualização da Informação

Após o advento da WWW

- ▶ Web como repositório universal de conhecimento
- ▶ Acesso livre
- ▶ Sem coordenação editorial
- ▶ Informação não-estruturada, não-categorizada, não-organizada
- ▶ Grande desafio da área de Pesquisa de Informação

Visão global

