

Bibliotecas digitais: novos desafios para os sistemas de recuperação de informação

**Paulo Quaresma
pq@di.uevora.pt
CITI.UE / CENTRIA.UNL**

Agenda

- Motivação
- Problemas
- Uma possível solução
- Conclusões

Motivação

- Bibliotecas digitais
 - “digital libraries” no google -> 5930000 páginas
 - California Digital library
 - New York Public Digital Library
 - National Science Foundation Digital Libraries (USA)
 - ...

Motivação

- Bibliotecas digitais
 - UNESCO – Greenstone project
 - Universidade de Waikato, Nova Zelândia
 - Projecto Gutenberg (1971-)
 - Livros “domínio público”
 - 10/2006 – 19000 itens, 50 novas entradas por semana
 - Google Open-library
 - Harvard, Stanford, Michigan, Oxford, NYPL, ...
 - ...

Motivação

- Bibliotecas digitais – (wiki)
 - Não há limitações de espaço físico
 - Disponibilidade 24h
 - Acessos múltiplos
 - Preservação e conservação
 - **“Melhor” Recuperação de Informação**
 - Redes de bibliotecas digitais
 - “Menor” custo

Motivação

- Bibliotecas digitais
 - **Texto**
 - Imagem
 - Video
 - Som
 - Música
 - ...



Motivação

- Recuperação de informação
 - Capacidade de seleccionar da biblioteca um documento/objecto **pretendido!**
 - Problemas:
 - Como comunicar a intenção?
 - Se a intenção for correctamente representada e comunicada, como processá-la? Isto é, como inferir qual o documento que satisfaz a intenção?

Problemas

- Como representar e comunicar a intenção dos utentes?
 - Bibliotecários
 - U: Procuo um livro que descreva os primeiros estudos arqueológicos na região de Évora.
 - B: Talvez o “Antiquitatum Lusitaniae” de André de Resende.
 - U: Mas queria a edição original.
 - B: Temos uma edição fac-simile. Vou buscar!

Problemas

- Como representar e comunicar a intenção dos utentes?
 - Bibliotecas digitais (versão 1)
 - U: Selecciona *arqueologia* no catálogo;
 - U: Selecciona *Évora* no catálogo;
 - U: Calcula a intersecção dos resultados;
 - U: “Procura” no resultado final...

 -
 - Provavelmente U encontra o que procura, após muitas tentativas e após seleccionar todos os livros que constam no resultado final!

Problemas

- Como representar e comunicar a intenção dos utentes?
 - Bibliotecas digitais (versão 2)
 - U: Selecciona “arqueologia *and* Évora” em texto livre;
 - U: Ordena os resultados por data.
 - U: “Procura” no resultado final
 -
 - Provavelmente U não encontra o livro que procura!

Problemas

- Sistemas de recuperação de informação baseados exclusivamente em
 - Texto livre
 - Catálogo
- são limitados e inadequados!

Problemas

- Pesquisas em texto livre:
 - Baseadas em reconhecimento de padrões (pattern matching)
 - Todas as palavras dos documentos são indexadas e usadas para responder às interrogações.
 - Incapacidade de representar as intenções dos utentes e o conteúdo dos documentos.

Problemas

- Pesquisas “orientadas” por catálogo:
 - Informação estruturada
 - Utilização de thesauri para representar o conhecimento e as suas relações.
 - Interacção inadequada com os utentes.
 - Dificuldade em representar as intenções dos utentes.

Possível solução

- Pesquisa pelo conteúdo semântico dos documentos:
 - Ontologia, que represente conhecimento e relações
 - Análise sintáctica e semântica dos documentos
 - Interpretação dos documentos (representação semântica), tendo em conta a ontologia que representa o conhecimento.
 - Interpretação da interrogação, tendo em conta a ontologia

Possível solução

- Pesquisa pelo conteúdo semântico dos documentos:
 - U: Procuero um livro que relate os primeiros estudos arqueológicos na região de Évora.
 - ...

Possível solução

- U: Procuro um livro que relate os primeiros estudos arqueológicos na região de Évora.
 - Ontologia:
 - contém os conceitos “arqueologia” e “estudos arqueológicos”, bem como as acções cuja execução representa um “estudo arqueológico”
 - contém os conceitos “local”, “cidade”, “região” e instâncias desses conceitos, como “Évora” (local, cidade e região), “S. Miguel da Mota” (local na região de Évora onde existe um santuário ao Deus Endovélico)

Possível solução

- U: Procuro um livro que relate os primeiros estudos arqueológicos na região de Évora.
 - “Primeiros estudos arqueológicos”
 - Evento (execução de uma acção) cuja acção associada representa um estudo arqueológico e cujo tempo é anterior a qualquer outro evento da mesma classe de acções.

Possível solução

- U: Procuro um livro que relate os primeiros estudos arqueológicos na região de Évora.
 - Ou seja, procura-se um evento de uma acção que represente um “estudo arqueológico”, cuja execução seja anterior a outros eventos de acções semelhantes, e que decorra num local da região de Évora.
 - Se o “Antiquitatum Lusitaniae” de André de Resende contiver este evento, então será seleccionado!

Exemplos “reais”

- CLEF – Cross Language Evaluation Forum:
 - Como morreu Jimi Hendrix?
- Texto:
 - Um patologista defendeu que Jimi Hendrix morreu de asfixia após ter ingerido álcool e uma dose excessiva de barbitúricos .
- Resposta:
 - De asfixia.

Exemplos “reais”

- CLEF – Cross Language Evaluation Forum:
 - Com quem se casou Michael Jackson?
 - A filha de Elvis Presley, Lisa Marie, 26 anos, confirmou o seu casamento há 11 semanas atrás com o cantor Michael Jackson, 35 anos, num comunicado enviado pela MJJ Productions, encarregada na segunda-feira das relações públicas do cantor para os principais media.
- Resposta: Lisa Marie.

Exemplos “reais”

- CLEF – Cross Language Evaluation Forum:
- Quantos desempregados há na Europa?
- A taxa de desemprego nos países industrializados deverá crescer 0,1 por cento até ao fim do ano, atingindo 8,6 por cento da população activa, ou seja, 33 milhões de pessoas, 23 milhões dos quais na Europa ocidental, sublinha um relatório da Organização Internacional do Trabalho (OIT) ontem divulgado em Genebra.
- Resposta: 23 milhões.

Conclusões

- Sistemas actuais de recuperação de informação em bibliotecas digitais são limitados e pouco adequados a representar as intenções dos utentes.
- A integração de metodologias próprias da Inteligência Artificial, tais como o processamento de Língua Natural e a representação de conhecimento, permitirá desenvolver sistemas mais poderosos e adequados às necessidades dos utilizadores.