



UiO : **Det humanistiske fakultet**
Diana Santos

Págico



Português mágico





Tarefa (genérica)

- Dada uma necessidade de informação
- Obter um conjunto de respostas cuja justificação esteja na wikipédia em português
 - Por meios automáticos
 - Ou "manualmente", para encontrar essa informação e a partir dela extrapolar como ajudar os utilizadores, e tentar avaliar a informação



Do ponto de vista do processamento computacional do português

- Sugerir a tarefa
- Entusiasmar investigadores na sua resolução
- Avaliar a wikipédia
- Criar recursos sobre o conteúdo lusófono
- Fazer investigação com a premissa de que estudar as coisas em português pode ter vantagens para o processamento em português



A nossa visão da RI e similares

- Desde 2003 envolvidos no CLEF, que é o forum da RI não-inglesa e multilingue
- Observámos muitas ideias interessantes e muitas tentativas de criar tarefas e soluções originais
- Mas também uma tendência para "nivelar" todo o processamento pelo inglês, que se torna a bitola e o teto
- Mesmo na nossa atividade, e nas tarefas que nós propusemos, passando pelo inglês ou usando recursos para o inglês era sempre melhor

Avaliações conjuntas só sobre o português



- Os problemas e as soluções têm de ser encontrados na nossa língua
- Os recursos têm de ser criados para a nossa língua
- As aplicações servem para a nossa língua (outros que as adaptem para outras línguas se estiverem interessados)



A questão da definição da tarefa

- Muita tinta correu sobre a questão da definição da tarefa de resposta a perguntas, para a delimitar da simples obtenção de documentos que possam conter a resposta
- Um dos problemas das perguntas é que não contêm geralmente a resposta...
 - De que cor é o céu?
- A forma mais comum de implementar um sistema de RAP é assumir que as palavras das perguntas co-ocorrem com a resposta



Mas a maior parte das necessidades de informação são mais complexas

- Validação/confirmação
 - É verdade que Aristóteles foi aluno de Platão?
 - É verdade que Sócrates acabou o curso num domingo?
- Comparação
 - O que apareceu primeiro, os patins de gelo ou de rodas?
 - Qual o sítio que produz mais sal? Aveiro ou Alcácer do Sal?
- Ver o que os outros dizem sobre um assunto
- Descobrir coisas



Mas a maior parte das necessidades de informação são mais complexas

- Onde é que há fiordes?
- Lista de sítios
- Onde é que há estalactites?
- Onde é que há restaurantes com 3 estrelas da Michelin?
- Qual a história da distribuição de água em Lisboa?
- Quais as percentagens de votação nas eleições europeias ao longo do tempo?



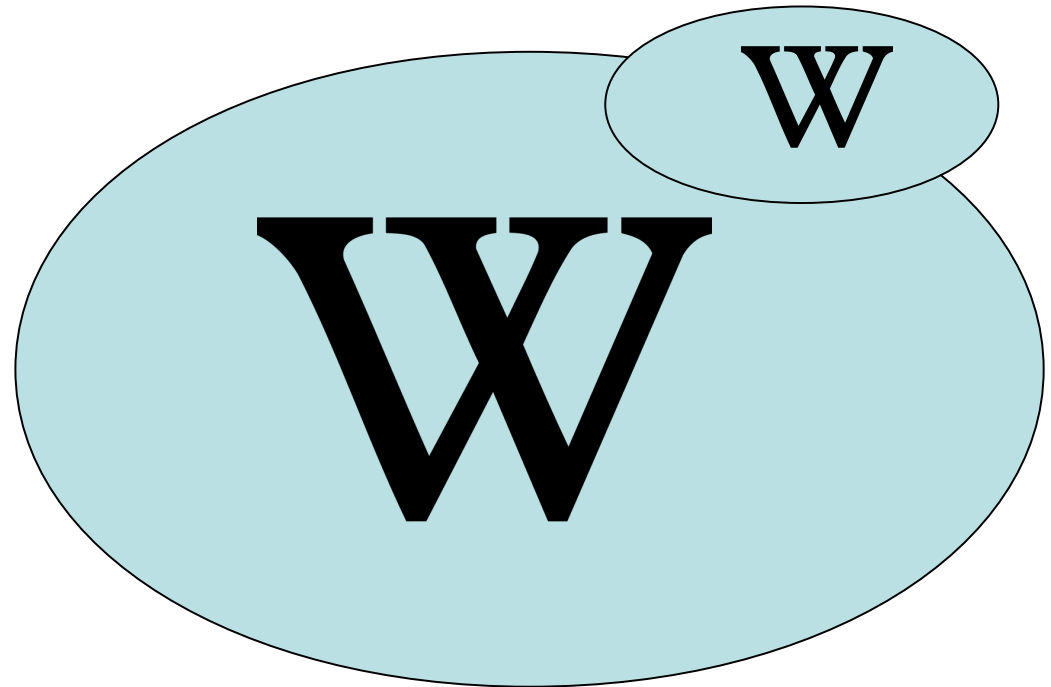
A única coisa que não interessa é a forma de especificar a pergunta

- Daí usarmos indiscriminadamente os termos tópico, pergunta, ou necessidade de informação, nas avaliações conjuntas que temos organizado
- A resposta pode ser evidente, surpreendente, ou estar implícita
- Sobretudo perguntas sobre definições, não existe (exceto em domínios científicos muito específicos) uma única resposta.
 - Quem é George W. Bush?
 - Quem foi Sócrates? O que é acometer?



Págico

- Em vez de olhar para a Wikipédia toda
- Concentramo-nos na em português





Primeiras perguntas

- Até que ponto é sobreposta?
- Até que ponto é dirigida pela estrutura da inglesa?
- Até que ponto é alinhada em vez de comparável?
- Casos relativos à cultura lusófona: em princípio aqueles em que é mais independente



Ligação da wikipédia com as elites culturais e com a sociedade

- Quantos intelectuais portugueses escrevem na wikipédia em portuguêsês?
- Quantas iniciativas políticas ou pedagógicas existem que usam a wikipédia como um parceiro? Ou que se constituam como alternativa à wikipédia?
- Quantos cientistas ou investigadores usam a wikipédia para publicitar as suas ideias e conhecimento? Quantos corrigem informações erradas ou tendenciosas?



Podemos confiar na wikipédia?

- Precisão
- Abrangência
- Fontes fiáveis
- Texto legível
- *Consistência*
- *Balanço*
- Atualidade
- Verificação pelos pares



Parcialidade(s)

- Históricas (os informáticos foram os primeiros)
- Políticas
- Culturais
 - Culturas periféricas têm os seus vultos menos divulgados que as dominantes
 - Aceitação por uma cultura dominante de um vulto de uma cultura periférica é essencial para a sua expansão



A minha experiência de ensino

- A cultura portuguesa está muito mal representada na wikipédia
- As páginas relativas a vultos internacionais estão muito mais preenchidas do que as nossas
 - Geralmente são traduzidas
 - Às vezes até estão incompreensíveis
- Do ponto de vista científico não há qualquer inovação



Tarefa da organização

- Criar um sistema (SIGA) e preparar a coleção
- Identificar tipos de necessidades de informação com resposta na coleção
- Escolher as respostas
- Avaliar as enviadas pelos participantes
- Definir critérios de avaliação
- Definir medidas
- Criar e disponibilizar recursos: Cartola