

technology
from seed

Do tópico às respostas: do processo humano à sua simulação

Luísa Coheur, INESC-ID / IST

Ângela Costa, INESC-ID/UNL



Instituto de Engenharia de Sistemas e Computadores
Investigação e Desenvolvimento em Lisboa
Laboratório de Sistemas de Língua Falada

- Tentar compreender os mecanismos usados por um humano aquando da sua participação no Páxico;
- Compreender a oferta, na área de PLN, de ferramentas capazes de simular esses mecanismos.

- Trabalho realizado no quadro de um mini-projecto da disciplina de Língua Natural do MEIC-Tagus (IST)
- 8 grupos de 23 alunos
- Objectivos pedagógicos:
 - Participação numa avaliação conjunta;
 - Abstracção das tarefas levadas a cabo na participação;
 - Estabelecimento de uma ponte entre as tarefas realizadas e as ferramentas disponíveis em PLN

- Política de “tentativa e erro”
 - Da query para os textos e destes para novas queries
- Muito ênfase dado à escolha de termos a usar na query
 - Expansão/processamento de termos
 - Curiosidade: vários trabalhos referem a distinção entre termos usados para encontrar páginas vs. termos usados para pesquisar nas páginas encontradas
- Quando considerar esgotada a fonte de informação?
 - Problema em aberto
 - Leva a consulta de fontes de fontes de informação externa

- Existem muitas ferramentas/recursos para ajudar a expansão das queries, eliminação de termos, etc.
- Mas... existem várias tarefas relativamente fáceis de realizar por um humano, mas muito difíceis de realizar pela máquina:
 - Capacidade do humano em extrair informação de queries mal conseguidas muito complicada de implementar;
 - Como fazer a máquina compreender que um texto tem ou não a informação necessária?
 - Como verificar as restrições de um tópico?
 - Quando se deve parar, ou seja, quando é que se deve assumir que um tópico não tem resposta?