Tecnologias da linguagem para o português

Diana Santos
Projecto Processamento computacional do português
http://www.portugues.mct.pt

Objectivos desta apresentação

- Dar a conhecer o projecto
- Descrever os catálogos criados no seu âmbito

Processamento computacional do português

- Projecto da responsabilidade do Ministério da Ciência e da Tecnologia
- Iniciado a 15 de Maio de 1998 no SINTEF (Oslo, Noruega)
- Tarefas principais
  - observação e planeamento da área
  - acções concretas na criação e disponibilização de recursos
  - consultadoria ao MCT na área

Observação da área

- Criação de vários catálogos sobre o que se encontra na rede (WWW)
- Análise dos dados oficiais do Observatório das Ciências e das Tecnologias
- Pedido aos investigadores para que complementem a informação
  - livremente, sob todas as formas
  - preenchendo um questionário na rede

Catálogos

- Recursos > 400 ponteiros
- Actores > 150 endereços (80 projectos)
- Publicações > 350 entradas

Dados oficiais

- Lista de projectos financiados
- Lista de bolsas
- Lista de doutorados

Efectuados como Contributos para um perfil por ocasião dos debates públicos sobre os perfis da Investigação Científica em Portugal, associados à preparação do *Livro Branco para o Desenvolvimento Científico e Tecnológico Português (1999-2006)*

Recursos

corpora (escrito, falado/transcrito, de fala)	23
léxicos e dicionários (gerais, especializados, enciclopédias)	144
ferramentas computacionais (ajuda à redacção, analisadores, tradutores, etc.)	75
material didáctico (língua materna, língua estrangeira)	58
"matéria prima" (literatura, jornais, outros textos, rádio e TV)	160

Actores

- Grupos, centros ou institutos
- Projectos (nacionais, europeus, internacionais)
- Associações, instituições e iniciativas oficiais
- Empresas
- Ensino

Conclusões sobre o resultado da catalogação

- A maioria dos projectos e recursos fica confinada aos grupos
- Há muito poucos resultados que tenham atingido a sociedade portuguesa
- Não existem mecanismos de avaliação ou comparação do que foi ou está a ser feito

Trabalho futuro

- Criação de uma interface amigável para a informação presente nos catálogos
- Disponibilização de ferramentas para várias tarefas básicas de processamento do português
- Lançamento de projectos colaborativos focando a avaliação, partilha e comparação de recursos

Observações finais

- O trabalho de catalogação foi descritivo e não normativo
- Houve pouquíssima participação da comunidade alvo na catalogação
- Não sabemos avaliar, de momento, o impacto real da nossa actividade

Agradecemos todas as sugestões e críticas que nos queiram fazer... projecto@informatics.sintef.no