

## Disponibilização de corpora pela WWW

Diana Santos

SINTEF Telecommunications and  
Informatics  
Diana.Santos@informatics.sintef.no

## Trabalho anterior

- Disponibilização de um corpus bósnio pela Web [www.tekstlab.uio.no/Bosnian/Corpus.html](http://www.tekstlab.uio.no/Bosnian/Corpus.html)
  - Desenho de uma interface para o CQP, o Corpus Workbench da Universidade de Stuttgart
  - Concepção de um serviço ajustado às necessidades da comunidade alvo
  - Reflexão sobre essa actividade:

"Providing access to language resources through the World Wide Web: the Oslo Corpus of Bosnian Texts", *Proceedings of the 1st International Conference on Language Resources and Evaluation* (Granada, 28-30 May 1998)

## E o português?

- É preciso pôr os recursos computacionais da nossa língua
  - + acessíveis
  - + utilizáveis
  - + "melhoráveis"
- É preciso criar uma rede de recursos portugueses

## Rede do português...

- Catálogo de endereços internet
- Acesso a corpora
- Acesso a dicionários e bases lexicais
- Utilização de ferramentas
- Respostas a perguntas sobre portuguesificação

## Acesso a corpora através da Web

- Concordâncias
- Distribuição
- Refinamento
- Anotação
- Relações entre palavras
- Comparação de corpora

## Em conclusão

- Muito trabalho já está feito
- Temos de o saber aproveitar
- É preciso aumentar a visibilidade do trabalho desenvolvido
- É preciso criarmos uma rede dos recursos

## Português computacional

- O português é uma das línguas mais faladas no mundo...

### MAS

- Quantos programas “funcionam” em português?
- Quantos sistemas e ferramentas computacionais foram pensados, de raiz, para a realidade portuguesa?