

## Colaboração entre a Linguateca e o RCAAP: primeiros passos

Fernando Ribeiro  
(fernando.ribeiro@fccn.pt)  
&  
Diana Santos  
(diana.santos@sintef.no)



www.linguateca.pt

## Linguateca

A Linguateca, um centro de recursos para o processamento computacional do português, tem como um dos seus objectivos a cooperação com outros projectos, neste caso com o RCAAP.

- Linguateca
- Linguateca – cooperação com o RCAAP
- Cooperação – trabalho futuro

www.linguateca.pt

2

## Linguateca – em poucas palavras

- 8,5 milhões de visitas ao nosso sítio
- AC/DC, CETEMPúblico, COMPARA ... Recursos consideráveis
- Morfolimpíadas A primeira avaliação conjunta para o português, seguida pelo CLEF e pelo HAREM
- Recursos públicos
- Investigação e colaboração
- Comparação e medição formal
- Uma língua, muitas culturas
- Cooperação usando a Web
- Não adaptar aplicações do inglês

www.linguateca.pt

3

## Linguateca – em mais detalhe

- Portal
  - catálogo de projectos, actores, recursos e ferramentas
  - fórum (notícias, bolsa de emprego e conferências)
  - ligações úteis (listas, informação técnica, revistas em português...)
  - um repositório de artigos ou ferramentas
- Um catálogo de publicações sobre a área e um sistema de gestão do mesmo (SUPeRB)
- Vários serviços ou sistemas na Web que dão acesso directo a recursos
  - Busca no AC/DC, COMPARA, Floresta
  - RAP no Esfinge
  - Corpógrafo: um ambiente para desenvolver terminologia e estudo de linguagens específicas
- Resposta a todas as perguntas que nos fazem!

www.linguateca.pt

4

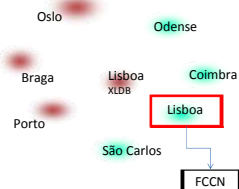
## Linguateca - um centro de recursos distribuído

- Projecto gerido pela FCCN, financiado pelo POSI, UMIC e MCTES

### Modelo IRA

- Informação
- Recursos
- Avaliação

www.linguateca.pt



www.linguateca.pt

5

## Linguateca - cooperação com o RCAAP

- Expressões de pesquisa inteligentes e com base no material existente
  - Não se encontra nenhuma entrada "Silva" no RCAAP mas sabemos que existe Silva no catálogo.
- Apresentar apenas um nome para cada autor:
  - Ao pesquisar-se um autor podem aparecer várias entradas diferentes para o mesmo autor.
    - Ex: pesquisar por "José Almeida" e aparecer apenas "José João Dias de Almeida".
      - Referência ao mesmo autor e não a autores diferentes

www.linguateca.pt

6

## Resposta inteligente ao utilizador

- expressões de pesquisa inteligentes e com base no material existente.

- Procurar por "José Amleida" em vez de "José Almeida"

The screenshot shows a search interface with a 'Pesquisar por' field containing 'José Almeida'. Below it are four filter categories: 'Tipo' (set to 'CD-ROM - Reconhecedor de Entidades'), 'Título', 'Autor', and 'Assunto'. The 'Autor' filter is currently selected.

Two error messages are shown side-by-side. Both start with a red exclamation mark icon and the text 'A pesquisa realizada não devolveu nenhum documento.' The left message includes a link to 'Sua pesquisa não encontrou resultados'. The right message includes a link to 'Verificar se o termo de pesquisa está correto'.

## Eliminação de entradas duplicadas de autores

- Apresentar apenas um nome para cada autor:

- Ex: Pesquisa por "José Almeida"

The screenshot shows a list of authors under the heading 'Autor'. The list includes names like 'Almeida, José Maria Fernando...', 'Almeida, José Domingos de (23)', 'Almeida, José B. (22)', 'Almeida, José, José César de (20)', 'Parafita, Manuel A. (14)', 'González-Pérez, José Manuel (7)', 'Almeida, José João (12)', 'Almeida, Augusto Henrique de... (11)', 'Almeida, J. J. (6)', and 'Almeida, José (4)'. Red boxes highlight several entries, indicating duplicates that should be removed.

## Cooperação – trabalho futuro

- Detecção automática de entradas repetidas (em diferentes repositórios)
- Possível aviso de auto-plágio (se duas entradas são demasiado parecidas)
- Obtenção de um índice de citações para o português
  - quem é mais citado em publicações de acesso aberto?
- Ajuda para desambiguação por temas:
  - José Silva - publica sobre medicina
  - José Silva - publica sobre filosofia
  - José Silva - publica sobre agronomia
- Citações cruzadas: que artigos citam este?
- Rede de citações: visualização por autores, e por artigos...

## Agradecimentos

- Ministério da Ciência, Tecnologia e Ensino Superior
- Agência para a Sociedade do Conhecimento, IP
- Fundação para a Computação Científica Nacional
- Programa Operacional Sociedade do Conhecimento

