

# Colaboração entre a Linguateca e o RCAAP: primeiros passos

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

# 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

# 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

# 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!

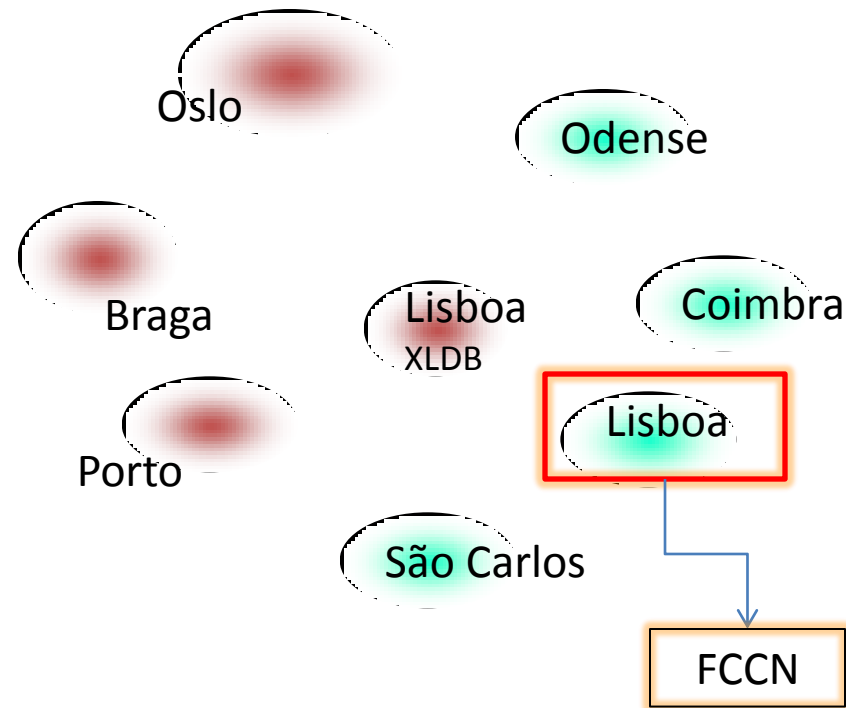
# Linguateca - um centro de recursos distribuído

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

## Modelo IRA

- Informação
- Recursos
- Avaliação

[www.linguateca.pt](http://www.linguateca.pt)



# Linguateca - cooperação com o RCAAP

- Expressões de pesquisa inteligentes e com base no material existente
  - Não se encontra nenhuma entrada “Amleida” no RCAAP mas sabemos que existe Almeida 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**

# Resposta inteligente ao utilizador

- expressões de pesquisa inteligentes e com base no material existente.
  - Procurar por “José Amleida” em vez de “José Almeida”

Tipo		Pesquisar por
	Título	RENA - Reconhecedor de Entidades
E	Autor	José Amleida
E	Assunto	

A pesquisa realizada não devolveu nenhum documento.



Tente melhorar a sua pesquisa.  
Verifique erros de linguagem.  
Use uma frase ou palavra menos específica.  
Use outras palavras com o mesmo significado.  
[Faça outra pesquisa](#)

A pesquisa realizada não devolveu nenhum documento.



Nenhum elemento no RCAAP contém o autor "José Amleida"  
Experimente [José Almeida](#) ... ou faça [nova pesquisa](#)

# Eliminação de entradas duplicadas de autores

- Apresentar apenas um nome para cada autor:
  - Ex: Pesquisa por “José Almeida”

The diagram illustrates the process of deduplicating author entries. It shows two panels of search results for 'Autor'.

**Left Panel (Initial Results):**

- Almeida, José Maria Fernand... (33)
- Almeida, José Domingues de (23)
- Almeida, José B. (20)
- Almeida Júnior, José Dias de (20)
- Parafita, Manuel A. (16)
- González-Méijome, José Manuel (15)
- Almeida, Tiago Augusto de, (13)
- Almeida, José João (12)
- Brandão, Augusto Henrique d... (11)
- Simões, Alberto (9)
- Almeida, J. J. (9)

**Right Panel (Deduplicated Results):**

- Almeida, José Maria Fernand... (33)
- Almeida, José João Dias de (25)
- Almeida, José Domingues de (23)
- Almeida, José B. (20)
- Almeida Júnior, José Dias de (20)
- Parafita, Manuel A. (16)
- González-Méijome, José Manuel (15)
- Almeida, Tiago Augusto de, (13)

An arrow points from the entry 'Almeida, José João (12)' in the left panel to the entry 'Almeida, José João Dias de (25)' in the right panel, indicating the deduplication process.



# 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

