

Alice no País das Maravilhas

PoloCLUP – Linguateca
Belinda Maia

Alice in Wonderland

- Alice wonders....
- Whether she is mad – or everyone else is...
- Yet she manages to ‘go with the flow’ and survive!

Is Linguateca Wonderland?

- Histórias para jovens???

- Lógica
- Linguagem
- Xadrez
- Competições
- Paródias
- Violencia...

Who's Who

in Liguatca / Wonderland?

- Alice?
- White Rabbit?
- Mad Hatter?
- Red Queen?
- Humpty Dumpty?
- Pack of cards?
-

Mais seriamente....

- A 'loucura' do PoloCLUP é tentar conciliar os interesses de PLN com a linguística mais tradicional, criando tecnologia 'user-friendly' e dirigida a projectos possíveis

PoloCLUP –Quando?

1. Novembro 2002 - Dezembro 2005
2. Janeiro 2006-Dezembro 2006
3. Janeiro 2007- Junho 2007
4. Junho 2007 – Julho 2008

Novembro 2002 - Dezembro 2005

- GE > Corpografo
- EXTEX
- METRA, BOOMERANG
- Avaliação de Tradução Automática
- EVAL > TrAva
- Entidades Nomeadas > REPENTINO e HAREM 2005
- SAGI - Sistema de Apoio à Geração de Interfaces

PoloCLUP - Como?

- Corpografo – um conjunto de ferramentas para
 - Criação e pesquisa de corpora pessoais
 - Criação de bases de dados de terminologia
 - Criação de bases de dados de lexicografia e fraseologia
 - Possibilidades para adaptação de critérios de análise dos dados ao pesquisador individual

Avaliação de Tradução Automática

- METRA
- Boomerangue
- Eval > TrAva

Janeiro 2007-Dezembro 2007

- Modificações ao METRA e Corpógrafo
- Trabalho de pesquisa por alunos

- Organização da Escola de Verão 2006
 - A 'loucura' do PoloCLUP juntou PLN com linguistas 'tradicionais'



Escola de Verão 2006 - Temas

- Questões actuais no processamento computacional da língua portuguesa
- Tradução automática: a tecnologia e a sua aplicação
- Análise sintáctica computacional: problemas e métodos
- Desenho, construção e utilização de corpora (e sua aplicação em ensino e extracção de conhecimento)
- Terminologia, tesouros e ontologias (e a sua aplicação em tradução, recolha e extracção de informação)
- Entrando em linha de conta com os utilizadores
- Avaliação

Janeiro 2007- Junho 2007


- Sem evolução nas ferramentas – responsável em sabática
- Pesquisa por alunos


Junho 2007 – Julho 2008

- Corpógrafo – $V_3 > V_4$
- Muitos melhoramentos e novas ferramentas

Corpógrafo V4

- Permite aos utilizadores:
 - Converter ficheiros nos formatos .pdf, .html, .ps, .doc, rtf, para formato de texto simples, .txt
 - Organizar os conteúdos de ficheiros em corpora
 - Pesquisar nos corpora utilizando expressões regulares
 - Obter concordâncias e ordená-las de acordo com o contexto à esquerda e à direita
 - Pesquisar nos corpora recorrendo a dicionários do NooJ para inglês, francês e português

- 
- Criar listas de N-gramas
 - Obter listas de candidatos a termos
 - Criar novas classificações para domínios específicos, relações semânticas e conectores de discurso
 - Alinhar textos em corpora paralelos e obter concordâncias paralelas
 - Visualizar as ocorrências de termos nos corpora
 - Criar e gerir bases de dados terminológicas multilingues

- 
- Criar e gerir bases de dados lexicais e de fraseologias multilingue
 - Importar e exportar bases de dados
 - Obter termos (ou itens lexicais) que co-ocorrem nos corpora para facilitar a obtenção de textos similares da internet
 - Associar objectos multimédia a termos
 - Identificar equivalentes de tradução em concordâncias em corpora comparáveis
 - Visualizar redes lexicais constantes numa base de dados

Resultados da 'loucura'

- Várias possibilidades para facilitar investigação
- Capítulos em livros, artigos e apresentações
- Varias dissertações de mestrado
- Doutoramentos em curso
- Muitos trabalhos pedagógicos
- e isso só na FLUP!
- Há muitos brasileiros inscritos no Corpógrafo
- METRA continua popular!

E agora?

- Alice acorda sempre para a realidade.....
- Será que pode sonhar outra vez?

LINGUATECA

- Como continuar o trabalho excelente feito :
 - Tomando em conta as mudanças na política educacional
 - Internacionais
 - Nacionais
 - Aproveitando as novas gerações e os seus conhecimentos da informática?
 - E mantendo um equilíbrio entre a realidade e o sonho?