

Pacote de Recursos do Port4NooJ

Este pacote pode ser descarregado de:

<http://www.linguateca.pt/Repositorio/Port4Nooj/>

e

<http://www.nooj4nlp.net/> - Resources - pt

Descrição dos Recursos

O Port4NooJ é um conjunto de recursos desenvolvido no ambiente linguístico **NooJ** para o processamento automático do português. Compreende um conjunto de dicionários e gramáticas, que podem ser utilizados em aplicações como a tradução automática do português para o inglês.

Fonte

O dicionário de cobertura larga resulta da conversão e revisão de entradas simples do dicionário de inglês-português do sistema de tradução automática **OpenLogos** (<http://logos-os.dfki.de/>). A conversão consistiu na substituição de cada representação numérica da **ontologia SAL** (<http://logosystemarchives.homestead.com/>) por mnemónicas SAL do Tutorial da Logos (**Logos Tutorial**) na aplicação web, **LearnLogos**. Estas mnemónicas nunca chegaram a ser implementadas no sistema da Logos. Algumas adaptações foram efectuadas, alguns conjuntos (sets) e subconjuntos (subsets) ainda não foram convertidos e a maior parte dos superconjuntos (supersets) dizem respeito ao inglês e não ao português.

Características novas

Para o módulo do português do NooJ, criámos paradigmas flexionais de raiz, limpámos entradas do dicionário da Logos que já estavam desactualizadas e acrescentámos novas propriedades, como a derivação (*DRV* - o nome *apresentação* deriva do verbo *apresentar*), anotações para nomes predicativos autónomos (*NPred*) e para nominalizações (*NPred+Nom*), entre outras.

Conteúdo do pacote

O pacote tem 3 pastas:

1. Lexical analysis/ - contém o dicionário de cobertura larga e uma amostra das regras flexionais e derivacionais
2. Syntactic analysis/ - contém algumas gramáticas

3. Projects/ - contém dois textos como amostra

Descrição dos ficheiros

1. *PT-Dict.dic* - um dicionário de cobertura larga com aproximadamente 40,000 entradas simples (lemas) que incluem: informação acerca do paradigma flexional; paradigma derivacional; informação sintáctico-semântica; verbo suporte; estrutura argumental e tradução em inglês, para cada entrada.
2. *PT-Morph.nof* - contém uma amostra de regras morfológicas de flexão e derivação. As regras flexionais formalizam paradigmas flexionais para as palavras simples. Por exemplo, *falamos* é a 1ª pessoa do plural do verbo *falar* no presente do indicativo. As regras derivacionais formalizam formas derivadas, como as nominalizações. Por exemplo, *citação* é uma nominalização derivada do verbo *citar*.
3. *PT-Contr.nom* - uma gramática morfológica para o processamento de contracções, como *das* = Prep (*de*) + Det (*as*) e *neste* = Prep (*em*) + DemPron (*este*).
4. *PT-Disamb-V-Pron.nog* - uma gramática sintáctica para anotar e desambiguar pronomes depois de verbos e antes de verbos que ocorrem a seguir ao pronome relativo *que*.
5. *PT-NE-PersonType.nog* - uma gramática para reconhecer e anotar entidades mencionadas do tipo PESSOA.
6. *PT2EN-Dates.nog* - uma gramática para reconhecer e traduzir para o inglês datas, tais como "*terça-feira, 27 de Fevereiro*".
7. *O texto da Declaração Universal dos Direitos Humanos*.
8. A obra *Viagens na Minha Terra* de Almeida Garret.

Licenças

Os ficheiros contidos no pacote de recursos do Port4NooJ são de acesso livre. A licença GPL da OpenLogos está também incluída no pacote.

Documentação

1. O tutorial introdutório do Port4NooJ
2. Port4NooJ Linguistic Resources Overview
3. Barreiro, Anabela, 2008. "Port4NooJ: Portuguese Linguistic Module and Bilingual Resources for Machine Translation". In *Proceedings of the 2007 International NooJ Conference* (Barcelona, Espanha, 7-9 Junho, 2007), Cambridge Scholars Publishing, 2008, pp. 19-47.

Como citar este pacote

Este pacote deve ser citado como "Pacote de recursos do Port4NooJ".

Autoria e agradecimentos

Este trabalho foi desenvolvido por Anabela Barreiro no âmbito de um programa de doutoramento financiado pela Fundação para a Ciência e Tecnologia, Programa POS_C através da bolsa SFRH/BD/14076/2003. Um agradecimento especial a Max Silberztein e Slim Mesfar pelo apoio técnico no NooJ, a Andrei Telenkov pela ajuda na extracção dos dados da OpenLogos e à LINGuateca pelo apoio no desenvolvimento, melhoria e disponibilização pública dos recursos.