

## LXService

### 1. BASIC INFORMATION

#### *1.1 Tool name*

LXService.

#### *1.2 Overview and purpose of the tool*

LXService is a Web Service that consists in a range of tools for Portuguese that have been developed for the processing of Portuguese. They were selected because they satisfy a number of features that are likely to make them more suitable for initial experimentation: They are fast, robust, the linguistic information in their output is well understood, and they perform at state of the art accuracy. They include the following individual tools, covering analysis and generation procedures:

**Constituency Parser** (see Silva et al., 2010): performs a syntactic analysis of Portuguese sentences in terms of their constituency structure. 88% f-score.

**Dependency Parser** (see Silva et al., 2010): allows the automatic parsing of sentences in Portuguese in terms of their grammatical functions. 86.8% accuracy.

**Named Entity Recognizer (number-based or name-based)** (see Ferreira et al., 2007): recognizes the expressions for named entities in Portuguese.

**POS tagger** (see Silva, 2007): assigns POS tags to tokens in context; 96.87% accuracy.

**Sentence chunker** (see Silva, 2007): detects and marks paragraph and sentence boundaries; 99.94% accuracy.

**Tokenizer** (see Silva, 2007): segments text into tokens, expands contractions, detaches clitic pronouns from verbs, etc.; 99.72% accuracy.

**Semantic Role Labeler**: generates syntactic trees with semantic labels; f-score of 82%.

The development of these web services started in the scope of the LT4eL-Language Technology for e-Learning project (available at [www.lt4el.eu](http://www.lt4el.eu)). They are supported by a range of language technology tools that have been developed in the past at the University of Lisbon in the scope of a number of previous projects.

#### *1.3 A short description of the algorithm*

Not applicable.

### 2. TECHNICAL INFORMATION

### *2.1 Software dependencies and system requirements*

It is an independent platform (Java).

### *2.2 Installation*

Upon request, an API will be delivered which permits to interact with the services (in Java) with a username and password.

### *2.3 Execution instructions*

Not applicable.

### *2.4 Input/Output data formats*

Not applicable.

### *2.5 Integration with external tools*

Not applicable.

## **3. CONTENT INFORMATION**

### *3.1 A test input file*

Raw text.

### *3.2 The output file*

It depends on the tool required.

**Constituency Parser:** S-expression

**Dependency Parser:** CoNLL 2007

**Named Entity Recognizer (number-based or name-based):** xml

**POS tagger:** raw text with POS tags

**Sentence chunker:** raw text with paragraphs and sentence marks

**Tokenizer:** tokenized raw text

**Semantic Role Labeler:** S-expression

### *3.3 Approximation of the time necessary to process the test input file.*

Not applicable.

## **4. ADMINISTRATIVE INFORMATION**

### *4.1 Contact person*

Name: António Branco

Address: Departamento de Informática NLX - Grupo de Fala e Linguagem Natural, Faculdade de Ciências da Universidade de Lisboa, Edifício C6,  
Campo Grande 1749-016 Lisboa

Position: Assistant professor

Affiliation: Faculty of Sciences, University of Lisbon

Telephone: +351 217 500 087

Fax: +351 217 500 084

E-mail: antonio.branco@di.fc.ul.pt

## 5. LICENSE

This tool is free for research purposes, with attribution and no redistribution or derivatives allowed. It will be available on the META-SHARE platform, only upon request to the authors.

## 6. RELEVANT REFERENCES AND OTHER INFORMATION

Branco, António, Francisco Costa, Pedro Martins, Filipe Nunes, João Silva, Sara Silveira, 2008, “LXService: Web Services of Language Technology for Portuguese”. In *Proceedings of the Sixth International Conference of Language Resources and Evaluation (LREC'08)*.

Ferreira, Eduardo, João Balsa and António Branco, 2007, “Combining Rule-based and Statistical Methods for Named Entity Recognition in Portuguese, TIL2007 - VWorkshop em tecnologia da Informação e da Linguagem Humana, Anais do XXVII Congresso da Sociedade Brasileira de Computação, pp.1615-1624.

Florabela Barreto, António Branco, Eduardo Ferreira, Amália Mendes, Maria Fernanda Bacelar do Nascimento, Filipe Nunes and João Silva, 2006. Open Resources and Tools for the Shallow Processing of Portuguese: The TagShare Project. In *Proceedings of the 5th International Conference on Language Resources and Evaluation (LREC'06)*.

Silva, João (2007). *Shallow Processing of Portuguese: From Sentence Chunking to Nominal Lemmatization*. MSc thesis, University of Lisbon. Published as Technical Report DI-FCUL-TR-07-16.

Silva, João, Rúben Reis, Patrícia Gonçalves, António Branco, 2010, “LX-Parser and LX-DepParser: Online Services for Constituency and Dependency Parsing”. In *Proceedings of the 9th International Conference on the Computational Processing of Portuguese (PROPOR'10)*, Porto Alegre.