

Lemmatizer for Portuguese

1. BASIC INFORMATION

1.1 Tool name

Lemmatizer for Portuguese

1.2 Overview and purpose of the tool

Based on the MXPOST part of speech tagger and UNITEX dictionaries for Portuguese, this tool produces the lemmas of the words of a text stored in a plain text file. The source code is also provided.

1.3 A short description of the algorithm

Based on the MXPOST part of speech tagger and UNITEX dictionaries for Portuguese, this tool produces the lemmas of the words of a text stored in a plain text file. The source code is also provided.

2. TECHNICAL INFORMATION

2.1 Software dependencies and system requirements

It is necessary java virtual machine.

2.2 Installation

Not applicable.

2.3 Execution instructions

Not applicable.

2.4 Input/Output data formats

Input

Media type: Text

Output

Media type: Text

2.5 Integration with external tools

Not applicable.

3. CONTENT INFORMATION

3.1 A test input file

Not applicable.

3.2 The output file

Not applicable.

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

It varies with the input size.

4. ADMINISTRATIVE INFORMATION

4.1 Contact person

Name: Thiago A. S. Pardo

Address: Avenida Trabalhador São-carlense, 400

– Centro 13566-590 São Carlos-SP Brazil

Affiliation: Instituto de Ciências Matemáticas e de Computação - USP

Position: Professor Doutor

Telephone: +55 (16) 3373-8170

Fax: +55 (16) 3373-9650

E-mail: taspardo@icmc.usp.br

5. LICENSE

This tool is free for research purposes under an [Apache License V2.0](http://www.apache.org/licenses/LICENSE-2.0) (available at <http://www.apache.org/licenses/LICENSE-2.0>). It is available on the META-SHARE platform.

6. RELEVANT REFERENCES AND OTHER INFORMATION

-