

WebJspell an online morphological analyser and spell checker

Rui Vilela

Universidade do Minho, Departamento de Informática
Campus de Gualtar 4710-057 Braga, Portugal
ruivilela@di.uminho.pt

Resumen: Webjspell es una herramienta multiusos para Internet destinada al análisis morfológico y corrección ortográfica de textos escritos en portugués. Además de estas funcionalidades provee: ejemplos de frases, tablas de conjugación verbal, sugerencia de palabras ante eventuales errores ortográficos y corrección ortográfica de páginas de Internet. En esta comunicación se describe las características de Webjspell y las posibles extensiones de sus técnicas a otras aplicaciones.

Palabras clave: corrección ortográfica, analizador morfológico.

Abstract: Webjspell is an Internet multipurpose tool for Portuguese morphological analysis and spell checking. It provides examples of phrases, frequencies, verbal conjugation tables, word suggestions, and Internet pages spell checking. This article describes Webjspell features, and results.

Keywords: spell checking, morphology analysis.

1 Introduction

People have compulsion for auto-evaluate and improve their written production. There is a wide range of available linguistics resources, paper or digital, helping all people to outshine their language knowledge.

All people, especially when they study foreign languages, have need for more online resources to leverage their language understanding, due to sparse and more expensive resources.

Webjspell was developed as solution for this problem, especially within the Portuguese language domain, making attainable a morphological analyser and a spell checker.

2 Webjspell

Webjspell was developed to spread the usage of the morphological analyser Jspell to a wider audience. Available online on <http://linguateca.di.uminho.pt/jspell>.

It was developed in collaboration of Natura Project¹ and Linguatca² to have a broader and more user-friendly interface. Development was made using Perl language and the available Jspell module. (Simões y Almeida, 2001)

Jspell and Portuguese dictionary were developed in 1994 by José João de Almeida

and Ulisses Pinto (Almeida y Pinto, 1995), based on Ispell spell checker for *UNIX* environment. Is an interactive command line application for analyzing mainly words in text files.

The Portuguese dictionary is currently used along other available open source applications, such as Firefox, Thunderbird, and OpenOffice. Along with diverse usage for different kinds of research projects.

Webjspell adds additional features, by using Jspell Perl interface. Beyond a new interactive interface, it uses public domain services and logging. On his foundations, it is divided in four services: morphological analysis, spell checking, Internet web pages spell checking, and word feedback or suggestion.

3 Morphological analyser

The morphological analyser, in figure 1, has a bigger notability than other available services. For each of the given words and languages, the program obtains a morphological and semantic classification.

Improvements were made over the original Jspell, such as: Verbose morphological classification; Inflected words stem from lemmas; Phrase examples from public corpora, Word frequencies; Suggestions; Feedback; Verb conjugation tables;

Further improvements are planned for extending some features, like external online

¹<http://natura.di.uminho.pt>

²<http://www.linguateca.pt>

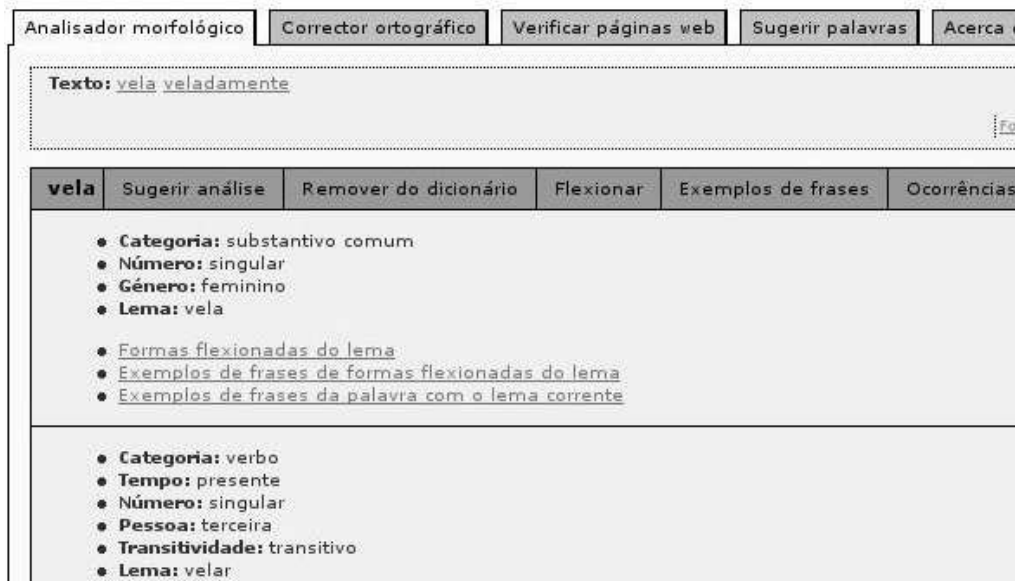


Figure 1: Morphological analysis

service usage, such as: language translation, word definition, and thesaurus capability.

3.1 Spell Checker

The spell checker aids the user to discover and fix misspelled words, with resource to word suggestion. Colours are used to mark errors, fixes, and also to identify foreign words.

Webjspell enhances some of the features of Jspell module, such has missing spaces, hyphens and in conversely way.

Further preferment can be implemented, like the use of patterns for common phonetical errors, better exploitation of Jspell morphological capabilities for finding simple grammatical errors, along with filtered suggestions, and duplicated word detection.

3.2 Web pages spell checker

It allows for a given Internet address, to search spelling mistakes, the program edits locally the page, and marks with colours the unknown and foreign words in other supported dictionaries.

3.3 Word suggestions

A interface that allows users to submit a wish list of words, that could be or not included in the dictionary.

4 Final considerations

The Webjspell results after some months on the wild, becomes worthy to analyze the ob-

tained feedback for self-improvement of diverse dictionaries.

Since the application was released, it has more than 2400 searches per month, along with an explosion of the number of word suggestions for the dictionaries, positively contributed for increasing the quality and precision of several dictionaries.

All words, especially the ones that Jspell is unable to identify are kept for later analysis. This method brings advantages in identifying typical user errors and new words. Assorted problems were mended, as much in features, as in interface, including the Perl interface and Jspell. Webjspell contributes to the dictionary development, on which depend several text processing applications.

Bibliografía

- Almeida, J.J. y Ulisses Pinto. 1995. Jspell – um módulo para análise léxica genérica de linguagem natural. En *Actas do X Encontro da Associação Portuguesa de Linguística*, páginas 1–15, Évora 1994.
- Simões, Alberto Manuel y José João Almeida. 2001. *jspell1.pm* – um módulo de análise morfológica para uso em processamento de linguagem natural. En *Actas da Associação Portuguesa de Linguística*, páginas 485–495.