

F-EXT-WS-2.0: A Web Service for Natural Language Processing

Eduardo N. Motta¹, Eraldo R. Fernandes¹, Ruy L. Milidiú¹

¹ Departamento de Informática, PUC-Rio, Rua Marquês de São Vicente 255, 22453-900,
Rio de Janeiro, Brazil
{emotta, efernandes, milidiu}@inf.puc-rio.br

Abstract. In this paper we present F-EXT-WS-2.0, a web service that provides Natural Language Processing capabilities to desktop and web client applications. The current version of the service provides Part-of-Speech Tagging, Noun Phrase Chunking, Named Entity Recognition, Chunking, Semantic Role Labeling and Clause Segmentation, for English and Portuguese texts. All these tools are built using Entropy Guided Transformation Learning, a state-of-the-art Machine Learning algorithm for such tasks. We also describe a demonstration setup for this service.

Keywords: Web service, natural language processing, machine learning, ETL.

1 Introduction

Natural language processing tasks such as Part-of-Speech Tagging (POS) are an important building block for applications that need to perform more complex text tasks, like extracting semantics from them.

F-EXT-WS-2.0 is a web service that provides Natural Language Processing capabilities and is built on top of the framework developed for F-EXT-WS, a web-based language processing service [1]. It allows desktop or web applications to submit online requests to obtain several linguistic features for input texts. These features can be used, for instance, by a semantic web application that mashes up company names obtained through named entity recognition (NER) and information on the company got from web sources, like president name, revenues etc.

The remainder of this paper is organized as follows. In section 2, we briefly describe the web service F-EXT-WS-2.0. In section 3, we show a demonstration setup.

2 NLP Web Service

All tasks exposed through F-EXT-WS-2.0 rely on the Entropy Guided Transformation Learning (ETL) algorithm. ETL is a machine learning strategy that combines Transformation Based Learning and Decision Trees [2].

2.1 System architecture

F-EXT-WS-2.0 is implemented using Simple Object Access Protocol (SOAP), an XML-based communication protocol and encoding format for inter-application communication. It allows platform independent applications to submit requests and get responses with texts annotated with linguistic features.

Architecture is designed to be flexible and expansible to accommodate more NLP tasks in the future.

2.2 NLP tasks

Currently, F-EXT-WS-2.0 provides five basic language processors for Portuguese and four for English. This set of processors corresponds to the training corpora that were available. The corresponding language processing tasks available in the F-EXT-WS-2.0 web service are shown in Table 1.

Table 1. Natural language processing tasks available per language.

Task	Portuguese	English
Part-of-Speech Tagging (POS)	•	•
Noun Phrase Chunking (NP)	•	
Named Entity Recognition (NER)	•	
Chunking (CK)	•	•
Semantic Role Labeling (SRL)		•
Clause Segmentation (CL)	•	•

3 Demonstration setup

The demonstration of the web service consists in presenting a client application submitting requests to the web service and displaying its corresponding results. F-EXT-WS-2.0 allows for multiple linguistic feature outputs, e.g., one can submit a text and get POS tags, NER tags and chunking using a single service call.

Client stubs to call the web service are available in Python, Java and C#. The demo application is a C# .NET application running under the Windows environment.

References

1. Fernandes, E. R., Milidiú, R. L., dos Santos, C.N.: Portuguese Language Processing Service. In: Proceedings of the Web in Ibero-America Alternate Track of the 19th International World Wide Web Conference. Madrid, April, 2009.
2. Milidiú, R. L., dos Santos, C. N., Duarte J. C.: Phrase chunking using entropy guided transformation learning. In: ACL-08: HLT, 2008, Columbus, Ohio, USA. 46th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies, 2008. p. 647-655.