

Quick Start on Collaborative Development of NLP tools with IXA Pipes

Rodrigo Agerri
IXA NLP Group
UPV/EHU
rodrigo.agerri@ehu.eus

Abstract

Looking at real NLP tools in a collaborative environment: IXA pipes

1 Introduction

Usar herramientas de PLN a través de las IXA pipes:
<http://ixa2.si.ehu.es/ixa-pipes>

IDEAS:

1. **Uso:** Bajar, compilar y usar las herramientas de IXA pipes: `ixa-pipe-tok`, `ixa-pipe-pos`, `ixa-pipe-nerc`, `ixa-pipe-chunk`, `ixa-pipe-parse`
 - (a) Baja cada uno de los módulos y compilar siguiendo las instrucciones de cada uno de los README.
 - (b) Compila cada módulo y aprende su funcionamiento con los textos de ejemplo proporcionados en la siguiente sección.
 - (c) Encadena las salidas de los diferentes módulos para obtener una visión general de la cadena de procesamiento.
 - (d) Para cada uno de los textos, detecta errores en cada nivel de anotación. Por ejemplo, compara las entidades nombradas automáticamente detectadas por `ixa-pipe-nerc` con las que tú anotarías.
2. **Desarrollo:** Usar un gestor de proyectos y un sistema de control de versiones para realizar, compartir y gestionar proyectos de software: por ejemplo, maven.apache.org y github.com:
 - (a) Crear un proyecto simple con Maven ponerlo en control de versiones de github e invitar a otro programador al repositorio.
 - (b) Fork un repositorio, realizar cambios y hacer un pull request.

3. **PLN:** Hacer un clone de los repos `ixa-pipe-tok` e `ixa-pipe-pos` e `ixa-pipe-nerc` y examinar su funcionamiento en detalle.
 - (a) Entrenar modelos para POS tagging, lematización y NERC usando los corpus públicos:
 - Universal Dependencies: <http://universaldependencies.org/>
 - CoNLL 2002: <http://www.cnts.ua.ac.be/conll2002/ner/>

2 Textos

- (1) The disappearance of York University chef Claudia Lawrence is now being treated as suspected murder, North Yorkshire Police said. However detectives said they had not found any proof that the 35-year-old, who went missing on 18 March, was dead. Her father Peter Lawrence made a direct appeal to his daughter to contact him five weeks after she disappeared. His plea came at a news conference held shortly after a 10,000 reward was offered to help find Miss Lawrence. Crimestoppers said the sum they were offering was significantly higher than usual because of public interest in the case.

(Example 1 of RTE-4)

- (2) American photojournalist James Nachtwey in a file photograph from May 18 2003 as he is awarded the Dan David prize in Tel Aviv for his outstanding contribution to photography. It was announced by Time magazine on Thursday, 11 December 2003 that Nachtwey was injured in Baghdad along with Time magazine senior correspondent Michael Weisskopf when a hand grenade was thrown into a Humvee they were traveling in with the US Army. Both journalists are reported in stable condition and are being evacuated to a US military hospital in Germany.

Caption 1470132 of the ImageCLEF-09 dataset.

- (3) Nothing special really. Comfortable and clean but very boring decor in comparison to other NH hotels. I stayed in NH in Brussels and Zurich and I really liked them because of their modern and stylish design and big rooms. This one was just like any other hotel. Basic rooms with basic and dull decor - bit disappointing. The customer service was average. The rate was very expensive and I still had to pay for Internet and 20 euros for breakfast!!! It was good but way overpriced! The best thing about the hotel was the location - city centre, 2min from a metro stop.

(review from OpeNER project)

- (4) La estrecha distancia entre Maduro y Capriles en los resultados electorales y la negativa de este último a reconocer el triunfo del chavismo en los

comicios hacen prever un escenario de tensión política y confrontación en los próximos días.

Esta misma noche se ha sentido un clima enrarecido en esta ciudad, Caracas, donde han volado rumores y conjeturas sobre el futuro más inmediato del país. Al contrario que en otras ocasiones, no ha habido un ambiente de gran celebración. En las filas del chavismo el triunfo de Maduro se ha sentido como una amarga victoria, más aún cuando Capriles ha rechazado los resultados oficiales.

Algunos grupos partidarios de Nicolás Maduro han salido a las calles, pero en sus rostros se percibía cierta decepción.

“Esto es una victoria amarga porque Chávez ya no está con nosotros y estamos por ello aún en pleno duelo y amarga porque los resultados no son los que esperábamos”, confiesa Marianela, una joven que ondea una bandera roja en el centro de Caracas.

“Pero como ha dicho Maduro, es momento de mejorar, de hacernos más fuertes, y que miremos los errores, las faltas y todo lo que queda por mejorar”, añade.

No muy lejos de allí varios vehículos tocan el cláxon y desde su interior salen voces de apoyo a Capriles. Cuatro peatones de mediana edad les saludan y sonríen. “La victoria es nuestra, no nos la van a robar”, gritan.

La negativa de Capriles a reconocer los resultados augura tensión y riesgo de inestabilidad política para las próximas semanas. El candidato de la oposición ha pedido el recuento del cien por cien de los votos, y ha dicho a Maduro que “el derrotado en el día de hoy es usted”.

- (5) Jeroen Dijsselbloem Eurotaldeko presidentearekin egin du bilera Varufakisek, eta haren aurrean nabarmendu du Greziako Gobernuak “egungo erreskate plana onartzen ez duen programa batengatik” aukeratu zutela herritarrek. Hala ere, azaldu du gobernuaren asmoa da Europako kidekin elkarlanean aritzea dela.

“Saiatuko gara gure kideak konbentzitzen irtenbide bat topatu behar dugula Europako interes komunak alde”, esan du Varufakisek.

Dijsselbloemek, berriz, gogorarazi du erreskate plana otsail amaierara arte luzatzekoa dela, eta gaineratu du: “Hori baino lehen, erabakiko dugu zer egingo den. Ez dago ondorioz oraindik”. Greziako zorrari buruzko konferentzia bat egitea proposatu dio Varufakisek; “konferentzia hori jada existitzen da eta Eurotaldea du izena”, erantzun dio.

Eurotaldeak Greziari laguntzeko konpromisoa duela adierazi du Dijsselbloemek, “Greziak bere hitza betetzen badu”. Azken urteotan “aurrerapen handiak” egin direla argudiatu du, “eta garrantzitsua da horiek ez galtzea”.