



UNIVERSIDADE DA CORUÑA

LENGUAJES NATURALES

CURSO 2012/2013

---

# My Little Trivial Player

Memoria de la Práctica

---

*Autores:*

Garabato Míguez, Daniel <daniel.garabato@udc.es>

López Beade, Vanesa <vanesa.lopezb@udc.es>

Valcarce Silva, Daniel <daniel.valcarce@udc.es> (Portavoz)

# Índice

<b>1. Arquitectura del sistema</b>	<b>2</b>
<b>2. Herramientas empleadas</b>	<b>3</b>
2.1. Toolkits . . . . .	3
2.2. Sistema de control de versiones . . . . .	3
<b>3. Manual de instalación y uso</b>	<b>4</b>
3.1. Instalación del sistema . . . . .	4
3.2. Uso del sistema . . . . .	4
<b>4. Diario de trabajo</b>	<b>5</b>
<b>5. Bibliografía</b>	<b>6</b>

## 1. Arquitectura del sistema

## **2. Herramientas empleadas**

### **2.1. Toolkits**

NLTK, Pattern.

### **2.2. Sistema de control de versiones**

Git y Bitbucket

### **3. Manual de instalación y uso**

#### **3.1. Instalación del sistema**

#### **3.2. Uso del sistema**

## 4. Diario de trabajo

A continuación se muestra un resumen del trabajo realizado a lo largo de la elaboración de la presente práctica. Las entradas se ordenan por su fecha cronológica y describen brevemente las decisiones y las acciones tomadas.

**2012/10/26**

En la primera reunión del grupo, hemos discutido sobre los primeros pasos a la hora de enfrentar la práctica. Tras un análisis de los diferentes *toolkits* que se nos presentan en el enunciado de la práctica, nos decidimos a usar NLTK por dos razones principales. El primer motivo es la gran cantidad de módulos que posee y su amplia documentación. En segundo lugar, porque Python nos parece un lenguaje muy cómodo para el desarrollo del proyecto.

Profundizando más en el desarrollo del trabajo, decidimos usar Git como sistema de control de versiones y apoyarnos en un repositorio privado de Bitbucket. Esto nos permitirá tener nuestro código bien organizado y documentado así como proporcionarnos un respaldo de los datos.

Por último, acordamos documentarnos más sobre el uso de Python en el procesamiento de lenguaje natural en general y con NLTK en particular. Para ello recurriremos a la bibliografía recomendada por los creadores del toolkit [1]. También optamos por estudiar las APIs de Google y de Bing para realizar consultas.

**2012/10/29**

Hemos instalado y configurado NLTK. Hemos estudiado utilizar el módulo Pattern (en Python) para la implementación de las consultas en los buscadores (Google y Bing). Hemos solicitado unas claves para poder utilizar las APIs de dichos buscadores.

Por otro lado, hemos comenzado a estudiar la formulación de la consulta (*query formulation*). Debemos eliminar aquellas palabras innecesarias (*stop-words*) por lo que utilizamos el diccionario de Porter ya integrado en NLTK. Consideramos mantener las comillas y los apóstrofes ya que dan mejores resultados en los buscadores. Optamos por no eliminar la interrogación final puesto que es irrelevante para ellos.

## 5. Bibliografía

### Referencias

- [1] S. Bird, E. Klein, E. Loper: *Natural Language Processing with Python — Analyzing Text with the Natural Language Toolkit*, O'Reilly Media (2009)