

Sesión de Prácticas 1

Introducción al Procesamiento del Lenguaje Natural

Para ver la situación real de la Ingeniería Lingüística en la actualidad, en esta primera sesión de prácticas vamos a probar diferentes herramientas y aplicaciones de PLN que se encuentran disponibles en la web.

Entra las direcciones que se indican a continuación, prueba la herramienta o la aplicación, y trata de responder a las preguntas que se formulan para cada una. Al final de la sesión de prácticas se hará una puesta en común de los análisis hechos.

1. Eliza. Uno de los primeros sistemas que intentó emular la conversación humana.

- Web: <http://www.manifestation.com/neurotoys/eliza.php3> (en inglés)
 - Por google podrás encontrar otras implementaciones de Eliza.
 - En español: <http://talk.to/psico>
- Prueba: Imagina que estás en la consulta del psicoanalista. Primero saludas, y luego inicia una conversación con alguna afirmación. Por ejemplo: “my sister says that I am silly”, y continúa la conversación.
- Cuestiones:
 - ¿Entiende algo ELIZA?
 - ¿Podrías determinar cómo funciona?
 - ¿Qué hace si no sabe qué le estás diciendo, es decir, si haces una intervención no esperada?
 - ¿Eres capaz de detectar alguna frase o patrón recurrente? ¿Cuál?
 - ¿Utiliza Eliza alguna fase de análisis lingüístico?
 - ¿Tiene alguna utilidad en la sociedad actual?
- Más información puedes consultar la wikipedia.

2. Resumen automático

- Web: <http://swesum.nada.kth.se/index-eng.html>
- Prueba: Os hemos facilitado dos noticias en formato texto (texto1.txt y texto2.txt). Primero léelos y luego cárgalos en la web.
- Cuestiones:
 - ¿Qué calidad tiene el resumen?
 - ¿Contiene la idea principal?
 - ¿Informa del contenido del texto original?
 - ¿Da datos superfluos?
 - ¿Podría pasar por la traducción de un humano?
 - ¿Qué errores tiene?
 - ¿Qué fases de análisis de PLN crees que utiliza y por qué?
 - ¿Qué utilidad puede tener actualmente?
- Más información: consulta los artículos disponibles en la web.

3. Sistema de Búsqueda de Respuestas:

- Web: <http://start.csail.mit.edu/>
- Prueba: Vamos a lanzarle preguntas al sistema a ver cuántas responde bien, cuántas falla y cuántas no es capaz de contestar. Puedes utilizar las preguntas del fichero “preguntas.txt”.
- Cuestiones:
 - ¿Qué tal funciona el sistema?
 - ¿Qué hace cuando no localiza la respuesta?
 - ¿Utiliza información lingüística? ¿Qué fases de PLN crees que utiliza?
 - ¿Qué utilidad puede tener actualmente?

4. Sistema de Recuperación de Información multilingüe.

- Web: <http://nlp.uned.es/wtb/> (buscar en *European Educational Repository*)
- Prueba: lanzar búsquedas en un idioma, a ver si realmente localiza documentos correcto en otro idioma. La consulta debe ser del ámbito educativo, pues los documentos son de este tema. Por ejemplo, puedes lanzar “cursos de francés”.
- Cuestiones:
 - ¿Localiza documentos en otro idioma correctos?
 - ¿Cómo crees que hace la traducción? ¿Qué tipo de recurso lingüístico utiliza?
 - ¿Qué utilidad tiene?
 - Comparado con Google, ¿qué ventajas tiene?

5. Analizador morfosintáctico:

- Web: <http://www.gedlc.ulpgc.es/>
- Prueba: En el fichero “oraciones.txt” tienes varias oraciones que puede lanzar al sistema.
- Cuestiones:
 - ¿Cómo crees que hace la desambiguación?
 - ¿Qué tal es? ¿Funciona bien?
 - ¿Cuántos errores detectas?
 - ¿Cómo se podría mejorar?
 - ¿Qué utilidades puede tener una herramienta como esta?

6. Léxico: WordNet

- Web: <http://nipadio.lsi.upc.edu/cgi-bin/wei4/public/wei.consult.perl>
- Prueba: WordNet es uno de los principales léxicos utilizados en PLN
- Cuestiones:
 - ¿Cuánta información aporta de cada palabra?
 - ¿Cómo relaciona los sentidos de una lengua y otra?
 - ¿Qué utilidad tiene?

7. Similitud semántica

- Web: <http://marimba.d.umn.edu/cgi-bin/similarity.cgi>

- Prueba: Entre otros usos, con WordNet se puede calcular la similitud semántica entre palabras según su posición en WordNet (cercanía o lejanía). Este programa hace el cálculo basado en varias fórmulas de similitud.
 - Calcula la similitud entre los verbos “walk” y “go”, y luego con nombres y adjetivos.
- Cuestiones:
 - ¿Cuánta tiene?
 - ¿Cambia mucho si se utilizan otras medidas de similitud?
 - ¿Cuál es la mejor?
 - ¿Qué utilidades puede tener saber la similitud semántica entre palabras?

8. Lectura de textos

Consulta los artículos que os hemos facilitado y contesta a las siguientes preguntas:

- ¿Qué fases tiene un sistema de PLN?
 - ¿Cuál es el objetivo central de la lingüística computacional o PLN?
 - Reconocimiento de habla ¿qué es? ¿qué problemas tiene y en qué línea de trabajos sigue actualmente?
 - Los sistemas de PLN deben tener cobertura y precisión. ¿Qué se entiende por estos conceptos?
 - Indica campos de investigación actuales: objetivos y problemas.
 - ¿En qué campos la TA funciona adecuadamente?
 - ¿Por qué se ha desarrollado la Recuperación de Información? ¿Utiliza actualmente técnicas lingüísticas? ¿Hacia dónde se dirige la recuperación de información?
- Puesta en común

5. Evaluación.

Se evaluará la realización de las pruebas y las intervenciones durante la puesta en común.