

Procesamiento del Lenguaje Natural

Práctica – Demostración de un etiquetador sintáctico basado en aprendizaje

1. Objetivo

El objetivo es poner en práctica conocimientos teóricos sobre el etiquetado sintáctico de textos.

2. Desarrollo de la práctica

Se pide efectuar una demostración de etiquetado sintáctico usando el paquete MXPOST. En particular, se ha de:

1. Descargar y formatear la colección de textos etiquetados sintácticamente usada para la CONLL Shared Task 2002, en castellano (Xavier Carreras, *train*).
2. Entrenar el etiquetador MXPOST con dichos datos.
3. Realizar una prueba sobre un fragmento breve de los propios datos de entrenamiento.
4. Presentar y discutir los resultados obtenidos.

3. Requisitos de la práctica

Se debe entregar una memoria **breve** de la práctica, que incluya la descripción detallada del proceso, y una discusión plausible de los resultados. La memoria será un archivo DOC o PDF. Los archivos se entregarán en un solo archivo comprimido ZIP terminado en OCX, llamado “grupoXXpractica03.zip.ocx”, adjunto a un mensaje de correo electrónico con la clave en el Asunto: “[PLN] Grupo XX – Práctica 03”. La práctica se entregará a lo sumo el día 8 de enero de 2007.

La evaluación de la práctica tendrá en cuenta:

1. El cumplimiento de los requisitos arriba mencionados.
2. La calidad de la discusión de los resultados.

Referencias

1. CONLL Shared Task 2002 - <http://www.cnts.ua.ac.be/conll2002/ner/>.
2. MXPOST - <ftp://ftp.cis.upenn.edu/pub/adwait/jmx/jmx.tar.gz>.