

Tarea del tema 4: Comparativa de etiquetadores estadísticos

En la siguiente página web:

<http://www-nlp.stanford.edu/links/statnlp.html>

en la sección "Part of Speech Taggers" puedes encontrar numerosos etiquetadores estadísticos. Muchos de ellos se basan en modelos distintos (HMMs, Support Vector Machine, etc.), utilizan distintos corpus de entrenamiento, sirven para distintos idiomas, etc.

En esta tarea debes comparar el comportamiento de al menos dos de ellos. Estúdialos, descríbelos (busca en la distribución y en la web detalles del modelo), y utilízalos para realizar el etiquetado de un pequeño texto (al menos 15 líneas), el mismo para ambos. Para ello asegurate que los etiquetados elegidos sirven para el mismo idioma. Debes elegir un texto en el que aparezcan palabras con más de una etiqueta léxica posible.

Después compara los resultados: etiquetas utilizadas por cada etiquetador y precisión del etiquetado. Para analizar la corrección puedes utilizar un texto de un corpus del que conozcas el etiquetado correcto. En otro caso tendrás que realizar el etiquetado correcto manualmente.

Documentación a entregar:

- Descripción de los etiquetadores seleccionados.
- Texto de prueba utilizado.
- Resultado del etiquetado con cada etiquetador seleccionado.
- Observaciones sobre la comparativa de los resultados.