

Assigment 2

Thamar Solorio , PhD - University of Houston - USA

Catedra Internacional de Ingeniería

Universidad Nacional de Colombia

Comparar exactitud de etiquetado POS de HMMS entrenados con diferentes corpora Corpora disponibles en NLTK:

- CONLL2000

- NPSCHAT

- BROWN

1. Separar para cada conjunto un conjunto de prueba y un conjunto de entrenamiento.
2. Entrenar modelos para corpora.
3. Evaluar cada modelo para cada conjunto de prueba (puede ser usando probabilidad o evaluando exactitud de predicción).
4. Analizar el resultado, en términos de robustes de los modelos a los cambios de corpora.
5. Otras opciones pueden ser analizar las curvas de aprendizaje
6. Entregar un minireporte con los resultados (1-2 paginas).
7. Fecha de entrega lunes 27 de Junio via dropbox Request . Entregar todos los archivos en un .zip, el archivo debe llamarse nlp-assign2-nombreapellido1-nombreapellido2.zip. Los grupos son de máximo 2 personas. El notebook se encuentra disponible en la página del curso.