

Introducción al Procesamiento de Lenguaje Natural
Diciembre de 2018

Consideraciones generales

- i) La prueba es sin material.
- ii) Escriba nombre y C.I. en todas las hojas.
- iii) Numere todas las hojas.
- iv) En la primera hoja, indique el total de hojas.
- v) Comience cada ejercicio en una hoja nueva.
- vi) Utilice las hojas de un solo lado.
- vii) Entregue los ejercicios en orden
- viii) El total de puntos es 40

Ejercicio 1 [12 puntos]

- a) ¿Qué se conoce por representación one-hot en el contexto de las representaciones vectoriales de palabras?
- b) Describa en términos generales en que consisten los métodos basados en predicción para la construcción de representaciones vectoriales de palabras.
- c) Indique en qué consiste la medida pointwise mutual information (PMI) y cómo puede utilizarse en los métodos basados en conteo.
- d) Considere el siguiente repertorio de oraciones:

El perro come hueso
El gato come pescado
El gato no come hueso
El perro come pescado
El gato no anda en bicicleta
Este perro anda en ómnibus
María come con Juan en el ómnibus
Juan anda en bicicleta
Juan anda en ómnibus
María anda en bicicleta
María anda en ómnibus

(i) Construya la matriz de co-ocurrencias palabra-palabra considerando como contexto de co-ocurrencia la oración completa. Considere para la matriz únicamente las palabras “come”, “anda”, “perro”, “gato”, “Juan”, “María”, “ómnibus” y “bicicleta”.

(ii) Grafique sobre los ejes correspondientes a “come” y “anda”, de acuerdo a las frecuencias de la matriz de la parte a), las palabras: “perro”, “gato”, “Juan”, “María”, “ómnibus” y “bicicleta”

Ejercicio 2 [6 puntos]

Indique si las siguientes proposiciones son Verdaderas o Falsas. Justifique en cada caso.

- a) La reducción β simplifica una fórmula del cálculo lambda aplicando los términos con abstracción funcional a constantes.
- b) El método SVM (Support Vector Machines) es un método secuencial que se utiliza para el etiquetado morfosintáctico.

- c) Naïve Bayes es un método de aprendizaje que genera clasificadores haciendo hipótesis de independencia, que no necesariamente se cumplen, entre los atributos de los ejemplos de entrenamiento.
- d) La medida *tf-idf* asociada a un término y un documento en recuperación de información, se define como la cantidad de ocurrencias de ese término en ese documento dividido la cantidad de documentos

Ejercicio 3 [10 puntos]

Considere una gramática G con las siguientes reglas:

- O → GV | GN GV
- GN → Det Nom | Nom | GN GP
- GV → V | V GN | V GN GP
- GP → Prep GN
- Det → el | la | los | las
- Nom → Carlos | caramelos | sandwiches | manzana | Salta
- V → como | come | salta
- Prep → con | de | en

- a) Aplique el algoritmo CKY para la entrada “*como sandwiches de manzana*” considerando las reglas de la gramática G. ¿Qué salida devuelve el algoritmo? Justifique.
- b) Escriba el árbol o los árboles de análisis sintáctico correspondiente(s) a la parte a).
- c) Realice las derivaciones de dos oraciones generadas por la gramática G que no contengan las mismas palabras y que no sean sintácticamente válidas en español. Explique porqué no lo son.

Ejercicio 4 [12 puntos]

Considere los siguientes ejemplos de grupos nominales:

- (i) *un gran jardín sombreado*
- (ii) *un jarrón rojo*
- (iii) *los verdes arbustos*

- a) Indicar la categoría gramatical de cada una de las palabras que aparecen en los 3 ejemplos anteriores.
- b) Escribir una gramática independiente de contexto que genere al menos los 3 ejemplos anteriores.
- c) Escribir las componentes de interpretación semántica asociadas a las reglas de la gramática de la parte 2.
- d) Realizar una derivación sintáctico-semántica para el ejemplo (i), efectuando las reducciones β que corresponda