

Introducción al Procesamiento de Lenguaje Natural Diciembre 2019

Consideraciones generales

- i) La prueba es sin material escrito.
- 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 60

Ejercicio 1 [8 puntos]

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

- a) Mediante el estudio de la morfología derivativa se puede deducir el género y el número de una palabra.
- b) Un tokenizador es un algoritmo que explora texto y devuelve cada palabra encontrada y su categoría gramatical.
- c) La gramática de dependencias consiste en un formalismo donde se le incorporan atributos a las reglas que indican los distintos rasgos de los constituyentes.
- d) El Penn Treebank es un corpus en el cual las oraciones han sido anotada en base a su estructura sintáctica

Ejercicio 2 [16 puntos]

Dada las siguientes reglas de una gramática libre de contexto

O → GN GV
GN → Nom | Det Nom | GN GP
GV → V | V GN | GV GP
GP → Prep GN
Det → el | la | los | las
Nom → Maria | Juan | comedor | estar | sala
Prep → del | del | en
V → sale | come | está | salir | estar | comer

- a) Aplique el algoritmo de Earley para la oración **"María está en el estar"** para verificar si puede ser analizada.
- b) Realice una derivación para otra oración que sea agramatical y explique por qué lo es.

Ejercicio 3 [16 puntos]

a) Describa brevemente (no más de diez líneas) una de las técnicas vistas en el curso para resolver el problema de desambiguación del sentido de las palabras (WSD - Word Sense Disambiguation en inglés).

b) Considere la siguiente gramática con anotaciones semánticas:

$o \rightarrow gn\ gv$	$o.sem = gn.sem(gv.sem)$
$gn \rightarrow det\ n$	$gn.sem = det.sem(n.sem)$
$gv \rightarrow v$	$gv.sem = v.sem$
$gv \rightarrow v\ gn$	$gv.sem = v.sem(gn.sem)$
$gv \rightarrow neg\ gv$	$gv.sem = neg.sem(gv.sem)$
$nprop \rightarrow Juan$	$nprop.sem = \lambda P . P(juan)$
$n \rightarrow perro$	$n.sem = \lambda x . perro(x)$
$n \rightarrow elefante$	$n.sem = \lambda x . elefante(x)$
$v \rightarrow ladra$	$v.sem = \lambda x . ladra(x)$
$v \rightarrow come$	$v.sem = \lambda P . \lambda y . P(\lambda x . come(y, x))$
$adj \rightarrow gris$	$adj.sem = \lambda P . \lambda x . P(x) \wedge gris(x)$
$neg \rightarrow no$	$neg.sem = \lambda P . \lambda x . \neg P(x)$
$det \rightarrow todo$	$det.sem = \lambda P . \lambda Q . \forall x P(x) \rightarrow Q(x)$
$det \rightarrow ningún$	$det.sem = \lambda P . \lambda Q . \forall x P(x) \rightarrow \neg Q(x)$
$det \rightarrow un$	$det.sem = \lambda P . \lambda Q . \exists x P(x) \wedge Q(x)$

Dibuje el árbol sintáctico y derive la expresión lógica asociada a la oración:

"Ningún elefante ladra"

Ejercicio 4 [20 puntos]

a) Dibuje el Triángulo de Vauquois, indicando qué representan los lados derecho e izquierdo del triángulo y ubicando dentro del triángulo los siguientes métodos de traducción automática: transferencia sintáctica, traducción directa, traducción basada en interlingua.

b) Mencione dos problemas del modelo booleano que son resueltos por el modelo vectorial. Explique brevemente cómo los resuelve.

c) Qué entiende por léxico afectivo y para qué tarea podría utilizarlo. ¿De qué forma?

- d) Explique en qué consiste el método GloVe para construir representaciones vectoriales de palabras.
- e) ¿En qué consiste el test de analogías para evaluar la calidad de los vectores de palabras?