

Introducción al Procesamiento de Lenguaje Natural Diciembre de 2013

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 40

Ejercicio 1 [3 puntos]

Mencione al menos tres cualidades que deberían encontrarse en un sistema de recuperación de información y explique brevemente en qué consisten.

Ejercicio 2 [5 puntos]

¿Para qué se utiliza la métrica BLEU? Escriba la fórmula y explique cómo se calcula dicha métrica.

Ejercicio 3 [11 puntos]

Considere la siguiente gramática con anotaciones semánticas:

$o \rightarrow gn\ gv$	$o.sem = gn.sem(gv.sem)$
$gn \rightarrow nprop$	$gn.sem = nprop.sem$
$gn \rightarrow det\ nom$	$gn.sem = det.sem(nom.sem)$
$nom \rightarrow n\ adj$	$nom.sem = adj.sem(n.sem)$
$nom \rightarrow n$	$nom.sem = n.sem$
$gv \rightarrow v$	$gv.sem = v.sem$
$nprop \rightarrow Carolina$	$nprop.sem = \lambda P . P(carolina)$
$n \rightarrow gato$	$n.sem = \lambda x . gato(x)$
$v \rightarrow salta$	$v.sem = \lambda x . salta(x)$
$v \rightarrow canta$	$v.sem = \lambda x . canta(x)$
$adj \rightarrow blanco$	$adj.sem = \lambda P . \lambda x . P(x) \wedge blanco(x)$
$det \rightarrow todo$	$det.sem = \lambda P . \lambda Q . \forall x P(x) \rightarrow Q(x)$
$det \rightarrow un$	$det.sem = \lambda P . \lambda Q . \exists x P(x) \wedge Q(x)$

Dibuje el árbol sintáctico y derive la expresión lógica asociada a las siguientes oraciones:

1. *Carolina canta.*
2. *Un gato blanco salta.*

Ejercicio 4 [12 puntos]

Considere la siguiente gramática:

O	\rightarrow	GN GV
GN	\rightarrow	Nom Det Nom Nom GP
GV	\rightarrow	V V GN GV GP
GP	\rightarrow	Prep GN
Det	\rightarrow	el la los las
Nom	\rightarrow	Juan cocina perro
V	\rightarrow	cocina cocinan come comen

Prep → en | a

- a) Aplique el algoritmo Earley a la oración *Juan cocina en la cocina*, e indique si la oración puede ser generada por la gramática (dibujando en tal caso el árbol sintáctico de la oración).
- b) La gramática G reconoce oraciones agramaticales en las cuales falla la concordancia en número o género. Dé dos ejemplos de estas oraciones. ¿Es posible reescribir G de modo que no las reconozca? ¿Qué ventajas tendría una gramática HPSG sobre la gramática libre de contexto G, en cuanto al chequeo de concordancia en número y género?

Ejercicio 5 [5 puntos]

Calcule la distancia de edición entre CALMA y CLAMAR (utilizando distancia de Levenshtein, es decir que todas las operaciones de edición valen 1), y muestre una lista de ediciones posible con esa distancia. Justifique.

Ejercicio 6 [4 puntos]

Se tiene un método de clasificación (que incluye un parámetro α que puede valer 1, 0.1 o 0), y se quiere utilizarlo para determinar si una oración dada es un insulto o no. Se cuenta con un corpus de 10.000 oraciones, anotada cada una de ellas con la clasificación esperada. Indique brevemente cuáles serían los pasos generales para construir un clasificador y evaluar su performance. **No** se pide que especifique cuáles serían los atributos utilizados.