

Introducción al Procesamiento de Lenguaje Natural

Diciembre de 2006

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.

Ejercicio 1 [6 puntos]

- a) Describa brevemente las dos hipótesis principales sobre la forma en que se realiza el procesamiento morfológico humano.
- b) Mencione dos ventajas de los transductores como herramienta para el desarrollo de analizadores morfológicos.

Ejercicio 2 [16 puntos]

Considere los **nombres** en el idioma ficticio Itio (ligeramente basado en el Esperanto), que se construyen según las siguientes reglas:

1. Comienzan con una raíz. Para este ejemplo consideraremos las siguientes: *hund, kat, bird, elephant* (en español: *perro, gato, pájaro, elefante*, respectivamente).
2. Una raíz por sí misma no es una palabra válida.
3. Todos los nombres tienen un sufijo *o*. Por lo tanto *elefanto* y *hundo* son palabras válidas.
4. Una palabra como *hundo* no tiene género marcado. Para marcarla como femenina, se agrega el sufijo *in* entre la raíz y el sufijo *o*. Por ejemplo *hundino* (perra) es una palabra válida.
5. Para marcar una palabra como diminutivo, se agrega un sufijo *et* entre la raíz y el sufijo *o*, por ejemplo *elefanteto* es una palabra válida. Para marcar un aumentativo, se agrega el sufijo *eg*.
6. Los sufijos femenino, diminutivo y aumentativo pueden co-ocurrir una cantidad arbitraria de veces.
7. Si una raíz termina en *t*, al agregarle el sufijo *in*, ésta se convierte en *v*, por cuestiones de pronunciación. Por lo tanto *gata* se escribe *kavino*.

- a) Describa brevemente los tres componentes principales de un analizador morfológico. Describa cada componente para un hipotético analizador morfológico para los nombres en Itio.
- b) Construya una expresión regular utilizando el álgebra de Xerox, que relacione un nombre en Itio con su análisis morfológico. Utilice las siguientes marcas léxicas para la salida: *+Noun* (Raíz), *+Aug*(Aumentativo), *+Dim*(Diminutivo), *+Fem*(Femenino). Por ejemplo, a la forma léxica *elefant+Noun+Fem* le corresponderá la palabra *elefanvino*.

Ejercicio 3 [36 puntos]

- a) Dé una definición de constituyente sintáctico.
b) Sea el conjunto de oraciones siguiente:

- Festejemos
- Festejemos en el tren
- El ángel exterminador es una película de Buñuel
- Los suinos gruñen
- El tren de Roma a Viena partió tardíamente
- El sobre está sobre la mesa de luz

- i) Modele este micro universo textual utilizando GLC.
ii) Indique al menos dos problemas que surgen al utilizar GLC para modelar lenguaje. Ejemplifique con la GLC construida en la parte anterior.
c) Sea la GLC siguiente:

Gramática		Léxico	
O	→ GV GN GV	N	→ vino dueño casa
GV	→ V GN V GN GP	Det	→ el la los
GN	→ Det N Det N GP	V	→ vino dice
GP	→ Prep GN	Prep	→ de a

- i) Aplique el algoritmo de Earley a la siguiente entrada: *vino el dueño de la casa*
ii) Escriba el árbol o los árboles de análisis sintáctico correspondiente(s) al punto anterior.

Ejercicio 4 [6 puntos]

- a) Indique 2 problemas que aparecen al representar en Lógica de Primer Orden enunciados en lenguaje natural.
b) Describa brevemente la aplicación de los métodos de aprendizaje automático supervisado y no supervisado para WSD. Mencione ventajas y desventajas de cada método.

Ejercicio 5 [30 puntos]

Sea la siguiente gramática aumentada con anotaciones semánticas:

o	→ gn gv	[o.sem = gv.sem (gn.sem)]
gn	→ det nominal	[gn.sem = < det.sem x nominal.sem(x) >]
gn	→ npropio	[gn.sem = npropio.sem]
nominal	→ n	[nominal.sem = λx isa(x,n.sem)]
nominal	→ nominal adj	
nominal	→ nominal pp	
gv	→ v	[gv.sem = v.sem]
gv	→ v gn	[gv.sem = v.sem (gn.sem)]
pp	→ prep gn	
v	→ lee	[v.sem = $\lambda x \lambda y \exists e$ leer (e) \wedge lee(e,y) \wedge leído(e,x)]
prep	→ de	[prep.sem = $\lambda x \lambda y$ de(y,x)]
prep	→ en	[prep.sem = $\lambda x \lambda y$ en(y,x)]

- a) Complete la gramática anterior con las 3 anotaciones semánticas que faltan. Estas anotaciones deben cubrir los casos en los que los adjetivos tienen una semántica interseccional y las preposiciones una interpretación relacional (tal como aparece en las reglas para “de” y “en”).
- b) Utilizando las reglas anteriores (completadas con las entradas léxicas necesarias) realice las derivaciones para las representaciones semánticas de :
- i. *Juan lee un libro interesante de Sartre.*
 - ii. *Juan lee un libro de un autor joven.*
- c) Exprese en lógica de primer orden las expresiones en *quasi logical form* obtenidas en el paso anterior.
- d) Discuta una posible extensión de la gramática anterior para que cubra oraciones como “*Juan lee un libro en el parque*”.

Ejercicio 6 [6 puntos]

Defina brevemente las medidas *Precision*, *Recall*, *E* y *F*. Explique para que se usan. La medida *E* lleva un parámetro que tiene un cometido especial. ¿Cuál es y para qué se utiliza?