

Introducción al Procesamiento de Lenguaje Natural

Diciembre de 2016

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 [8 puntos]

- i) ¿Qué es un corpus paralelo? ¿Para qué se utiliza?
- ii) Indique cuáles son los tres niveles de alineación que puede tener un corpus paralelo. ¿Cuál de estos niveles de alineación es más difícil de conseguir?
- iii) Dibuje las alineaciones a nivel de palabra entre los siguientes pares de oraciones. ¿Qué problema presenta el uso de funciones de alineación para representar las alineaciones de estas oraciones?
 1. a) *La casa es muy pequeña*
 b) *The house is tiny*
 2. a) *La casa es enorme*
 b) *The house is very big*

Ejercicio 2 [10 puntos]

- i) Describa brevemente (no más de diez líneas) **una** de las medidas de similitud entre pares de palabras vistas en el curso.
- ii) Considere la siguiente gramática con anotaciones semánticas:

o → gn gv	$o.sem = gn.sem(gv.sem)$
gn → nprop	$gn.sem = nprop.sem$
gn → det n	$gn.sem = det.sem(n.sem)$
gv → v	$gv.sem = v.sem$
gv → v gn	$gv.sem = v.sem(gn.sem)$
gv → neg gv	$gv.sem = neg.sem(gv.sem)$
nprop → Juan	$nprop.sem = \lambda P . P(juan)$
n → perro	$n.sem = \lambda x . perro(x)$
n → elefante	$n.sem = \lambda x . elefante(x)$
v → ladra	$v.sem = \lambda x . ladra(x)$
v → come	$v.sem = \lambda P . \lambda y . P(\lambda x . come(y, x))$
adj → gris	$adj.sem = \lambda P . \lambda x . P(x) \wedge gris(x)$
neg → no	$neg.sem = \lambda P . \lambda x . \neg P(x)$
det → todo	$det.sem = \lambda P . \lambda Q . \forall x P(x) \rightarrow Q(x)$
det → 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: “*Un perro no ladra*”.

Ejercicio 3 [14 puntos]

Considere la siguiente gramática, similar a la vista en el curso:

O	→	GV GN GV
GN	→	Det Nom Nom GN GP
GV	→	V V GN V GN GP
GP	→	Prep GN
Det	→	una un el la
Nom	→	Juan sopa leche perro lente lentes sal tomo
V	→	tomo toma sal
Prep	→	con sin

1. Aplique el algoritmo CKY para la entrada “*un perro con lentes toma sopa sin sal*”, usando la gramática G. ¿Qué salida devuelve el algoritmo?
2. Dibuje los árboles sintácticos posibles para la oración.
3. ¿Es posible obtener, a partir de la aplicación del algoritmo, todas las subsecuencias de palabras de la entrada que tienen estructura de oración? En caso afirmativo, lístelas.
4. Mencione una oración generable por la gramática que no sea sintácticamente válida en el idioma español. ¿Cómo se podría modificar la gramática o el formalismo para enfrentar estos casos?

Ejercicio 4 [4 puntos]

Los adjetivos en el idioma ficticio Itio están compuestos por una raíz (por ejemplo: “xax”, que quiere decir “pálido”, y “ses”, que quiere decir “ligeramente desviado hacia la derecha”), un sufijo opcional “a” (indicando género femenino), y otro sufijo (obligatorio), que vale “x” si el adjetivo es usado en forma positiva, y “v” si es usado en forma negativa. Por cuestiones ortográficas, en Itio no está permitida la duplicación de la letra “x”. Algunos ejemplos de palabras y sus correspondientes análisis:

xaxax	xax+Fem+Pos
xax	xax+Masc+Pos
xaxv	xax+Masc+Neg
sesav	ses+Fem+Neg

Implemente un analizador para los adjetivos de Itio utilizando el álgebra de expresiones regulares vista en clase.

Ejercicio 5 [4 puntos]

Considere un corpus de entrenamiento conformado por los siguientes documentos:

Texto	Clase
Un gran juego	Deporte
Un juego limpio pero olvidable	Deporte
La elección finalizó	No deporte
Muy limpio partido	Deporte
Fue una elección limpia	No deporte

Utilice un clasificador Naïve Bayes con un modelo Bag of Words para estimar la clase del documento cuyo texto es: “el juego finalizó temprano”. Justifique sus resultados.