

Introducción al Procesamiento de Lenguaje Natural Diciembre de 2014

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 [9 puntos]

- i) ¿Qué es un corpus paralelo? ¿Para qué se utiliza?
- ii) Defina cuáles son los niveles de alineación que puede tener un corpus paralelo. ¿Cuál de estos niveles de alineación es más difícil de conseguir?
- iii) Explique brevemente en que consiste el Modelo Vectorial propuesto por Salton y mencione al menos 3 problemas que a su juicio pueden plantear inconvenientes a la hora de aplicarlo con una base de documentos
- iv) ¿Qué es la selección de modelo en el contexto de un método de clasificación? Mencione y explique un método utilizado para realizarlo.

Ejercicio 2 [12 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:

$o \rightarrow gn\ gv$	$o.sem = gn.sem(gv.sem)$
$gn \rightarrow nprop$	$gn.sem = nprop.sem$
$gn \rightarrow det\ nom$	$gn.sem = det.sem(nom.sem)$
$nom \rightarrow n\ adj$	$nom.sem = adj.sem(n.sem)$
$nom \rightarrow n$	$nom.sem = n.sem$
$gv \rightarrow v$	$gv.sem = v.sem$
$nprop \rightarrow Juana$	$nprop.sem = \lambda P . P(juana)$
$nprop \rightarrow Carlos$	$nprop.sem = \lambda P . P(carlos)$
$n \rightarrow niño$	$n.sem = \lambda x . niño(x)$
$v \rightarrow salta$	$v.sem = \lambda x . salta(x)$
$v \rightarrow llora$	$v.sem = \lambda x . llora(x)$
$adj \rightarrow pequeño$	$adj.sem = \lambda P . \lambda x . P(x) \wedge pequeño(x)$
$det \rightarrow todo$	$det.sem = \lambda P . \lambda Q . \forall x P(x) \rightarrow Q(x)$
$det \rightarrow un$	$det.sem = \lambda P . \lambda Q . \exists x P(x) \wedge Q(x)$

Dibuje el árbol sintáctico y derive la expresión lógica asociada a la oración: “*Todo niño pequeño llora*”.

Ejercicio 3 [12 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	→	el la los las
Nom	→	Juan café leche perro tomo
V	→	tomo corro
Prep	→	con

1. Aplique el algoritmo CKY para la entrada “tomo café con leche”, usando la gramática G.
2. Dibuje los árboles sintácticos posibles para la oración. ¿Es posible obtener esta información del resultado del algoritmo? Justifique.
3. ¿Cuál de los dos árboles le parece más adecuado para el ejemplo? ¿Cómo podría mejorarse el algoritmo de análisis para seleccionar el mejor?

Ejercicio 5 [7 puntos]

Considere los adjetivos en el idioma ficticio Itio:

- Los adjetivos tienen una raíz (el morfema principal), diferentes sufijos, y un posible prefijo. Para este ejercicio, consideraremos solamente las raíces: *kapik* (profundo), *diak* (alto), y *sipia* (ligeramente inclinado).
- Uno de los sufijos (opcional) indica el género: *-x* (masculino), *-o* (neutro). Los adjetivos son femeninos cuando no existen sufijos, y todos los adjetivos admiten los tres géneros.
- Otro sufijo, que lo sigue, que, si está presente, indica plural: *-s* (plural)
- Para indicar la certeza del adjetivo se lo precede de un prefijo opcional: *yii-* (con seguridad), *moo-* (increíble), *pro-* (probablemente)
- Por cuestiones ortográficas, el sufijo correspondiente a plural se convierte en *-es* cuando sigue a una consonante.

a) Describa brevemente los tres componentes principales de un analizador morfológico. Describa cada componente para un hipotético analizador morfológico para los adjetivos en Itio.

b) Construya una expresión regular utilizando el álgebra de Xerox, que relacione un adjetivo con su análisis morfológico. Utilice las siguientes marcas léxicas para la salida: *+Adj* (marca de sustantivo), *+Masc* (masculino), *+Fem* (femenino), *+Neutro* (neutro), *+Singular*, *+Plural*, *+Seguro*, *+Incr*, *+Prob+*. Suponga que las únicas raíces son las mencionadas, pero diseñe la solución de modo que pueda ampliarse fácilmente.

Ejemplos de adjetivos en Itio:

kapikx (profundo) → *kapik+Adj+Masc+Singular*
yiiikapikx (seguramente profundo) → *+Seguro+kapik+Adj+Masc+Singular*
prokapikx (probablemente profundo) → *+Prob+kapik+Adj+Masc+Singular*
moodiak (increíblemente altas) → *+Incr+diak+Adj+Fem+Plural+Incr*
sipias (ligeramente inclinadas) → *sipia+Adj+Fem+Plural*