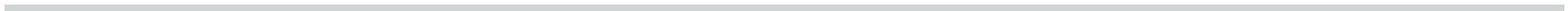


Introducción al Procesamiento de Lenguaje Natural

Grupo de PLN - InCo

Distancia de Mínima Edición



Distancia de Mínima Edición

- Encontrar una noción de distancia entre palabras
 - Encontrar la palabra más "próxima"
 - Por ejemplo, para autocorrección
 - Francia/Croacia
 - cena/pena/cana/ana
 - inelefantemente/indefectiblemente
 - Entre todas las posibles palabras del diccionario, sugerir la más cercana
-

Distancia de Mínima Edición

- DME (Minimum Edit Distance):
mínimo número de operaciones de edición (inserción, borrado, sustitución) necesarias para transformar un string en otro

 - Método: Programación dinámica
-

Distancia de Mínima Edición

Ejemplo:

* * I T A * * L I A

F R * * A N C * I A

i i d d i i d

1 1 1 1 1 1 1

DME = 7

Distancia de Mínima Edición

- También se puede utilizar en la evaluación de casos de evaluación en la comparación de posibles secuencias de palabras
- Ejemplo:

El viaje * en avión salió * * demorado

El viaje de * avión salió con retraso *

i d i i d

$$DME = 5$$

Distancia de Mínima Edición

- En el método original, cada operación de edición tiene costo 1
 - En 1966 Levenshtein propone operaciones con distinto costo
 - inserción = 1
 - borrado = 1
 - sustitución = 2
 - Es importante para obtener ese mínimo, que las palabras estén lo más “alineadas” posible
-

Distancia de Mínima Edición

Ejemplos:

C R O A C I A

F R A N C I A

S S S DME=6

I T A L * I A

F R A N C I A

F R A N C I A

I T A L * I A

S S s i DME = 7

S S s b DME = 7

T E M A

T O D A

S S DME = 4

Distancia de Mínima Edición

- Sean $X \rightarrow$ palabra₁ de largo n
 $Y \rightarrow$ palabra₂ de largo m
 - Se define una matriz \mathbf{d} ($n+1 \times m+1$)
 - Se inicializan la fila y columna 0
 - $d(0,0) = 0$
 - para i de $1 \dots n$ $d(i,0) = i$
 - para j de $1 \dots m$ $d(0,j) = j$
-

Distancia de Mínima Edición

Algoritmo:

para cada i de 1 a n

para cada j de 1 a m


$$d(i,j) = \min \begin{cases} d(i-1, j) + 1 \\ d(i, j-1) + 1 \\ d(i-1, j-1) + \begin{cases} 0 & \text{si } X(i) = Y(j) \\ 1 & \text{si } X(i) \neq Y(j) \end{cases} \end{cases}$$

Distancia de Mínima Edición


Ejemplo: calcular la DME entre TEMA y TODA

	#	T	E	M	A
#	0	1	2	3	4
T	1	0	1	2	3
O	2				
D	3				
A	4				

operaciones de
inserción



operaciones de
borrado



Distancia de Mínima Edición

Ejemplo: calcular la DME entre TEMA y TODA

	#	T	E	M	A
#	0	1	2	3	4
T	1	0	1	2	3
O	2	1	2	3	4
D	3	2	3		
A	4				

Distancia de Mínima Edición

Ejemplo: calcular la DME entre TEMA y TODA

	#	T	E	M	A
#	0	1	2	3	4
T	1	0	1	2	3
O	2	1	2	3	4
D	3	2	3	4	5
A	4	3	4	5	4

→ DME(x,y)

Distancia de Mínima Edición

Ejemplo: calcular la DME entre TEMA y TODA

	#	T	E	M	A
#	0	1	2	3	4
T	1	0	1	2	3
O	2	1	2	3	4
D	3	2	3	4	5
A	4	3	4	5	4

→ DME(x,y)

Referencias

J.Martin & D.Jurafsky. Speech and Language Processing.
- Tercera Edición - Capítulo 2
