



## **Proyecto 5: Generación y evaluación de oraciones mediante modelos de lenguaje N-grama y medición de perplejidad**



01

**Corpus**

[infopankki.en-es.es](http://infopankki.en-es.es)

# Contenido del corpus

El contenido no está disponible en el idioma elegido.

Por favor seleccione otro idioma.

finés

El contenido no está disponible en el idioma elegido.

Por favor seleccione otro idioma.

finés

El contenido no está disponible en el idioma elegido.

Por favor seleccione otro idioma.

finés

El contenido no está disponible en el idioma elegido.

Por favor seleccione otro idioma.

finés \_ sueco

El contenido no está disponible en el idioma elegido.

Por favor seleccione otro idioma.

finés \_ sueco

El contenido no está disponible en el idioma elegido.

Por favor seleccione otro idioma.

finés \_ sueco

El contenido no está disponible en el idioma elegido.

Por favor seleccione otro idioma.

finés \_ sueco

El contenido no está disponible en el idioma elegido.

Por favor seleccione otro idioma.

finés \_ sueco

Esta asociación realiza labores sociales, de formación y de información en Finlandia.

Ayuda al Refugiado asiste a los refugiados y a los inmigrantes en materias como integración, vivienda o creación de organizaciones propias.

Apoyo a los refugiadosfinés \_ sueco \_ inglés

Asistencia jurídica para los refugiadosfinés \_ sueco \_ inglés

Hay muchos países que mantienen algún tipo de representación diplomática en Finlandia.

Puede tratarse de embajadas o de consulados.

Las embajadas se encuentran en Helsinki, mientras que algunos países mantienen además consulados en otras ciudades.

Póngase en contacto con los representantes diplomáticos de su país cuando tenga que resolver asuntos relacionados con las autoridades de su país.

Es posible que a través de la embajada o el consulado pueda resolver asuntos como los que se mencionan a continuación:

Obtener un pasaporte nuevo.

Obtener un certificado de nacimiento.

Solicitar la nacionalidad de su país de origen, en caso de que la haya perdido o haya renunciado a ella.

Declarar el nacimiento de su hijo ante las autoridades de su país de origen, en caso de que su hijo haya nacido en Finlandia.

Solicitar la nacionalidad de su país de origen para su hijo, en caso de que su hijo haya nacido en Finlandia.

Registrar su matrimonio a efectos de su validez en su país de origen, en caso de que lo haya contraído en Finlandia.

Votar en las elecciones de su país de origen.

Tenga en cuenta que no todas las misiones diplomáticas extranjeras ofrecen los mismos servicios.

No todos los asuntos mencionados más arriba podrán ser tratados forzosamente en todas las legaciones.

Los servicios ofrecidos por la representación diplomática de su país de origen dependerán de las respectivas leyes nacionales.

Le darán más información en la embajada o el consulado de su país.

No todos los países mantienen representación diplomática en Finlandia.

En ese caso, frecuentemente le atenderá una legación de su país de origen situada en algún país vecino de Finlandia.





02

# Interpolación

Probabilidad no 0

# Interpolación

$$\hat{P}(w_n|w_{n-2}w_{n-1}) = \lambda_1 P(w_n) + \lambda_2 P(w_n|w_{n-1}) + \lambda_3 P(w_n|w_{n-2}w_{n-1})$$

"el perro corre rápido"

"el gato salta alto"

"el perro come carne"

$$P^{\wedge}(\text{carne} | \text{gato, come}) = \lambda_1 P(\text{carne}) + \lambda_2 P(\text{carne} | \text{come}) + \lambda_3 P(\text{carne} | \text{gato, come})$$

$$P^{\wedge}(\text{carne} | \text{gato, come}) = 0.3 \cdot 0.0833 + 0.3 \cdot 1 + 0.4 \cdot 0$$

$$P^{\wedge}(\text{carne} | \text{gato, come}) = 0.025 + 0.3 + 0 = 0.325$$

$$P(\text{gato}) = 1/12 = 0.0833$$

$$P(\text{come}) = 1/12 = 0.0833$$

$$P(\text{carne}) = 1/12 = 0.0833$$

$$P(\text{come} | \text{gato}) = 0$$

$$P(\text{carne} | \text{come}) = 1/1 = 1.0$$

$$P(\text{carne} | \text{gato, come}) = 0$$

'gato come carne'



03

# Backoff

Probabilidad no 0

# Backoff

$$S(w_i | w_{i-N+1:i-1}) = \begin{cases} \frac{C(w_{i-N+1:i})}{C(w_{i-N+1:i-1})}, & \text{if } C(w_{i-N+1:i}) > 0 \\ \lambda S(w_i | w_{i-N+2:i-1}), & \text{de lo contrario.} \end{cases}$$

"el perro corre rápido"

"el gato salta alto"

"el perro come carne"

$P(\text{gato}) = 1/12 = 0.0833$

$P(\text{come}) = 1/12 = 0.0833$

$P(\text{carne}) = 1/12 = 0.0833$

$P(\text{come} | \text{gato}) = 0$

$P(\text{carne} | \text{come}) = 1/1 = 1.0$

$P(\text{carne} | \text{gato}, \text{come}) = 0$

'gato come carne'

$S(\text{carne} | \text{gato}, \text{come}) = 0.4 \times P(\text{carne} | \text{come}) = 0.4 \times 1.0 = 0.4$





04

# Perplejidad y MLE

Capacidad de predicción y  
probabilidad condicional



# Perplejidad Y MLE

$$\text{perplexity}(W) = \sqrt[N]{\prod_{i=1}^N \frac{1}{P(w_i | w_1 \dots w_{i-1})}}$$

```
corpus = [  
    ['anita', 'lava', 'la', 'tina'],  
    ['tina', 'es', 'verde', 'grande'],  
    ['casa', 'anita', 'es', 'grande'],  
    ['gato', 'de', 'anita', 'camina', 'en', 'casa'],  
    ['ginebra', 'es', 'un', 'pais', 'grande']  
]
```

$P(\text{lava}|\text{anita}) = \text{Frecuencia de (anita,lava)} / \text{Frecuencia de anita}$

$P(\text{es} | \text{tina}) * P(\text{verde} | \text{es}) * P(\text{grande} | \text{verde}) * P(\text{anita} | \text{casa}) * P(\text{es} | \text{anita}) * P(\text{grande} | \text{es})$   
 $* P(\text{de} | \text{gato}) * P(\text{anita} | \text{de}) * P(\text{camina} | \text{anita}) * P(\text{en} | \text{camina}) * P(\text{casa} | \text{en}) * P(\text{es} |$   
 $\text{ginebra}) * P(\text{un} | \text{es}) * P(\text{pais} | \text{un}) * P(\text{grande} | \text{pais}) \approx 0.0041$

Perplejidad total =  $(1 / 0.0041) ^ (1/23) \approx 1.27$

Bigrama	Probabilidad
(anita, lava)	0.3333
(lava, la)	1.0
(la, tina)	1.0
(tina, es)	1.0
(es, verde)	0.3333
(verde, grande)	1.0
(casa, anita)	1.0
(anita, es)	0.3333
(es, grande)	0.3333
(gato, de)	1.0
(de, anita)	1.0
(anita, camina)	1.0
(camina, en)	1.0
(en, casa)	1.0
(ginebra, es)	1.0
(es, un)	0.3333
(un, pais)	1.0
(pais, grande)	1.0



# 05 | CÓDIGO :)