

Análise computadorizada de textos hispánicos

TRABALLO PARA O EXAME ORAL

TEXTO DEL TRABAJO: El Quijote

ENVÍO: el trabajo puede enviarse por email a pablo.gamallo@usc.es. El plazo límite de envío son las **14h del día 8 de Febrero** de 2011.

PROCEDIMIENTO: contesta a cada ejercicio describiendo brevemente los pasos que has dado para obtener el resultado requerido. En anexo, escribe los scripts que han sido utilizados para cada ejercicio.

1. Cuenta las ocurrencias de las siguientes palabras:

- Quijote
- Sancho
- Dulcinea
- Rocinante

2. Cuenta las ocurrencias de las palabras que cumplan las siguientes condiciones:

- acaba en “ble”
- comienza por “sobre”
- comienza por “in” y acaba en “mente”

Y escribe de mayor a menor frecuencia las palabras que cumplan cada una de estas tres condiciones.

3. Cumple el texto de El Quijote la Lei de Zipf? Razona la respuesta.

4. Compara la riqueza léxica de El Quijote con la de otro texto, teniendo en cuenta que, en clase, asociamos la riqueza léxica a la media entre tokens y formas.

5. Cuenta y escribe de mayor a menor frecuencia todos los nombres propios.

6. Escribe de mayor a menor frecuencia todos los lemas que cumplan las siguientes condiciones:

- adverbios que aparecen inmediatamente a la izquierda de cualquier adjetivo.
- nombres que aparecen inmediatamente a la derecha del verbo “quemar”.

7. Escribe de mayor a menor frecuencia todos los lemas que cumplan las siguientes condiciones:

- nombres modificados por los adjetivos “bueno” y “malo”.
- nombres en función de objeto directo de los verbos “quemar” y “buscar”.

Para realizar estas búsquedas puedes utilizar la gramática por defecto de DepPattern o escribir una gramática personal. En este último caso, coloca en anexo las reglas de la gramática.