

## TEMA 2 – HOJA DE EJERCICIOS V

Se proponen diferentes ejercicios relacionados con el **análisis semántico**. Se deberá utilizar WordNet de NLTK (<https://www.nltk.org/api/nltk.corpus.reader.wordnet.html>).

1. Utilizando WordNet, obtener los diferentes sentidos de la palabra “fight”. Se quiere obtener diferente información:
  - a. Saber cuántos sentidos diferentes tiene esta palabra.
  - b. Para cada sentido diferente hay que imprimir su nombre, la definición, los diferentes lemas asociados al mismo sentido y, también, ejemplos de uso.

A continuación, se muestra parte de la salida que se pide:

```
Número de sentidos de 'fight': 9

Sentido 1:
Nombre: battle.n.01
Definición: a hostile meeting of opposing military forces in the course of a war
Ejemplo(s): ['Grant won a decisive victory in the battle of Chickamauga', 'he lost his romantic ideas about war when he got into a real engagement']

-----
Sentido 2:
Nombre: fight.n.02
Definición: the act of fighting; any contest or struggle
Ejemplo(s): ['a fight broke out at the hockey game', 'there was fighting in the streets', 'the unhappy couple got into a terrible scrap']

-----
Sentido 3:
Nombre: competitiveness.n.01
Definición: an aggressive willingness to compete
Ejemplo(s): ['the team was full of fight']
```

2. Haz lo mismo que en el ejercicio anterior, pero obteniendo los *synsets* solo cuando la categoría de la palabra sea un verbo. ¿Qué ha cambiado?
3. Imprime los lemas de los sinónimos de “fight” cuando la categoría gramatical sea un nombre, sin distinguir por sentidos.
4. ¿Qué ocurre si se imprimen los lemas de los sinónimos de “fight” cuando sea un nombre, pero de forma diferenciada para cada sentido de la palabra? ¿Cuáles son las diferencias con la salida del ejercicio anterior?
5. Obtener todos los sentidos de “bank” cuando es un nombre. ¿Cuántos sentidos tiene? Buscar las siguientes relaciones para todos los sentidos de la palabra:
  - a. Buscar los hiperónimos
  - b. Buscar los hipónimos

Un hipónimo concreta el significado de su hiperónimo, de esta manera “mesa” es más específico que “mueble” y “escritorio” es un tipo particular de “mesa”.

6. Es posible calcular la similitud semántica entre dos *synsets* y para ello WordNet tiene varias medidas de similitud semántica disponibles.

Hay que elegir diferentes conceptos y mostrar la similitud entre ellos, utilizando medidas variadas. A continuación, se presentan ejemplos de conceptos para los que se quiere medir su similitud:

dog – cat, vehicle – car, dog – car, whale - eagle

Como cada palabra anterior puede tener varios sentidos, lo razonable es medir la similitud semántica entre sentidos concretos de estas palabras. Por ello, se elegirá el primer sentido del listado de *synsets* para cada palabra, que suele coincidir con ser el sentido más frecuente (MFS, *Most Frequent Sense*).

Se pide calcular la similitud semántica entre los pares anteriores con diferentes medidas:

- a. Medidas basadas en *path*, donde se mide la longitud que hay en el camino que lleva de un concepto a otro y cuánto más corto sea el camino, más similitud hay.
  - b. Medidas basadas en el contenido, de forma que cuanta más información comparten dos conceptos, más similares serán. Estas medidas combinan una taxonomía (hipérónimos/hipónimos) y una estimación de frecuencia en el corpus (information content). Es necesario utilizar wordnet\_ic para calcular esta similitud.
7. Se quiere hacer desambiguación del sentido de una palabra (WSD). Una estrategia muy simple (y poco precisa) para desambiguar palabras consiste en seleccionar un *synset* de WordNet al azar sin tener en cuenta el contexto en el que se encuentra la palabra a analizar. Aunque esta estrategia no es de mucha utilidad en la práctica, permite comprender la importancia de utilizar métodos más sofisticados.

En este ejercicio se pide desambiguar la palabra “watch” en la frase “You must watch this film”, seleccionando para ello el primer *synset* identificado por WordNet. Muestra la definición del *synset*. ¿Es la acepción correcta de la palabra?

8. En muchas ocasiones, el significado de una palabra polisémica puede determinarse a partir de su categoría gramatical. Por ejemplo, la palabra “watch” adquiere distintas acepciones dependiendo de si funciona como verbo o sustantivo.

En este ejercicio, se pide desambiguar la palabra “watch” en función de la categoría gramatical que presenta en dos frases diferentes. Específicamente se pide:

- a. Identificar la categoría gramatical de la palabra “watch” dentro de las frases “You must watch this film” y “He gave me a watch as a present”.

- b. A partir de la categoría gramatical, seleccionar el primer *synset* obtenido en WordNet.
  - c. Observar el resto de *synsets* que Wordnet proporciona al especificar la categoría gramatical. ¿Ha mejorado la búsqueda de la acepción correcta? ¿Es suficiente?
9. Lo más razonable a la hora de desambiguar el sentido de las palabras es tener en cuenta el contexto en el que se encuentran. Uno de los primeros algoritmos desarrollados para abordar este problema es el algoritmo de Lesk. Dada una palabra polisémica y el contexto en el que aparece, el algoritmo de Lesk devuelve el *synset* con el mayor número de palabras coincidentes entre la frase en la que se encuentra la palabra a desambiguar y las distintas definiciones de los *synsets* que puede poseer una palabra.
- a. Se pide usar el algoritmo de Lesk para desambiguar la palabra “bank” en las frases “*I went to the bank to deposit my money*” y “*The land along the river bank has vegetation*”. Muestra la definición del *synset*. ¿Son las acepciones correctas de la palabra?

Nota: <https://www.nltk.org/howto/wsd.html>

- b. En el apartado anterior se ha podido ver cómo no siempre está desambiguando el sentido correcto. Se piden dos cosas:
  - Comparar para cada frase el sentido más frecuente de WordNet para cada palabra a desambiguar (suele ser el primer sentido) con el que se obtiene con la desambiguación con el algoritmo Lesk. Analizar el resultado, ¿cuándo se acierta más?
  - Modificar lo anterior para tener en cuenta la categoría gramatical de la palabra “bank”, de forma que se haga la desambiguación cuando es un nombre. ¿Cómo han variado los resultados?
- c. Dados varios pares de frases para desambiguar palabras, comparar la salida del algoritmo Lesk con el sentido más frecuente, es decir, se quiere saber cuántas veces coinciden en el sentido que desambiguan. Analizar quién se acerca más a una desambiguación correcta.
  - Realiza este análisis sin tener en cuenta la categoría gramatical de las palabras.
  - Realiza el análisis considerando la categoría gramatical de la palabra en cada caso.

Probarlo con las siguientes sentencias:

```
examples = [  
    ("bank", "The fisherman sat on the bank of the river."),  
    ("bank", "She went to the bank to apply for a loan."),  
    ("plant", "The workers shut down the plant."),  
    ("plant", "She watered the plant in the garden."),  
    ("light", "Turn on the light."),  
    ("light", "This suitcase is very light."),  
    ("bat", "The bat flew out of the cave."),  
    ("bat", "He hit the ball with a bat."),  
    ("interest", "He has a strong interest in music."),  
    ("interest", "The bank pays interest every month.")  
]
```