

1. 2 aplicaciones que tiene el análisis léxico en el PLN
Reconocimiento y categorización de palabras.
2. ¿Para qué sirve la tokenización?
Sirve para realizar el reconocimiento de palabras por clase específica.
3. Explica la diferencia entre un token y un type.
Sirve para realizar el reconocimiento de palabras por clase específica. Un token es cada una de las formas que aparecen en un texto sin importar el número de apariciones, mientras que un type son las distintas formas de palabras diferentes.
4. ¿En qué consiste el conteo de palabras?
Consiste en segmentar el texto en distintos elementos usando tokenización con el fin de llevar un análisis en un futuro
5. ¿Cómo podemos saber el número de veces que se repite una palabra en el corpus?
Podríamos hacer una tokenización del texto, posteriormente contamos el número de apariciones que tuvo el token que deseamos buscar.
6. ¿Qué es y qué papel juegan los lexicones computacionales en el PLN?
Los lexicones son la base del trabajo en PLN, ya que al ser estas bases de datos que guardan sintáctica y semántica estas facilitan los análisis léxico.
7. ¿Para qué sirven los N-gramas en PLN?
Ya que los N-gramas son secuencias de palabras, estas buscan hacer las determinaciones estadísticas y la elaboración de predictores de texto.
8. ¿Qué son las colocaciones?
Son secuencias de palabras que ocurren con una frecuencia mayor a la normal en la lengua.
9. ¿En qué consiste la cuantificación de las colocaciones?
Consiste en 4 principales elementos que son conteo de palabras cercanas, conteo de palabras a la izquierda, conteo de palabras a la derecha y conteo de concordancias.