

Técnicas de Procesamiento del Lenguaje Natural en Sistemas de Recuperación

El procesamiento del Lenguaje Natural (PLN) permite mejorar la precisión de la recuperación de información. Se debe tener en mente que, en la mayoría de los buscadores, las consultas son vistas como un conjunto de cadenas de texto que deben coincidir, en el mayor grado posible, con las cadenas de texto de un documento para ser este considerado relevante.

El análisis del lenguaje permite mejoras a este sistema, ya que se puede reconocer la categoría gramatical de los términos de la consulta o documento (p.e. no es lo mismo Camino, como verbo, sustantivo o como nombre propio). Además las reglas de flexión, tanto de género como de número, pueden ser aplicadas de forma más pertinente si se tiene en cuenta la categoría gramatical.

Una aplicación de determinar la categoría gramatical, es la posibilidad de dar menos peso (o ninguno) a las palabras vacías. Una palabra vacía es un término que tiene menor poder discriminante que otros. Por ejemplo, un sustantivo tiene generalmente mayor poder discriminante que un determinante o una preposición (así “sobre” como sustantivo es más discriminante que como preposición, aunque obviamente depende de la consulta).

Al igual que ocurre con la categoría gramatical se puede extender el razonamiento a fragmentos más complejos del texto. Reconociendo frases, patrones u oraciones, o resolviendo las anáforas y otros referentes entre sentencias.

Por último se analizan algunas aplicaciones software de procesamiento del lenguaje natural, junto con comparativas de su aplicación. También se muestra una herramienta de minería de datos, con el fin de poder aplicarlo para estimar el peso de los términos en casos concretos.

