

Procesamiento del Lenguaje Natural para recuperación de Información

Introducción a la recuperación de documentos del Web

El **Procesamiento del Lenguaje Natural para la Recuperación de Información**, (*PLN* , o *NLP* ; *Natural Language Processing*), es una subdisciplina de la Inteligencia Artificial y la rama ingenieril de la lingüística computacional. El **Procesamiento del Lenguaje Natural** se ocupa de la formulación e investigación de mecanismos eficaces computacionalmente para la comunicación entre personas o entre personas y máquinas por medio de lenguajes naturales . El **Procesamiento del Lenguaje Natural** no trata de la comunicación por medio de lenguajes naturales de una forma abstracta, sino de diseñar mecanismos para comunicarse que sean eficaces computacionalmente, que se puedan realizar por medio de programas que ejecuten o simulen la comunicación. El lenguaje natural sirve sólo de medio para estudiar estos fenómenos.

Componentes

Los principales componentes de un sistema de **Procesamiento del Lenguaje Natural para la Recuperación de Información** son:

- Análisis morfológico: Análisis de las palabras para extraer raíces, rasgos reflexivos, etc.
- Análisis sintáctico: Análisis de la estructura sintáctica de la frase mediante gramáticas.
- Análisis semántico: Extracción del significado de la frase, y la resolución de ambigüedades.
- Análisis pragmático: Análisis del texto más allá de los límites de la frase, determinando los antecedentes referenciales de los pronombres.
- Planificación de la frase: Estructurar cada frase del texto con el fin de expresar el significado adecuado.
- Generación de la frase: Generación de la cadena lineal de palabras a partir de la estructura general de la frase.

Aplicaciones

Las principales tareas de trabajo en el **Procesamiento del Lenguaje Natural para la Recuperación de Información** son:

- Síntesis del discurso
- Análisis del lenguaje
- Comprensión del lenguaje
- Reconocimiento del habla
- Síntesis de voz
- Generación de lenguajes naturales
- Traducción automática
- Respuesta a preguntas
- Recuperación de la información
- Extracción de la información