**Unidad 3 – Procesamiento del lenguaje natural**

**Analisis del lenguaje natural con expresiones regulares**

1. **Introducción**

A través del análisis sintáctico del lenguaje natural con expresiones regulares se puede pueden hacer tareas como detectar sesgos. A través de regesxy nltk se pueden detectar palabras claves y descubrir donde y con que frecuencia se usan.

1. **Interpretación y emparejamiento**

Una expresión regular es un patrón que describe una cierta cantidad de texto. Existen varios métodos en Python para expresiones regulares

* .complie(): toma una regex y genera un objeto para que luego se pueda utilizar con strings.
* .match(): evalúa si encuentra una coincidencia al principio del texto dado. Devuelve un objeto o None

1. **Búsqueda**

La funcion .search() buscara en todo el texto.

La funcion .findall() busca todas las ocurrencias y devuelve una lista

1. **Etiquetado de parte del discurso (Part-of-Speech Tagging)**

El etiquetado gramatical es el proceso de identificación y etiquetado de la parte del discurso de las palabras.

Este proceso se puede hacer a través de la librería nltk con la función pos\_tag

**Reconocimiento de entidades**

El reconocimiento de entidades o NER (Named Entity Recognition) consiste en identificar y clasificar entidades en un texto. Estas entidades normalmente hacen referencia a sustantivos como nombre de personas, lugares, organizaciones, fechas, etc…

Librerias como NLTK, Spacy o Stanza se pueden utilizar para este reconocimiento de entidades. Con la librería spacy puedes hacer .label\_ para ver que etiqueta corresponde a cada entidad