

Análisis de Personajes Mitológicos

Sergi Fornés

Trabajo Final de Máster

El objetivo de este trabajo es crear una base de datos de personajes mitológicos a partir de **Wikipedia**. Una vez hecho esto, se pretenden analizar los artículos de estos personajes y las relaciones entre ellos haciendo uso de algoritmos de NLP y Network Analysis.

Guión

Entender la mitología es entender al ser humano. Diferentes culturas han creado mitos para responder cuestiones complejas de manera sencilla. En estos mitos aparecen personajes sobrenaturales y las relaciones que mantienen con los seres humanos.

En los mitos aparecen todo tipo de personajes, desde diosas del amor hasta monstruos devoradores de humanos. Dada esta diversidad aparecen preguntas interesantes, ¿diferentes culturas tienen personajes similares? ¿existen relaciones entre las características de diferentes personajes? ¿y entre las características de los mismos personajes?

Los objetivos de este trabajo son:

1. Crear un conjunto de datos de personajes mitológicos de diversas culturas.
2. Analizar el conjunto de datos.

Creación del conjunto de datos. El conjunto de datos se obtendrá utilizando librerías de web-scraping (Beautiful Soup y Selenium) de Python.

Todos los datos se extraerán de la enciclopedia digital **Wikipedia**. Hay que señalar que cualquier persona puede editar esta enciclopedia, por lo que es posible que el conjunto de datos tenga errores. Sin embargo, Wikipedia es una de las plataformas de contenido libre más completas de internet.

Existen artículos en Wikipedia de prácticamente todos los personajes mitológicos. Estos artículos están clasificados en categorías similares, por lo que se navegará a través de estas categorías para obtener el texto disponible de los personajes. Además se añadirá información adicional de cada personaje en función de la categoría en la que se encuentre, como la mitología, la región, el continente, o el tipo de personaje (deidad, héroe, monstruo, ...).

Por otro lado, los artículos de los personajes mitológicos están relacionados entre sí mediante enlaces,

así que también se extraerá una matriz con estas relaciones.

Análisis del conjunto de datos. Se realizará un análisis exploratorio de los datos y se usarán técnicas de análisis de grafos sobre la matriz de relaciones para crear un índice de importancia de los personajes.

La variable más importante de los personajes es el texto extraído de Wikipedia, así que también se aplicarán diversos algoritmos de NLP para su análisis.