

Resumen

En el presente trabajo se llevó a cabo el desarrollo de un sistema informático para la normalización de entidades biomédicas basado en búsquedas semánticas multilenguaje, sobre la base de datos médica UMLS, donde se aplicó técnicas de procesamiento de lenguaje natural. El propósito fundamental radica en potenciar la comprensión entre los términos médicos, para mejorar la calidad de diagnósticos, tratamientos y decisiones clínicas. La implementación del sistema se llevó a cabo siguiendo una metodología ágil de desarrollo de software, que permitió un proceso organizado, flexible y enmarcado en un plazo determinado. El alcance del proyecto es llegar a construir un sistema que permita buscar e identificar entidades biomédicas dentro de una base de datos, por lo cual se implementó el desarrollo bajo una arquitectura y frameworks actuales. Para el desarrollo del proyecto se han implementado dos interfaces, la entrada de texto y la visualización de los resultados generados por el sistema. La implementación de estas interfaces se llevó a cabo en conjunto con la construcción de diversos módulos desarrollados en Python. Estos módulos, permitan trabajar de manera conjunta con los datos proporcionados por la base de datos UMLS donde contiene todos los datos normalizados y relacionados a conceptos médicos. También se utilizaron modelos Transformer para la vectorización de los datos los cuales fueron utilizados para la implementación de la búsqueda semántica.

Palabras clave: Búsqueda semántica, Normalización de datos, búsqueda de diagnóstico médico.

Abstract

In the present work, the development of a computer system for the normalization of biomedical entities based on multilanguage semantic searches was carried out, on the basis of UMLS medical data, where natural language processing techniques were applied. The fundamental purpose is to promote understanding between medical terms, to improve the quality of diagnoses, treatments and clinical decisions. The implementation of the system was carried out following an agile software development methodology, which allowed an organized, flexible and framed process within a specific period. The scope of the project is to build a system that allows searching and identifying biomedical entities within a database, for which the development was implemented under current architecture and frameworks. For the development of the project, two interfaces have been implemented, the text entry and the visualization of the results generated by the system. The implementation of these interfaces was carried out in conjunction with the construction of various modules developed in Python. These modules allow working together with the data provided by the UMLS database where it contains all the standardized data and related to medical concepts. Transformer models were also used for the vectorization of the data, which were used for the implementation of the semantic search.

Keywords: Semantic search, Data normalization, medical diagnosis search.