



Universidad del Azuay

Departamento de Posgrados

**Magister en Sistemas de Información, mención Inteligencia de
Negocios**

**Recomendador cuantitativo basado en
minería de texto y técnicas de procesamiento
de lenguaje natural, para identificar la
similitud de conceptos entre los documentos
obtenidos de dos componentes de un
mashup web educativo y el concepto del
componente origen**

Autora:
Mayra J. Espinoza Guarango

Directora:
Irene P. Cedillo Orellana

**Cuenca – Ecuador
2023**

DEDICATORIA

A mi familia, que son mi motor y me dan las raíces firmes que me permiten poder volar.

AGRADECIMIENTOS

Doy gracias a la vida, que me ha dado la dicha de caminarla con maravillosas personas. A mi amado Esposo que me alienta a dar nuevos pasos, a mi Hija del corazón que nos llena de luz y nos regala su amorosa paciencia, a mi Hermanita que me impulsa por nuevos horizontes, a mis cuatro Padres por su gran sabiduría.

Un especial agradecimiento a mi tutora, que con su cordialidad y sus acertados ejemplos ha logrado impartir sus conocimientos y brindarme una guía durante este proceso.

Resumen

La presente investigación, plantea un modelo que permite optimizar las recomendaciones de textos en el desarrollo de los mashup web educativos. Se centra en la comparación de textos extraídos de diferentes servicios web. Estos son analizados con técnicas de minería de textos y procesamiento del lenguaje natural. Con el objetivo de identificar la similitud que existe entre ellos y poder recomendar la mejor opción de lectura. Con el fin de agilizar el proceso de investigación de académicos y aprendices, para así aportar a un aprendizaje eficaz.

En el desarrollo de la investigación se extraen documentos de texto de los componentes de un mashup web educativo. Con estos datos se aplica la metodología CRISP-DM, que permite sistematizar el proceso de minería y la aplicación de técnicas de PLN. Lo que lleva al cumplimiento de los objetivos planteados.

Esta investigación puede ser parte complementaria a los procesos de desarrollo de los mashup web. Puesto que, apoyaría en el contenido ya generado por el mashup, para analizarlo y recomendar al usuario final el documento que aporte información relevante y relacionada al tema de su búsqueda.

Palabras clave

Mashup web educativo, Knowledge Discovery in Databases (KDD), Knowledge Discovery in Texts (KDT), CRISP-DM, minería de textos, procesamiento del lenguaje natural (PLN), modelo generativo Latent Dirichlet Allocation (LDA), procesamiento de textos, limpieza de textos, Python.



Ing. Priscila Irene Cedillo Orellana Ph.D
Directora del trabajo de titulación



Ing. Mayra Jadira Espinoza Guarango
Autora

Abstract

This research proposes a model that allows optimizing text recommendations in the development of educational web mashups. It focuses on the comparison of texts extracted from different web services. These are analyzed with text mining and natural language processing techniques identifying the similarity that exists between them and being able to recommend the best reading option and facilitate the research process of academics and apprentices, contributing to more efficient learning. While developing this research, text documents were obtained from the components of an educational web mashup. The CRISP-DM methodology is applied with this data, which allows the systematization of the mining process and the application of PLN techniques that lead to the fulfillment of the set objectives. This research can be a complementary part of the web mashup development processes. It will support the content already generated by the mashup, analyzing it, and recommending to the end user a document that provides relevant information related to the topic of their search.

Key words

Educational Web Mashup, Knowledge Discovery in Databases (KDD), Knowledge Discovery in Texts (KDT), CRISP-DM, Text Mining, Processing Natural Language (PNL), Latent Dirichlet Allocation (LDA) Generative Model, text processing, cleanup text, Python.



Ing. Priscila Irene Cedillo Orellana Ph.D
Degree Project Director



Ing. Mayra Jadira Espinoza Guarango
Author

Translated by:



Ing. Mayra Jadira Espinoza Guarango