

RESUMEN

El procesamiento del lenguaje natural es un campo que combina las ciencias computacionales como el aprendizaje automático con la lingüística aplicada, con el fin de lograr el procesamiento asistido por computador de cierta información expresada en lenguaje corporal para llevar a cabo ciertas tareas que imitan comportamientos inteligentes. Por esta razón, y en base a los procesos de aprendizaje que tiene el ser humano, se puede llegar a implementar el reconocimiento de texto utilizando las computadoras y el potencial analítico de algoritmos que forman parte de la herramienta Weka, minimizando el error en la clasificación de texto. Para ello, el desarrollo del prototipo propuesto en el trabajo de titulación parte de un estudio de características que contiene la web, en base a los recursos seleccionados se llegó a la construcción de un web crawler que permite la extracción y almacenamiento de texto por género musical de artistas latinoamericanos. Seguidamente, se muestra un análisis de los algoritmos que forman parte de la clusterización de datos indicando los porcentajes de instancias correctamente agrupadas. Para finalizar con el proceso de desarrollo, se elabora una página web que contiene el resultado del estudio que se llevó a cabo dentro de la investigación, logrando brindar recomendaciones en base a las canciones más populares de cada género musical.

PALABRAS CLAVE

- **PROCESAMIENTO DE LENGUAJE NATURAL**
- **APRENDIZAJE AUTOMATICO**
- **WEKA**
- **CLUSTERIZACION**
- **CRAWLER**

ABSTRACT

Natural language processing is a field that combines computational science such as machine learning with applied linguistics, in order to achieve computer-assisted processing of certain information expressed in body language to carry out certain tasks that mimic intelligent behavior. For this reason and based on the learning processes that the human being has, text recognition can be implemented using computers and the analytical potential of algorithms that are part of the Weka tool, minimizing the error in the classification of text. For this, the development of the prototype proposed in the titration work is based on a study of features contained in the web. Based on the selected resources, a web crawler was built that allows the extraction and storage of text by musical genre. of Latin American artists. Next, an analysis of the algorithms that are part of the clustering of data indicating the percentages of correctly grouped instances is shown. To finish the development process, a web page is created that contains the results of the study that was carried out within the research, making recommendations based on the most popular songs of each musical genre.

KEYWORDS:

- **PROCESSING OF NATURAL LANGUAGE**
- **AUTOMATIC LEARNING**
- **WEKA**
- **CLUSTERIZATION**
- **CRAWLER**