

Classification of Resources in an e-Library Using Machine Learning Algorithms

MAC Akmal Jahan ¹ and Roshan G Raga ^{1,2}

Faculty of Applied Sciences, South Eastern University of Sri Lanka.

¹Faculty of Engineering, University of Peradeniya.

*Corresponding Author: jahana@seu.ac.lk

Library is the heart of a university and students spend a large amount of time in library in search of knowledge. The trend of reading resources in printed materials such as books, journals and other research publications is gradually changing. Since it is an uneasy and time-consuming process, students are interested in soft materials such as e-journals, ebooks and other web based resources. Nowadays, in a library most of the resources in digital form are stored without any classification. They are not categorized or utilized by the users since it does not have any proper way to access or find appropriate material when the users' queries applied. Even though there are a lot of manual ways to access text based materials or resources in a library, they cannot be applied to the digital resources since it needs some kind of text mining and machine learning.

This project addresses this issue through a closed domain question answering system for a resource pool in an e-library. As the initial step, the project uses a narrowed down search space by processing the abstracts of the resources.

More than 300 abstracts are extracted along with their title and pre-processed. 75% of the data are used as training sets and the remaining are used for testing. Different machine learning techniques such as classification and clustering are applied with this large collection of textual data using Weikato Environment of Knowledge Analysis (WEKA) and their performance metrics and error rates were compared. The most suitable machine learning technique and the mode of testing for the textual data were selected and applied for training models as the solution for the classification problem of the electronic resources.

Keywords: WEKA, Machine learning, Cross-validation, Clustering