



INFORMATICS INSTITUTE OF TECHNOLOGY

In Collaboration with

UNIVERSITY OF WESTMINSTER (UOW)

BEng/BEng.(Hons) in Software Engineering

Final year Project 2017/2018

Research paper Classification using Keyword Clustering

By

2014119 Akshina Fernando

(UoW No: w1534107)

Supervised By

Cassim Farook

Submitted in partial fulfilment of the requirements for the
BSc (Hons) Software Engineering Degree
Department of Computing

May 2018

©The copyright for this project and all its associated products resides with
Informatics Institute of Technology

Abstract

This research is to classify a research paper using keyword clustering. The objective is to reduce the time consumption in manual classification of research papers. Classifying of a research paper is achieved by introducing an automated way to generating keyword sections and finding the respective domain. This takes away the need to scan the entire research. In other words, the manual classification approach is succeeded using a keyword extraction model and an unsupervised clustering model. This research is based on unsupervised learning and clustering model.

A comprehensive literature review and an online survey of knowledge resources were conducted to gather requirements to design a solution. Natural language Processing and Machine Learning are used in implementing the solution. Finally, the evaluation of the test results of the project show the possibility that this novel approach can be used to automatically classify research papers. In conclusion, this research makes a key note for identifying new approaches to keyword clustering and extraction.

Subject Descriptors:

Machine learning

 Unsupervised learning

 Cluster analysis

 Machine learning algorithms

 Feature selection

Artificial Intelligence

 Natural Language Processing

 Information extraction

Keywords: Information extraction, Unsupervised Learning, Cluster Analysis, Feature Selection