A SEMANTICS-BASED CLUSTERING APPROACH FOR SIMILAR RESEARCH AREA DETECTION: A CASE STUDY OF NIGERIAN UNIVERSITIES

BY

ADIGUN, EMMANUEL BUKUNMI

(16PCG01361)

A DISSERTATION SUBMITTED IN THE DEPARTMENT OF COMPUTER AND INFORMATION SCIENCES, TO THE SCHOOL OF POSTGRADUATE STUDIES, COVENANT UNIVERSITY, OTA, OGUN STATE.

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE

AWARD OF MASTER OF SCIENCE DEGREE IN COMPUTER SCIENCE.

JUNE, 2018

ACCEPTANCE

This is to attest that this dissertation is accepted in partial fulfillment of the requirements for the award of Masters of Science degree in Computer Science in the Department of Computer and Information Science, College of Science and Technology, Covenant University, Ota, Ogun State.

Mr. Philips John Akinokhai ----------

Secretary, School of Postgraduate Studies *Signature and Date*

Prof. Samuel Wara ----------

Dean, School of Postgraduate Studies *Signature and Date*

DECLARATION

I hereby declare that Adigun, Emmanuel with matriculation number **16PCG01361**, carried out this research entitled "**A Semantics-based clustering approach for similar research area detection: A case study of Nigerian Universities**". The project is centered on an original study in the department of Computer and Information Sciences, College of Science and Technology, Covenant University, Ota, under the supervision of Dr. (Mrs) Marion Adebiyi. Concepts of this research project are results of the research carried out by Adigun, Emmanuel and ideas of other researchers have been fully recognized.

Signature:----------

Date:----------

CERTIFICATION

This is to certify that this research entitled "**A Semantics-based clustering approach for similar research area detection: A case study of Nigerian Universities"** was carried out by **Adigun Emmanuel** with matriculation number **16PCG01361** under our supervision and approved by us:

Dr. (Mrs) Marion O. Adebiyi ----------

Supervisor *Signature and Date*

Prof. Victor C. Osamor ----------

Head of Department *Signature and Date*

External Examiner

Signature and Date

DEDICATION

I dedicate this project to God Almighty for His sufficient grace, wisdom and knowledge given to me throughout my Master's Degree Programme.

ACKNOWLEDGEMENTS

I would like to use this opportunity to thank the following individuals whose immense contributions have made this work come to a reality.

First, I want to thank my supervisor, Dr.(Mrs) M.O. Adebiyi, for her insight and encouragement in guiding me through to the completion of this work.

Secondly, I want to thank the Head of Department, Computer and Information Sciences, Prof. V.C. Osamor, as well as the PG coordinator, Prof. A.A. Azeta for giving me the special privilege to pursue my passion and interest.

Thirdly, I wish to thank the members of faculty, Prof. A. A. Adebiyi, Dr. (Mrs) I.T. Afolabi, and Dr. (Mrs) O.O. Oladipupo, who have contributed their insights, expertise and time to make this work a reality.

Also, I will like to thank my colleagues, Opeyemi Makinde, Onosu Dafe, Ogese Mark, Abasifreke Bassey and Deji Osanyin for their continuous support throughout the duration of my program.

I want to especially thank the Chancellor and Management of Covenant University, for giving me the platform and the platform to pursue God's agenda for my life in academics.

I will not fail to mention my loving parents, Dr. and Dr. (Mrs) Thomas Adigun, for their support and setting high standards for me to emulate in my academic pursuit. I will like to thank my special siblings, Tomi Adigun and Fiyin Adigun for their encouragement and support.

Finally, I thank my heavenly Father and Creator who has led and guided me as I worked through this research.

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

The place of research collaborations is indispensable in coming up with research publications. The task of detecting similar research areas is crucial to the development and furtherance of research. Prominent and rookie researchers alike are predisposed to seek existing research publications in a research field of interest before coming up with a thesis. The manual process of searching out individuals in an already existing research field is cumbersome and time-consuming. Besides, it tends to not capture publications with keywords that do not match a keyword query which results in inaccurate results. From extant literature, automated similar research area detection systems have been developed to solve this problem. However, most of them use keyword matching

techniques which do not sufficiently capture the implicit semantics of keywords thereby leaving out some research articles. In this work, we have proposed a similar research area detection framework to address this problem. The aim of this study is to develop a semantics-based clustering method for similar research area detection. This study employs a number of techniques such as Ontology-based pre-processing, Latent Semantic Indexing and K-Means Clustering to develop a prototype similar research area detection system, that can be used to determine similar research domain publications. However, traditional document clustering techniques suffer from high dimensionality and data sparsity problems. In a bid to solve these problems, a domain ontology is used in the preprocessing stage to weight concepts and determine semantically similar concepts, while Latent Semantic Analysis is used as the topic modelling technique in order to capture the implicit semantic relationship between terms in the text corpus. To test our framework, publications from a number of Nigerian University faculties were randomly selected and used as the dataset for our clustering model. A proof-of-concept implementation was developed using the Python programming language. From the evaluation of our system, we were able to derive more accurate clustering results as a result of the integration of ontologies in the pre-processing stage in comparison with documents that were not pre-processed with the ontology.

CHAPTER ONE
INTRODUCTION

1.1 BACKGROUND OF THE STUDY

The recent surge in the number of publications, scientific journals, books and conference