OBAFEMI AWOLOWO UNIVERSITY

DEVELOPMENT OF A COMPUTATIONAL SYSTEM FOR INTEGRATING USAGE INTO DOCUMENT INDEXING

By

LUKMAN ADEWALE AKANBI
M.Sc. (Computer Science), Ife

A THESIS SUBMITTED TO
DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING
FACULTY OF TECHNOLOGY
OBAFEMI AWOLOWO UNIVERSITY, ILE-IFE

IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE AWARD OF DOCTOR OF PHILOSOPHY IN COMPUTER SCIENCE

2014

OBAFEMI AWOLOWO UNIVERSITY ILE-IFE, NIGERIA

HEZEKIAH OLUWASANMI LIBRARY

POSTGRADUATE THESIS

AUTHORISATION TO COPY

Author: AKANBI Lukman Adewale

Title: DEVELOPMENT OF A COMPUTATIONAL SYSTEM FOR INTEGRATING USAGE INTO DOCUMENT INDEXING

Degree: Ph.D. (Computer Science)

Year: 2014

I, AKANBI Lukman Adewale, hereby authorise the Hezekiah Oluwasanmi Library to copy my thesis in part or whole in response to request from individuals and or organisations for the purpose of private study or research.

_____

Signature of Author and Date

OBAFEMI AWOLOWO UNIVERSITY

CERTIFICATION

I, Lukman Adewale Akanbi with registration number TP08/09/R0033 in the Department of Computer Science and Engineering, Faculty of Technology, Obafemi Awolowo University certify that this is an original research carried out by me under the supervision of:

Supervisor: _____
Prof. E. R. Adagunodo

Co-Supervisor: _____
Prof. A. A. David

Head of Department: _____
Dr. A. I. Oluwaranti

DEDICATION

This work is dedicated to glory of Almighty Allah, by whose permission the work have been completed.

ACKNOWLEDGEMENT

All glorification and adoration are due to Allah, the Almighty, who out of His mercy, made the dream of having a Doctorate degree became reality. I am most grateful to my employer, the Obafemi Awolowo University, Ile-Ife for giving me the opportunity and the required supports. My sincere appreciation goes to the Heads of Department of Computer Science and Engineering, past and present for their support and advice throughout the course of the research work. I am particularly grateful to all the members of staff of the Department, for their supports and holding forth for me when I had to travel.

My profound gratitude goes to my supervisors, Professors Emmanuel Rotimi Adagunodo and Amos Abayomi David, for their time and resources. I appreciate your fatherly role and the sacrifices you have to make throughout the course of the work. I pray God almighty reward you with good in manifolds. The effort of Prof. David at securing grant that made it possible for me to visit France twice during the course of the work is highly appreciated. His hosting me in France and making my staying in France very rewarding is highly appreciated.

My appreciation goes to the French Government through the French embassy in Nigeria, for supporting this work through the provision of grant that facilitated my visit to France in 2012 and 2013. I also appreciate my colleagues in France, Toyin Oguntunde and Bunmi Akere for making me feeling at home in France.

I am most grateful to Dr. B. S. Afolabi for his immense contribution towards the success of this work. The advice and inspiration from Dr. O. A. Odejobi is highly appreciated. I am also grateful to Drs. S. I. Eludiora, A. O. Ajayi, F. O. Asahiah, S. A. Bello, Mrs. K. C. Olufokunbi, Mrs. D. F. Ninan, Mr. J. A. Hassan and other members of the CISRG for their supports during the

course of the research work. Dr. AdulWaheed Bamgbade is highly appreciated for his professional support through timely review of the grammatical structure of the thesis.

My friends, AbdulAzeem Ewenla, Lukman Olawoyin, Azeez Adebisi, Taofeek Olusesi, Mikail Farinde and others are appreciated for their love and care. I thank you all. The contribution of Mahmood Oyewo in coding at the initial stage of the work is highly appreciated.

I am particularly grateful to my loving wife, Mrs. Adebimpe Monsurah Akanbi and my beautiful daughter Rodiat Adeola Akanbi for their patient and support throughout the course of this research work. I cannot appreciate you enough. I love you. My appreciation goes to other members of my family for their understanding and support in the course of this study.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

The study formulated a model that augments document with usage, designed, implemented and evaluated a system based on the model. This is with the view of enhancing the quality and quantity of useful documents that are returned during document search operation.

Attribute Value Pair technique of data abstraction in document annotation and vector model technique of Information Retrieval were used to formulate the document usage model. Unifying Modelling Language (UML 2.0) was used to design the Competitive Intelligence based Document Usage Creation and Exploration (CIDUCE) system. The prototype was implemented with the use of PHP and MySQL technology. Data on document usage was collected through questionnaire administration and guided interview from 20 selected postgraduate students (M.Sc. and Ph.D.) in various departments in the Faculty of Technology. Ninety-nine (99) documents and twenty (20) decision problems were extracted from the questionnaire and used to populate the database of the system. Document recall rate, a function of the similarity measure between identified relevant documents by the respondents and their decision problems (i.e. research problems) was used to evaluate the system.

The results showed that the usage-based document index consistently produce high recall rate, that is, identified high number of relevant documents at different retrieval thresholds than the keyterm-based index. For example, at the retrieval thresholds of 0.20, 0.30, 0.40, 0.50, 0.60, 0.70 and 0.80, the keyterm-based index has 47.47, 27.27, 14.14, 9.09, 2.02, 1.01 and 0.00% recall rates, respectively as compare with the usage-based index with recall rate of 100.00, 100.00, 100.00, 100.00, 100.00, 91.92 and 61.62%, respectively. These recall rates at different thresholds translated to 47, 27, 14, 9, 2, 1 and 0 documents, respectively in the keyterm-based index and 99, 99, 99, 99, 99, 92 and 62 documents, respectively in the usage-based index.

The study concluded that in an information seeking process, there are usually documents in the document collection space whose index may not contain terms in the users query but which are very relevant to users' need.