

UNIVERSITI TEKNOLOGI MARA

**KEYWORD BASED PROFILE
CREATION USING LATENT
DIRICHLET ALLOCATION,
DOMAIN DICTIONARY AND
DOMAIN ONTOLOGY**

NOR ADZLAN BIN JAMALUDIN

Thesis submitted in fulfilment
of the requirements for the degree of
Master of Science


Faculty of Computer and Mathematical Sciences

June 2016

AUTHOR'S DECLARATION

I declare that the work in this thesis was carried out in accordance with the regulations of Universiti Teknologi MARA. It is original and is the results of my own work, unless otherwise indicated or acknowledged as referenced work. This thesis has not been submitted to any other academic institution or non-academic institution for any degree or qualification.

I, hereby, acknowledge that I have been supplied with the Academic Rules and Regulations for Post Graduate, Universiti Teknologi MARA, regulating the conduct of my study and research.

Name of Student : Nor Adzlan bin Jamaludin
Student I.D No. : 2011589147
Program : Master of Science
Faculty : Faculty of Computer and Mathematical Sciences
Thesis Title : Keyword Based Profile Creation Using Latent
Dirichlet Allocation, Domain Dictionary And
Domain Ontology
Signature of Student :.....
Date : June 2016

ABSTRACT

Expert Finding is a field in information retrieval that focuses on finding an expert based on several criteria. Some of the methods that have been applied for expert finding include statistical, machine learning and ontology-based methods. Profile creation is one of the steps or tasks that are required in expert finding, which is the process of capturing and representing the details of experts and users which later can be used for retrieval. An issue that is faced for profile creation in expert finding is that the profiles being created are focused on the details of the experts but not on the users who are searching for these experts. This research explores a profile creation model that creates domain specific keyword-based profiles of users using Latent Dirichlet Allocation, domain dictionary and domain ontology from bookmarks. The domain of agriculture is selected as the case study for this research. The model is implemented in a form of a prototype and is evaluated by comparing how similar the prototype created profiles with manually built ones. From the results and analysis of the research, it is concluded that the method can successfully create domain specific profiles. The significances and contributions of the research include the application of LDA in user profiling, the proposed model, model prototype and the results and findings of the experiments conducted throughout the research.

TABLE OF CONTENTS

	Page
CONFIRMATION BY PANEL OF EXAMINERS	ii
AUTHOR'S DECLARATION	iii
ABSTRACT	iv
ACKNOWLEDGEMENT	v
TABLE OF CONTENTS	vi
LIST OF TABLES	ix
LIST OF FIGURES	x
LIST OF EQUATIONS	xiii
LIST OF ABBREVIATIONS	xiv
CHAPTER ONE: INTRODUCTION	
1.1 Research Motivation	1
1.2 Research background	2
1.3 Problem Statement	3
1.4 Research Questions	4
1.5 Research Objectives	5
1.6 Research Significance	6
1.7 Scope and Limitations of the study	6
1.8 Thesis Organization	6
CHAPTER TWO: LITERATURE REVIEW	
2.1 Introduction	8
2.2 Related Works	8
2.3 Information Retrieval	9
2.3.1 Ad-hoc Retrieval	10
2.3.2 Information Filtering	11
2.3.3 Differences between Ad-hoc and Information Filtering	12
2.4 Classical Information Filtering Methods	13
2.4.1 Content-Based Filtering	13

CHAPTER ONE

INTRODUCTION

1.1 RESEARCH MOTIVATION

In general, an expert can deliver more practical knowledge than a document can do as people with expertise can apply their general experience to larger classes of problems, while documents are inclined to remain focused on a rather tight problem context or completely ignore such context.

In Malaysia, farmers consult MARDI to obtain agricultural information related to their agriculture produce and products. The inquired items include seedlings, pest & disease management and more. When the farmer contacts MARDI, the attending MARDI officer will refer the farmer to a knowledgeable research officer who can help him. However, the process of finding a knowledgeable research officer who can assist the farmer can be difficult. A typical scenario in MARDI will see the farmer's inquiry being transferred between departments and research officers until it reaches a research officer who is willing to help the farmer. Even when the research officer helps the farmer, the information provided to the farmer may not be the best that can be offered. Agriculture is a vast body of knowledge consisting of three main fields which are horticulture, fishery and animals. Most researchers focus on and specialize only on several products in one of the three fields. Additionally, difference various aspects of a product for example in nutritional requirements or pest and diseases that affect each product can be significant even for similar products. This problem that is faced by farmers is called expert finding.

The research attempts to solve the problem of expert finding in agriculture domain by approaching it as an information filtering task. Furthermore, this task is divided into three sub-tasks which are profile creation, query analysis and query-profile matching. This thesis focuses on the task of profile creation and attempt to create a profile for agriculture researchers to assist in the expert finding task in the domain of agriculture.