Contents

Certificates i-iii
Declaration iv
Acknowledgements v
Abstract vi
Organization of the Dissertation vii

Chapter 1 Introduction 1
Web Service 1
How a Web Service is structured and published? 3
WSDL (Web Service Description Language) 4
Information Retrieval 6
Motivation behind Information Retrieval 6
Information Retrieval versus Data Retrieval 7

Chapter 2 Types of Information Retrieval Models 8
Set Theoretic Models 8
Boolean Retrieval Model 8
Fuzzy Set Retrieval Model 8
Algebraic Model (Vector Space Model) 9
Term Frequency and Weighting 9
Document Frequency and Inverse Document Frequency 9
Tf-idf Weighting 9
Document Vector 10
The Vector Space Model for Scoring 10
Probabilistic Model 11
Language Model 12

Chapter 3 Related Works on Web Service Discovery 13
A Probabilistic Semantic Approach for Discovering Web Services 13
PLSA Introduction 13
Probabilistic Semantic Discovery Approach 14
Probabilistic Semantic Matching Algorithms 14
Efficiently Finding Web Services Using a Clustering Semantic Approach 15
Overview of CPLSA Approach 15
Eliminating Irrelevant Services from Service Collection 15

Chapter 4 Language Modeling Concept 17

Chapter 5 Background of Proposed Model & Related Works 20
Purpose of the Project 20
Overview of the Language Model 20
Selected Language Models to be used for Smoothing Purpose 20
1. Jelinek-Mercer Language Model 20
2. Topic Language Model 21

Proposed Model 22
Principle of Jelinek-Mercer Language Model 23
Principle of Clustering based Topic Language Model 23
The Expectation Maximization Algorithm 24
Input, Procedure, and Output of the EM Algorithm 25
The k-means Clustering Algorithm 27
Basic Measures for Text Retrieval 28
Precision, Recall & F-score 28

Chapter 6 Work Accomplished 29
The Information Collection Stage 29
The Data Processing Stage 29
The Synonym Derivation Stage 29
The TF Weighting Stage 30
The Term Document Matrix Generation Stage 30
The Language Model Training Stage 31
Expectation Maximization Algorithm in case of Jelinek-Mercer LM 31
Clustering & Expectation Maximization Algorithm in case of Topic LM 37

Chapter 7 Experimental Result 38
Experimental Dataset 38
Experiment Done 38
Performance Results for Jelinek-Mercer Language Model 40
Performance Results for Topic Language Model 44

Conclusion & Future works 46

References 47