ONTOLOGICAL SOLUTION FOR SRI LANKAN LEGAL SYSTEM

Parana Palliya Guruge Dinesh Asanka

(139109R)

University of Moratuwa, Sri Lanka.
Electronic Theses & Dissertations
www.lib.mrt.ac.lk

Degree of Master of Science in Artificial Intelligence

Faculty of Information Technology

University of Moratuwa

Sri Lanka

December 2015

ONTOLOGICAL SOLUTION FOR SRI LANKAN LEGAL SYSTEM

Parana Palliya Guruge Dinesh Asanka

(139109R)

University of Moratuwa, Sri Lanka.
Electronic Theses & Dissertations
www.lib.mrt.ac.lk

Thesis submitted in partial fulfillment of the requirements for the degree of Master of Science in Artificial Intelligence

Department of Computational Mathematics

University of Moratuwa

Sri Lanka

December 2015

Declaration

I declare that this dissertation does not incorporate, without acknowledgment, any material previously submitted for a Degree or a Diploma in any University and to the best of my knowledge and belief, it does not contain any material previously published or written by another person or myself except where due reference is made in the text. I also hereby give consent for my dissertation, if accepted, to be made available for photocopying and for interlibrary loans, and for the title and summary to be made available to outside organization.

Name of Student | Signature of Student

PPG Dinesh Asanka | Date:

University of Moratuwa, Sri Lanka.
Electronic Theses & Dissertations
www.lib.mrt.ac.lk

Supervised by | Signature of Supervisor

Prof. Asoka S. Karunananda | Date:

Acknowledgements

I would like to express my sincere gratitude to my supervisor Prof. Asoka Karunananda for guiding me throughout the project. He has given his fullest corporation to me whenever I sought for advice. Also, Prof. Karunnanda's teaching techniques helped me to consider problems in different avenues.

I would like to thank Dr. Subha Fernando for her kind co-operation as a course coordinator during the last stage of the research.

Since this research is legal domain specific domain, I had to acquire lot of knowledge on the legal domain. I was able to get support from two prominent lawyers. Mr. Weerasinghe and Mr. Madhurapperuma gave me lot of advices and showed me the direction for the research. So, I would like to pay my gratitude to both of these professionals for helping me dispute their busy schedule.

I must also thank all the members of the lecture panel. During the period of one year and three months, these valuable lectures helped me to think differently.

University of Moratuwa, Sri Lanka.
Electronic Theses & Dissertations
www.lib.mrt.ac.lk

For this research, I had to refer to many books and research papers as reference. I would like to thank all the authors of those publications.

Also, my batch mates helped me lot in many different ways. I would like to pay my gratitude for them. They made my life enjoyable during the course period.

I would like to place my gratitude to my loving parents and wife for always encouraging for me on higher studies. Specially wife and two kids sacrificed lot on behalf on me which words might not be enough to describe.

I would like to thank Ms. Madhushika Lewke Bandara for supporting me during the project by providing samples and instructions. Her immense support and comments gave me the courage to enhance the project features.

Shamil Saleem was my one of the responders to verify questions who is also a law student. Shamil offered me lot of support during the project not only by answering the question but also by helping in many legal aspects.

Also, I would like to appreciate Mr. Dimuthu and Ms. Nadeeka from the non-academic staff for support that they provided throughout the period of the course. I would like to extend my gratitude to members of other departments such as exam department.

Last but not least, I would like to thank all my colleagues and others who are not mentioned, for all the support extended to me. My co-workers helped me by covering office on calls and other work so that I will be able to focus much more on the research and the MSc. Therefore, without their dedication, the project would not have been successful.

University of Moratuwa, Sri Lanka.
Electronic Theses & Dissertations
www.lib.mrt.ac.lk

Abstract

Body of knowledge in legal domain is complex, unstructured, interconnected in many ways and constantly being updated. This has resulted in hindering general public to access the legal domain, and creating gap between society and the legal system. Legal professionals also find it difficult to refer to case with adequate depth and the breadth. This research has been conducted to develop Ontology for legal domain with a particular emphasis on constitutional law in Sri Lanka. The research has focused on 1978 constitute including the 19th amendment which was passed in 2015 May.

Semi-Automated mechanism was used to construct ontology. After ontology is constructed, subject matter expert has corrected the Ontology. End users has the option of querying the Ontology for simple text as that is the current procedure in the legal domain. End users have the option of execute queries against the ontology. Protégé has been selected as the development environment for propose ontological system. To enhance performance of the system, relational database management system was used to implement Ontology. This system can be used by Lawyers and students by running queries to get relevant answers to their questions. In addition to the answers to the question, this system will provide the appropriate legal act and any other relevant legal cases. Natural Language Processing and Text Mining techniques were used to identify user queries. Ontology has questions and user questions are matched against those questions. Each question is mapped to one or many incidents, content and cases. Also, Agent technology is used to extract updated legal documents content and updated to legal Ontology. To enhance the performance of the knowledge base, relational database management system was used. Two verifications were used. First, verify whether users question and knowledge base questions are correctly matched. Secondly, time taken to answer and correctiveness are taken as the parameters. When all the scenarios are considered ranking relevancy is at least than 60% percent and in some scenarios it is 100 percent.

University of Moratuwa, Sri Lanka.
Electronic Theses & Dissertations
www.lib.mrt.ac.lk

Contents

Abstract **vi**

Chapter 1 Introduction **1**

1.1 Prolegomena 1

1.2 Aim and Objectives 1

1.3 Background and Motivations 2

1.4 Problem in Brief 4

1.5 Novel Approach to Legal Research 5

1.6 Structure of the Thesis 5

1.7 Summary 5

Chapter 2 Computational Legal Domain - State-of-the-Art **7**

2.1 Introduction 7

2.2 Basic Sri Lankan Legal System 7

2.3 Constitution Law 9

2.4 Researches in the Legal Domain on Other Countries 10

2.5 Ontology Construction Methodology in Legal and Other Domain 12

2.6 Question and Answering System 15

2.7 Summary 18

Chapter 3 Incorporated Technologies **19**

3.1 Introduction 19

3.2 Ontology 19

3.3 Natural Language Processing 22

3.4 Multi Agent Technologies 28

3.5 Summary 29

Chapter 4 Novel Approach to Legal Domain Search **30**

4.1 Introduction 30

4.2 Hypothesis 30

4.3 Inputs 30

4.4 Output 30

4.5 Process 31

4.6 Non Functional Requirements 31

4.7 Users of the System 32

4.8 Summary 32

University of Moratuwa, Sri Lanka.
Electronic Theses & Dissertations
www.lib.mrt.ac.lk

Chapter 5 System Design of FAHoL **33**
5.1 Introduction 33
5.2 Ontology Construction 33
5.3 Ontology Design 34
5.4 Basic Queries 35
5.5 Executing Legal Queries 36
5.6 Matching Questions 37
5.7 Summary 38
Chapter 6 Implementation of the Proposed System **39**
6.1 Introduction 39
6.2 Challenges 39
6.3 Basic Search 40
6.3.1 Basic Queries 40
6.4 Identifying Classes for Ontology 41
6.5 Ontological Construction 41
6.6 Question Matching 42
6.7 Finding Answers 47
6.8 Summary 47
Chapter 7 Evaluation **48**
7.1 Introduction 48
7.2 Ontology Content 48
7.3 Evaluation of Basic Search 49
7.4 Question Matching Capability 50
7.5 Finding Answers 53
7.6 Performance 54
7.7 Summary 54
Chapter 8 Conclusion & Further Work **55**
8.1 Introduction 55
8.2 Further Work 55
8.2.1 Improving Knowledge Base 55
8.2.2 Improving Technical & Research Area 56
8.3 Conclusion 56
8.4 Summary 56
References **58**

University of Moratuwa, Sri Lanka.
Electronic Theses & Dissertations
www.lib.mrt.ac.lk

Appendix A Script to Retrieve Simple Search **63**
Appendix B Code to Get Question Type **65**
Appendix C Code to Find Similar Question **67**
Appendix D Code to Find Global Terms **70**
Appendix E Calculation of Cosine Distance and TF-IDF **71**
Appendix F Question Type Ranking **73**
Appendix H Example of a Question **75**
Appendix I Accuracy of SOUNDEX **76**
Appendix J E-R Diagram **83**
Appendix K Overview of Full Text Search Architecture of SQL Server **84**
Appendix L Interfaces of HeainOnline Legal Systems **85**
Appendix M Configuration of Full Text Search in SQL Server **86**
Appendix N User Interface of Proposed System **88**

University of Moratuwa, Sri Lanka.
Electronic Theses & Dissertations
www.lib.mrt.ac.lk

List of Figures

Figure 2.1: Ontological Construction Method Proposed by Akmar 14
Figure 2.2: The Constitution Ontology 15
Figure 2.3: Ontological-Based Question Answering System 16
Figure 2.4: Flow Chart for Question Matching 17
Figure 3.1: The Component Diagram of Minerva 21
Figure 5.1: Semi-Automated Ontology Construction for Legal Domain 34
Figure 5.2: Ontology Design 35
Figure 5.3: Question Answering Architecture 37

University of Moratuwa, Sri Lanka.
Electronic Theses & Dissertations
www.lib.mrt.ac.lk

List of Tables

Table 3-1: Text Representation and Enabled Analysis 22
Table 3-2: Different Measures and Its Properties 24
Table 3-3: Methods of Calculating Term Frequencies 25
Table 3-4: Algorithms to Find Similar Terms 26
Table 3-5: Algorithms to Find Similar Terms 27
Table 3-6: Logic for SOUNDEX 28
Table 6-1: Question Type and Their Counts 43
Table 6-2: Excluded Terms 44
Table 6-3: Abbreviations 45
Table 6-4: Similar Terms 46
Table 7-1: Output Count for Different form of word “Person” 49
Table 7-2: Different form of word “Person” in Same Article 49
Table 7-3: Questions Returned for Scenario 1 51
Table 7-4: Questions Returned for Scenario 4 52
Table 7-5: Questions Returned for Scenario 5 52
Table 7-6: Performance Difference between Protégé and RDBMS 54

University of Moratuwa, Sri Lanka.
Electronic Theses & Dissertations
www.lib.mrt.ac.lk