ONTOLOGICAL SOLUTION FOR SRI LANKAN LEGAL SYSTEM

Parana Palliya Guruge Dinesh Asanka

(139109R)

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)

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:

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.

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.

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.

Contents

List of Figures

List of Tables

Introduction

1.1 Prolegomena

In the past century, there are a major improvements in the artificial technology techniques as scientists and researchers were looking for avenues to improve the effectiveness of usage of the technology. Artificial Intelligent (AI) techniques are used to resolve many complex real world problems which will be difficult to solve with legacy or traditional techniques. With the increase in popularity of AI, numerous AI techniques including Multi Agent System (MAS), Artificial Neural Network (ANN), Genetic Algorithms, Expert Systems, Data Ware Housing, Data Mining, Natural Language Process (NLP) and Ontology are identified. From the above list of AI techniques, Ontology techniques are provided for effective solutions for large number of knowledge base systems [6]. Ontology techniques are used to solve some real world problems such as Military, Legal etc.

Having recognized the legal system as an inherently complex system, this research has been conducted to implement ontology system for legal domain in Sri Lanka. In this connection, this chapter presents aim and objectives, background and motivation, problem in brief, novel approach to legal domain and structure of the thesis.

1.2 Aim and Objectives

The aim of this project is to develop ontology for legal domain in Sri Lanka. In order to reach this aim, following objectives are identified.

1. Critically study about Legal System in Sri Lanka.

2. Study about the current computational models of legal domains in other countries.

3. Critically review how ontology has been used in legal domains and other domains to solve problems.

4. Provide basic searching facility to search legal text.

5. Design and develop ontology system for Sri Lankan 1978 constitution by semi-automated process.

6. Provide answers to legal queries.

7. Evaluate the ontology by using known legal cases.

1.3 Background and Motivations

There were many legal contradictions and controversies have taken place in Sri Lanka. In 2005, there was a debate about when to hold Presidential Election. Main opposition parties in Sri Lanka argued that it should be hold in 2005 taking into account that previous election was held in 1999. However, then government challenged that it should be held in 2006 taking into account that then president, Her Excellency Chandrika Bandaranayaka Kumaranathunga took her oath in 2000 as her second term [44] citing the third amendment of the 1978 constitution. In the 1978 constitution, when the current president, re-elected as the president, he or she has the option of not taking the oath immediately after the president election according to the third amendment. However, opposition argue that President took the oath just after the election held in 1999. Since there was a contradiction, Supreme Court decided to take decision considering the obligation of the constitution. This was a practice taken by Karnataka courts in India in which when there is a contradiction, main objective of the constitution should be considered. This debate open the authors' interest about the legal domain in Sri Lanka.

Also, in 2001 when United National Front government led by then Prime Minister Hon. Ranil Wickramasinghe came to power, first time after 1977, powers of ministry of defense was vested to the different personal other than the president. In 2001, then President, Her Excellency Chandrika Bandaranayaka Kumarathunga agreed to appoint Mr. Thilak Marapana as the defense minister. However, 2002 President changed her stance by requesting opinion of the judiciary. As judiciary gave their judgment in favor of President then again this legal debate came to the society.

Again in 2014, another battle between main three institutes of President, Parliament and Judiciary over removing then Chief Justice Mrs. Sriyani Bandaranayake has become a major topic in the society which question lot of legal contradiction. According to 1978 Constitution president has the power to appoint chief justice when the post is

vacant but powers to remove of chief justice is vested with the parliament. However, during the process of removal of the chief justice, lot of contradictions were prominent between the judiciary institute and the parliament.

This interest has become more prominent during the 2015 presidential election. In that election, there were lot of arguments about the powers of the President, Parliament Privileges etc. During this election, some of close allies of then president, His Excellency Mahinda Rajapakshe challenged that in the event that president lose the election, he has the privilege to be in the office until 2016. Even lawyers from president side argued for this cause. However, third amendment which was enacted in December 1982, clearly state that in the event of a different person wins the elections other than the current president, current president term is ending as soon as election commissioner announces the president election results. This shows how difficult to find answers to legal queries even for a people who are actively engaged in the politics for something which has clearly stated in the constitution.

In 2015 June, there was another controversial constitution legal debate arises during a political debate in Derana Television, Wada Pitiya (වාද පිටිය) [51] with regards to the appointment of Prime Minister Hon. Ranil Wickramasinghe by the President, His Excellency Maithreepala Sirisena. When the Prime minister was appointed, former Prime Minister Hon. D.M. Jayarathne was the prime minister. According to the constitution 43.3, The President shall appoint as Prime Minster the Member of Parliament who in his opinions is most likely to command the confidence of Parliament [11]. In this sense, according to senior lawyer Mr. Srinath Perera, this action taken by president is correct. However article 47 indicates, The Prime Minister, a Minister of the Cabinet of Ministers, any other Minister or Deputy Minister shall continue to hold office throughout the period during which the Cabinet of Ministers continues to function under the provisions of the Constitution unless he is removed by a writing under the hand of the President, resigns his office by a writing under his hand addressed to the President or ceases to be a Member of Parliament [11]. So constitution specialist Mr. Udaya Prabath Gammanpila argued that when the new prime minister was appointed, it was not vacant either of three provision provided in the constitution.

There are couple of researches done outside of Sri Lanka on Legal domain ontology construction. Both were done in Italy as Italy is historically have contributed much to

the domain of legal by means of Roman law. Many research papers were published by Aldo Gangemi who is full professor at LIPN, University Paris and associate researcher at Italian National Research Council, Rome.

1.4   Problem in Brief

Legal research is the process of finding the law that governs activities in human society. It involves activities in human society. It involves locating both the rules which are enforced by the state and commentaries which explains or analyzes these rules. This procedure is necessary in legal practice to determine both the impact of past actions and the implications of contemplated actions. Though, researches and lawyers find the sources needed to predict, how courts will act, and to persuade them to take particular action.

Applying the law to a particular situation requires expertise in legal analysis. A lawyer is able to analyze a factual situation and apply legal doctrines developed in an analogous or similar situations. These doctrines may conflict with each other, making it necessary to weigh their importance and relevance to a particular problem. It is this expertise that most of the law school curriculum seeks to develop.

Finding the proper law to apply, on the other hand, requires expertise in legal research. Skill in the research process requires, understanding which resources to consult for each situation, as well as the purposes and relative value of the resources available, experienced legal researchers know which sources are authoritative or useful for what purposes.

Legal research involves the use of a variety of printed and electronic sources. The printed sources include court decisions, statutes, administrative document and scholarly commentaries. Computer databases containing these and other materials have dramatically changed the nature of legal research and improved its effectiveness. The volume and variety of legal literature continues to grow making the research choice of tools and tactics for each problem more difficult.

A thorough understanding of available legal resources, both published and computerized is necessary. There are as many procedures as there are problems and no single approach can work every time. Most of the tool available are search which has

naïve semantic features. However, those tools do not provide answers to user queries instead it provides text and semantic searching.

1.5   Novel Approach to Legal Research

Ontology method was proposed as the knowledge presentation technique for the domain of legal system in Sri Lanka. Ontology is defined as partial specification of conceptual vocabulary used for formulating knowledge-level theories about a domain of discourse [10]. As said in the previous section, legal domain has lot of documents such as constitutions, acts, gazettes, ordinances, cases etc. special storage techniques are used. Ontology can be used to store these information.

When Legal information is stored in Ontological structure, it has the benefit of integrate with the other Ontologies as and when necessary.

Also, Natural Language Processing techniques are used to refine queries asked by end users as there can be different types of questions asked by end users.

1.6   Structure of the Thesis

Rest of the thesis is organized as follows. Chapter 2 critically review the domain of legal system in Sri Lanka and current computational implementation to solve legal domain highlighting the current issues and limitation defining the research problem. Chapter 2 also contains the other domains where ontology is used. Since this research also deals with question matching, that areas also covered in the same chapter. Chapter 3 describes essentials of Ontology technology showing it's relevant to solve the legal domain. Also same chapter contains essentials of natural language processing with respect to this research. Chapter 4 presents the novel approach to Legal System with Ontology system, presents Chapter 5 is the design of Ontology and design of the system. Chapter 6 contains implementation of the system. Chapter 7 reports on evaluation of the new solution with the real world legal cases. Chapter 8 concludes the outcome of the research with the note on future work.

1.7   Summary

This chapter describes the full picture of the whole research project showing background, research problem objectives, hypothesis and novel approach. The chapter provides the motivation for this research with the many legal cases and debates.

Next chapter contains the current legal domain work with respect to technology. Also, next chapter discusses about the Ontology development on legal domain and other domains.

Computational Legal Domain - State-of-the-Art

2.1 Introduction

To start with the research, it is important to find out what is the current state of the research in Sri Lanka as well as in other countries. As legal domain is very vast and complex, special techniques needs to be considered. Before moving into the technical analysis and the implementation of the legal domain, it is required to analysis the structure of the legal domain in Sri Lanka.

Literature survey was carried out in four aspects. They are,

- Basic Sri Lankan Legal System

- Research carried out in the legal domain on other countries

- Ontology development on legal domain and other domains

- Question Answering Systems

2.2 Basic Sri Lankan Legal System

Sri Lankan regal system divided into four areas depending on the legal areas [1]. They are;

- The Constitutional Law – The Constitution of the Democratic Socialist Republic of Sri Lanka was introduced in 1978, replacing the 1972 constitution. This Constitution is an amalgamation to the Westminster model with the French presidential system. There is an executive President with executive powers but also a Prime Minister and a Cabinet. Parliament is the unicameral legislative system which is mainly responsible finance. 1978 Constitution has been amended by nineteen times up to now. In this research, Sri Lanka Constitution, 1978, including up to the eighteen amendment (excluding 12th amendment which was not enacted) [11] were used to build the Ontology.

- Civil Offences and Common Law - Offences against one person or group of people by one person or group of people.

7

- Criminal Offences and Law – Criminal law and the law of evidence are statutory. The early statutes were copies of Indian enactments which were drafted for use in British India. The Criminal procedure code, 1898 which was a copy of the Indian enactment was replaced by the administration of Justice Law, 1973, which in turn was replaced by the coded criminal Procedure Act, 1979. The influence of English Law is apparent and over-powering in the Criminal Procedure Code, 1898 and continues, but to a lesser degree [7]. Any offence against the society. Murders, robberies, kidnaps, rapes etc are considered to be Criminal Offences and Law [1].

- Commercial and Mercantile Law – Principles adopted from England, which repealed the Roman Dutch law, apply in relation to bills of exchange, sales of goods, partnership, companies, insolvency, banks and banking, maritime matters, carriage of goods and insurance.

Many different systems of law have been used to develop of the law of Sri Lanka. These are Sinhalese Law, Buddhist Law, Hindu Law, Tesawalamai Law, Islamic Law, Mukkuvar Law, Roman Dutch Law and English Law [7].

- Kandiyan Law

The Sinhalese legal system was developed over many centuries. Local and indigenous customs, canonical writings, practices and rites of Buddhism, Hindu Laws and customs, Sakyan and Mauryan customs, other Indian customs and the laws and customs which the original Sinhalese brought with them to Sri Lanka and had in varying degrees influenced the laws of Sinhalese. According to the Balasingham, those families who have long lived rooted to the soil of any provinces where Kandiyan Law prevailed and speak the language and follow the customs there prevalent, may be regarded as Kandiyans [9]. Though the provinces are not definitely defined for the Kandiyan Law, The Kandyan Marriage and Divorce Act of 1952 refers to the Central, North Central, North Western, Sabaragamuwa and Uva provinces [7]. Kandyan Law today has restricted application and applies in some respect in the law of persons, property and succession. The Kandyan law being customary law was mainly concerned with such matters like slavery, caster, land tenure, law relating to family relations and intestate succession to property [7].

- Thesawalame Law – The Tamil word Thesawalme means the Custom of the Land [7] which means that Thesawalem law predominantly based on Land. The Thesawalame applies to all persons who came within the description of the Malabar inhabitants of the province of Jaffna [7]. Though this says Jaffna is a province, it is a district not a province. During the dutch period the term Jaffna province is applies to the Jaffna peninsula including the islands.

  However, certain sections of the Thesawalame Law apply to lands in the Northern Province of Sri Lanka irrespective of the race or nationality of the owner.

- Muslim Law (Islamic Law) - Muslim law is based on Holy Quran and Muslim law is pure personal law and it does not have a territorial requirement. It is difficult to suggest that the legal principles of Islamic jurisprudence would be applied for Sri Lankan Muslims. Currently, there are three legislations on Muslim law in Sri Lanka [16].

  - Muslim Intestate Succession Ordinance No 10 of 1931.

  - Muslim Marriage and Divorce Act No 13 of 1951.

According to Abdul Rahman vs. Ussan Umma (1916), Islamic law applies only so far as it is consistent with the usages of the Mohammedans in Sri Lanka. The Muslim law applied in Sri Lanka contains usages and customs peculiar to local Muslims and which do not form a part of the Islamic system of jurisprudence.

2.3 Constitution Law

The Constitution of the Democratic Socialist Republic of Sri Lanka has been the constitution of the Sri Lanka since its declaration by the National State Assembly on 7 September 1978. It is Sri Lanka's second republican constitution and its third constitution since the country's independence in 1948. As of September 2010, it has been formally amended with eighteen amendments out of which seventeen amendments were passed as 12th amendment was not passed [26]. 1978 constitute of Sri Lanka contain 25 chapters.

2.4 Researches in the Legal Domain on Other Countries

Databases are excellent sources of good quality information. There are the most useful databases for English law. Several online databases are available for legal domain such as HeainOnline, Justic, Kluwer Arbitration, Lexis and WESTLAW [3].

HeinOnline: HeinOnline is a legal database providing a considerable collection of United States and international legal research materials in an online, fully-searchable, image-based format. This means all the documents in PDF format are just as they look as if in the original print [5].

HeionOnline provides subscribed users options of searching for different parameters. It provides the option of searching for text or phrase for content or title or author or state or country or author. HeionOnline has the option of searching in different sections such as Articles, legislation, editorials, comments, cases, notes, decisions, reviews and editorials. It also has the option of searching using the law enacted dates.

Justis: A legal database covering The Law Reports, Weekly Law Reports, English Reports, Daily Cases, State Trials, Justices, United Kingdom Statutes and Statutory Instruments. Includes repealed legislation.

Kluwer Arbitration: Offers access to the full-text of primary and secondary materials in the field of international commercial arbitration.

HeainOnline, Justic, Kluwer Arbitration, Lexis and WESTLAW has rich contents or legal material. However, searching is not intelligent as all the system are searching for simple searching.

Lexis: Full text law database containing scores of thousands of reported and unreported cases from the following jurisdictions:- England and Wales, 1900, Scotland, 1944, Northern Ireland, 1945, European Court of Justice, all cases; Also includes the full text of all in-force United Kingdom legislation, European Union legislation. Also covers the full text of fifty United Kingdom law journals, over 700 United States law journals, twenty five Canadian law journals and twelve Australian law journals.

Westlaw: Full text database of United Kingdom, European Union and United States legal materials. United Kingdom coverage includes all the cases from the Law Reports back to 1865. Also contains the full text of cases from nineteen other series of United

Kingdom and European Union law reports, as well as the full text of all United Kingdom and European Union legislation. Use Westlaw can be used to find specific cases or legislation, cases, journal articles or legislation on a particular area of law and e-copy of legal commentaries and practitioner texts.

Brief details of cases matching the party names you entered are displayed on the search results screen. Official transcripts of judgments are available for almost all cases along with a case analysis in the Metropolitan University, Westlaw system [4].

In a book written by Morris L Cohen, Kent C. Olso [2] indicates how legal research is done. This book is mainly on United States Law. In this book most important areas for this research is Legal Research Methodology, Preparing for Trial, How to Find Case Law Using the Digests, Finding Statutes, How to Research American Constitutional Law, Preparing for Trial. Since the proposed research is used to find answers to legal domain questions, it is essential to know what is the mechanism or real world approach to how the legal searches are done.

In Sri Lanka there is a web site called LawLanka [50] which gives users options of searching by providing key words.

It has following features,

- Acts of Sri Lanka
- Consolidated Legislative Enactments
- Amendment Acts from 1981 till date
- Sri Lanka Law Reports
- All New Law Reports Text Version
- Incorporated Amendment Act Navigation
- Section to Cases Navigation
- Case to Sections Navigation
- Powerful Text Search engine
- New Acts in Sinhala & Tamil Versions.

After searching in the above screen, following are the output in three different categories, Act and Ordinances, Consolidated Legislative Enactments and Case Law Reports.

Also, this system has the option of viewing the related cases [50] which is essential in legal domain as cases will be refereed as legal entities later. This system was studied to find out how should be user interface for the proposed system.

Constituteproject [25] is a project implemented to analyses worldwide constitutes. In this project there are options to compare different constitute with different versions. In this project, Amendment, Culture and Identity, Elections, Executive, Federalism, International Law, Judiciary, Legislature, Principles and Symbols, Regulation and Oversight, Right and Duties are identified as classes.

For each topic, there are sub topics as well. Those sub topics will have multiple topics which will create a hierarchical structure. For example, under Executive topic there can be multiple sub topics such as Cabinet, Executive Independence and Power, Head of Government, Head of State, Military and Structure of Executive etc. Under the sub-topic of Cabinet, there are another set of sub topics such as Cabinet removal, Cabinet selection, Eligibility for cabinet, Establishment of cabinet or ministers, Power of cabinet etc. This data was used to build the ontology for the proposed system as it provides base of the legal topics.

By using these topics, ontology can be mapped. However, as research is done on Sri Lanka constitute, Sri Lankan Constitute was included. For example, Federalism class is replaced with Provincial Council.

2.5 Ontology Construction Methodology in Legal and Other Domain

Research paper by titles Overview of Methodologies for Building Ontologies by Fernández López, M. [49], indicates building the ontology has three steps.

- Ontology capture, which means:

  - Identification of key concepts and relations in the field of interest, which is the scope. It is important to center on the concept such as, but not the words representing them.

  - Construction of precise unmistakable classification text for such concepts and relationships.

  - Identification of terms to refer to such perceptions and relations.

The authors use a middle out approach to perform this step and acclaim that rather than looking for the most all purpose or the most specific concepts as key concepts, the most important concepts be identified, which will then be used to obtain the remainder of the hierarchy by generalization and specialization.

- Coding.

Involves explicitly representing the knowledge acquired in step 2.1 in a formal language.

- Integrating existing ontologies.

During either or both of the capture and coding processes, there is the question of how and whether to use ontologies that already exist. Key concepts identification is helpful for the proposed research as it is key task in the proposed research. However, this research paper does not provide steps to identify the key concepts.

R. Subhashini's survey paper on ontology building methodologies describes about different domain Ontology is used [10]. Natural disaster management system, medicine, military intelligence, cooking, agriculture, auto mobiles are the some of those areas. Though the legal domain is not discussed in this research, it was helpful to acquire knowledge of ontology definitions.

Swartout et. al divides ontology into two classes which are domain ontology and theory ontology [12]. Domain ontology deals with the formal description of the classes and the association between classes while theory ontology tends to be abstract and slighter [14]. The proposed research is domain ontology. As this research paper indicates, legal domain needs define classes and their associations.

The majority of existing ontologies have been generated manually. Generating ontologies in this manner has been the normal approach undertaken by most ontology engineers. However, this procedure is very time consuming, error-prone, and poses technical hitches in maintaining and updating ontologies. For this reason, researchers are looking for other replacements to produce ontologies in more efficient and effective ways [14]. Main difficulty in generating Ontology in automatic or semi-automatic method is identifying the classes [12].

Maizual Akmar Ismail has proposed five step iterative process for Ontology construction [13] as shown in Figure 2.1.
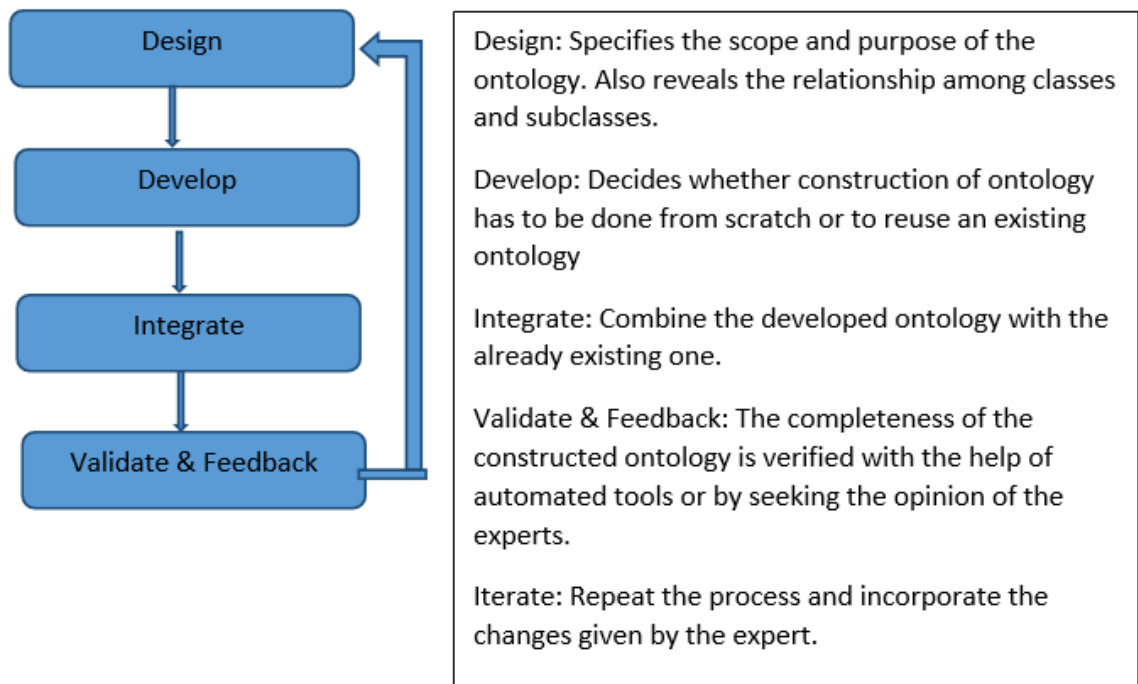
Figure 2.1: Ontological Construction Method Proposed by Akmar

When the Akmar model is considered for this research, at the develop stage, it is identified that ontology needs to be built from the scratch as there is no existence of Ontology for Sri Lankan legal domain. As discussed before, at the validate step several contradiction cases needs to be identified. In this research, there is no integration step as integration was identified as a future work for this research area.

Research paper titled Constitute: The World's Constitutions to Read, Search, and Compare [48] by Zachary Elkinsa et all describes importance of Ontology driven constitution as shown in Figure 2.2.
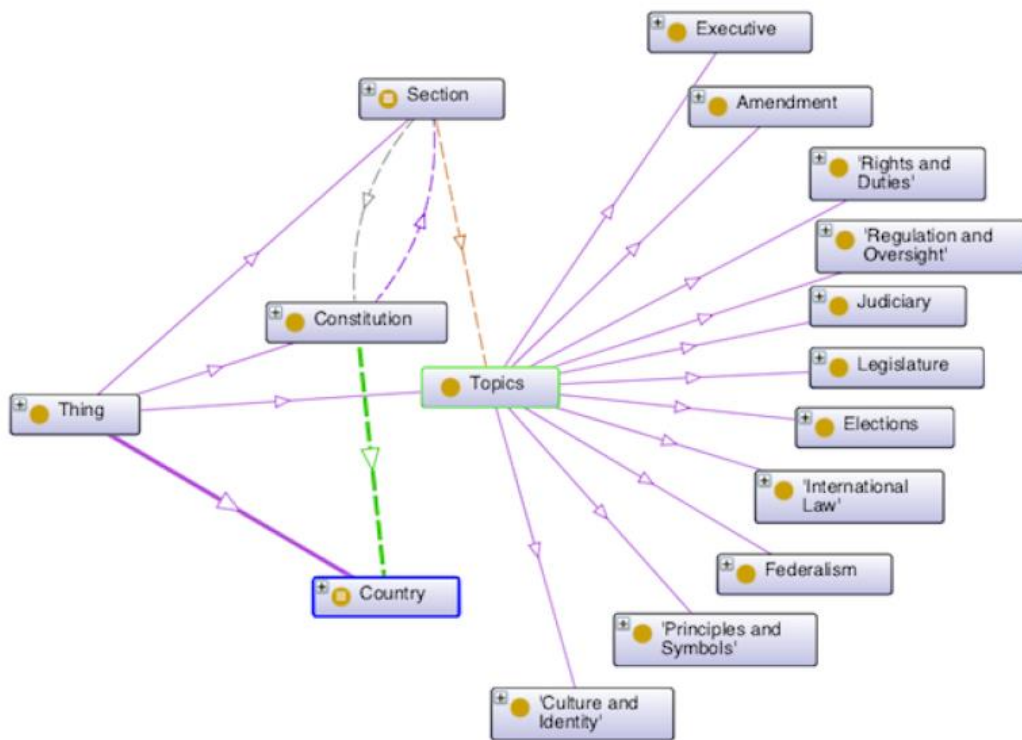
Figure 2.2: The Constitution Ontology

However, this research does not indicate usage of this ontology. Nevertheless, this research was helpful to build the ontology as it has basics for legal topics and sub topics. However, all this topics needs to be verified to suit to Sri Lanka context. For example, Federalism is not a legal topic in Sri Lanka. This needs to be verified with Provincial council which is the closest in Sri Lanka.

2.6   Question and Answering System

In the research paper about Ontology based Question Answering System, an architecture was proposed [27]. In this research paper, it shows the path for ontology lookup for 4 types (concept completion, example, enablement, and goal orientation). For each of the type, they have observed diverse search paths. For example, after searching a class for questioned entity, three kinds of paths are possible. The path selection is determined by query type. In determining search path, we should also consider the type of questioned entity (class instance). In this paper, it is can be seen

that the path for concept completion type is divided with two paths according to the result of class or instance discrimination. It architecture is shown in Figure 2.3.
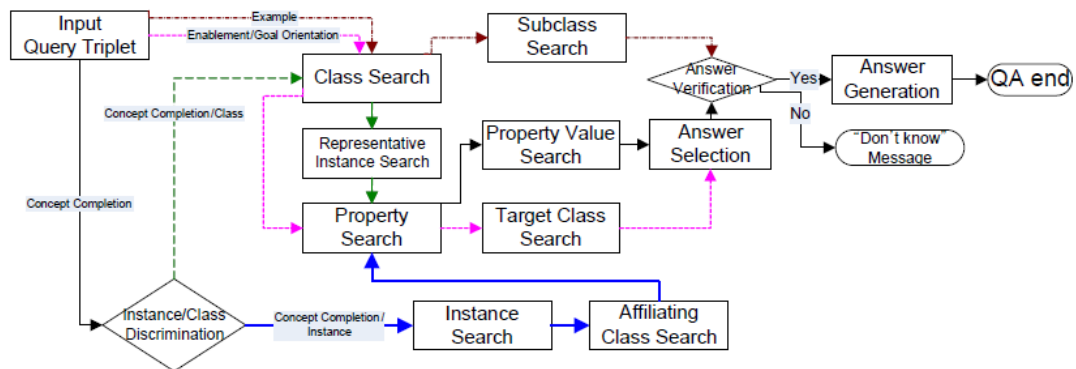
Figure 2.3: Ontological-Based Question Answering System

Above research is very close to the proposed research as this uses Ontological knowledge base. To get an answer from the above system, questions should be in a given format. However, in the proposed system, questions are entered by the users therefore, it is difficult to set a format or structure to questions.

In another short paper named, Rich Lexical Knowledge based question and answering system for Ubiquitous Knowledge Service, it describes a simple domain specific architecture for ontological question answering system [28].

Above question and answering system is based on two sources of knowledge which interact:

- Lexical data and in particular lexical semantics, and lexical inference.

- The domain data as represented by the rich conceptual functions, i.e. Rice Ontology, Some general purpose knowledge, useful for answering questions.

In another research paper by Lakshmi Palaniappan, Dr. N. Sambasiva Rao, they have proposed simple architecture. In the Proposed architecture, dining domain was selected to illustrate. User query will be passed through annotator of named entities, then queries will be checked whether it is synchronized, after test by reasoner, the answer will be searched in the text pool and it is retrieved. Instead of going to database query, template process queries faster. Because of this approach reusability is enhanced largely.

In this research, another important mechanism is to match the questions in the Ontology with the question that users are entering. Research paper by Jing Yu and others published a research paper [30] is to identify same question from question bank. Vector space model (VSM) is a common method for measuring test questions similarity in massive item bank system proposed by Salton in 1970s [31]. It is relatively old algorithm which was used in measuring text similarity. Though this algorithm is easy to be applied, it has ignored the relations among words in the documents and only uses words frequency. In this research paper, better solution was proposed as shown in Figure 2.4.
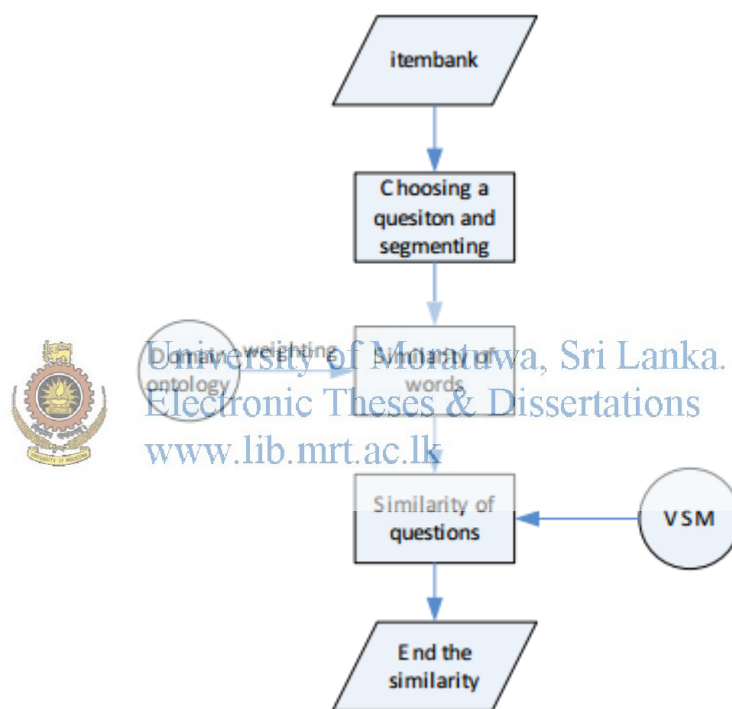
Figure 2.4: Flow Chart for Question Matching

However, as indicated by authors themselves, this method is suitable for simple question search. In the area of legal domain, questions are highly complex and it needs to improve the above method.

Also, in another research [33] by the researcher of this research, PPG Dinesh Asanka, cosine distance was used to identify using TF-IDF techniques. It has few cleaning steps to get more accurate data.

In this research paper, it has shown metric to find closely matching document using cosine distance so that users can identify the closely matching documents with the provided rankings.

Diagonal should have a value of Cosine Distance to 1 as it is comparing same document. Red cells are the documents which are closely identical documents. So in the given matrix, red cells indicate that they are closely identical documents, while white cells are the documents which are loosely identical documents. Other cells show, how closely those documents are matching with each other. This research is somewhat close to the proposed question matching technique. Since it is needed to match questions in this research, same techniques such as TF-IDF and cosine distance will be used. However, in this research simple term technique is used. Because of this, error rate is high. This method is quite fair for document searching as document is having large number of words. In the proposed research, it is supposed to match questions. Questions does not have large number of words. Hence, semantic searching needs to be included. Also, this research does not cater for spelling errors. In the proposed research, since users are entering questions, there is a high chance of spelling errors which needs to be considered.

2.7   Summary

In this chapter, literature review was done by giving prominent to four areas. Initially, basic Sri Lankan law was identified in detail. Since this research is focusing on constitute law of Sri Lanka, 1978 constitute was analyzed with up to the 18th amendment. Then the technical review was done with respect to current state of research in the other world and ontology construction techniques. Then the Ontological solutions were analyzed as this research is using Ontological knowledge base. Since this research is on answering questions, lastly review was done in the question answering system. By analyzing these literature, it was identified that all these areas needs to be amalgamated for the proposed research.

Next chapter discusses about the technologies used for this research, mainly Ontology and Natural Language Processing.

Incorporated Technologies

3.1 Introduction

This chapter discusses about the main artificial intelligence techniques that will be used during this research. Since this research is mainly using Ontology technology, major emphasis will be given to the Ontology. Apart from using Ontology, Natural Language Processing Techniques needs to be used to some extent. Hence NLP techniques will be discussed accordingly. To improve, distributive of the proposed system, agent technology was used minimally. Some of multi agent techniques are used to update the legal knowledge base. Hence multi agent technologies are also discussed in short.

3.2 Ontology

Ontology is an important evolving discipline that has the huge potential to improve information organization, management and understanding. Gruber defines an Ontology as an explicit specification of a conceptualization, i.e. a view onto a system is described by a standardized terminology and relations between entities. In many cases, there is no single ontology, but many different and concurrent ones [24]. Ontology describes the common vocabulary for researchers who need to share information in the area. It includes machine-interpretable definitions of basic concepts in the area and the relationship between them [20]. It has a crucial role in ensuring, on the basis of content access, compatibility, communication and providing a new level of service to the next generation of web converting a Semantic Web [6]. An ontology explicitly and formally defines the concept and relations that may exist in a given domain. A common vocabulary and frame is provided in an ontology so that researches can share, test and modify the conceptualization [19].

There are few definitions for Ontology by various scientists as stated in the following.

T. R. Gruber explained that "An ontology is a recognized explicit description of a shared conceptualization" [40]. M. Uschold, M. Gruninger: "An ontology is a shared understanding of some domain of interest." [41]. H. Herre et al.: "Formal Ontology is the science that is concerned with the systematic development of axiomatic theories

describing forms, modes, and views of being at different levels of abstraction and granularity" [42]. G. Pickert: "An Ontology is a six-tuple, O:=(C, R, F, G, H, A)" [43].

With:

C: A set of concepts

R: Set of binary relations on C (Relationships between concepts)

F: Function connecting symbol sets to sets of terms

G: Function connecting symbol sets to sets of relations

T: Taxonomy for the partial ordering of C in generic concepts / narrower concepts

A: Set of axioms (constraints): statements with elements of C and R.

Ontologies are the basis for computer applications to communicate among each other and with humans. They even can help humans to communicate among each other [39]. Ontologies are always concerned with a specific domain of interest, for example tourism, biology or law. Ontologies, thus, form a basis for understand. In case of legal domain, there are lot of assumptions are made. Making the assumptions underlying the implementation of the domain allows you to easily change these assumptions, if our knowledge of the changes in the domain. Rigid assumptions about the world of coding in the programming language code makes these assumptions not only hard to find and understand, but also difficult to change, especially for someone with no programming experience. Moreover, explicit specification of domain knowledge are useful for new users who need to know what the conditions in the means [20]. This is the one of the main reasons to build ontology for Sri Lankan legal domain.

The issues pertaining to ontology producing, mapping and maintenance are serious areas that need to be learnt.

Ontology is a complex multi-disciplinary field that draws on the knowledge of the organization of information, natural language processing, information extraction, artificial intelligence, knowledge representation and acquisition. Ontology is gaining recognition and is advertised as a new technology that has great potential to improve the organization, management and understanding, in particular, where the ontology provides a common base of common understanding of specific areas that can be transmitted between humans and the application system, it has a significant impact on

the regions involved in a huge number of heterogeneous and distributed information through a computer [6].

Ontology learning began to emerge as a sub field of ontology technology in the rapid growth of web documents and best practices shared information retrieval, machine learning, natural language processing and artificial intelligence communities. Most of the existing ontologies were obtained manually. Generating ontology so the normal approach was implemented most engineers' ontology. However, this process is very time consuming, prone to error, and creates a problem in maintaining and updating the ontology more effective and efficient ways.

Protégé 4.3.0 (Build 304) is used for the research project to build ontologies. However, since Sri Lanka constitute contain lot of documents and classes protégé may not be better solution when performance is considered. Research paper by Stijn Heymans and others [45], suggests different approaches to accumulate ontology to the relational database management system structures. Figure 3.1 shows those approaches in relational databases by investigating an improved triple store, namely Minerva [46].
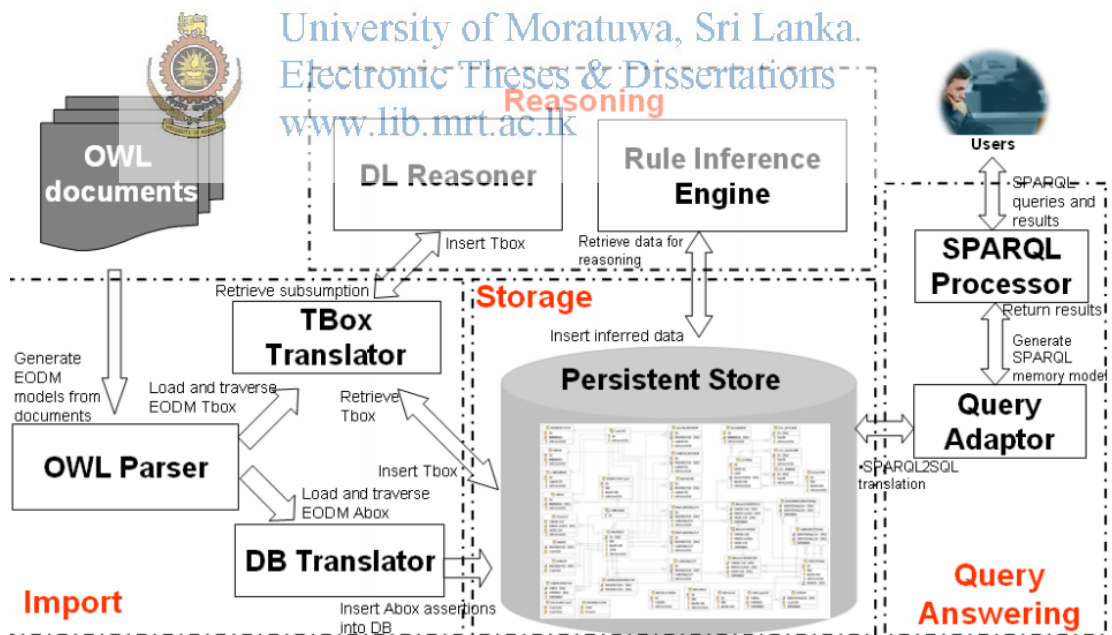

Figure 3.1: The Component Diagram of Minerva

Above diagram of Minerva, which is included of Import Module, Inference Module, Storage Module (viz. an RDBMS schema) and Query Module. The import module consists of an OWL parser and two translators. The function of the translator is twofold,

one is to fill all asserted axioms into a Description Logic (DL) reasoner and the other is to obtain inferred results from the DL reasoner and insert them into the database.

Another research paper by Carmen Martinez [47] indicates the importance of using database technology to store Ontology information. This research paper also emphasis that for large Ontologies, it is efficient to use database technology as databases has their own capabilities such as indexes, database statistics etc.

3.3 Natural Language Processing

Natural language processing (NLP) has been noticed in the field of computer science in the past few years. Lots of attempts have also been made in the field of NLP to convert the natural language processing text into computer programs. This conversion system has been so wonderful that the NLP identifies conditionals, loops and other aspects in the procedural programming. It then builds the program from natural language text.

In this research, NLP will be used mainly to identify end user queries. To match the questions of users to the knowledge base queries, engine will raise questions from the end user. For both of these NLP will be used.

In NLP there are different text representation and depending on the text representation analysis is different as shown in the Table 3-1 [52].

Table 3-1: Text Representation and Enabled Analysis

Text Representation	Enabled Analysis	Examples of Application
String	String Processing	Compression
Words	Word relation analysis Topic analysis Sentiment analysis	Thesaurus discovery Topic and Opinion Related applications
+ Syntactic Structures	Syntactic graph analysis	Stylistic analysis Structure based feature extraction
+ Entities & Relations	Knowledge graph analysis	Discovery of knowledge and opinions about specific entities

	Information network analysis	
+ Logic predicates	Integrative analysis of scattered knowledge Logic inference	Knowledge assistant for biologists

String text representation is the most general representation and it is robust. String representation mostly used for applications like compression. Words representation used for lot of analytics techniques such as Topic analysis. Word representation is quite robust and relatively common. However, word representation is costly with respect to the performance. Next three presentations are additional presentations to either string or words representations. Syntactic Structures are used to structural analysis of statements. Entities and relations can be used to find the associations of topics etc. logic and knowledge base integration are used in Logic predicates. So in this research, it is obvious that words representations is used. Also to match questions Entities & Relations is used while to use legal ontology Logic predicates is used.

In this research, before the question and answer, basic semantic search needs to be performed. For that purposes, SQL Server 2014 full text search option is used. Architecture of the full text search is included in Appendix K Overview of Full Text Search Architecture.

Term frequencies can be used to calculate the distance between the documents. There are several ways to calculate distance between two objects. There are Euclidean distance, Cosine distance, City-block (Manhattan) distance, Chebychev distance and Mahalanobis distance [35]. Usually, Cosine distance method is used to calculate the distance between two documents [34]. Apart from above distance measures, there are other measures like Chi-square, Lift, AllConf, Jaccard, Cosine, Kulczynski, MaxConf etc [36].

Table 3-2: Different Measures and Its Properties

Measure	Definition	Range	Null-Invariant
$\chi^2(A,B)$	$\sum_{i,j=0,1} \frac{(e(a_i b_j)-o(a_i b_j))^2}{e(a_i b_j)}$	$[0, \infty]$	No
$Lift(A,B)$	$\frac{s(A \cup B)}{s(A) \times s(B)}$	$[0, \infty]$	No
$AllConf(A,B)$	$\frac{s(A \cup B)}{max\{s(A), s(B)\}}$	$[0, 1]$	Yes
$Jaccard(A,B)$	$\frac{s(A \cup B)}{s(A)+s(B)-s(A \cup B)}$	$[0, 1]$	Yes
$Cosine(A,B)$	$\frac{s(A \cup B)}{\sqrt{s(A) \times s(B)}}$	$[0, 1]$	Yes
$Kulczynski(A,B)$	$\frac{1}{2}\left(\frac{s(A \cup B)}{s(A)} + \frac{s(A \cup B)}{s(B)}\right)$	$[0, 1]$	Yes
$MaxConf(A,B)$	$max\left\{\frac{s(A)}{s(A \cup B)}, \frac{s(B)}{s(A \cup B)}\right\}$	$[0, 1]$	Yes

i – i^{th} document

j – j^{th} document

A, B – Two different set of document attributes.

A U B - Number of attributes in of Document A and B

s – Size

max (s(A), s(B)) – maximum of size of attributes of document A or B.

There are several methods to calculate Term frequencies as shown in table 3-3.

Table 3-3: Methods of Calculating Term Frequencies

Method	Calculation	Description
Word Count	Y=X	Simply count the words
0/1 Bit Vector	IF Exists 1 else 0	Ignore Counts
Log Vector	Y = Log (1+X)	Avoid high number of frequencies weighs as high counts are penalized little bit.
BM25	$Y = \frac{(k+1)X}{X+k}$	Upper value is k+1 where k is a defined value to suite the data set. When k is very large Y = X.

For this research, word count is selected as the term frequency method.

Calculating Cosine distance for documents using Term frequencies may not be accurate as there can be common terms across the documents.

Hence, Term Frequency Inverse Document Frequency (TF-IDF) for each document each term is calculated using following formula.

$$TFIDF = TF * Log\left(\frac{N}{n}\right)$$

N - Total number of documents

n – Number of document Term Frequency contains in the documents [33].

Cosine Distance (CD) [33] is widely used to calculate the distance between two documents and shown in below formula.

$$d_c = \frac{\sum_{k=1}^{T} d_{ik}d_{jk}}{\sqrt{\sum_{k=1}^{T} d_{ik}^2 \sum_{k=1}^{T} d_{jk}^2}}$$

d_c – Cosine Distance

i,j – Documents

T – Number of Documents

In this research, there is a need to match the knowledge base terms and terms entered by users. There can be miss-spelled words entered by users. There are several algorithms to find similar terms as shown in Table 3-4.

Table 3-4: Algorithms to Find Similar Terms

Algorithms	Purpose
Soundex	Simple character substitution algorithm.
Refined Soundex	Slight enhancement to the original Soundex
Cologne Phonetic	Same as Soundex but the encoding is to suit German pronunciation.
NYSIIS	Starts to consider combinations of letters and their changing role depending on their relative positions within a word.
Daitch-Mokotoff	Intended to cater for Jewish and Eastern European names.
Caverphone	Intended to cater for New Zealand names. Largely a character substitution algorithm
Metaphone	Sophisticated algorithm sensitive to both combinations of letters and their relative positions in a word.
Double-metaphone	Returns two phonetic codes for a given input to allow for multi-lingual pronunciations.

Also, there are properties of each algorithms and depending on those attributes better algorithm to be selected [52].

Table 3-5: Algorithms to Find Similar Terms

Algorithm	Character Substitution	Start Of Word	End Of Word	After Vowel	Combinations	Case Sensitive	Input < Output
Soundex	Y	Y					
Refined Soundex	Y	Y					
CoognePhonetic	Y	Y					
NYSIIS	Y	Y	Y	Y	Y		
Daitch-Mokotoff	Y	Y		Y	Y		Y
Caverphone	Y	Y	Y	Y		Y	
Metaphone	Y	Y	Y	Y	Y		Y
Double-Metaphone	Y	Y	Y	Y	Y		Y

The last column indicates that the number of substituted characters can exceed the number of characters being replaced. Since there are only selected words in the legal domain SOUNDEX algorithm is selected as term matching technique.

SOUNDEX has a fairly simple rule set which is stated below.

Always retain the first letter

For all other letters up to the first non-alphabetic character replace them with numbers as in the table 3-6 [53].

Table 3-6: Logic for SOUNDEX

Number	Letter
0	A, E, H, I, O, U, W, Y.
1	B,F,P, V
2	C, G, J, K, Q, S, X, Z
3	D, T
4	L
5	N, M
6	R

- If any adjacent numbers occur in pairs or more only retain the first number
- Strip out the zeroes.
- If longer than four characters then truncate to 4 characters
- If shorter than four characters then pad out with zeroes.

Since Microsoft SQL Server 2014 is supporting SOUNDEX function, it was decided to use SOUNDEX function for this research.

3.4   Multi Agent Technologies

Usage of Multi Agent Technology is minimal in this project. Main use of Multi Agent technology is for couple of things. Multiple agents will be used to identify incremental changes of the legal documents, so that it can be updated to the legal ontology. Secondly, it will be used integrate multiple ontologies.

Multi Agent Technologies are used for many application as it has many features which can be utilized to improve the features of the system. Since MAS has the capability of communication and distribution [17], in this system MAS are used to integrate different low areas to update to Ontology.

3.5   Summary

Ontology. NLP technology and Multi Agent Technology areas were discussed in this chapter as they are the man techniques used for this research. Ontology is the main technology used but NLP and MAS also used to support to deliver the end user requirements. Term count is selected as the term frequency method whereas cosine distance was selected to measure the distance between the user question and the question in the question bank.

Approach to this research is discussed in the next chapter in detail.

Novel Approach to Legal Domain Search

4.1   Introduction

In the previous chapters, current practices of legal search were identified. This chapter discusses about the approach, authors have made to solve the issues of legal domain search. In this chapter, users inputs are discussed. System output is also discussed. Also, the process which makes the input to more readable output are also discussed. Apart from those details, non-functional requirements such as performance and scalability is etc. also discussed.

4.2   Hypothesis

Preliminary legal domain queries can be answered by using legal domain ontology.

4.3   Inputs

After carrying out few interviews with prominent lawyers, it was decided to provide basic searching options to the end users. According to lawyers, they start the legal search with the basic search. So in that sense, basic text should be taken as inputs.

Legal domain queries of the users are the input the proposed system. Since the queries are submitted in plain text, natural language processing techniques (NLP) needs to be used. If the question cannot be identified error message will be returned. That question will be store in ontology so that later this question can be added to knowledge base so that knowledge base can be improved further.

4.4   Output

Relevant answers will be provided to the user. Apart from the answer to the submitted query, relevant legal clauses as well as the previous legal cases will also be provided to the end users. Also, during the interviews, it was discovered that some incidents are also considered as part of a legal answers.

For the basic search, relevant clauses of the constitution, should be displayed to the end user.

4.5 Process

In this research, construction of Legal Ontology is semi-automated process to avoid user errors. After the Ontology is constructed, subject matter experts (SME) should verify whether the Ontology is correct. When it was verified, it is available for queries.

When the queries are passed, query will be analyzed using NLP techniques to identify the context. When the correct question is matched, answer to the question, relevant constitute contents, relevant cases and relevant incidents are matched.

Also, there are similar words used in legal domain. For example, in the Sri Lankan context, head of state, head of cabinet and president is referred to a single personal. So, those equal words needs to be processed. Also, most of the time users tend to user abbreviations. For example, PM is stand for Prime Minister. Therefore, those when users enter PM, system needs to match the correct word. Also, not all the time users will enter the correct spelling. Though there is no spelling checker included however actions should be taken to minimize the errors.

4.6 Non Functional Requirements

Since the legal domain will change time to time, it should be able to add or modify existing domain. Also, since court cases and incidents will also become part of laws, those court cases and incidents also to be included to the Legal Ontology.

Since some queries might have errors, system should be able to automatically correct them. Also, system have option of saving queries, so that when another user requested for a similar type of query, system should be able to retrieve the previously executed queries.

Ontologies should be able to update as and when the laws are amended. This modification is done incrementally which means when a law amendment is completed, ontology should update only the modification not the entire ontology.

Query output should be return to end user efficiently. Since there content of law domain is very large, output should process parallel to increase performance of queries.

Since constitute and cases has lot of contents, database technologies are used to improve performance. Also, primary keys, foreign keys, constraints and indexes are included to improve performance. Foreign keys and constraints are also used to improve data

integrity so that garbage data will not be able to save. Also, when the database design is considered, correct data types are used to considering the performance and the usability.

4.7 Users of the System

This system mainly aims at people who needs to solve basic questions of their needs. The system may not provide comprehensive answers to all the questions raised by end users. So the system is helpful for users who are looking to get basic idea about legal system. Also, this system may help professional lawyers so that they can understand what areas to look for.

4.8 Summary

This chapter discussed about the approach to design and implement the legal domain search for Sri Lanka. Input and outputs are discussed with the process. Also, this chapter includes users of the system and non-functional requirements of the system. Next chapter, discusses the system design of the proposed system.

System Design of FAHoL

5.1   Introduction

Since this system is mainly intend to provide domestic users to search for the legal advices before consulting an expert. Therefore this system is named as FAHoL (First Aid Home Lawyer).

This chapter discusses about the system design aspect of FAHoL system. Design of the system has adhered to the approach discussed in the previous chapter. In this chapter, design of the system is divided into, ontology construction, ontology design, basic queries and question matching.

5.2   Ontology Construction

To avoid or minimize human errors and make the ontological construction high performance, it was decided to follow semi-automated process to for ontology construction. After the Ontology is constructed, subject matter experts (SME) needs to manually verify the Ontology.

Time to time, there should be agent named Incremental Legal Content Extractor Agent to identify the legal domain changes. This is a scheduled agent and it also has the option of executing as administrator needs to execute. Since there can be different types of law documents like constitution, acts, incidents and legal cases this agent can be distributed multiple sub agents.
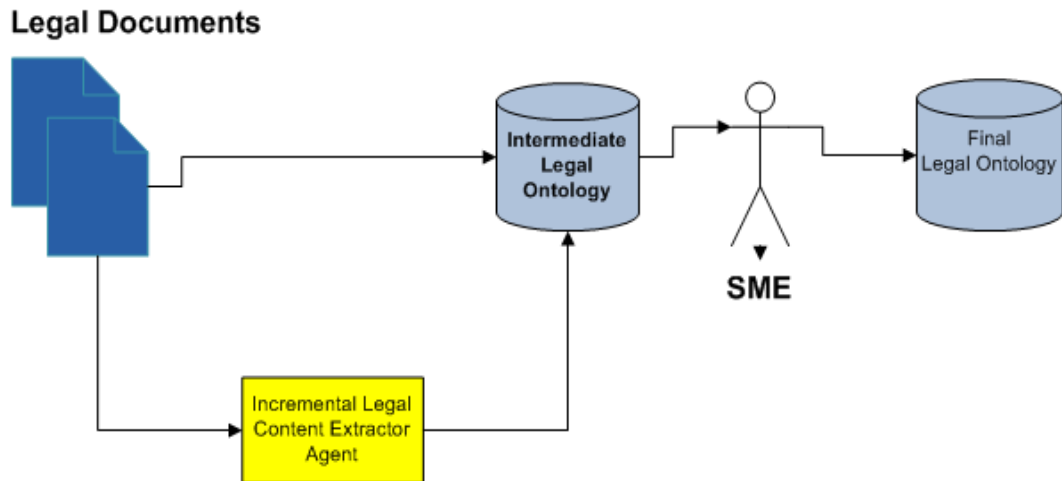
Figure 5.1: Semi-Automated Ontology Construction for Legal Domain

Figure 5.1 shows the design sketch of the semi-automated ontology updation. Legal documents needs to be in some defined format so that Content Extractor Agent can read. Then agent will update the ontology which will be read by a legal matter expert to finalize the Legal Ontology.

5.3 Ontology Design

Ontology is designed to with different legal content classes. In addition to classes, there will be questions and questions are linked to answers and sections, cases and incidents. So in the process part, questions need to be matched. In case, questions cannot be matched, it should be logged so that expert can analysis and update the knowledge base. Basic Ontology design structure is shown in the Figure 5.2.

Figure 5.2: Ontology Design
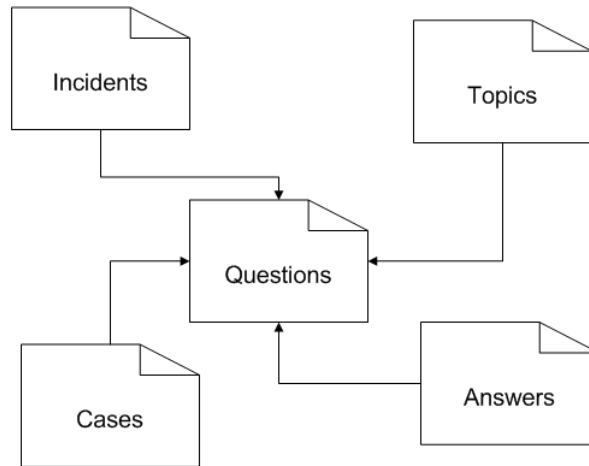
As shown in figure 5.2 questions will be linked to Answers, Cases, Topics and Incidents. Example is shown in Appendix E. That example shows how removal of members of parliament is linked to different types of contents.

Typically, one question will have one answer. However, for a question there can be multiple cases and incidents. Also, it might need to link to multiple topics in the constitution.

As one question can have multiple cases, similarly single case may be referred to multiple question. This means there is many to many relationship between cases and question. Also, there is many to many relationship between contents and questions. As content can be referred by multiple questions. Similarly, questions and incidents relationship is also a many to many.

5.4 Basic Queries

When the survey was carried out among the lawyers and law students, it was identified that they first do a basic search. However, in most cases, they do the word or simple text search. Since they do not have sophisticated techniques or tools, currently they are doing a simple text search only.

Currently, they do not have the options of getting the results for semantic searching. For semantic searching, Full Text Search option of Microsoft SQL Server 2014 is amalgamated to this this solution. Full-text queries perform linguistic searches against text data in full-text indexes by operating on words and phrases based on rules of a

particular language such as English or Japanese. Full-text queries can include simple words and phrases or multiple forms of a word or phrase. A full-text query returns any documents that contain at least one match. A match occurs when a target document contains all the terms specified in the full-text query, and meets any other search conditions, such as the distance between the matching terms [21].

To achieve the above goal, all the legal documents are stored in the database and searching options are will be provided to the end users.

5.5 Executing Legal Queries

User Interface will be provided to users to key in their legal queries. These queries validated with NLP techniques. As discussed in the chapter 3, Cosine distance technique is used. In case queries are invalid, user will be notified with the error. With that user has the option of saving the question, so that it will be revealed later to improve the questions. Existing questions in the knowledge base are listed out in Appendix A.

When the query is valid, using NLP and Ontology techniques relevance legal argument will be identified and returned to the user. Along with the legal answers, relevant legal documents, legal cases and legal incidents will be returned to the users. Figure 5.3 shows system flow and connections for the proposed regal research system.
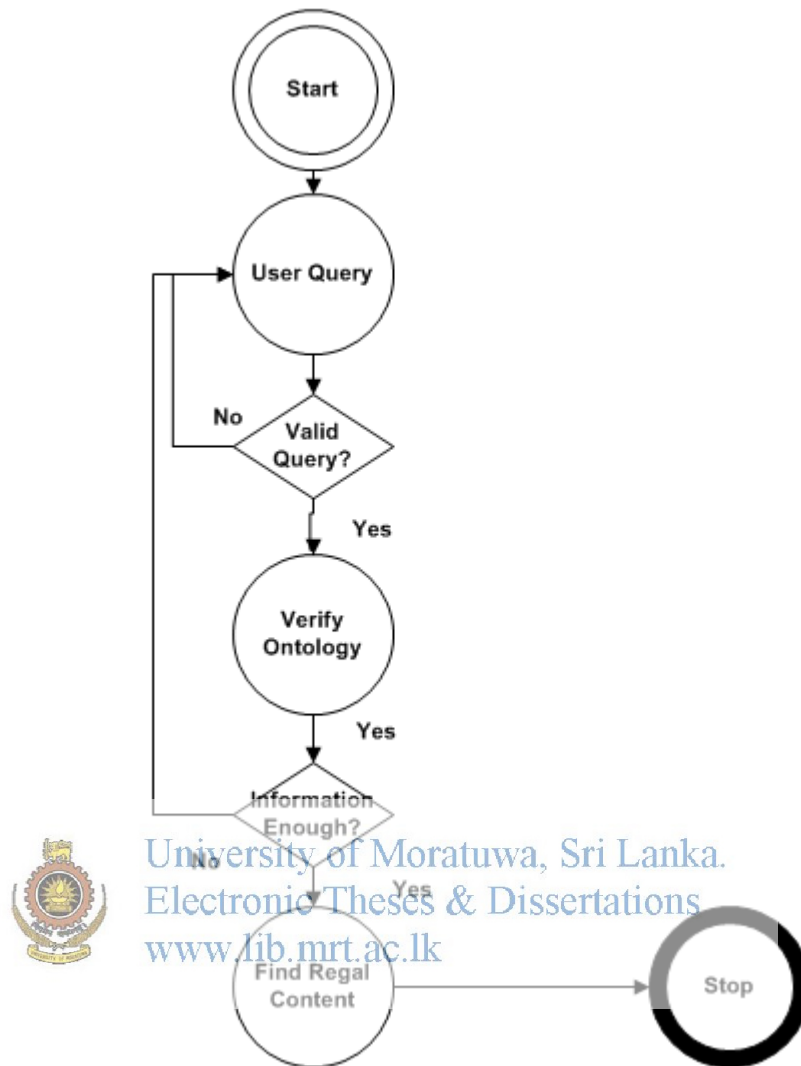
Figure 5.3: Question Answering Architecture

5.6   Matching Questions

Question matching is done in two steps. First, all questions in the ontology needs to be analyzed and second step is to match the user question with the existing questions.

In the first step, there are three sub steps. Initial step will capture global terms for all question with the frequency. Next step is to find out, terms for each question which is called term lookup. Third step is to compute TF-IDF for each question.

When the user enter question, first it needs to check for spelling. After the spell checking, closest questions will be listed. To find the closest question cosine distance is used.

However, question type also needs to be considered. There can be same terms used in multiple questions. Only different between those questions are objective of the question. That is which, what, how many etc. Every question in the question bank, will have a questions type. So when user enters the question, apart from the term matching, there needs to have a question type matching as well.

Since there are two rankings, one from terms and another ranking from question type. Question type ranking and term ranking is multiply to obtain the final ranking.

Final Ranking = Term Ranking * Question Type Ranking$^{0.5}$

n - is decided by doing few experiments as shown in Appendix F.

Typically, users use abbreviations and similar words. For example, most people uses MP for Members of Parliament and PM for Prime Minister. Also, users use similar words. For example, Presidents is Head of State and Head of Cabinet. So when users enter President or Head of State both times it is referring to the same person or object. So Ontology is used to store those information which is described in implementation chapter later.

5.7   Summary

This chapter discussed about the system design of the FAHoL. In this chapter, ontology construction and the ontology design is discussed. Also, usage of full text search technology is also discussed to support basic inquires. Finally, most complex query executions are discussed. Those queries support abbreviations and similar words. In the next chapter, implementation of FAHoL design which was discussed in this chapter.

Implementation of the Proposed System

6.1 Introduction

Up to now, input and output were discussed and then the detail discussion has been carried out on the design of FAHoL in varies stages. After the design, it is important to discuss the implementation of the said system. Since there are various options of implementation best suited options has to be selected in order to system is viable to use.

In this chapter, detail discussion has taken place about technologies used for implement the FAHoL.

6.2 Challenges

When it comes to legal domain, especially in Sri Lanka lot of Roman Dutch terms are used are Sri Lanka law is inherited from Roman Dutch Law. Since Natural Language Processing techniques are used in this research, those terms needs to be excluded.

For example, in some questions there are terms like *nulla poena sine lege*. Since "nulla poena sine lege" is non English terms, those needs to be excluded from the knowledge base. In case of legal domain, there are lot of latin and italin words where they should be either removed or replaced from the knowledge bases.

Another challenge that was faced during the implementation of this system is the volume of the content. Since volume of the legal domain is extremely huge, databases are needed to store legal content. Though protégé is great tool to simulate ontology however when it comes to storing of large documents performance will decrease [46] [47]. Hence, suitable database technology needs to be selected. Microsoft SQL Server 2014 Enterprise edition is selected as the database technology for this product has it has built in features to support large content.

6.3 Basic Search

Basic search option is to facilitate the basic search options for keyword and the semantic searching. Initially, entire constitution is saved in the database in the table structure given in the Entity Relationship diagram is in Appendix F:

In the Constitute_Chapters, all the chapter titles are stored to later reference. For each chapter all the contents are and the sub chapter titles are also stored in the Constitute_Details table.

To enable search for the law content, Full Text Search functionality is used. English language is enable to details column. Since Full Text Search does not support multiple language at same time. This is another reason to eliminate Roman Dutch words from the knowledge base. Details of configuring of Full Text search is included in Appendix M.

6.3.1 Basic Queries

As identified in the surveys and the interviews conducted typically lawyers search for exact match or textual match. CONTAINS key word search for the exact match. You can use CONTAINS, CONTAINSTABLE, FREETEXT, or FREETEXTTABLE to search for all the different tenses and conjugations of a verb or both the singular and plural forms of a noun or for synonymous forms of a specific word [22].

The RANK column returns a rank value for each row that indicates how well the row matched the selection criteria. The higher the rank value of the text or document in a row, the more relevant the row is for the given full-text query. Note that the different rows can be ranked identically. You can limit the number of matches to be returned by specifying the optional top_n_by_rank parameter.

With the results found in the survey, it was decided to implement basic application to support basic searching capabilities. This includes searching on legal contents and law cases.

As shown in Appendix P, search is made for the exact content. In the example given above, search was carried out for Persons key word and results were found at subsection 12.1, 12.4 and 22.5.

Similarly, same key word was searched for the Full Text Search option as shown in Appendix P. In this, more results we received than that of previous search. This is due to the fact that in the full text search option system searches for word Person, Persons etc.

6.4 Identifying Classes for Ontology

One of the main challenging task is to identify the classes for the legal ontology. GATE tool [23] developed by The University of Sheffield, Department of Computer Science, used to identify the classes needed for the Ontology. Since it is difficult to identify classes manually, GATE tool was used to identify classes.

GATE tool has some tedious configurations to be done. Some configurations are not straight forward. When the constitution content are analyzed from the GATE tool, it provides the classes for the Ontology which can be used to create the Ontology. However, it cannot be automated since, manual intervention is needed.

6.5 Ontological Construction

As said in the Chapter 5, protege tool was used to build the ontology. Answers, Question, Cases and Topics are the main classes in the legal domain ontology. Classes like area, constitution, row type are the classes identified from the GATE tool and from the constitute project [25].

There are sub classes in each Topic. For example, Principles and Symbols is a sub-class of Topics. State Definition and Symbols is a sub-class of Principles and Symbols. National anthem, national capital, National flag and National motto are sub-classes of State Definition and Symbols can be included in the Ontology for better results and information.

There are additional properties or annotation are added. For example, labels and descriptions are added. Label is added to user to get an idea about the class and the description is to tell the details of that National Anthem is added. However, it was identified during the surveys and interviews, that there are some key words attached to the every topic. For example, "Sri Lanka Matha (ශ්‍රී ලංකා මාතා)" or "Namo Namo matha (නමෝ නමෝ මාතා) " are mostly referred to the national anthem. So those key words are stored in the legal ontology so better searching capabilities.

Also, these topics are linked to questions. In the previous example of National Anthem topic, there is a question linked to it named anthem. Question Text of the question will have the entire question. This question will be matched with the user question. In the Answer class, answer test annotation is created. Questions are mapped to cases, incidents and answers.

There are around thousand questions in the law question bank. These questions are taken from standard question database and then modified to Sri Lankan context. For example, some countries like United Kingdom has vested special legal powers on church. However, in Sri Lanka 1978 constitute special priority is given from to the Buddhism from the chapter 2 [11]. Also, Sri Lanka federalism is not in place instead we have a different institutions called provincial councils and local government councils. So the questions are changed accordingly.

Questions are mapped to cases and answers. For cases, there is a label and date of the judgment (year) and brief description about the case.

Incidents are important elements as those have become examples when it comes to legal issues. There are lot of incidents such as amendments, deaths of MPs, resignation of MPs etc.

These incidents are linked to questions and each incidents has properties such as dates and descriptions etc.

6.6 Question Matching

As discussed in Design chapter, it was decided to use Cosine distance technique to match questions. Since Question bank is very large, for the better performance, it is better to use database to store the questions.

Since questions are in the Ontology, Question Update Agent is used to synchronize between Ontology and the database table. Question and Header is stored in the Question table. Cosine distance between users entered question and stored questions needed to be calculated. However, to improve performance by avoiding calculating the cosine distance component for stored questions, it is calculated and stored in advance so that performance is improved. In this, squired summation of TF-IDF of each term is calculated for each question and stored in the table.

As discussed in the design chapter, question type needs to be derived from the question. There are three types of questions types.

1. First word of the question. Example. Which provisions are not amendable? If questions are start with Does, Who, What, Are, Do, Is, When, Which, Whose, Can, Should, Was terms are falling into this category.

2. For some questions, you need to get the first two words to get an understanding of the question type. For example, how much, how many, under what, under which, in which, in what are the terms falling into this category.

3. Some questions question type determine at the middle which can be found after the comma (,).

Above implementation is done using a script in Appendix B. There are 675 questions in the question and 28 question types.

Table 6-1: Question Type and Their Counts

Question Type	Count	Question Type	Count	Question Type	Count
Are	39	At what	1	By what	4
Can	3	Do	7	Does	224
For how	1	For which	3	How are	4
How does	7	How is	3	How long	1
How many	11	How often	1	In how many	4
In what	2	Is	28	Should	2
To what	2	To which	1	To whom	1
Under what	6	Under which	1	Was	1
What	210	Which	28	Who	78
Whose	1				

At the initial stages, some questions were not able to assign to a question type. This was due to the way the question was formed. To support the above question type assigning questions were re-phrased. For example, questions like State the difference between provincial council and the local government councils was modified to what is the difference between provincial council and the local government councils. With this, 100% correctness was achieved for question type assigning.

GlobaleTerms will have Term for all questions. Since there are lot of words conduction words etc, those needs to be eliminated. Since these words should be configurable, separate table ExcludeTems is used. Table 6-2 contains excluded terms.

Table 6-2: Excluded Terms

A	All	An	Any	are	as	at	Be	Bit	But	By
Can	Do	Does	Due	Ex	For	From	Has	Have	ie	If
In	Is	It	Many	No	Nor	Not	Of	One	Or	Than
That	The	There	This	To	Use	We	Were	With	Year	you

When terms are captured, excluded terms are excluded. Also, frequency for each term is stored. As discussed in the Chapter 2 and Chapter 3, IDF is calculated for each term.

There can be similar terms as for example, amend, amendable, amending, amendment, amendments are same word with different formations.

SOUNDEX algorithm is identified to find the similarities of similar words as indicated in the Technology chapter.

SOUNDEX function is used in Microsoft SQL Server 2014 [37] to identify the SOUNDEX value of each term and stored it in the database.

In Appendix H, it can be seen that, for amend, amendable, amending, amendment, amendments same SOUNDEX value, A553 is used. Also, frequency is updated to number of soundex_values not mere term frequencies.

Details of SOUNDEX values is analyzed in Appendix F. According to the Appendix, only 15% of terms are mismatched. Therefore, it can be concluded the SOUNDEX function is much suitable. Also, there is a chance of users entering incorrectly

44

misspelled words. For example, instead of *Constitution* users might enter *Constitutien.* When SOUNDEX values of these two terms are analyzed it is C523. This means *Constitution* and *Constitutien* are treated as same words. This means that spelling mistakes are corrected automatically because of the usage of SOUNDEX function in SQL Server 2012. Appendix D Code to Find Global Terms shows the code to extract global terms.

Next step is to identify terms for each question and IDF and the TF-IDF values. For this, Microsoft SQL Server Integration Service (SSIS) 2014 is used and configured. This data flow task is called Term Lookup. From the SSIS Extract lookup, term, frequency and QuestionID are populated. IDF and TFIDF and Soundex_Values are calculated later and stored in TermFrequency table. Appendix E Calculation of Cosine Distance and TF-IDF shows the implementation of calculating TFIDF and cosine distance component.

As it can be observed, SOUNDEX value is used calculate TF-IDF and the cosine component.

Next step is to find the equal mapping questions for the question entered by the user. Appendix C Code to Find Similar Question shows script to find the closest questions matching the question entered by the user using cosine distance measure. Apart from the question, topic and the ranking is also displayed so that users can get an indication about the closeness of the returned questions.

As discussed in the Design chapter, abbreviations and similar words are used by users when raising questions. Table 6-3: Abbreviations show abbreviations which are used in the Ontology.

Table 6-3: Abbreviations

ID	Abbreviation	Description
1	CL	Common Law
2	Cx	Constitution
3	J	Judge
4	PL	Public Law

5	SRL	Sri Lanka
6	PM	Prime Minister
7	MP	Member of Parliament
8	Const	Constitution
9	MPs	Members of Parliament

When the users are enter questions, first those words are queried in the Global Terms. If it is not found in the global terms, then those are queried in the abbreviations. In the Ontology both MP and MPs are used mainly due to the fact those two has different SOUNDEX values. MP SOUNDEX value M100 while MPs has M120 as the SOUNDEX value. Since they have different SOUNDEX values both terms need to be included to the Ontology.

Similar to abbreviations, similar words also needs to be considered. Following table contains list of Similar Terms which are in the ontology.

Table 6-4: Similar Terms

ID	Abbreviation	Description
1	President	Head of state
2	President	Head of Cabinet
3	Commander	Head of state

Approach to amalgamate Similar Terms to question matching is as same as abbreviations techniques.

Next step is to provide proper user interface to the end user to get relevant Appendix P:, it will displays most relevant ten questions with the topic.

Ten closely matched questions will be displayed with the highest final rank at the top. In the practical scenario, it is not necessary to display the Cosine Distance Rank, Question Type Rank and the Final Rank. However, for research evaluation purpose those values are listed.

When user entering questions there can be spelling or grammar errors. Since SOUNDEX mechanism is used, effect of the spelling mistakes can be avoid. With respect to the grammar, in this mechanism word is not considered. Therefore, effect of both spelling and grammar mistakes are reduced.

When the right question is matched, user has the option of viewing the details such as answers, cases, incidents and relevant constitution references by clicking the Get Reference button. In case, user is unable to find the relevant question, user can log the question by clicking the No Matching Question button so that subject matter expert can include a new question or technical resources can verify whether there is an issue with any algorithms used.

6.7 Finding Answers

Another important step in the research is to find out the answers to the mapped users query. Every question is attached to Answers, Constitution references, Incidents and Cases. So when question is selected as shown in Appendix P:

Basic idea is taken for above user interface by using LawLanka website [50]. Advantage of the above user interface is that users can easily retrieve all the details for a given question. For each tab, users can view the details by clicking the relevant record in the list view. All these details are saved in Ontology.

6.8 Summary

In this chapter, implementation of the legal search tool was discussed in detail. Ontology generation and their properties are discussed in detail. Also, the implementation of basic search option is also discussed. Detail discussion was taken place for query matching as it is the heart of this research. Also, user interface of the system was discussed as it will provide the better usage of the system.

Next chapter discuss about the evaluation of the system which will decide whether the proposed system can be used by the target users.

Evaluation

7.1 Introduction

Since this research deals with more subjective topics than objective, evaluation is much needed to verify the techniques used in this research. In this chapter, various evaluations are carried out on the various part of the research to validate the research outcomes. Mainly research outcome was evaluated Ontology Content, Question Matching Capability, Finding the Answers and performance aspects.

7.2 Ontology Content

As said in previous chapters in several times, legal content is very vast and complex. After building the ontology, it has to be verified. Since legal domain is a specific domain where you need subject matter experts to verify. However, there is a knowledge or technical gap in subject matter experts (in this case lawyers) between their expert knowledge and the technical knowledge. Therefore, verifying the ontology has become a tedious and challenging task.

Another difficult task was that find out specific subject matter experts. In the legal domain, there are lawyers for different areas. Though this research is on legal domain, constitution is selected as proof of concepts. Therefore, legal experts are needed in the constitution sector. To analyze this ontology, constitution lawyers were selected. They were given the option to ask any questions from the researcher and researcher has verified whether those contents exist in the ontology correctly. There is no quantitative evaluation on this and qualitative assessment was done on the Ontology content.

After every step of evaluation, new set of contents were added. Incident class was added to the ontology during one of the evaluation step. Similarly, many contents and questions were also added.

Ontology evaluation is also done during the question matching evaluation as well. This means, ontology evaluation was done throughout the research. So it can be said the after end of all the steps of the research, high quality ontology was created.

7.3   Evaluation of Basic Search

Basic search has two features, content search and basic search. Though this is a simple search, output if this was evaluated.

"Person" word was used to evaluate basic search functionalities. When it was inquired from the lawyers, in case they want to search for "Person" they will use different form of "Person" such as "Person", "Persons" and "Personal".

Following is the list for output count for each of the different form of "Person"

Table 7-1: Output Count for Different form of word "Person"

Form of the word "Person"	Count of the output
Person	56
Persons	5
Personal	3

When the full text search was done for word called "Person", it returned 59 records. However, according to the above table it should be 64 (56 + 5 + 3). When it was analyzed further, it was due to the fact that same sub sections contain multiple types of different form of "Person".

For example, In Chapter 4 (Language), Sub Section,Languages   of   administration article 22.5 contains different forms of Person and Persons.

In the above section, it can be observed that both person and persons words exits. Table 7-2 contains different form of "Person" in the same article or sub section.

Table 7-2: Different form of word "Person" in Same Article

Chapter Number	Chapter Name	Article ID	Form of Words
4	Language	22.5	Person Persons
11	The Legislature - Procedures and Powers	76.3	Person Persons
3	Fundamental Rights	13.2	Person Personal

3	Fundamental Rights	13.3	Person
			Personal
3	Fundamental Rights	13.4	Person
			Personal

Since there are five instances where there are multiple forms of "Person". Because of that it can be conclude that full text matching is correct.

7.4 Question Matching Capability

Heart of this research is ability to match the user queries with the knowledge base or ontology questions. If the matching success is improved then automatically success of the project is also increased. So it is needed to evaluate question matching capability of the project. Question matching can be divided into three areas such as normal question matching, questions with abbreviations and questions with similar words.

Scenario 1: How to amend the constitution

When the above question is entered, following are the matching questions.

Table 7-3: Questions Returned for Scenario 1

Question	Is it Relevant
Who approves amendments to the constitution?	Yes
Which provisions are not amendable?	Yes
What are the details of the amendment proposal process?	Yes
How is the head of government selected?	No
What are the details for the amendment approval process?	Yes
How is the head of state selected?	No
How many executives are specified in the constitution?	No
What proportion of the vote is needed to approve a constitutional amendment?	Yes
How are members of the second chamber selected?	No
Does the constitution provide for at least one procedure for amending the constitution?	Yes

Out of the ten question returned, seven questions are relevant to the question and no questions in the question bank was missing in the list. This means there is 70% of accuracy whereas the ranking is concerned it is 67% (37 / 55) is the ranking relevancy.

Scenario 2: How to remove members of parliament?

When the above question is raised only one question is retrieved. So, there is 100% matching both with the number of questions and the ranking is concerned.

Scenario 3: How to remove mps?

51

Scenario 3 the same question as Scenario 2 with introducing on mps instead of members of parliament. However, same question is returned.

Scenario 4: What are the provisions to remove head of state?

Following are the questions returned for the above question.

Table 7-4: Questions Returned for Scenario 4

Question	Is it relevant?
Are there provisions for dismissing the head of state?	Yes
Who can propose a dismissal of the head of state?	Yes
What are the details of the process to remove judges?	No
Who are the electors for the head of state?	No
Are there provisions for dismissing the head of government?	Yes
Who may remove the chief of the central bank?	No

Out of the six questions returned, three questions are relevant to the question and no questions in the question bank was missing in the list. This means there is 50% of accuracy whereas the ranking is concerned it is 62% (13 / 21) is the ranking relevancy.

Scenario 5: What are the provisions to remove president?

Following are the questions returned for the above question. This scenario has similar word to scenario 4.

Table 7-5: Questions Returned for Scenario 5

Question	Is it Relevant?
Are there provisions for dismissing the head of state?	Yes
Who can propose a dismissal of the head of state?	Yes
What are the details of the process to remove judges?	No
Who are the electors for the head of state?	No
Are there provisions for dismissing the head of government?	Yes

Out of the five questions returned, three questions are relevant to the question and no questions in the question bank was missing in the list. This means there is 60% of accuracy whereas the ranking is concerned it is 67 % (10 / 15) is the ranking relevancy.

When all the scenarios are considered ranking relevancy is at least than 60% percent and in some scenarios it is 100 percent.

7.5 Finding Answers

Evaluating answers and other relevant cases and incidents to user entered queries is a qualitative measure rather than a quantitate measure. Three lawyers were participated for this evaluated like in the case of ontology evaluation. Also, one member from general public also participated to this exercise in order find out whether this product is helpful for the general public.

So the basic question was given to lawyers and law students. Following three question were asked from subject matter experts and law students and the results of the system also compared with the answers.

Question 1: Are there provisions for dismissing members of parliament?

Appendix E: Example of a Question has detailed answer to this question which has the answer, constitution details, cases and incidents.

When this was asked from lawyers and law students, all answered correctly. However, no one knew that there are nine reasons to vacate his post. No lawyer was able to come up with all the reasons and one lawyer was able to come up with four reasons. Other two lawyers came up with less number of reasons. Anyway, lawyers and students were not able to point to the exact sub section which is very much needed. Also, lawyers were not able to list out the relevant cases and incidents where are FAHoL was able to provide accurate data.

Question 2: Are amendments to the constitution attached to the document as a separate part or incorporated?

Though the answer to this question is Yes, one law student and a person from general public said no. System provided the correct answer to this along with the correct incidents and cases.

Question 3: When former president appointed as a member of a parliament, will he loose presidential benefits?

This is a question raised recently due to the current political development whereas the former president Mahinda Rajapakshe selected to as a member of parliament from the Kurunagala District in 2015 general election. No answer can be derived from the constitution for this question but it is needed to refer to Presidents Entitlements Act, No. 4 of 1986. It says for his life time former president will get the benefits irrespective of his current position. Since FAHoL does not support acts, system could not provide the correct answers to this question.

So when the ability of the system to find answer is measured with respect to sample questions it can be observed that correct answer is obtained with much correct examples and constitution references.

7.6 Performance

As said in few places, SQL Server 2014 was introduced to improve the performance over protégé. So the performance was measured which is listed in Table 7-6: Performance Difference between Protégé and RDBMS.

Table 7-6: Performance Difference between Protégé and RDBMS

Parameter	Protégé	RDBMS
Question Matching	5-9 Secs.	Less than 1 Sec
Finding Answers	2 Secs.	Less than 1 Sec

As database has the capabilities to use index and statistics features, with the ontology was moved the database, performance is much faster.

7.7 Summary

Few ways of evaluating this research was identified. Since Ontology construction was very important and qualitative evaluation was done. When the system provides answers, different ranking were identified to perform the quantitative evaluation. .After evaluating the system with different aspects, it can be concluded that designed system can be used at acceptable level with respect to features and performance.

Conclusion & Further Work

8.1 Introduction

When the research is carried out, due to many reason like resources and time, it will be difficult to achieve its all goals. Every research leads to few more research and this research is also no difference. Since computational techniques are new to legal domain, this research will open lot of avenues for other research.

8.2 Further Work

Further work can be divided into mainly two parts. Those two parts are improving the knowledge bases and improving the technical and research area of the system.

8.2.1 Improving Knowledge Base

1978 Sri Lankan Constitution was selected as proof of concepts. However, legal domain contains lot of documents. Most of the legal issues arises not mainly because of constitution but with other acts. For example, provincial councils were brought to the Sri Lanka from 13th amendment. However, provincial councils are enacted by provincial council act. During the "divinaguna" controversy debate in 2014 was due to intermediate list which is in provincial council act. Also, in 2006, former President Her Excellency Chandrika Bandaranayake Kumaranathunga Vs President secretory Mr. Lalith Weerathunga in decision was made from Presidents Entitlements Act, No. 4 of 1986 not by referencing the constitution itself. This case indicates that it is essential to include acts and ordinances also into the knowledge base or to the ontology.

Proposed system should have the option of comparing two different legal domains. For example, when new legal clauses are introduced typically law professionals compare the law with different countries. When the "Regulate the seed and planting act" was enacted in 2003, there were comparisons made to other acts around the world [15].

Ontology is integrated to other Legal Ontologies so that in case Sri Lankan legal system does not provide any support for the queries, laws of common wealth countries will

prevail. For example, in the judgment to when to hold the presidential election in 2005, common wealth countries law was used.

8.2.2 Improving Technical & Research Area

In the current system, user question needs to be matched to the question to get the answer and other related contents. This research can be further improved by providing an option automatically find answer even if there is no matching question.

Though there are additional features introduced with SOUNDEX function and also system has the capability of matching with abbreviations and equal terms. However, system does not have the ability of matching questions semantically. If that can be adopted to the proposed system, system will be further enhanced. For example, when user enters query which includes, "person" system should be able to understand that citizen and person are equal in semantic. Currently, only the words in local ontology will be matched but system should be extended to get semantic words from defined third party Ontology.

8.3 Conclusion

Research was divided into multiple sections. Main task and the challenging task was to critically analyze the current legal system in Sri Lanka. After the challenging task was accomplished, next step was to implement 1978 Constitution of Sri Lanka in Ontology. This was accomplished using protégé and SQL Server 2014. Next task was to match user queries. This was accomplished using cosine distance and newly introduced question type. Finally, answers and relevant information was retrieved to user by reading the Ontology.

In the approach chapter, hypothesis was defined as preliminary legal domain queries can be answered by using legal domain ontology. Since FAHoL has the capability of finding answers to the preliminary legal domain questions. Though this research has provided its main objectives, there are lot of areas to improve by technically and well as content wise.

8.4 Summary

There were several aims and objectives of this research as stipulated in the Chapter 1. To start the research, it was very much required to study the legal system in Sri Lanka which was concluded. To proceed with the research, it was required to study about the

current computational models of legal domains in other countries. Since Ontology was identified as the proposed technique then it was reviewed how ontology has been used in legal domains and other domains to solve problem. As lawyers and students who are the intent users of this system, starts the searching with basic text search, it was decided to provide basic searching facility to search legal text using Full Text search. Then Ontology was developed ontology for Sri Lankan 1978 constitution by semi-automated process using SQL Server 2012. Then the ability of the system to provide answers to legal queries and relevant queries were measured.

By analyzing all the aims and objectives, it can be said the research has achieved its objectives effectively.

References

1. Yasawardane A., නීතිය නොදැනීම සමාවට කරුණක් නොවේ. (Ignoranta Juris Haud Excusat) , 2011, Saranga Publication , ISBN 978-955-684-005-6, pp, 55-70.

2. Cohen M.L, Olso K.C., "Legal Research - in a nut shell", 2003, ISBN-13: 978-0314147073, pp. 65-78.

3. LAW – DATABASES, Queen Mary, University of London http://www.library.qmul.ac.uk/subject/law/databases, Accessed on 2014-05-06.

4. Westlaw: quick user guide, London metropolitan university, Library Services, https://metranet.londonmet.ac.uk/fms/MRSite/psd/ls/subjectGuides/law/Quick%20guide%20to%20using%20Westlaw.pdf, Accessed on 2014-05-06.

5. HeinOnline, Law Journal Library, http://heinonline.org/HeinDocs/LawJournalLibrary.pdf , Accessed on 2014-05-06.

6. Ding, Y., Foo, S., Ontology Research and Development, Part I – A Review of Ontology Generation, Journal of Information Science, 2002, 28(2), 123-136.

7. Coray, L.J.M., "An Introduction to the Legal System of Sri Lanka", 2003, ISBN 995-8733-18-0, pp. 150-158.

8. "HeinOnline's Getting Started Webniar", https://www.youtube.com/watch?v=VGq7ZsZlsMc,Marcie Baranich, HeinOnline Support 2009-04-15, Accessed on 2014-05-19.

9. Balasingham, K., Laws of Ceylon, 1929, Sweet and Maxwell, London. Chapter 15 & 17.

10. Subhashini, R., Akilandeswari, J., A Survey on Ontology Construction Methodologies, Volume 1 www.ijecbs.com, Vol 1, Issue 1, January 2011.

11. 1978 Sri Lanka Constitute including 19th amendment, Government Press, 2015.

12. Swartout, B., Patil, R., Knight, K., 1996, "Towards Distributed Use of Large-Scale Ontologies", AAAI Technical Report, SS-97-06.

13. Ismail, M.A., Yaacob, M., Kareem, S.A., "Ontology Construction: An overview, Malaysia", 2006.

14. Ding Y., Foo S., 2001, "Ontology research and development. Part 1 – Review of ontology generation", Journal of Information Science, pp. 123-136.

15. Mihalcea, R.L iu H., Lieberman, H., 2006. "NLP (Natural Language Processing) for NLP (Natural Language Programming)", In Proceedings of 7th international conference on Computational Linguistics and Intelligent Text Processing, pp: 319-330.

16. Ismail, Z., The Muslim Law of Sri Lanka, http://www.scribd.com/doc/27552670/The-Muslim-Law-of-Sri-Lanka, Accessed on 2014-09-06.

17. Rzevski, G., Skobelev, P., "Emergent Intelligence in Large Scale Multi-Agent Systems", International Journal of Computers, Issue 4, Volume 1, 2007.

18. Furbach, U., Schoon, I., Stolzenburg, F., "Cognitive Systems and Question Answering", Computer Systems, Artificial Intelegence, 18 November 2014.

19. Gruber, T.R., 1993, "A Translation approach to portable ontology specifications", Knowledge Acquisition 5 (2), 199-220

20. Noy, N.F., McGuinness, D.L., "Ontology Development 101: A Guide to Creating Your First Ontology", Stanford University, Stanford, CA, 94305.

21. Full-Text Search (SQL Server), SQL Server 2014, MSDN, http://msdn.microsoft.com/en-us/library/ms142571.aspx, Accessed on 2014-12-04.

22. Query with Full-Text Search, MSDN, https://msdn.microsoft.com/en-us/library/ms142583.aspx, Accessed on 2014-12-04.

23. General Architecture for Text Engineering, GATE, Access 2014-12-10.

24. Gruber, T. R., "A Translation Approach to Portable Ontology Specifications. Knowledge Acquisition", 5(2):199–220.

25. Constitute Project, https://www.constituteproject.org/, Accessed 2014-12-30.

26. Jayakar, T., Sri Lankan Constitutes, From Solsberry to 1978 18th Amendment, 2011 January.

27. Lee, S., Ryun, P., Choi, K., "Ontology-based Question Answering System", Korea Advanced Institute of Science and Technology.

28. Kawtrakul, A., Thunkijjanukij, A., Khantonthong, N., "Rich Lexical Knowledge based Q&A System for Ubiquitous Knowledge Service".

29. Palaniappan, L., Rao, N.S., "An Ontology-based Question Answering Method with the use of Query Template", International Journal of Computer Applications (0975 – 8887), Volume 9– No.9, November 2010.

30. Yu, J., Li, D., Hou, J., "Similarity Measure of Test Question Based on Ontology and VSM", The Open Automation and Control Systems Journal 2014, 6, pp 262-267.

31. Galton S., Buckley, C., "Term-weighting approaches in automation text retrieval", Information Processing and Management, vol. 24, no. 5, pp. 513-523, 1988.

32. Asanka, P.P.G.D, "Finding Similar Text Files using Text Mining", 8th IEEE ICCSE 2013 Colombo.

33. Russel M.A, Whiz-Bang, A., "Introduction to TF-IDF" in Mining the Social Web, New Delhi, O'Reilly Media, Inc., 2011, p. 201.

34. Hand, D., Mannila, H., Smyth, P., "Retrieval by Content," in Principles of Data Mining, New Delhi, Prentice Hall of India Private Limited, 2007, p. 456–464.

35. Hair, J.F., Black, W.C., Babin, B.J., "in Multivariate Data Analysis", New Delhi, Pearson Education in South Asia, 2012, p. 599.

36. Han, J., Pattern Discovery in Data Mining, University of Illinois at Urbana-Champaign, Course Period 2015-Feb-09 – 2015-Mar-08.

37. SOUNDEX (Transact-SQL), MSDN, https://msdn.microsoft.com/en-us/library/ms187384.aspx?f=255&MSPPError=-2147217396, Accessed on 2015-02-20.

38. "The Soundex Indexing System". National Archives and Records Administration. 2007-05-30. Accessed 2015-02-24.

39. Busse, J., Humm, B., Lubbert, C., Actually, What Does "Ontology" Mean? A Term Coined by Philosophy in the Light of Different Scientific Disciplines, Journal of Computing and Information Technology - CIT 23, 2015, 1, pp. 29–41.

40. Gruber, T.R., A translation approach to portable ontologies. Knowledge Acquisition, 5(2), 1993, pp. 199–220,

41. Uschold, M., Gruninger, M., Ontologies: Principles, Methods and Applications. Knowledge Engineering Review, 11(2), June 1996.

42. Herre H., General Formal Ontology (GFO):A Foundational Ontology Integrating Objects and Processes. Technical Report, 8, University of Leipzig, 2006.

43. Pickert, U., Einfuhrung in Ontologien, ¨http://www.dbis.informatik.hu-berlin.de/dbisold/lehre/WS0203/SemWeb/artikel/2/Pickert Ontologien final.pdf, Accessed on 2014-10-12.

44. "Sri Lanka's Presidential Poll Due This Year: Court", Arab News, http://www.arabnews.com/node/272102, 2005-08-25, Accessed 2015-01-12.

45. Heymans, S., Ma, L., Anicic, D., Ontology Reasoning With Large Data Repositories, IBM China Research Lab,

46. Zhou, J., Ma, L., Liu, Q., Zhang, L., Yu, Y., 2006, Minerva: A Scalable OWL Ontology Storage and Inference System. In Proceedings of the 1st Asian Semantic Web Conference, volume 4185 of Lecture Notes in Computer Science, pages 429-443, Springer.

47. Martinez-Cruz, C., Ignacio, J. Blanco,M. Amparo Vila, Ontologies versus relational databases: are they so different? A comparison, Springer Science+Business Media B.V., 2011.

48. Zachary Elkinsa, James Meltonb, Tom Ginsburgc, Constitute: The World's Constitutions to Read, Search, and Compare, Journal of Web Semantics, 2013.

49. Fernández López, M., Overview Of Methodologies For Building Ontologies, IJCAI-99 workshop on Ontologies and Problem-Solving Methods (KRR5) Stockholm, Sweden, August 2, 1999.

50. Law Lanka, A-Z guide to Laws of Sri Lanka and Reported Cases, http://www.lawlanka.com/, Accessed on 2015-03-12.

51. Ranasinghe, R., Perera, S., Gammanpila, U.P., De Silva, H., Wada Pitiya, Derana Televsion, Producer Sanka Amarajith De Silva, 2015-06-17 10:30PM – 12:00 AM.

52. Zhai, C.X., Week 1: NLP, Text Representation, and Word Association Mining, Text Mining and Analytics, Department of Computer Science, University of Illinois at Urbana-Champaign., www.coursera.org.

53. Poole D., Soundex - Experiments with SQL CLR, SQLServerCentral.com, http://www.sqlservercentral.com/articles/Programming/101836/, 2013/09/12, Accessed on 2015/06/12.

54. Poole, D., Soundex - Experiments with SQLCLR Part 2, SQLServerCentral.com, http://www.sqlservercentral.com/articles/soundex/120628/,, 2014/12/30, Accessed on 2015/06/12.

Script to Retrieve Simple Search

```
SELECT C.ChapterNumber
       ,C.ChapterName
       ,D.Subsection_ID
       ,D.Subsection
       ,D.Details
FROM [dbo].[Constitute_Details] D
INNER JOIN [dbo].[Constitute_Chapters] C ON D.ChapterID = C.ChapterNumber
WHERE CONTAINS (Details, 'religion')
GO


SELECT C.ChapterNumber
       ,C.ChapterName
       ,D.Subsection_ID
       ,D.Subsection
       ,D.Details
FROM [dbo].[Constitute_Details] D
INNER JOIN [dbo].[Constitute_Chapters] C ON D.ChapterID = C.ChapterNumber
WHERE CONTAINS (Details, 'religion AND caste')
GO

SELECT C.ChapterNumber
       ,C.ChapterName
       ,D1.Subsection_ID
       ,D1.Subsection
       ,D1.Details
FROM [dbo].[Constitute_Details] D1
INNER JOIN [dbo].[Constitute_Chapters] C ON D1.ChapterID = C.ChapterNumber
WHERE FREETEXT (Details, 'religion')
GO


SELECT KL.RANK
       ,FL.Subsection
       ,FL.Subsection
       ,FL.Details
FROM [dbo].[Constitute_Details]  AS FL
   INNER JOIN
   FREETEXTTABLE([dbo].[Constitute_Details] , Details,
          'religion') AS KL
   ON FT_TBL.Subsection_ID = KL.[KEY]
ORDER BY KL.RANK DESC
GO

SELECT *
```

```
FROM Constitute_Details AS FT_TBL INNER JOIN
  FREETEXTTABLE (Constitute_Details,
   Details,
  '(language NEAR religion)'

 ) AS KL
  ON FT_TBL.Subsection_ID = KL.[KEY]
WHERE KL.RANK > 2
ORDER BY KL.RANK DESC;
GO
```

Code to Get Question Type

```sql
CREATE FUNCTION [dbo].[fn_get_questiontype]
(
@question VARCHAR(max)
)
RETURNS VARCHAR(15)
BEGIN
DECLARE @questiontype VARCHAR(15), @finalQuestion VARCHAR(15),
@lngthstring INT;
DECLARE @commansepques VARCHAR(max)
DECLARE @T TABLE (words VARCHAR(50))

INSERT INTO @T VALUES ('Does'),('Who'),('What'),('Are')
                                ,('Do'),('Is'),    ('When'),('Which')
                                ,('Whose'),('Can'), ('Should'),('Was')


SET @questiontype =  SUBSTRING(@question,1, CHARINDEX(' ' ,@question)-1)

IF @questiontype = ANY(SELECT words from @t)
   SET @finalQuestion = @questiontype
ELSE
 BEGIN
   IF ((@questiontype = 'How') OR (@questiontype = 'Under') OR (@questiontype
= 'For')
       OR (@questiontype = 'By') OR (@questiontype = 'To')
       OR (@questiontype = 'In')  OR (@questiontype = 'At'))
       AND CHARINDEX(',',@question,1) = 0
        BEGIN
                SET @lngthstring = LEN(@questiontype)
                SET @finalQuestion =   @questiontype + ' ' +
SUBSTRING(@question,@lngthstring+2,
                                        CHARINDEX(' '
,@question,@lngthstring+2)  - @lngthstring - 1)
        END
   ELSE IF CHARINDEX(',',@question,1) > 0
       BEGIN
           SET @commansepques = LTRIM(SUBSTRING (@question,
CHARINDEX(',',@question,1) + 1,
                                LEN(@question) -
CHARINDEX(',',@question,1)- 1 ))
                SET @questiontype =  SUBSTRING(@commansepques,1,
CHARINDEX(' ' ,@commansepques)-1)
                IF @questiontype = ANY(SELECT words FROM @t)
```

```
                        SET @finalQuestion = @questiontype
            ELSE
            BEGIN
               IF (@questiontype = 'How') OR (@questiontype = 'Under')
OR (@questiontype = 'For')
                     OR (@questiontype = 'By') OR (@questiontype = 'To')
                     OR (@questiontype = 'In')  OR (@questiontype = 'At')
                  BEGIN
                           SET @lngthstring =
LEN(@questiontype)
                           SET @finalQuestion =   @questiontype
+ ' ' + SUBSTRING(@commansepques,@lngthstring+2,

    CHARINDEX(' ',@commansepques,@lngthstring+2)  - @lngthstring - 1)
                  END
               ELSE
               BEGIN
                  SET @questiontype =   SUBSTRING(@question,1,
CHARINDEX(' ',@question)-1)
                  SET @lngthstring = LEN(@questiontype)
                  SET @finalQuestion =   @questiontype + ' ' +
SUBSTRING(@question,@lngthstring+2,
                                          CHARINDEX(' '
,@question,@lngthstring+2)  - @lngthstring - 1)
                  END
            END
         END

         IF @finalQuestion = 'In How'
          SET @finalQuestion = 'In how many'
END
 RETURN @finalQuestion
 END
```

Appendix C

Code to Find Similar Question

```
CREATE PROC [dbo].[usp_Find_Similar_Question]
@Question Varchar(150) , @n int = 0.5
AS
BEGIN
SET NOCOUNT ON

DECLARE @Q_CD FLOAT
DECLARE @Question_Type VARCHAR(50), @Question_Question_Type
VARCHAR(50)
DECLARE @Question_ID INT, @Question_Type_Ranking INT
DECLARE @RecCount INT
DECLARE @Term VARCHAR(100), @Description varchar(150)
DECLARE @NewQuestion VARCHAR(MAX)

CREATE TABLE #Que
(Term VARCHAR(100),
Frequency INT,
Soundex_Value varchar(10),
IDF FLOAT,
TFIDF FLOAT)

CREATE TABLE #FirstQue
(Term VARCHAR(100),
Soundex_Value varchar(10)
)

INSERT INTO #FirstQue
(Term,Soundex_Value)

SELECT value,SOUNDEX(value) FROM (
SELECT value,Count(1) TF FROM [dbo].[SplitWords](@Question)
WHERE value NOT IN (SELECT Term FROM ExcludedTerms)
GROUP BY value
) P


SELECT Term,b.[Description] INTO #Temp1 FROM #FirstQue a
INNER JOIN Abbrivations b
ON A.Term = b.Abbrivation
WHERE Soundex_Value NOT IN
(SELECT Soundex_Value FROM GlobalTerms
)
```

```
SELECT @RecCount = COUNT(1) FROM #Temp1
SET @NewQuestion = @Question
WHILE (@RecCount > 0)
BEGIN
        SELECT Top 1 @Term = Term, @Description =Description
    FROM #Temp1

        SET @NewQuestion = REPLACE(@NewQuestion,@Term,@Description)

    DELETE FRom  #Temp1 Where  Term = @Term
        SELECT @RecCount = COUNT(1) FROM #Temp1
END

INSERT INTO #Que
(Term,Frequency,IDF)

SELECT value,TF,IDF FROM (
SELECT value,Count(1) TF FROM [dbo].[SplitWords](@NewQuestion)
Where value NOT IN (SELECT Term FROM ExcludedTerms)
GROUP BY value
) P
INNER JOIN GlobalTerms GT
ON GT.Term = P.value

UPDATE #Que
SET TFIDF = IDF * Frequency,Soundex_Value = SOUNDEX(Term)
SELECT @Q_CD = SUM(TFIDF * TFIDF) FROM #Que

SELECT @Question_Type = dbo.fn_get_questiontype(@Question)

SELECT TOP 100 T.QuestionID,
    MAX(Q1.Question) Question,100* (SUM(T.TFIDF * Q.TFIDF) )
/ ( MAX(Q1.CosineDistanceQuestionComponent) * @Q_CD ) CD ,
 MAX(Q1.Header) Header,MAX(Q1.QuestionType) Question_Type
INTO #Temp
 FROM TermFrequency T
INNER JOIN #Que Q
ON T.Soundex_Value = Q.Soundex_Value
INNER JOIN Question Q1
ON Q1.ID = T.QuestionID
GROUP BY T.QuestionID
ORDER BY 3 DESC

ALTER TABLE #Temp ADD Question_Type_Ranking INT ,Final_Ranking FLOAT;



SET @RecCount = 0
```

```sql
SELECT @RecCount = COUNT(1) FROM #Temp WHERE
Question_Type_Ranking IS NULL

WHILE (@RecCount > 0)
BEGIN

  SELECT Top 1 @Question_ID = QuestionID,
  @Question_Question_Type = Question_Type FROM #Temp
  WHERE Question_Type_Ranking IS NULL

  IF @Question_Question_Type = @Question_Type
   BEGIN
    SET @Question_Type_Ranking = 3
   END
  ELSE IF CHARINDEX(@Question_Type,@Question_Question_Type,1) > 0
           OR CHARINDEX(@Question_Question_Type,@Question_Type,1) >
0
  BEGIN
  SET @Question_Type_Ranking = 2
      END
  ELSE
   BEGIN
   SET @Question_Type_Ranking = 1
      END

  UPDATE #Temp
  SET Question_Type_Ranking =@Question_Type_Ranking
  WHERE QuestionID = @Question_ID

  SELECT @RecCount = COUNT(1)
  FROM #Temp
  WHERE Question_Type_Ranking IS NULL
END

UPDATE #Temp SET Final_Ranking = POWER( Question_Type_Ranking,@n ) *
CD

SELECT Top 10 QuestionID,Question,Header,ROUND(CD,4) CD
,Question_Type_Ranking,ROUND(Final_Ranking,4) Final_Ranking
FROM #Temp
ORDER BY Final_Ranking DESC

END
```

Code to Find Global Terms

```sql
SET NOCOUNT ON
TRUNCATE TABLE GlobalTerms
CREATE TABLE #t (Term VARCHAR(150))

DECLARE @Question VARCHAR(MAX)
----------------------------------------------------------
DECLARE @MyCursor CURSOR
SET @MyCursor = CURSOR FAST_FORWARD
FOR
SELECT Question FROM   Question
OPEN @MyCursor
FETCH NEXT FROM @MyCursor
INTO @Question
WHILE @@FETCH_STATUS = 0
BEGIN
SET @Question = REPLACE(@question,'?','')
SET @Question = REPLACE(@question,')','')
SET @Question = REPLACE(@question,'(','')
SET @Question = REPLACE(@question,':','')
SET @Question = REPLACE(@question,'/',' ')
SET @Question = REPLACE(@question,'\',' ')
SET @Question = REPLACE(@question,'"','')

INSERT INTO #t (Term)
SELECT value FROM dbo.SplitWords(@Question)

FETCH NEXT FROM @MyCursor
INTO @Question
END
CLOSE @MyCursor
DEALLOCATE @MyCursor


INSERT INTO GlobalTerms (Term)
SELECT DISTINCT term FROM #t
WHERE Term NOT IN (SELECT Term FROM ExcludedTerms)

DROP TABLE #t
```

Calculation of Cosine Distance and TF-IDF

```sql
SET NOCOUNT ON
TRUNCATE TABLE GlobalTermSoundex

UPDATE [dbo].[GlobalTerms]
SET Soundex_Value = SOuNDEX([Term])

UPDATE [dbo].[TermFrequency]
SET Soundex_Value = SOuNDEX([Term])

 INSERT GlobalTermSoundex
 (Soundex_Value,TermCount)
 SELECT Soundex_Value, SUM([Score]) Score
 FROM [GlobalTerms]
 GROUP BY Soundex_Value

 UPDATE  G
 SET   G.Score = P.C
 FROM
 (SELECT Soundex_Value Soundex_Value, SUM(Frequency) C FROM
TermFrequency
 GROUP BY Soundex_Value) P JOIN [GlobalTerms] G
 ON G.Soundex_Value = P.Soundex_Value

DECLARE @N int

SELECT @N  = COUNT(1) FROM Question

SELECT @N

UPDATE  A
SET A.IDF= B.IDF
FROM
(SELECT Term,Soundex_Value,LOG(@N/[Score]) IDF FROM
[dbo].[GlobalTerms])
B
JOIN TermFrequency A ON
A.Soundex_Value = B.Soundex_Value


UPDATE [dbo].[TermFrequency]
SET [TFIDF] = Frequency * [IDF]
```

```
UPDATE [dbo].[GlobalTerms]
SET IDF = LOG(@N/[Score])

UPDATE Q
SET Q.CosineDistanceQuestionComponent = TF.CDQC
FROM
(SELECT QuestionID,SUM (TFIDF * TFIDF) CDQC FROM
[dbo].[TermFrequency]
Group By QuestionID) TF
JOIN Question Q ON TF.QuestionID = Q.ID
```

Appendix F

Question Type Ranking

Obtaining a best power value of the Question Type Ranking

FR = QTRn * CDR

FR – Final Ranking
QTR – Question Type Ranking
CDR – Cosine Distance Ranking

Several Questions were searched for different n values.

Q1.How does the Constitution treat the status of colonies, and territories?
N = 0.5

Question ID	Question	Header	CDR	QTR	FR
195	Sample Question 195	General	0.9715	3	0.9715
197	Sample Question 197	General	0.7491	1	0.7491
112	Sample Question 112	Executive	0.4788	2	0.4788
297	Sample Question 297	Legislature	0.4411	1	0.4411
219	Sample Question 219	General	0.4112	2	0.4112
196	Sample Question 196	General	0.3958	1	0.3958
139	Sample Question 139	Executive	0.3895	1	0.3895
14	Sample Question 14	Criminal Procedure	0.3857	3	0.3857
173	Sample Question 173	Executive	0.3803	1	0.3803
249	Sample Question 249	General	0.3765	1	0.3765

N = 1

Question ID	Question	Header	CDR	QTR	FR
195	Sample Question 195	General	0.9715	3	2.9145
14	Sample Question 14	Criminal Procedure	0.3857	3	1.1571
15	Sample Question 15	Criminal Procedure	0.3219	3	0.9656
112	Sample Question 112	Executive	0.4788	2	0.9577
219	Sample Question 219	General	0.4112	2	0.8223
197	Sample Question 197	General	0.7491	1	0.7491
640	Sample Question 640	Race, Ethnicity,	0.2437	3	0.7312

Question ID	Question	Header	CDR	QTR	FR
		and Language			
656	Sample Question 656	Race, Ethnicity, and Language	0.2437	3	0.7312
203	Sample Question 203	General	0.331	2	0.662
77	Sample Question 77	Elections	0.301	2	0.6019

N = 2

Question ID	Question	Header	CDR	QTR	FR
195	Sample Question 195	General	0.9715	3	8.7436
14	Sample Question 14	Criminal Procedure	0.3857	3	3.4713
15	Sample Question 15	Criminal Procedure	0.3219	3	2.8968
640	Sample Question 640	Race, Ethnicity, and Language	0.2437	3	2.1935
656	Sample Question 656	Race, Ethnicity, and Language	0.2437	3	2.1935
112	Sample Question 112	Executive	0.4788	2	1.9153
219	Sample Question 219	General	0.4112	2	1.6447
203	Sample Question 203	General	0.331	2	1.3239
77	Sample Question 77	Elections	0.301	2	1.2038
574	Sample Question 574	Rights	0.2592	2	1.0368

When all the three instances are compared, there are two close questions.

- How does the Constitution treat the status of colonies, protectorates, and territories?
- What specific colonies or protectorates does the Constitution mention?

So the correct questions were identified when n = 0.5

PN: Actual question was removed to avoid plagiarism.

Appendix H Example of a Question

Incident	Dates	References
Death of Colombo District MP. Mr. CV. Gunarathne	7-Jun-2000	Sri Lanka Cabinet Minister of Industries development, Minister Mr.Gunaratne and his wife were killed.
Death of National List MP. Mr. C.V. Gunarathne	12-Aug-2005	
Resignation of National List MP Mr. Samarasinghe	30-Jun-2008	Mr. H. M. Wasantha Samarasinghe resigned to contest the North Central provincial council election
Mr, Mahipala Herath resigned	7-May-2004	Mr. Mahipala Herath resigned to contest the Sabaragamuwa provincial council elections.
Mrs. Chandrika Kumaratunga resigned to take up presidency.	12-Nov-1994	
Mr. Sarath Fonseka removed as he didn't participate parliament for three months	2010	
S,B. Dissanayake removed as he didn't participate parliament for three months	2006	

Are there provisions for dismissing members of parliament?

Dilan Perera against Dr. Rajitha Senarathne, 1999

Accuracy of SOUNDEX

A130	A153	A516	A162	A223	A225
Abide About	Appointing Appointment	Approval Approve Approved Approves Approving	Affairs Appears	Accused Association	Accession Assign Assigned
A231	**A235**	**A252**	**A312**	**A313**	**A323**
Active-duty Activities	Academic Actions	Against Agencies Agency	Advise Advisory	Adopted Adoption	Adequate Attached
A351	**A353**	**A362**	**A535**	**A553**	**A623**
Attempt Attempts	Attendance Automatic	Address Authorize Authorized	Anthem Anything	Amend Amendable Amended Amending Amendment Amendments	Arrest Arrested
A632	**A652**	**B220**	**B230**	**B255**	**B300**
Artists Arts	Arms Arranged Arrangement Arrangements	Basic Basis	Based Best	Becoming Beginning	Body Both
B520	**B636**	**C134**	**C200**	**C325**	**C420**
Bank Being	Birthright Borders Broader	Capital Capitalism	Case Choose	Citizen Citizens Citizenship	Class Clause
C430	**C436**	**C452**	**C512**	C513	**C522**
Called Child	Children Cultural Culture	Calling Challenge Colonies	Campaigns Compose Convicted	Competitive Confederal	Census Consist
C514	**C515**	**C516**	**C523**	**C524**	**C525**
Compulsory Conflict	Compensation Convene	Chamber Chambers Confront	Consider Considered Constituent Constitution Constitutional	Consult Consults Council Counsel	Commission Conscience Conscientious Consumer

			Constitutionality Constitutions		
C526	**C530**	**C532**	**C533**	**C536**	**C553**
Censorship Concerning	Cannot Committee	Committees Conduct	Committed Conditions	Central Contrary Controls Counter Countries Country	Commander Commanders Comments Communities
C630	**C632**	**C635**	**C655**	**D000**	**D125**
Cohort Court	Cohorts Courts	Certain Coordinated Credential	Criminal Criminals	Day Die	Division Divisions
D153	**D225**	**D236**	**D246**	**D252**	**D255**
Defendants Definition Depending	Decision Decisions	Distributed District Districts	Declaration Declarations Declare	Dismiss Dismissal Dismissed Dismissing	Document Documents
D261	**D300**	**D320**	**D365**	**D526**	**E123**
Described Describes	Date Death Duty	Deities Duties	Determination Determinations Determine Determined Determines	Democracy Democracies Democratic	Effect Episodes
E214	**E215**	**E216**	**E223**	**E231**	**E236**
Explicit Explicitly	Expansion Expense	Expiration Express Expression Expropriate Expropriation	Executive Executives	Establish Established	External Extradition Extraordinary
E255	**E423**	**E536**	**F400**	**F420**	**F530**
Economic Examine	Elected Election Elections Electoral Electors Electronic	Enter Entered Entirety Entry	Feel Full	False Files Flag	Find Found
F540	**F552**	**F600**	**F623**	**F625**	**F654**
Family Final	Finance Financial	Fair Free	First First-level	Foreign Foreigner	Formally Formula

	Financing		Forced Fraction		
G165	**G564**	**G653**	**H430**	**I353**	**I512**
Government Governmental Governments Governor	General Generally	Grand Grant Granted Granting Grounds Guarantee	Health Healthy Held	Identified Identify	Investigate Investment
I514	**I516**	**I523**	**I524**	**I526**	**I531**
Implementation Implemented Inviolable Involved Involvement	Important Imprisoned Imprisonment	Institution Institutions Instruments	Include Included Including	Incorporate Incorporated Incrimination	Independence Independent Individual Individually Individuals
I532	**I533**	**I536**	**J232**	**L100**	**L140**
Indicted Indicting Indigenous Integration	Initiate Initiates Initiative Initiatives	Interesting International Interpellate Interpellation Interpreter Intervention Introduction	Justice Justices	Levy Life	Label Level
L142	**L200**	**L224**	**L230**	**L240**	**L522**
Levels Libelous	Laws Law	Legislate Legislation Legislative Legislators Legislature Legislatures	Last Least List	Legal Likely Local	Language Languages
M230	**M300**	**M500**	**M516**	**M523**	**M535**
Most Must	Media Mode Motto	Man Men Money	Member Members Membership	Minister Ministers	Mention Mentioned
M600	**N354**	**N364**	**N400**	**N500**	N516

Marry More	National Nationals	Natural Naturalization Naturalized	Newly New	Name Non-derogable Non-discrimination	Number Numbered Numbering
N553	**O155**	**O163**	**O500**	**O520**	**O625**
Nominated Nominates Nomination Nominations	Opinion Opinions	Operation Operations Opportunity Overridden Override Overthrow	On Own	Once Ones	Organic Organizations Organs Original
P142	**P230**	**P420**	**P600**	**P612**	**P613**
Public Published	Passed Post	Place Please Policy	Power Pre-trial	Previous Privacy Profession Proposal Propose Proposes Provision Provisions Purposes	Private Prohibit Prohibited Prohibition Provide Provided
P616	**P620**	**P622**	**P623**	**P625**	**P626**
Preference Preferred Property Proportion	Powers Press	Process Processes Prosecution Prosecutor	Presides Procedure Procedures Proceed	Person Personality Presumption	Prescribe Preservation Progress
P630	**P632**	**P652**	**P653**	**P654**	**R120**
Part Party	Participation Particular Parties Parts Protect Protected Protecting Protection Protectorates	Principle Principles	Print Promote	Promulgated Promulgation	Reviews Revoke
R142	**R162**	**R163**	**R215**	**R223**	**R225**
Replaced Replacement Replacing	Representation Representative	Referred Reporting	Responsibilities Responsibility	Registry Reject	Recognize Resign

	Representing		Responsible		
R230	**R232**	**R236**	**R243**	**R261**	**R263**
Rest Right	Right Rights	Restart Restraint Restrict Restriction Restrictions	Regulate Regulations Regulatory	Reserve Reserved	Record Recruitment Regard Regarding Regards Required
R400	**R420**	**S100**	**S121**	**S123**	**S124**
Role Rule	Release Rules	Safe Shape	Specific Specifically Specified Specify Specifying	Subsidiary Substantive	Special Specialized
S153	**S162**	**S163**	**S200**	**S225**	**S235**
Spending Subnational	Sovereignty Suffrage	Separate Separation Support	Sex Size Such	Secessions Secession	Section Sections System
S242	**S253**	**S263**	**S300**	**S320**	**S332**
Socialism Socialist	Second Second-level Sequentially	Secret Security	Said Set	Seats Stage	States Status
S353	**S362**	**S410**	**S423**	**S436**	**S500**
Standard Standing	Strike Structure	Self Self-determination	Selected Selecting Selection	Shelter Solidarity	Same Sine Some
S546	**S610**	**S612**	**T120**	**T200**	**T230**
Similar Similarly	Serve Survey	Serves Service	Topic Topics Types	Take Tax These Those Took Twice	Text Thought
T623	**T630**	**T632**	**T636**	**T652**	**U525**
Theorists Throughout	Third Third-level Trade Treat Tried Truth	Towards Treaties	Territorial Territories Territory Torture	Terms Transfer Transitional Translated	Unconstitutional Unconstitutionality
U536	**V300**	**V320**	**V352**	**W200**	**W336**
Under Understand	Veto Void Vote	Vetoes Votes	Vetoing Voting	Was Which Whose	Withdraw Withdrawal

			81		
Understand s Unitary					
W430	**W635**				
Wealth Would	Writing Written				

Accuracy of SOUNDEX
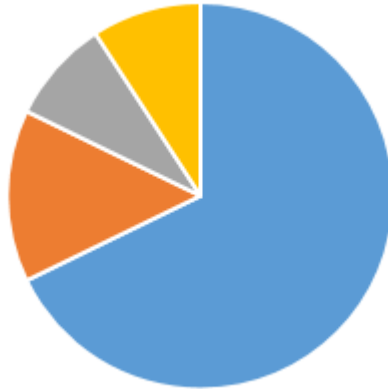
■ Single Word   ■ Not Matched   ■ Perfectly Mathced   ■ Partity Matched

Accuracy of SOUNDEX
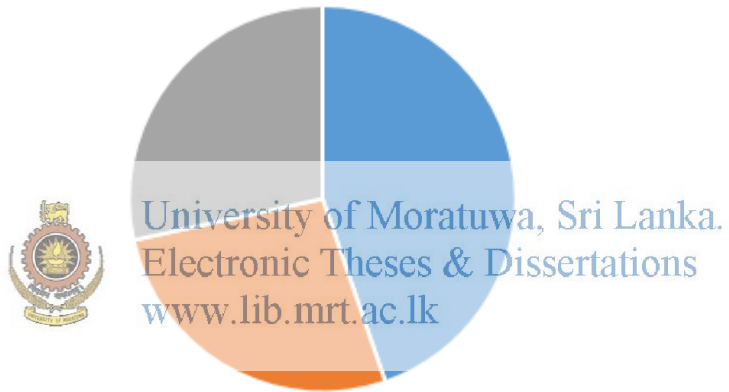
■ Single Word   ■ Not Matched   ■ Perfectly Mathced

E-R Diagram
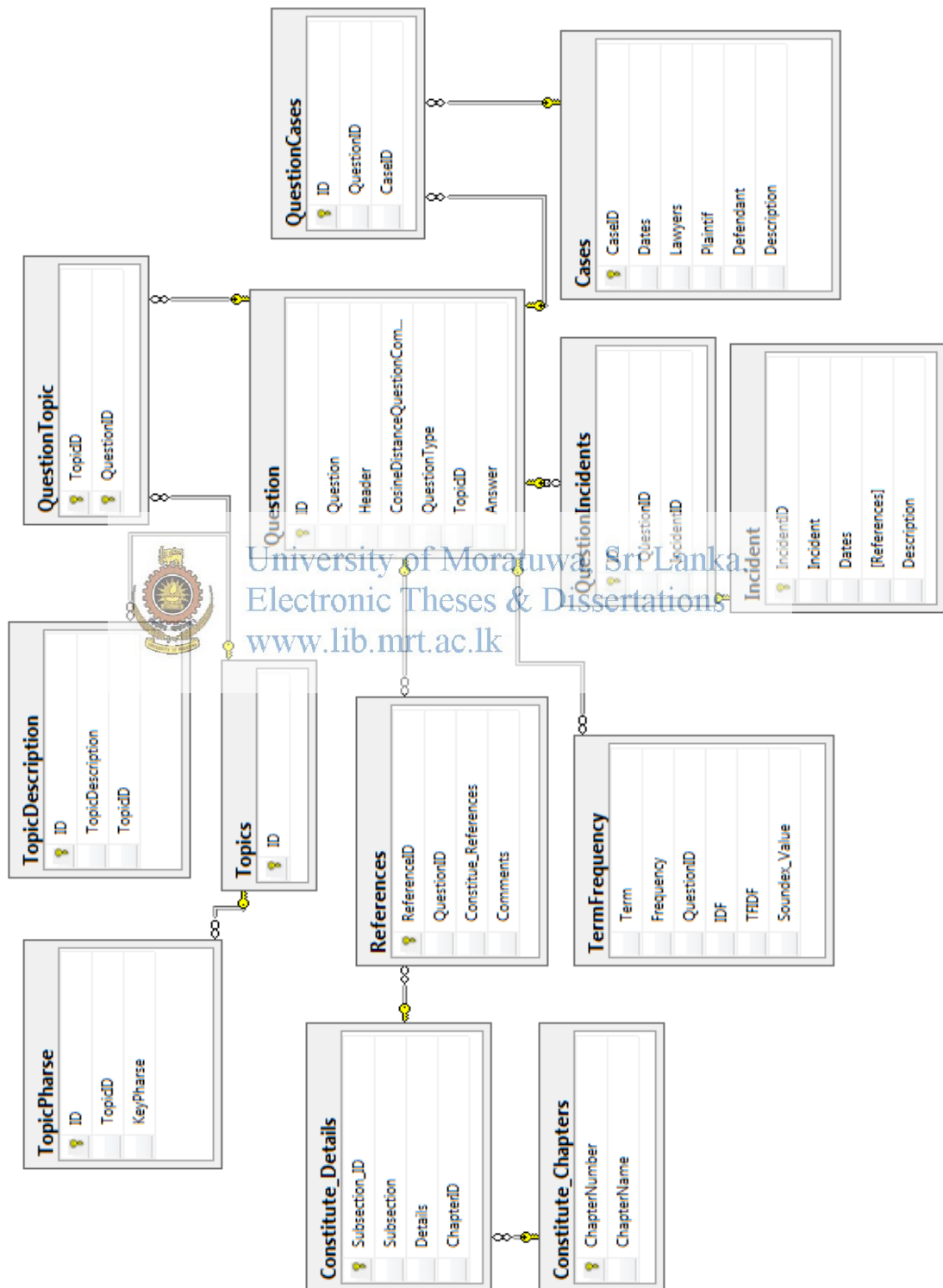
Overview of Full Text Search Architecture of SQL Server

Interfaces of HeainOnline Legal Systems

Configuration of Full Text Search in SQL Server
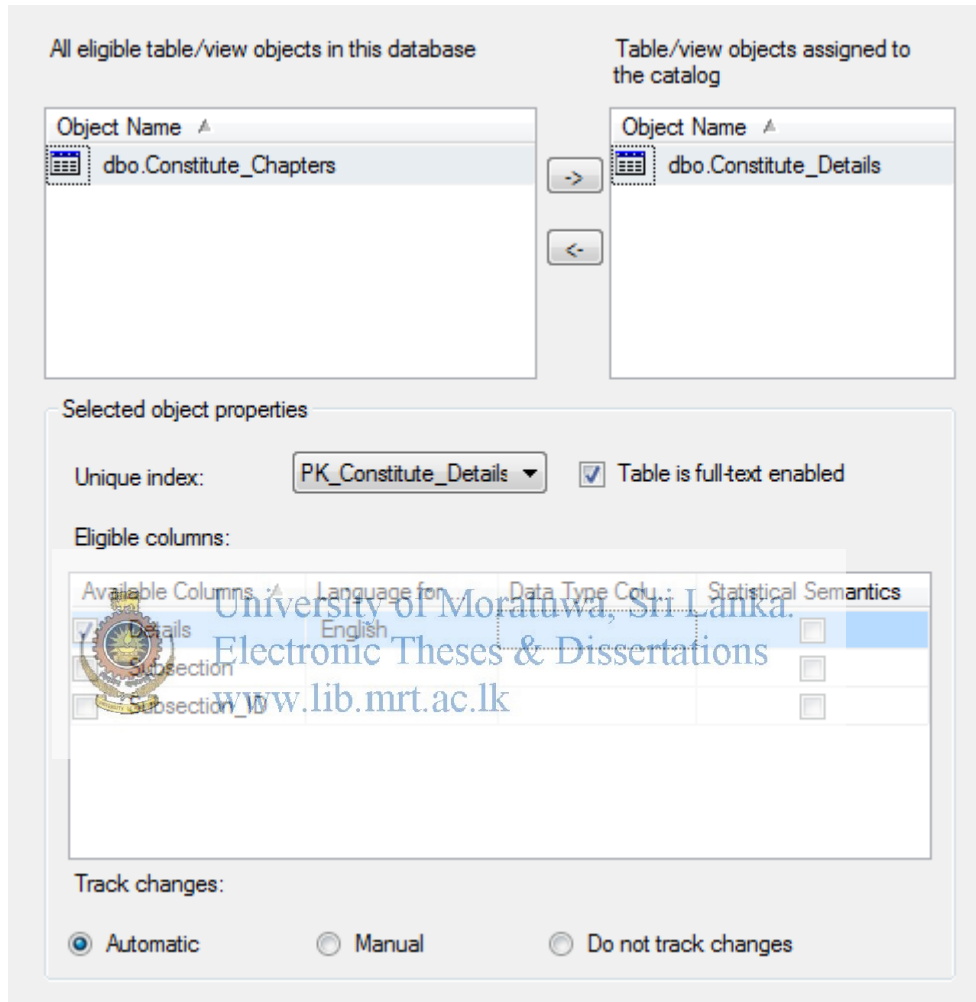
Enabling Full Text Search

In the above configuration automatic track change option is selected.

Since the legal content is changing with the amendments and new judgments, full text needs to be rebuild with a given schedule as shown in below figure.

Scheduling Full-Text Search Rebuilding

As shown in the above figure, every day full text search is rebuilding. However, end user has the option of modifying the schedule depending on the need.

Accent Sensitive is also enabled as shown in figure below.

Enabling Accent Sensitivity

User Interface of Proposed System

EasyQuestionSearch

Question:

Does the constitution provide for at least one procedure for amending the constitution

Get Matching Question

Question	Label	CD Rank	QT Rank	Final Ra...
Does the constitution provide for at least one procedure for amending the constitution?	Amending	2.0655	3	2.0655
Who approves amendments to the constitution?	Amending	1.0912	1	1.0912
Which provisions are not amendable?	Amending	1.0199	1	1.0199
What are the details of the amendment proposal process?	Amending	0.6858	1	0.6858
What languages does the constitution list as national?	Race, Ethnicity...	0.6713	1	0.6713
What are the details for the amendment approval process?	Amending	0.6521	1	0.6521
Does the constitution provide for a right of privacy?	Rights	0.6303	3	0.6303
What proportion of the vote is needed to approve a constitutional amendment?	Amending	0.5557	1	0.5557
What languages does the constitution list as official?	Race, Ethnicity...	0.5481	1	0.5481

Get References   No Matching Questions

QuestionReferences

Does the constitution provide for at least one procedure for amending the constitution?

Answers | Constitution Reference | Incidents | Cases

Chapter Number	Chapter Name	Sub Section	Sub Section Name	Description
12	THE LEGISLATURE - Ame...	82.1	Amendment or repeal ...	No Bill for the amendment of any provision of the Constitution...
12	THE LEGISLATURE - Ame...	82.2	Amendment or repeal ...	No Bill for the repeal of the Constitution shall be placed on th...
12	THE LEGISLATURE - Ame...	82.3	Amendment or repeal ...	If in the opinion of the Speaker, a Bill does not comply with th...
12	THE LEGISLATURE - Ame...	82.4	Amendment or repeal ...	Notwithstanding anything in the preceding provisions of this ...
12	THE LEGISLATURE - Ame...	82.5	Amendment or repeal ...	A Bill for the amendment of any provision of the Constitution ...
12	THE LEGISLATURE - Ame...	82.6	Amendment or repeal ...	No provision in any law shall, or shall be deemed to, amend, r...
12	THE LEGISLATURE - Ame...	82.7	Amendment or repeal ...	In this Chapter, "amendment" includes repeal, alteration and ...

No Bill for the repeal of the Constitution shall be placed on the Order Paper of Parliament unless the Bill contains provisions replacing the Constitution and is described in the long title thereof as being an Act for the repeal and replacement of the Constitution.