AUTOMATED KNOWLEDGE UNDERSTANDING AND RECOGNITION ASSISTANT (AKURA)

Project ID: 17-026

Project Proposal Report

Bachelor of Science Special (honors) In Information Technology

Department of Software Engineering

Sri Lanka Institute of Information Technology

Sri Lanka

March 2017

AUTOMATED KNOWLEDGE UNDERSTANDING AND RECOGNITION ASSISTANT (AKURA)

Project ID: 17-026

Project Proposal Report

Authors:

Name	Registration Number
Nilesh Jayanandana	IT14001826
Nipuna Herath	IT13104504
Sameera Piyasundara	IT14063442
Rishanthakumar Rasarathinam	IT14087820

Supervisor: Mr. Darshika Niranjan Koggalahewa

Bachelor of Science (Honors) in Information Technology

Department of Software Engineering

Sri Lanka Institute of Information Technology

Sri Lanka

March 2017

Declaration of the candidate & Supervisor

We declare that this is our own work and this proposal does not incorporate without acknowledgement any material previously submitted for a degree or diploma in any other university or Institute of higher learning and to the best of our knowledge and belief it does not contain any material previously published or written by another person except where the acknowledgement is made in the text.

Name	**Student ID**	**Signature**
H.H.N.C.Jayanandana	IT14001826	
H.P.N.H.Herath	IT13104504	
H.M.S.Piyasundara	IT14063442	
R.Rishanthakumar	IT14087820	

The above candidates are carrying out research for the undergraduate Dissertation under my supervision.

Signature of the supervisor: Date:

Abstract

This research involves addressing the difficulty in automating the integration of knowledge models extracted by Natural Language Processing and Understanding through software means in a dynamic machine environment. This document proposes and describes of a generic framework, which would enable automating knowledge understanding, learning and integration into a generic model, which could then be used to develop a domain specific product by means of simply giving unstructured textual data as the raw input. The approach to this project involves in coming up with a simulation of Natural Language Processing and Understanding, which is similar to humans, by using an ontological approach with a learning aspect and representing the extracted data into a knowledge base which could later be retrieved in an automated way. This will be done by semantic analysis and identification of relationships between tokenized streams of information, extraction of the said semantics and mapping the semantics taken to a heterogeneous knowledge model which will later be used on knowledge retrieving concepts on the existing heterogeneous dynamic knowledge model. The generic framework proposed in this document will allow users to develop a constantly evolving knowledge model by processing unstructured human readable forms of data and extraction and integration of the said data without any human interaction. This framework will allow organizations to develop a well-organized, meaningful and a well-structured domain specific product according to the requirements they please.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF ABBREVIATIONS

Abbreviation	Description
RDF	Resource Description Framework
API	Application Programming Interface
OWL	Web Ontology Language
NLP	Natural Language Processing
SRG	Semantic Resource Graph

LIST OF TABLES

1. INTRODUCTION

1.1. Background

Natural language processing and learning concepts, knowledge extraction concepts, heterogenic ontology integration concepts, knowledge retrieval concepts and self-learning concepts are vast growing areas in today's world. These concepts have been researched for many years and yet the accuracy of the applications based on these concepts are not upto the expected level. Therefore, many research projects are still being carried out in order to achieve better accuracy. Currently, there are several tools and models, which have been developed by using these concepts such as NLTK library, GATE from Stanford University, coreNLP, etc.

The proposed research will focus on natural language processing (NLP) and understanding concepts, as Human Language understanding is required. NLP is a combination of computer science, computational linguistics and artificial intelligence. The main part of NLP is to have good corpus/corpora, which means a huge set of words, which can be used to extract information from the source, that NLP actions performed. As per the definition mentioned above for ontology models, to have a truly intelligent model/system there should be a way to capture the knowledge, process it, reuse it, and communicate it out to the world. To cater the said problem, ontologies are defined as explicit specification of conceptualization [26]. In our context, the ontologies will play a major role with context understanding and information mapping processes.

In the information mapping process, the main objective is to create disposable ontology models or ontology models that can be mapped and integrated into a primary ontology model which can be called as knowledge base or the brain of the system. The techniques of mapping information will be discussed in the latter part of the document. After the mapping of meaningful information to the discussed ontology models it should be validated and verified that the new knowledge model ontologies which are heterogenic does not exist in the primary model before it absorbs their knowledge into the core of the system.

Since these co-ontology models or the disposable ones are made for each and every data chunk, it will be difficult to map those into the parent model manually. So, that the mapping process of knowledge models to the parent model should be automated dynamically. The final outcome will be a strong ontology knowledge model, which has the ability of representing or communicating on behalf of several domains, which has been learnt from its own. The core idea of the base knowledge model is to provide a generic platform where it can be used as a business model, which generates more power and more revenue to the corresponding party.

Ontologies are models of the entities of interest to a concept and the relationships among those entities. Ontologies consist of Classes which are types of entities and usually in a class there can be class-subclass hierarchies. Ontology models consist of following properties which are designating relationships among entities (members of classes). Usually there exists property-sub property hierarchies as well. When it comes to knowledge representation techniques, RDF: Resource Description Framework provides a graph based data model or framework for structuring data as statements about resources. Each statement is called a triple which would contain a subject, predicate and an object. The subject of a statement is called a resource. The predicate is called a property and the object is called a value. Objects of RDF triples should be either URIs or literals. There are multiple machine readable syntaxes for RDF triples namely, RDF/XML, Notation 3/ N3, Turtle, N-Triples, etc. There exist many software systems for editing ontologies as well. Out of them proteage by Stanford university and TopBraid Composer [53] by Top Quadrant stands out. Table 1.1 displays a comparison between semantic network and an ontology and how they standout. Semantic network is a set of linked data available where everything is linked with each other where relevant. This has given rise to another concept called semantic web which focuses on bringing meaningful information to the web with the help of URIs. [55]

In contrast with machine learning and deep learning algorithms and techniques, ontology driven approaches have been favorable and efficient on resource consumption as neural networks deployed in deep learning and machine learning techniques tend to consume a lot of processing power which results in costly initiation and maintenance. Ontologies require nothing of the sort as it does not consume processing power as much and SPARQL is a very efficient technique for processing and retrieving information from a knowledge base modelled into an ontology.

Table 1: Semantic Network vs. Ontology Comparison

Features	Semantic Network	Ontology
Unique Name Assumption	If two objects have different names, they are assumed to be different.	There is no assumption on whether objects are the same or different unless there is an explicit statement about specifying the relationship.
Open vs. Closed World Assumption	Nothing can be entered into it until there is a place for it in the corresponding template.	Anything can be entered into ontology unless it violates one of the constraints.
Assertion vs. Classification	Defining facets on a slot at a class, or defining a constraint on a slot at the top level, makes a statement about all instances of those describing necessary conditions for instances of that class.	There are effectively two kinds of statements about classes: a) those that are true of all individuals in a class, and b) those that is collectively necessary and sufficient to define the class.
Ability to define Rules	Rules can be applied when implementing semantic net using PROLOG.	No rules can be applied when implementing ontology using OWL.
Expressive Power	There are no restrictions on relationships and property characteristics.	It allows some restrictions; anonymous classes; necessary and sufficient conditions; expressions

Source: [54]

Heterogenic Ontology Integration is an ongoing research area where multiple researchers try to automate ontology integration. Homogenous Integration has been achieved for a known structure of an ontology model, but for an unknown set of heterogenic ontology structures, a proper automated theory is still under research. OWL API is a Java API and reference implementation of creating, manipulating and serializing OWL technologies. Knowledge retrieval from a knowledge base can be done through querying the knowledge base by using query technologies such as SPARQL, RDQL, nRQL, Jena2 etc. Ontology query languages were developed to query the information defined by the ontology languages and reasoning systems. The information needs may be expressed in different queries because of different user perspectives, background knowledge, terminological habits and vocabulary [52] "An Ontology-based Framework for Knowledge Retrieval". Knowledge retrieval plays a major role in commercial systems that are build using ontologies, because there should be a proper output that can be displayed to the user in human understandable or readable format.

1.2. Literature Survey

1.2.1. Already existing systems

The framework proposed in this document doesn't exist in any form of research yet with all four components explained below integrated together as one. Therefore, the summary of the information we extracted from multiple research papers will be included in the below sections of this document. However, there are multiple existing implementations of the product we are going to develop as the proof of concept using the proposed platform will be mentioned here.

Google Cards / Google Now

Google Cards provide a summary of information in form of cards for movies, shows, actors, etc., by using a machine learning approach. The summary includes a brief description, related personnel, customer reviews and ratings.

Wolfram Alpha

Wolfram Alpha, is a computational knowledge engine which uses machine learning techniques to targets almost every domain and tries to display a summary of the data, which is not very accurate for the time being. [47]

1.2.2. Natural language processing and data extraction

We found quite a few researches in the field of Natural Language Processing and Data Extraction. In [14] "A Programmable Implementation on Information Extraction and Categorization", a journal published in International Journal of Multimedia and Ubiquitous Engineering, the research suggests using a "Domain Dictionary" as the best technique for Information Extraction because it accesses the core part of the word pattern and analyzes the theoretical properties of the word. The three techniques they compare are Stemming, Domain Dictionary and Execution List. Here, the group categorizes stemming into two sub techniques as Derivational and Inflectional stemming where Derivational Stemming derives a new word by simply changing its grammar. They describe Inflectional stemming by quoting "When the normalization is confined to regularizing grammatical variants such as singular/plural or past/present, it is referred to inflectional stemming" in [16] "Text Mining Application Programming", a journal published by Manu Konchady. In the example *verify-verified-verifies, walk-walked-walks*, in both the cases, all the words will be treated as 'verify' and in the second example, will be treated as 'walk'. The next technique they considered is "Domain Dictionary" method. This method simply uses a Knowledge Base which consists of a collection of 'feature terms'. The dictionary structure is further divided into three categories namely, Parent Category, Sub-category and Word. The 'Parent Category' is a set of words, which are unique, and 'Sub-category' may inherit multiple parent categories and then the words inherit the 'Sub-Category'. The last of the techniques they considered is "Exclusion List", where an exclusion list or an unwanted words list is maintained separately containing words such as the, a, an, if, off, on etc. and extracting only the 'wanted words'.

Another study, "Automated Concept Extraction from Plain Text carried out by a group in Michigan State University" [18] which follows the "bag-of-words" concept which suggests ignoring all the sense of the whole sentence but focusing on the relationship of occurrences of words in the sentence. Here, they suggest identifying the relationships of different words in a text based on a lexical database and "identifying groups of these words which form closely tied conceptual groups". The knowledge base that they have used is a WorldNet, which is a network of binary related nouns and verbs. The relationships extracted are then used to create a graph called a Semantic Relationship Graph (SRG). The SRG is then used to distinguish each concept, which will appear in the text.

1.2.3. Ontology driven Information extraction

Ontology based information extraction is a vast area which has been discussed over a long period of time by several researches and still it is in the research field to be explored. While gathering the information via the journals and individual research parties, we have found that ontology and information extraction are two processes which are interconnected by the applications and the context of the applications. Unlike in traditional Information Extraction Process Ontology based Information Extraction is able to link the newly extracted entity to the formal ontology. And also OBIE does not require various algorithms for Information Extraction. Ontologies will guide them towards efficient Information Extraction.

Different systems built in this area mainly differ from Ontology Extraction components, Ontology Update processes etc. OBIE is a process which will output the most efficient and relevant information through Information Extraction process. Among many tools that facilitate this GATE [28], UIMA [29], sProUT [30], SOBA [31], Text-To-Onto [32], and OntoX [33], Stanford coreNLP [41] plays a major role as most of them are open-source and freely available. Among these open source tools GATE and coreNLP plays a major role in this context and both the tools are Java related or at least JVM based tools. The main sectors of this area which has been discussed in several researches are data extraction, extraction of meaningful information, understanding the natural language and context of it and mapping knowledge to ontology models manually. The important part of this ontology based information extraction was identifying or understanding the context of the sentences.

After extracting the raw data from the world-wide web (Twitter, Facebook) through natural language processing operations such as tokenization, POS tagging, etc. the extracted data will be arranged to a meaningful manner which can be called as meaningful information. Then from that meaningful information we have to identify the context of each sentences/phrases. The process of identifying the context called as the natural language understanding. To perform natural language understanding process or context understanding they have used the Stanford coreNLP toolkit. After the context understanding process the flow will be continue through the mapping process of knowledge into the ontology models. This process was done manually by feeding extracted knowledge to ontology models. Currently they were using only one ontology model at a time and therefore it can be only applied to a specific domain [38] “Towards Partial Completion of the Comprehensive Area Exam”.

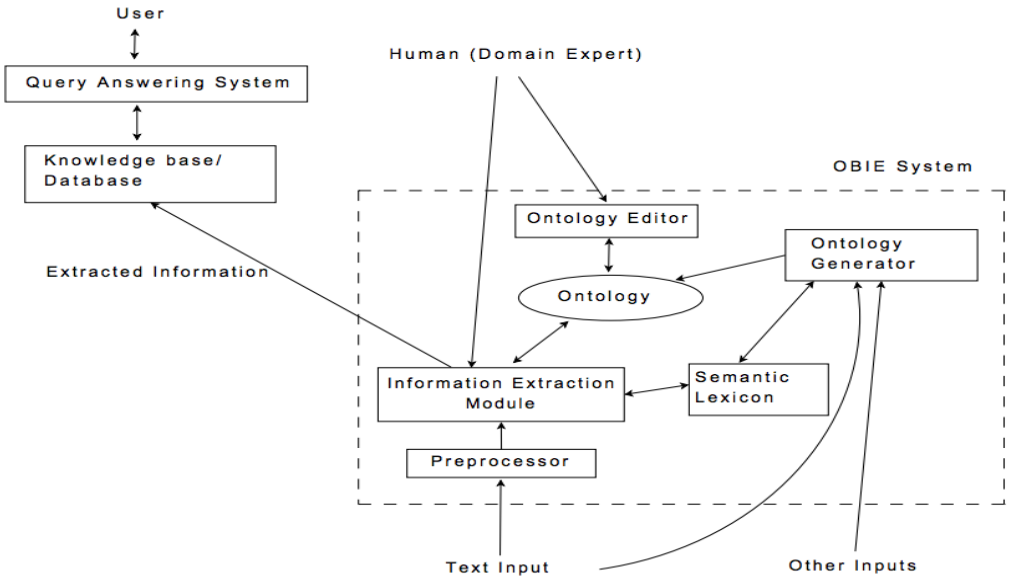

Figure 1.1: The general architecture of an OBIE System

According to the researches OBIE definition is highlighted as a system that processes unstructured or semi-structured natural language text through a mechanism guided by ontologies to extract certain types of information and presents the output using ontologies. Most of the information extraction techniques used by these OBIE systems have been adopted. Following techniques are the main techniques which were followed by the researches [38],

- Linguistic rules represented by regular expressions - Capturing certain information by specifying regular expressions.
- Gazetteers - This relies on finite-state automata just like linguistic rules but recognizes individual words or phrases instead of patterns
- Classification Techniques
- Construction of partial parse trees
- Analyzing HTML/XML tags
- Web based search

From the above discussed research reviews, they all have tried to come up with an ontology model which focusses on a specific domain and some of their future works have mentioned that they are planning to enhance their OBIE systems to cater multiple domains without targeting on a specific domain and also to use multiple ontology models in their information understanding and mapping processes.

1.2.4. Ontology based knowledge mapping and integration

There were a number of researches that was promising in the field of ontology integration or in this case mapping or expansion. In [22], "Semi-Automatically Mapping Structured Sources into the Semantic Web", a journal published in 2012 by the developers of Karma Framework explores expanding an ontology using existing databases or data stores with a semi automatic approach. As the first step, they assign semantic types, which involves mapping each column of the source to a node in ontology. This is not automated and is a user-guided process where the user guides the system on assigning the types using a GUI provided. In the next step A graph is constructed that defines space for all the mappings between source and ontology. Third step refines the graph based on user input. The graph is constructed so that the mapping between source and ontology can be computed using Steiner tree algorithm [23]. In the final step a formal specification will be generated and the data will be converted using the said specification.

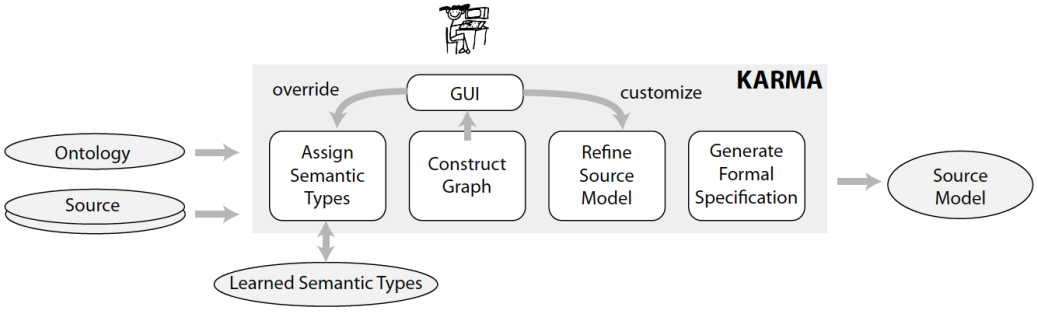

Figure 1.2: Karma process to model structured sources, Source [22]

According to Choi [24] "A survey on ontology mapping", ontology mapping can be classified into three categories:

1) Mapping between an integrated global ontology and local ontologies,
2) Mapping between local ontologies
3) Mapping on ontology merging and alignment.

In particular, the third category is used as a part of ontology merging or alignment in an ontology reuse process which explores further in [25] "An Approach to Ontology Integration for Ontology Reuse", which mentioned below have proposed a framework for ontology reuse.

The researchers Enrico and Antonio in [25] "An Approach to Ontology Integration for Ontology Reuse", have approached ontology integration for ontology reuse with implementing the following components in their framework. Reference model retrieval function block which is responsible for retrieving the reference models corresponding to the domain of interest. In order to search for proper data models, it is needed to identify the knowledge domain and the related subdomains covering the specific topic under study. The second component of the framework is the Reference Model Reconciliation and Normalization function block. It is responsible for adapting the collected reference models to a common representation format. The main component of the framework is the Reference Model Matching function block. It is responsible for obtaining an alignment which a set of correspondences between the matched entities from the reference models. The matcher involves three types of matching operations: string, linguistic and extended linguistic matching. The fourth component of the framework is the Reference Models Merging or Integration function block. It is responsible for integrating the selected input models into a global, richer and consistent view, abstracting the local conceptualizations of the input models themselves. As the case study, they have selected food domain specifically to the industrial production of food and has yielded successful results.

In [39] "Geo-ontology Integration Based on Category Theory", Ling Wang explore ontology integration between semantics solving the problem of semantic interoperability. They introduce Category Theory, which is a highly abstract mathematical theory to realize the integration of ontologies by abstracting ontologies into category objects and screening heterogeneousness of ontologies through the morphisms of the objects. The following challenges were identified on geo-ontology merging which are Different terms are used to refer to the same semantic, the same term is used to refer to different semantic and the same ontology is classified differently in different category systems.

The new ontology produced by merging two heterogeneous ontologies covers both similar and different concepts, also keeps the original structure and level, making sure that the merged result is complete, exclusive and minimum. Meanwhile through the integration of ontologies by abstracting ontologies into category objects and screening heterogeneousness of ontologies through the morphisms of the objects, category theory offers a systematic approach to the integration of heterogeneous geographical information, a tool, research approach to computer theory.

Yanhui Lv and Chong Xie in [40] "A Framework for Ontology Integration and Evaluation" have proposed a framework for ontology integration based on ontology similarity measures including syntactic similarity and graph structural similarity. Syntactic similarity focuses on the intended meaning of the concepts. This is discovered by rules such as Porter Stemming, Levenshtein, Prefix and Suffix, Substring, and Metaphone algorithm. These rules are explained in detail in [40]. Graph structural similarity addresses the aspect of how to organize the concepts while the semantics is agreed to be identical. The approach has advantages over the existing approaches in that, it first finds the places where ontologies overlap instead of computing similarity score between entities one by one, which can reduce the burden of maintenance of the alignments. After integrating, they tailor the integrated ontology through similarity measures and check the consistency of the consequential ontology.

In [43] "Research on high-speed railway ontology integration method Based on Semantic Relationships" explores integration of multiple majors' domain ontology. This is also a domain specific integration as the research focuses on railway systems. Research [43] uses Protégé, and gives three different ways to the ontology alignment of multiple majors' domain.

1. Single Ontology Approach
2. Multiple Ontology Approach
3. View of Ontology Integration

Single Ontology Approach is using one ontology as the top ontology and provide a shared vocabulary for related semantics, the primary ontology will be closely related with all of the information sources. In Multiple Ontology Approach, each information source has its own domain ontology. However, in the case that multiple ontology approach is lacking with public vocabulary, the ontologies from different data sources are difficult to compare. In order to solve this problem, ontology mapping need to be constructed. View of ontology integration is similar with multiple ontology approach. Each information source has its own domain ontology, describes the semantic by its own ontology, as shown in figure. Establishing a shared ontology view in the top level (core terminology Collection), covering the domain of basic elements, is convenient for comparison between single ontology.

Although the above research is on ontology semantic relationship and ontology integration, but the research that considered from the systematic angle on multiple majors' domain ontology is absent to some extent. Fisnik Dalipi, Florim Idrizi, Eip Rufati, Florin Asani in their research, "On Integration of Ontologies into E-learning Systems" [44] tries to implement a better solution for organizing and visualizing didactic knowledge. Their paper aims at proposing a model which is focused on integrating ontological principles with e-learning standards. They have developed a prototype model that is integrated with an ontology, which gives a semantic representation of learning contents by adding semantic notations to each learning resource. They have used the Resource Description Framework (RDF) as the underlying technology in achieving their target on the specified domain of e-learning.

1.2.5. Knowledge retrieval and representation

There were lots of researches done in the area of knowledge retrieval from ontology, in order to improve the accuracy level of retrieving the required knowledge. In [3] "Ontology-Based Semantic Retrieval for Education Management Systems", Lijun Tang and Xu Chen proposed ontology based semantic retrieval approach and framework for education management system. They presented some rules for constructing domain ontology from the education management system and also the semantic annotation method of the constructed ontology was given. On this basis, they provided a brief ontology-based semantic retrieval algorithm. In this they were targeted on semantic retrieval, which can retrieve information fully and precisely, based on the knowledge understanding and knowledge reasoning.

"Ontology Based Information Retrieval System for Academic Library" [4], proposed a system that overcomes the limitation of keyword-based query handling systems and capable extracting relevant information instead of giving list of answers. They used Jena API for mapping of SPARQL with RDF database and retrieving the relevant information. Protege editor is used for creating ontology in RDF data format for Academic Library. This system is domain specific but in future they are targeting to apply this method for different domains also.

“Ontology Semantic Approach to Extraction of knowledge from Holy Quran” [5], presents an abstract representation semantic search system for Holy Quran knowledge using a combination of the natural language processing approach and semantic technology. The semantic search approach is used to retrieve more precise relevant verses relating to the query by the use expressed in the complete natural language expression. In [6], “An Ontology-Based Retrieval System for Mammographic Reports” a novel system based on text pre-processing techniques and a modeled medical knowledge, using an improved radiological ontology. This system can be used for text reports retrieval and classification, integrating it in radiological RIS/PACS systems, enabling the user to do advanced queries, search for similar pathological cases, rare cases, statistics extraction and so on.

In [49] “Comparison of Question Answering Systems Based on Ontology and Semantic Web in Different Environment”, they use a semantic search methodology for retrieving answers from ontology model. Where this semantic search is used to improve the accuracy of the search by understanding the intent of the user and the meaning of the terms in the searching sentence. Semantic Search uses semantics to produce relevant searching results. Here they propose the Graph Matching Algorithm for query matching with the ontology using Spread Activation Algorithm. Graph Matching Algorithm is used to search the concepts in the repositories. Spread Activation is a method for searching the nodes in the ontology as in semantic manner.

In [50] “Querying Ontologies: Retrieving Knowledge from Semantic Web Documents”, An application which provides a handy tool to retrieve knowledge from an OWL 2 Ontology. Designed with GWT (development toolkit for building and optimizing complex browser-based applications), the client side code (interface) runs faster and makes the procedure of querying a knowledge base sufficiently user friendly. By using SPRQL-DL as the query language, it complies with the W3C’s recommendations and takes advantage of the advanced query capabilities of Pellet. SPAQL-DL proves to be a query language that can support the advanced expressivity of OWL 2. Pellet itself proves to be quite effective in supporting this expressivity. Thus, the whole application performs knowledge retrieval effectively.

A Keyword retrieval method based on thesaurus and ontology, aiming at combining the general retrieval system and domain ontology effectively to solve varied problems between user's keywords and words used by domain experts to describe the ontology, to improve the efficiency and accuracy of retrieval was proposed in [51] "Keywords Retrieval Based on Ontology Inference". In this a retrieval model combining the traditional retrieval method and keywords search based on ontology reasoning method is used. The model consists of three parts: Information Retrieval System, Synonym Processor System and Ontology Retrieval Service System. Distance measure is used on the results to control the results in the availably range in order to limit the result collection not to overexpansion.

According to the above researches and other refereed research papers, most of them were done based on specific domain. The main thing they were lacking is building generic framework which can be adopted to set of domains which will lead to better reusability rather than implementing specific models to targeted domain.

1.3. Research Gap

Natural language processing and data extraction

According to the literature survey we conducted, we found a number of knowledge extraction methodologies that are still under research. One of the main differences of the research undergoing compared to existing researches is the usage of dynamic textural sources. The currently existing researches only addresses how to extract data from a limited data set which is already existing and fed from a static resource. In one of the researches carried out [42] the textural resource they use are PDFs. In the aforementioned research paper, the PDFs are broken into triple-store data after pre-processing and are represented in a hyper-graph that consists of collections of binary relations of 'triples'. But the approach we will be taking will be referring to differently structured multiple sources of data to feed the knowledge base with information, while looking for information that do not exist in the knowledge base yet. The approach includes fetching date, understanding the data according to the context, purify and pre-process the data for meaningful data extraction. For an instance, a set of user reviews will be given percentages according to the positivity or the negativity of the review.

In the literature survey conducted, Behrang QasemiZadeh [45] comes up with an approach for text mining and extraction from scientific or research papers. But he only uses Natural Language Processing for his approach. But what we will be researching on is using a combination of Natural Language Processing, Semantic Analysis and Understanding and automation of the combination. Another research published by H. Mima [46] suggested a knowledge management approach using ontology-based similarity calculation but in our research, we will be using Natural Language Processing also apart from Ontology based knowledge extraction.

Meaningful Information Extraction into Ontologies

According to the research paper reviews of ontology based information extraction and context understanding it was noticed that ontologies are widely used by knowledge representation systems and is a sub section of artificial intelligence which is related to cognitive science in [38]. Furthermore, the word semantic web is highlighted in several research papers featuring ontology models. Semantic web is an approach to bring meaning to the unstructured data in web. Ontologies are said to play a vital role in the backbone of the semantic web. Ontology based information extraction (OBIE) has been researched from several researchers throughout the last few years to gain the quality of ontologies and to increase accuracy level. These OBIEs' will be different from the traditional information extractions systems by the following sectors,

- Process unstructured or semi structured natural language text
- Present the outcome using ontologies
- Use an information extraction process guided by an ontology

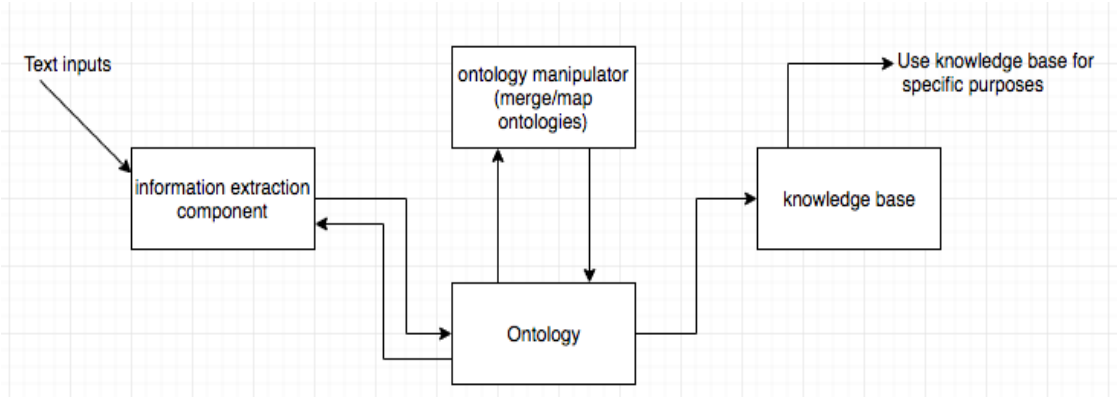

Figure 1.3: Ontology based information extraction

Figure 1.3 displays the process of learning an ontology model by manipulating the information or data retrieved by an external source. Hence it is understood that the learning part of this ontology model will be done manually or will depend on a specific domain which would not be dynamic. Therefore, taking into account the issue mentioned above, the main objective of this component will be to overcome the static environment of this ontology based information extraction and provide a realistic dynamic feature to it. Therefore, OBIE component which will be developed in the proposed framework would implement a dynamic learning model which will focus on being generic rather than specific for any given domain.

Automated ontology mapping and integration

According to the literature survey conducted, ontology mapping and integration is a topic which is still under heavy research. Out of the findings, many researches have tried multiple approaches and gained fruitful results only under specific set of domains. One of the generic platform approaches for the ontology integration was addressed in the Karma Framework [22] "Semi-Automatically Mapping Structured Sources into the Semantic Web". But the approach taken by the said framework requires human assistance in mapping two or more data sources together into a knowledge model. There exists no such proper system to automatically map two or more ontologies of same or different domains irrespective of their attributes and structure with less or no human interactions in a dynamic machine environment. Some researchers have identified particular approach techniques in mapping two ontologies of the same domain such as the case in [25] "An Approach to Ontology Integration for Ontology Reuse". They have explored the techniques and approaches in their paper so that someone could better enhance them and finish what they have started off with.

The generic platform we are going to implement will consider the above factors and take account to existing theories and algorithms at hand such as the Category Theory implemented in [39] "Integration between semantics solving the problem of semantic interoperability." which was focused on geo-ontology domain, [43] "Research on high-speed railway ontology integration method Based on Semantic Relationships" which focused on railway domain and improve them in ways such that they would support generic approaches rather than specific domains.

Knowledge retrieval and representation

According to the findings, what the existing knowledge retrieval systems are lacking is proper communication among the other module components in order to continuously learn and acquire new knowledge. The existing systems are only able to give a valuable response to the user if it contains the required information only. Where this frustrates the user when they continuously experience this kind of situation. First of all, it is important to identify the unavailability of user-required information in a considerable accuracy level and with less time processing. Then there should be a communication path where it provides necessary information to the data extraction model to retrieve the required information from the relevant sources and it should continuously update the knowledge base. Where this will make the user to experience new information gathered by continuous learning feature of the framework. Therefore, in order to have this continuous learning and dynamic knowledge representation in the ontology, the knowledge retrieval model should be able to provide sufficient information to the data extraction model to acquire the relevant information from the provided sources.

Research Problem

In the digital era, web is considered to be one huge knowledge base of trillions of unprocessed, unstructured chunks of data which stays dormant and unused for the eternity of time which instead could be better used for extracting knowledge and retrieving meaningful information to serve a higher purpose. Currently, there exists no solution or framework that would enable unstructured textual data processing and extracting knowledge into a machine-readable format and enable users of the framework to build domain specific projects on top of the framework. Many existing solutions are either domain specific or need human intervention to realize its goals rather than be automated and dynamic. The proposed framework will address the difficulty in simulating the human process of natural language learning and understanding, knowledge extraction, knowledge integration and expansion, applying retrieval and answer interpretation in a Dynamic machine environment and provide endpoints and functionality to build a domain specific application on top if it efficiently.

2. OBJECTIVES

2.1. Main Objective

The main objective of this project is to provide a generic framework that would fully automate the knowledge extraction process from unstructured textual data and the integration of the said knowledge into a larger machine-readable model in a dynamic machine environment. The generic nature of the proposed framework should allow any domain specific product to be developed on top of it. Currently there exists no such generic framework at commercial / production level that would facilitate the features mentioned above. This framework is intended to make development of any domain specific product which relies heavily on unstructured textual data and dynamic knowledge models easier and efficient by providing interfaces and built in functions to support the required functionality.

2.2. Specific Objectives

The system proposed in this document will be a generic framework providing abstract functionality and certain endpoints consisting of the following four separate components, which will exhibit the properties low coupling and high cohesion.

Natural language processing and data extraction

This component involves in providing an automated approach for processing unstructured textual data in a generic way which will be provided by multiple dynamic data sources (Preferably APIs). Furthermore, this will enable and explore improving existing text tokenization and trimming techniques effectively and will focus on providing an output of meaningful structured dataset in an efficient way.

Automated ontology driven knowledge extraction out of semi-structured data

This component will derive an automated approach to extract a knowledge model represented by an ontology from a given set of textural data. This will explore understanding semantics of a given dataset and deriving concepts and relations between them which would be extracted and represented in an ontology.

Knowledge mapping and integration of heterogeneous ontologies

The main objective of this component is to automate integration or merging of two or more heterogeneous ontologies into a single primary ontology from learned content by the system which will be constantly evolving and expanding as the system evolves. Due to the generic feature of the proposed framework, the structure and relationships of the primary ontology will be dynamic as opposed to having a structured ontology.

Knowledge retrieval and representation of acquired knowledge in human understandable manner.

This component's main focus is to retrieve the information from the knowledge base and represent upon user's requests. Representation could be done through derived analytics, statistics, recommendations etc.

3. METHODOLOGY

3.1. System Overview

AKURA is a self-automated generic framework which is capable of processing the unstructured textual data given as input, extract knowledge from it and then expand or update its own existing knowledge model which could be used to derive meaningful information with less or no human interaction. All the algorithms in the proposed framework would be generic and API endpoints will be provided for integration with any domain specific product. However, as a proof of concept, AKURA will be integrated into a product review system that would extract domain specific data and process them with the proposed generic framework and then provide statistics to the user regarding the knowledge harnessed for a certain product using its customer reviews on multiple sites.

System will have 4 major components as,

1. Natural language processing and data extraction
2. Ontology driven information extraction
3. Ontology driven knowledge mapping and integration
4. Knowledge retrieval and representation

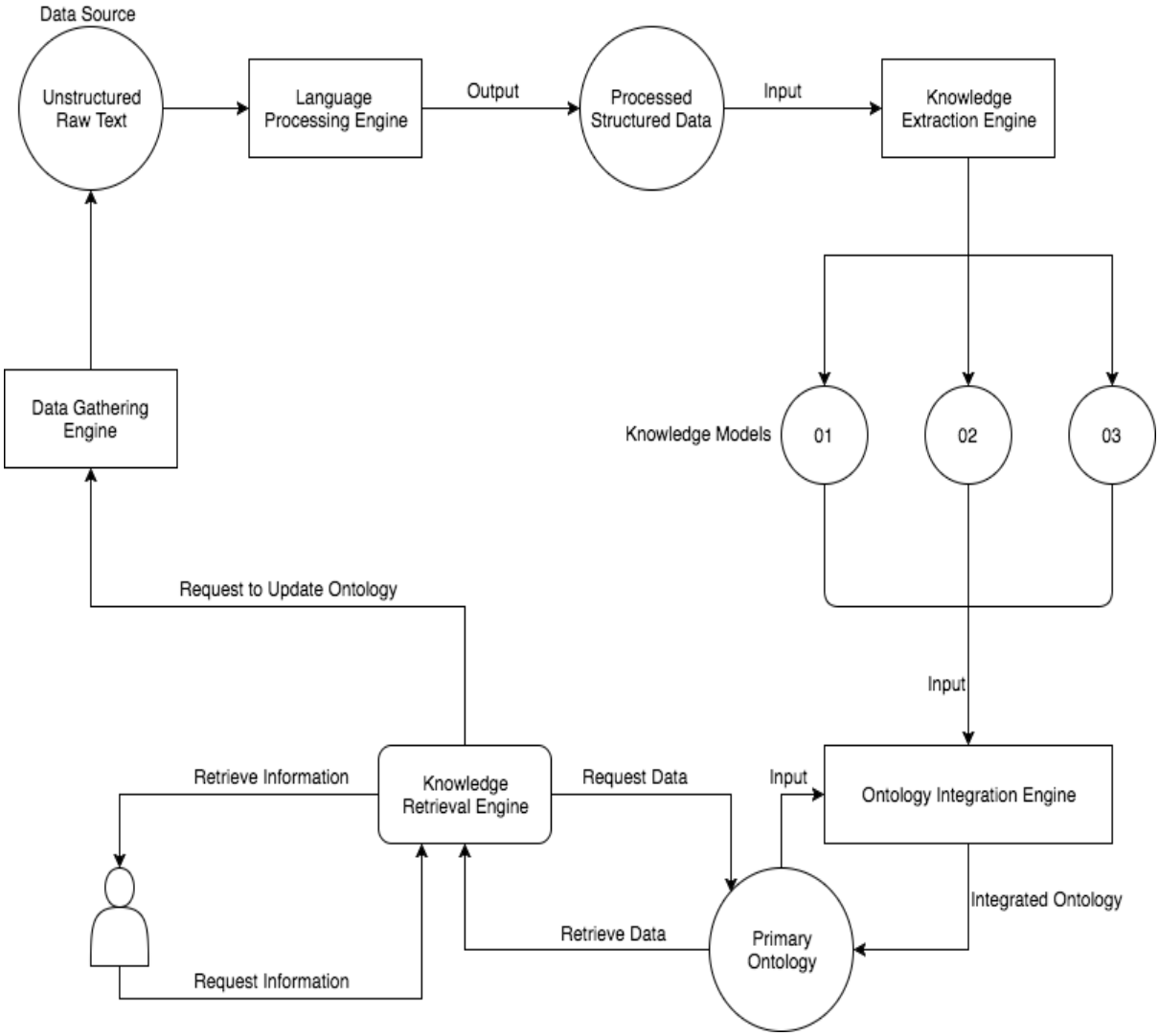

Figure 3.1: System Architecture

3.2. System Components

3.2.1. Natural language processing and data extraction

This component is the entry point of the project. This is the part, which gathers all the information that will be persisted in the knowledge base after a series of processing. The data retrieval will be done in many ways such as Public APIs of web communities and web crawling.

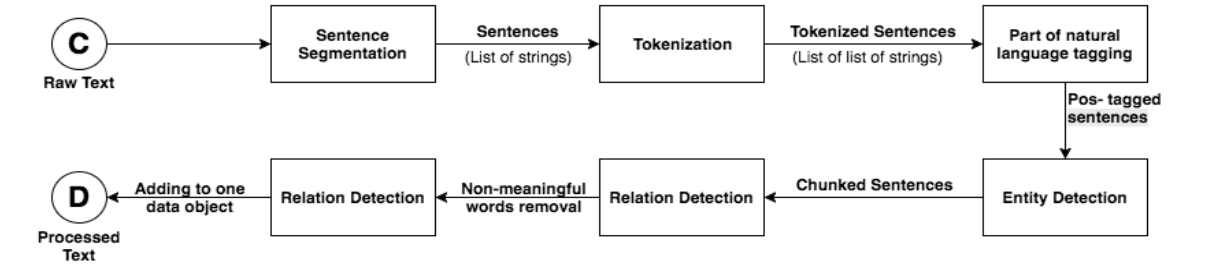

Figure 3.2: Gathering data from web sources

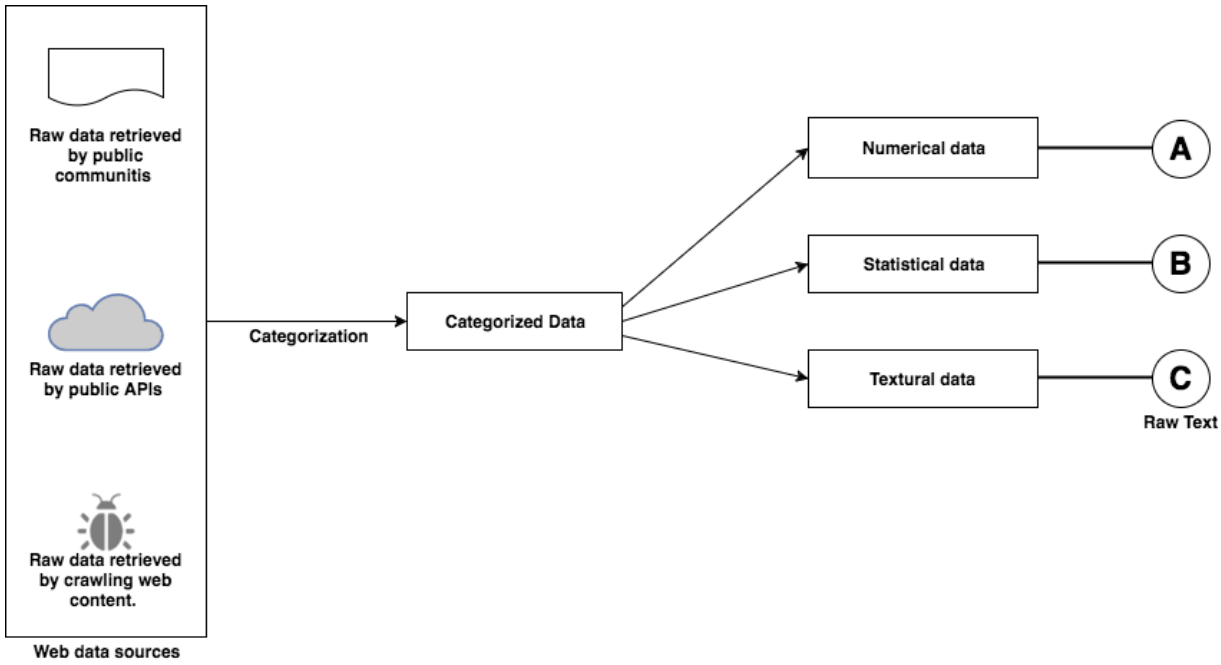

Figure 3.3: Processing strings

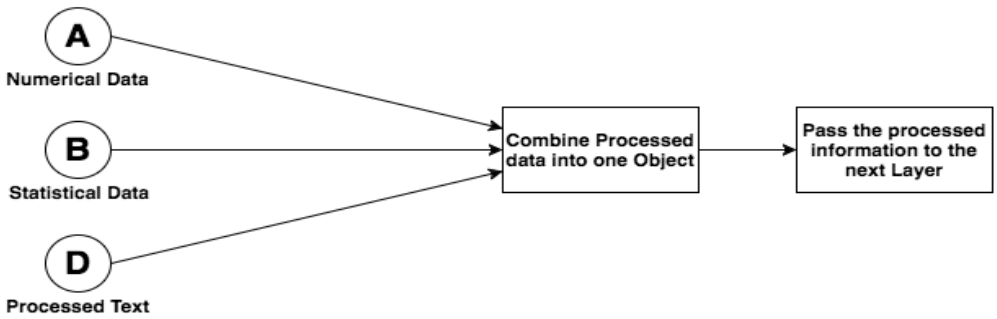

Figure 3.4: Combining processed strings

The data extraction will be done as a combination of multiple steps. First the raw data will be retrieved in various ways such as from public APIs of popular web services, using web crawlers or publicly available textural sources. Then the retrieved data will be categorized into different types based on what field the framework is configured. Once the data is categorized into different types, the categorized data will be again categorized into subcategories as numerical data, statistical data and textural data that needs further refining. Once the numerical and statistical data have been extracted, the remaining textural data (User reviews, Comments) will be further refined in order to remove unnecessary data such as emoji's, words with mistook spellings etc. Then the further refined textural data will be added into the previously defined categories using Open Source Natural Language Processing Tools such as Natural Language Toolkit(NLTK) to understand the context and decide the category. The technologies that we will be using are mostly Open Source projects. As this component heavily involved in data crawling, text processing and understanding, following Open Source libraries will be used.

Natural Language Toolkit (NLTK): Open Source Python framework for text mining, data scraping and sentiment analysis.

Apache Nutch is an Open Source data crawler developed by Apache Software which can be used for extracting data from websites which do not have a specific structure and does not provide public APIs.

Scrapy is an open Source Python framework that was developed for web scraping. This will also be used alongside NLTK to maximize the interoperability of the frameworks that we use.

3.2.2. Ontology driven information extraction

Ontologies are capable of representing data in such a way that the relationships are identified easily by providing a common ground for data representation. Another advantage of using an Ontology based approach in order to support Dynamic Knowledge Acquisition is that ontologies are easily scalable. Hence, new relationships can be added to Ontological models easily. This component is directly responsible in understanding the context of semi structured data derived from section 3.2.1 and deriving heterogenic ontology models out of it. The methodology is described below.

Information Extraction

Information Extraction is the procedure of processing unstructured or semi structured data, that is guided by ontologies in order to extract different types of information and provide an output using ontologies. In this process, obtaining high accuracy is vital and also very difficult. We will be using the approach mentioned in journal [12], where they claim to have obtained 95% accuracy, in order to achieve this.

Understanding the context

The problem of identifying what a pronoun or a noun phrase refers to, known as Anaphora Resolution and Coreference Resolution, which is the task of identifying all the expressions that refer to the same entity are recognized as a great challenge when it comes to NLP. We will be using Stanford coreNLP [11] implementation to achieve both Anaphora Resolution and Coreference Resolution in the proposed system.

Building an Ontology of a new domain

Step 1: Theme Concept Identification

Theme Concept refers to the overall topic of a text. It is unarguable that identifying the topic or what the text is about is what contributes the most towards the accuracy of the ontology model. As an example, if the text describes a person called Mary, we should identify Mary as the theme concept. Once this is identified, attaching the attributes and assigning values for them can be interpreted. The procedure of identifying the theme concept is as follows.

- Extracting tokens that are tagged as nouns (through sentence parsing, tokenization and parts-of-speech tagging) into a set. These are identified as concepts.
- Constructing the subjectList containing the subjects identified through triplet extraction (done by using StandfordNLP [11]).
- Form the MaxOccurConcepts set that contains the concepts that occur in the text maximum number of times.
- Identifying the theme concept can be found as the intersection of the above-mentioned sets as follows [12], $Themeconcept = Concept \cap subjectList \cap MaxOccurConcepts$. According to this theme concept is what occurs in the text most number of times and also is the subject in one or more sentences.

Step 2: Domain Ontology Identification

In order to decide on suitable domain ontology, we must identify the domain the Theme Concept is specific to. For this we use two rules as Explicit mention rule and Implicit lexicon match rule. Explicit mention rule is strings that appear in the context are names for the class itself. If Explicit Mention Rule is not satisfied Implicit lexicon match rule approach is used. Assuming that ontology of the specific domain exists, a string matching is performed on a domain ontology lexicon that is derived from class attribute names. Then a semantic lexicon is created for each domain as a set that contains the attribute names, class names and their relation names with associated weights. Here, weights are assigned as to express our confidence level in determining the domain the data refers to. Number of matches of the semantic lexicon is found for each domain with the text and the domain of the text is decided.

Step 3: Extraction of Ontology Attribute Values

As the next step, we should extract the values and map them into attributes concerning the domain we have identified. This will be achieved via pattern matching.

Step 4: Ontology Update

The extracted attribute-value pairs should be added to the Ontology. In order to do this, they are first converted into a RDF (Resource Description Framework) format and appended. We will be using Apache Jena for this purpose. [13]. Apache Jena is an open source Semantic Web framework which can be used to manipulate RDF graphs through an API [9]. Each graph will be represented as a "model".

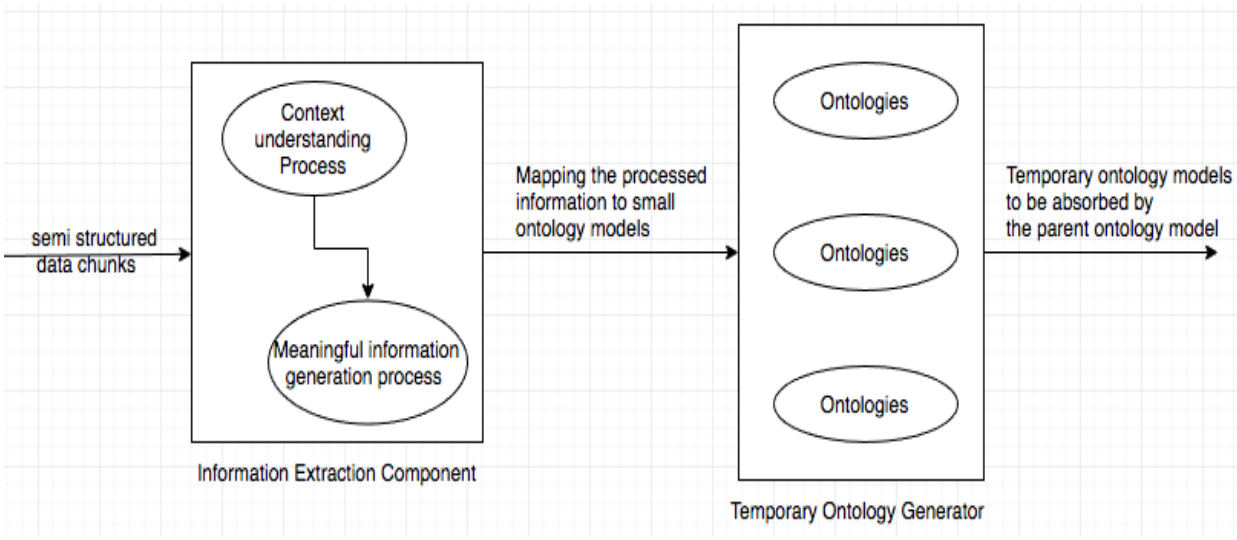

Figure 3.5: Information extraction

3.2.3. Ontology based knowledge mapping and integration

Ontology integration is a complicated process done either by hand, semiautomatic tools or full automatic tools. The framework proposed in this document aims to automate the ontology integration and mapping process in a generic way which will be applicable to multiple domain segments in a dynamic machine environment. There are many definitions and concepts of ontology expansion or integration. Definition given in [15] "The process of finding commonalities between two different ontologies A and B and deriving a new ontology C that facilitates interoperability between computer systems that are based on A and B ontologies. The new ontology C may replace A or B, or it may be used only as an intermediary between a system based on A and system based on B. Depending on the amount of change necessary to derive C from A and B, different levels of integration can be distinguished"

The concept of integration means anything ranging from integrations, merges, mapping, extending, approximation, unified views and as described in [17]. When an application is built using one or more ontologies, ontology integration is achieved through ontology matching and ontology alignment. Ontology matching concept is used to find contact between ontologies by finding the semantic similarities between the ontologies. Semantic similarity measures play a consequential role in text cognate area and application in ontology matching. [48]. There are three approaches for Ontology based data integrations. A single ontology is used as a global reference model in the system. This is the simplest approach as it can be simulated by other approaches. [20] The second approach is the multiple ontology approach in which contains multiple ontologies, each modeling an individual data source, are used in combination for integration. Though, this approach is more flexible than the single ontology approach, it requires creation of mappings between the multiple ontologies. Ontology mapping is a challenging issue and is focus of large number of research efforts in computer science [20]. The approach, hybrid approach involves the use of multiple ontologies that subscribe to a common, top-level vocabulary. The top-level vocabulary defines the basic terms of the domain. Thus, the hybrid approach makes it easier to use multiple ontologies for integration in presence of the common vocabulary. [20]

In the proposed platform, rather than having a static ontology with specific sets of predefined classes and attributes, a dynamic ontology is implemented which will map, expand and evolve as more knowledge is extracted from sections 3.2.1 and 3.2.2 as the system progresses. Single ontology approach will be used for ontology integration which in turn will enables the proposed framework to maintain a single global knowledge model which would undergo dynamic changes with time.

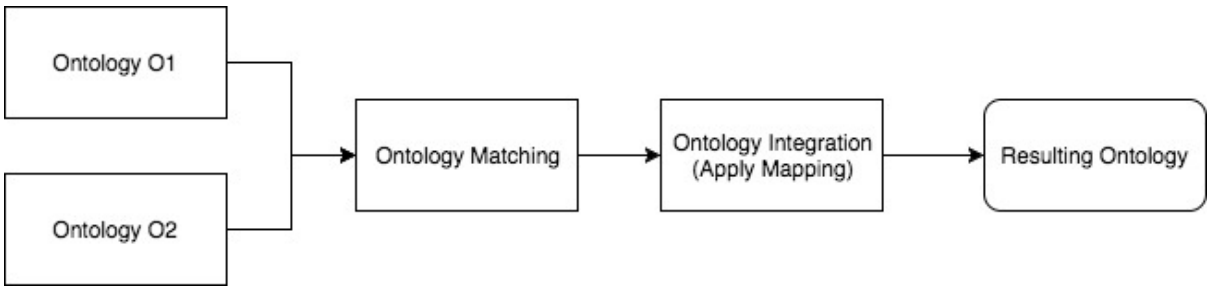

Figure 3.6: Ontology Integration Process

As shown in the above diagram, multiple ontologies created in section 3.2.2 would be merged together for a single knowledge model and then mapped into the main ontology model of the system. This approach will be far more efficient, as opposed to mapping each and every local ontology models received from section 3.2.2 individually to the global model. The ontology-matching component will explore the semantic nature of the concepts defined which will be utilized to measure the homogeneity between two or more ontologies. As shown in the figure below, there are multiple techniques when it comes to ontology mapping.

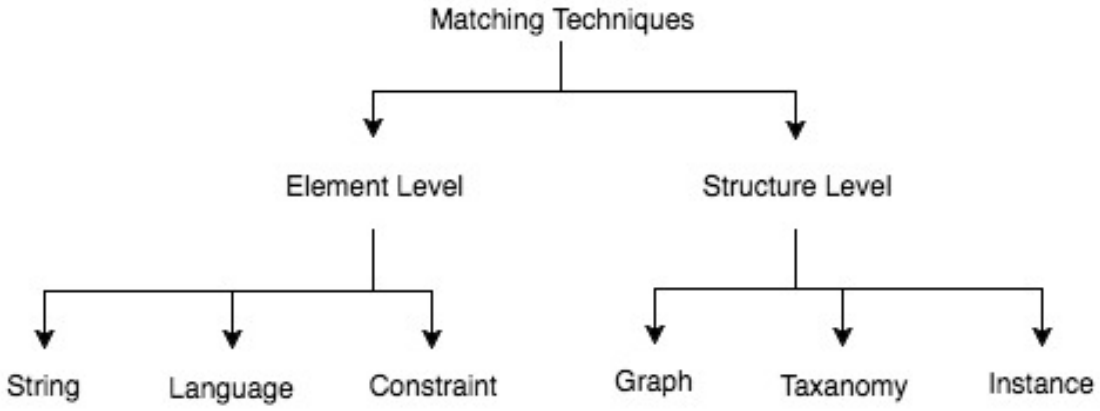

Figure 3.7: Ontology Matching Techniques

The proposed framework will explore element level techniques rather than structure level techniques as the system is built in a generic form, which would explore having dynamic structures. After ontology matching component is completed, ontology mapping or integration will take place exploring the heterogeneity of the given ontologies and dynamically expanding the primary ontology according the structure of the knowledge base.

Protégé is a free, open-source platform that provides a growing user community with a suite of tools to construct domain models and knowledge-based applications with ontologies. This will be used to configure and model the main knowledge base of the proposed system.

3.2.4. Knowledge Retrieval and Representation

After creating ontology, it is also important to retrieve the knowledge from the knowledge base and represent useful information to the user. There should be a proper knowledge retrieval model, which will take care the part of retrieving the relevant information according to the business model and the requirement. In our model in addition to retrieving the relevant information, if there is a situation where the required information is not included in the knowledge base, then there should be a sufficient way to acquire the relevant knowledge from specific sources and update or expand the existing knowledge based on newly learnt content. In order to acquire the relevant needed information there should be a proper notifying model, which is responsible to notify the information extraction model to extract the new information based on the query.

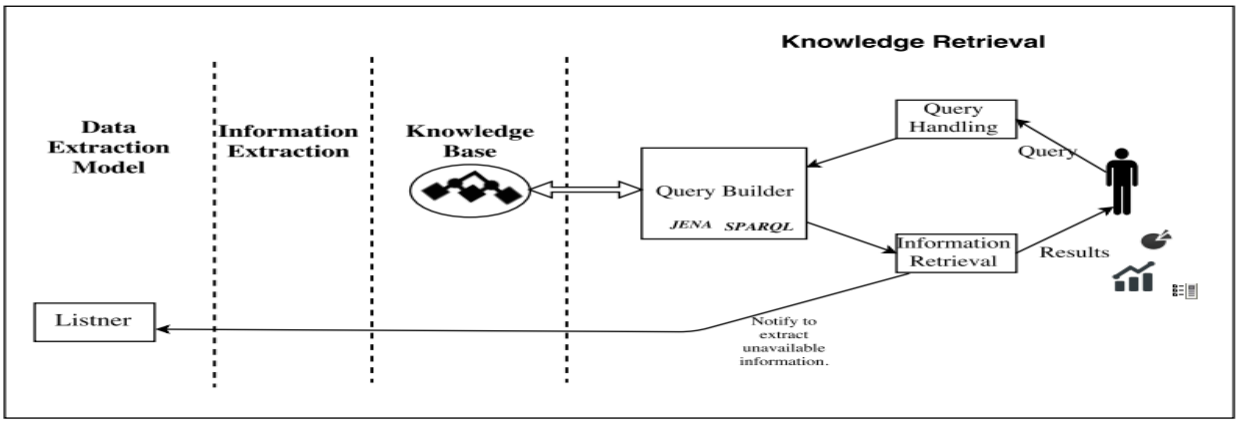

Figure 3.8: Structure of knowledge retrieval model

As shown in figure 3.8, there will be an interface exposed to the user where user can interact with the system and make queries according the business model. Then those queries will be passed to the query handling part in order to get the relevant information from the user's query in order to build the query that will be used to interact with the knowledge base and retrieve required information. After retrieving the information, that information will be represented to the user in accordance with the business model requirements. If the queried information is not available then, regarding that information a notification will be send to the data extraction model in order to extract the relevant information from sources. Because of this feature the knowledge base will be continuously updated and required information will be retrieved.

SPARQL is an RDF query language, that is, a semantic query language for database, able to retrieve and manipulate data stored in Resource Description Framework(RDF) format. SPARQL allows users to write queries against what can loosely be called "key-value" data or, more specifically, data that follows the RDF specification of the W3C. The entire database is thus a set of "subject-predicate-object" triples. [8]

Apache Jena is an open source Semantic Web framework for Java. It provides an API to extract data from and write to RDF graphs. The graphs are represented as an abstract "model". A model can be sourced with data from files, databases, URLs or a combination of these. A Model can also be queried through SPARQL. [9]

3.3. Gantt Chart

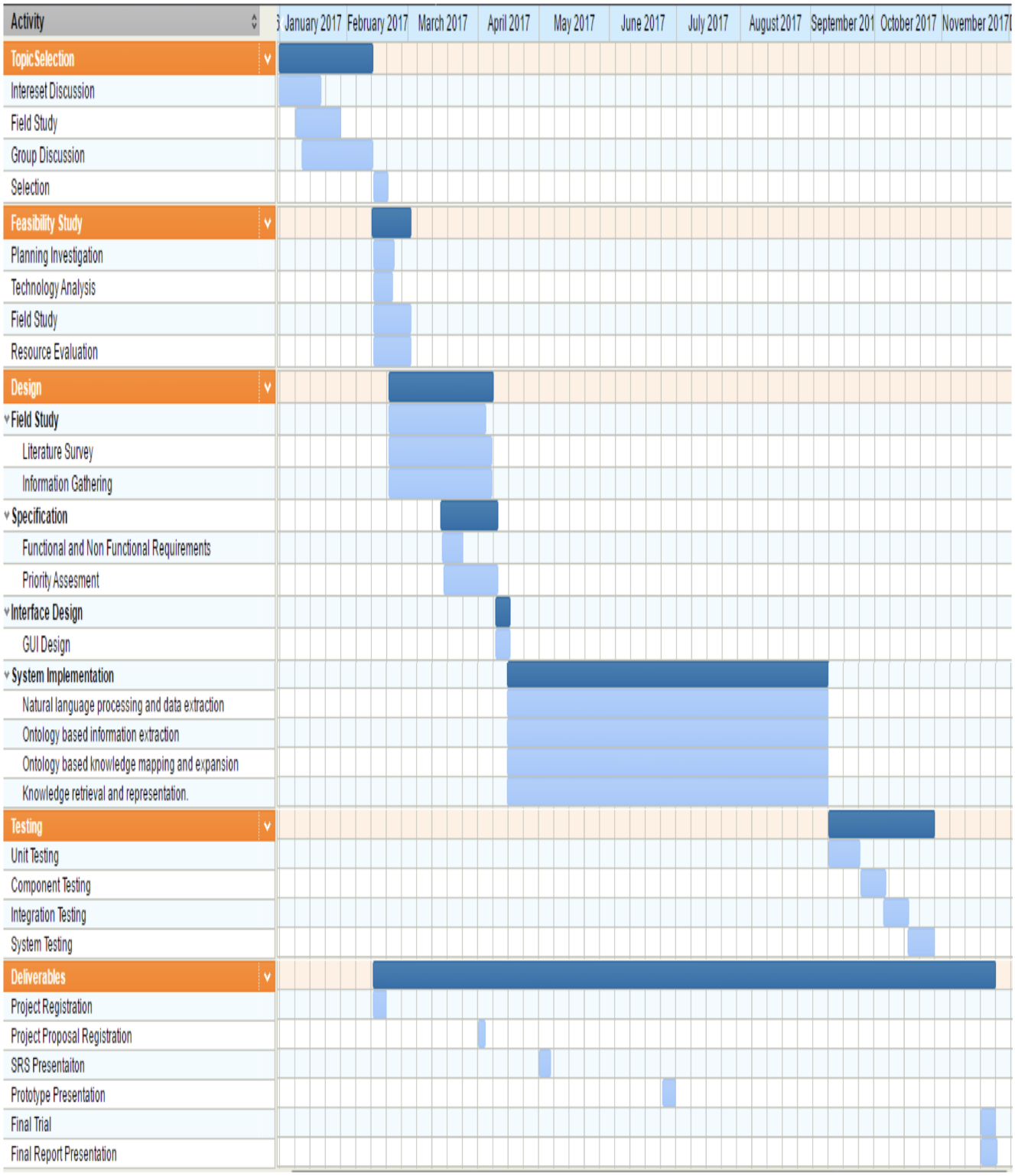

Figure 3.9: Gantt chart

4. PERSONAL & FACILITIES

In this section the workload assigned to each member is described. The project group has four members and the workload is assigned to all four members with equal proportion so they can work on the proposed system with equal effort and focus.

Table 2: Work breakdown chart

Member	**Component**	**Task**
H.H.N.C.Jayanandana IT14001826	• Heterogenic ontology mapping and integration.	• Requirement Gathering • Requirements Analysis & feasibility study • Ontology and Architectural Design • UI & UX • System Testing • System Deployment • Maintenance • Documentation
H.P.N.H.Herath IT13104504	• Natural Language processing and data extraction. (Knowledge Procurement)	• Requirement Gathering • Requirements Analysis & feasibility study • Ontology and Architectural Design • System Testing • System Deployment • Maintenance Documentation

H.M.S.Piyasundara IT14063442	• Ontology based information extraction. (Knowledge Understanding)	• Requirement Gathering • Requirements Analysis & feasibility study • Ontology and Architectural Design • System Testing • System Deployment • Maintenance Documentation
R.Rishanthakumar IT14087820	• Applying the acquired knowledge by targeting a relevant business model. (Knowledge Retrieval and Representation)	• Requirement Gathering • Requirements Analysis & feasibility study • Ontology and Architectural Design • UI & UX • System Testing • System Deployment • Maintenance Documentation

5. REFERENCES

[1] 'Issues in Anaphora Resolution', [On-line]. Available: http://nlp.stanford.edu/courses/cs224n/2003/fp/iqsayed/project_report.pdf [Accessed: 10-March-2017]

[2] 'A Guide to Creating Your First Ontology', [On-line]. Available: http://protege.stanford.edu/publications/ontology_development/ontology101.pdf [Accessed: 10-March-2017]

[3] Lijun Tang and Xu Chen, “Ontology-Based Semantic Retrieval for Education Management Systems” in Journal of Computing and Information Technology - CIT 23, 2015, 3, 255–267 (2015)

[4] Amol N. Jamgade and Shivkumar J. Karale, “Ontology Based Information Retrieval System for Academic Library” in IEEE Sponsored 2nd International Conference on Innovations in Information, Embedded and Communication systems (ICIIECS) 2015.

[5] Aliyu Rufai Yauri, Rabiah Abdul Kadir, Azreen Azman, Masrah Azrifah and Azmi Murad, “Ontology Semantic Approach to Extraction of knowledge from Holy Quran”, Faculty of Computer Science and Information Technology, Universiti Putra Malaysia IEEE Publications. (2012).

[6] Albert Comelli, Luca Agnello, Salvatore Vitabile, “An Ontology-Based Retrieval System for Mammographic Reports” in 20th IEEE Symposium on Computers and Communication (ISCC) (2015).

[7] ‘RDF Query Language’, [On-line]. Available: https://en.wikipedia.org/wiki/RDF_query_language [Accessed: 10-March-2017]

[8] ‘SPARQL’, [On-line]. Available: https://en.wikipedia.org/wiki/SPARQL [Accessed: 10-March-2017]

[9] ‘Jena (framework)’, [On-line]. Available: https://en.wikipedia.org/wiki/Jena_(framework) [Accessed: 10-March-2017]

[10] ‘Ontology’, [On-line]. Available: http://tomgruber.org/writing/ontology-definition-2007.htm [Accessed: 10-March-2017]

[11] ‘CoreNLP’, [On-line]. Available: http://stanfordnlp.github.io/CoreNLP/coref.html [Accessed: 10-March-2017]

[12] S. R. R. Raghu A, "Ontology guided information extraction from unstructured text," International Journal of Web & Semantic Technology, vol. 4, no. 1, p. p19, 2013.

[13] ‘Ontology’, [On-line]. Available: https://jena.apache.org/documentation/ontology/ [Accessed: 10-March-2017]

[14] Atika Mustafa, Ali Akbar, Ahmer Sultan, Knowledge Discovery using Text Mining: A Programmable Implementation on Information Extraction and Categorization, International Journal of Multimedia and Ubiquitous Engineering.

[15] ‘Top Level Categories’, [On-line]. Available: http://www.jfsowa.com/ontology/toplevel.htm [Accessed: 10-March-2017]

[16] Text Mining Application Programming by Manu Konchadi. Published by Charles River Media. ISBN: 1584504609

[17] M. Keet, “Aspects of Ontology Integration”, 2004.
[18] Automated Concept Extraction from Plain Text. Boris, GARAGe Michigan State University, East Lansing MI 48824.
[19] Handbook on Ontologies, edited by Steffen Staab, Rudi Studer.
[20] ‘Ontology based data integration’, [On-line]. Available: https://en.wikipedia.org/wiki/Ontology-based_data_integration [Accessed: 10-March-2017]
[21] ‘The OWL API’, [On-line]. Available: http://owlapi.sourceforge.net/ [Accessed: 10-March-2017]
[22] Craig A. Knoblock, Pedro Szekely, Jose Luis Ambite, Aman Goel, Shubham Gupta, Kristina Lerman , Maria Muslea , Mohsen Taheriyan , and Parag Mallick. Semi-Automatically Mapping Structured Sources into the Semantic Web (2012)
[23] ‘Steiner Tree Problem’, [On-line]. Available: https://en.wikipedia.org/wiki/Steiner_tree_problem [Accessed: 10-March-2017]
[24] N. Choi, I.-Y. Song, and H. Han. A survey on ontology mapping. ACM Sigmod Record, 35(3):34–41, 2006.
[25] An Approach to Ontology Integration for Ontology Reuse, Conference Paper · July 2016 [On-line]. Available: https://www.researchgate.net/publication/307546407 [Accessed: 10-March-2017]
[26] ‘Ontologies’, [On-line]. Available: https://www.obitko.com/tutorials/ontologies-semantic-web/ontologies.html [Accessed: 10-March-2017]
[27] ‘Towards the Development of a Framework for Socially Responsible Software by Analyzing Social Media Big Data on Cloud Through Ontological Engineering’, [On-line]. Available: http://www.sciencedirect.com/science/article/pii/S187705091500527X [Accessed: 10-March-2017]
[28] Wimalasuriya D.C., Dou D., Ontology-based information extraction: An introduction and a survey of current approaches. Journal of Information Science, 36, (2010), No. 3, 306
[29] Wimalasuriya D.C., Dou D., Components for Information Extraction: Ontology-Based Information Extractors and Generic Platforms, CIKM’10, Canada, (2010)
[30] Bontcheva K., Tablan V., Maynard D., Cunningham H., Evolving GATE to meet new challenges in language engineering, Natural Language Engineering, 10, (2004), 349- 373
[31] Ferrucci D, Lally A (2004) UIMA: An Architectural Approach to Unstructured Information Processing in the Corporate Research Environment. Natural Language Engineering, 10 (3- 4), 327–348.

[32] Drozdzynski W., Becker M., Krieger H.-U., Piskorski J., Schäfer U., Xu F., SProUT (Shallow Processing with Unification and Typed Feature Structures) (2002), [On-line]. Available: http://sprout.dfki.de [Accessed: 10-March-2017]
[33] Buitelaar P., Siegel M., Ontology-based Information Extraction with SOBA. In: Proceedings of the 5th International Conference on Language Resources and Evaluation, Italy, (2006)
[34] Maedche A., Staab S., The Text-To-Onto Ontology Learning Environment. In: Software Demonstration at the Eighth International Conference on Conceptual Structures, SpringerVerlag, Berlin, (2000)
[35] Yildiz B., Miksch S., OntoX - a method for ontology-driven information extraction. In: Proceedings of the 2007 International Conference on Computational Science and Its Applications, Springer, Berlin, (2007)
[36] T. Gruber, "Ontolingua: a translation approach to providing portable ontology specifications", Knowledge Acquisition, 5(2), pp. 199-220, 1993
[37] 'Development of an Ontology', [On-line]. Available: http://www.saitm.edu.lk/fac_of_eng/RSEA/SAITM_RSEA_2013/imagenesweb/14.pdf [Accessed: 10-March-2017]
[38] 'Ontology Based Information Extraction', [On-line]. Available: https://www.cs.uoregon.edu/Reports/ORAL-200903-Wimalasuriya.pdf [Accessed: 10-March-2017]
[39] Ling Hu, Jianxiong Wang, Geo-ontology Integration Based on Category Theory. In: 2010 International Conference on Computer Design And Applications (ICCDA 2010)/
[40] Yanhui Lv, Chong Xie. A Framework for Ontology Integration and Evaluation. In: 2010 Third International Conference on Intelligent Networks and Intelligent Systems
[41] 'The Stanford CoreNLP Toolkit', [On-line]. Available: https://nlp.stanford.edu/pubs/StanfordCoreNlp2014.pdf [Accessed: 10-March-2017]
[42] Rishabh Upadhyay, Akihiro Fuji. Semantic Knowledge Extraction from Research Documents In: Proceedings of the Federated Conference on Computer Science and Information Systems.
[43] Xincheng Liu, Hui Du, NianNian Liu. Research on high-speed railway ontology integration method Based on Semantic Relationships.
[44] Fisnik Dalipi, Florim Idrizi, Eip Rufati, Florin Asani. On Integration of Ontologies into E-learning Systems. In: 2014 Sixth International Conference on Computational Intelligence, Communication Systems and Networks.

[45] Behrang QasemiZadeh, "Towards Technology Structure Mining from Scientific Literature", 9th International Semantic Web Conference, ISWC 2010, Shanghai, China, November 7-11, 2010

[46] Mima, H., Ananiadou, S. & Matsushima, K. (2004) Design and Implementation of a Terminology-based literature mining and knowledge structuring system, in Proceedings of international workshop on Computational Terminology, CompuTerm, Coling, Geneva, Switzerland.

[47] 'Making the world's knowledge computable', [On-line]. Available: https://www.wolframalpha.com/about.html [Accessed: 10-March-2017]

[48] Computational Intelligence in Data Mining - Volume 3

[49] S.Kalaivani and K.Duraiswamy, "Comparison of Question Answering Systems Based on Ontology and Semantic Web in Different Environment"(2012).

[50] Gerasimos Tzoganis, Dimitrios Koutsomitropoulos and Theodore S. "Querying Ontologies: Retrieving Knowledge from Semantic Web Documents".

[51] Guowei Chen and Pengzhou Zhang. "Keywords Retrieval Based on Ontology Inference". (2012)

[52] Xiaohui Tao, Yuefeng Li, Ning Zhong , Richi Nayak "An Ontology-based Framework for Knowledge Retrieval" (2005).

[53] 'TopBraid Composer', [On-line]. Available: http://www.topquadrant.com/products/ [Accessed: 10-March-2017]

[54] Abdel-Bdel, Badeeh M. Salem, Marco Alfonse. "Ontology versus Semantic Networks for Medical Knowledge Representation" 12th WSEAS International Conference on COMPUTERS, Heraklion, Greece, July 23-25, 2008

[55] Steven J Miller. "Introduction to Ontology Concepts and Terminology" University of Wisconsin-Milaukee, Lisbon, Portugal September 2, 2013