TOWARDS KNOWLEDGE MODELING AND MANIPULATION TECHNOLOGIES: A SURVEY

BIMBA ANDREW THOMAS
Faculty of Computer Science and Information Technology, University of Malaya, 50603 Kuala Lumpur, Malaysia.
bimba@siswa.um.edu.my

NORISMA IDRIS
Faculty of Computer Science and Information Technology, University of Malaya, 50603 Kuala Lumpur, Malaysia.
norisma@um.edu.my

AHMED AL-HUNAIYYAN
Computer & Information Systems Department, Public Authority for Applied Education and Training, Kuwait.
hunaiyyan@hotmail.com

ROHANA BINTI MAHMUD
Faculty of Computer Science and Information Technology, University of Malaya, 50603 Kuala Lumpur, Malaysia.
rohanamahmud@um.edu.my

AHMED ABDELAZIZ
Faculty of Computer Science and Information Technology, University of Malaya, 50603 Kuala Lumpur, Malaysia.
ahmedaziz@siswa.um.edu.my

SULEMAN KHAN
Faculty of Computer Science and Information Technology, University of Malaya, 50603 Kuala Lumpur, Malaysia.
sulemankhan1984@yahoo.com

VICTOR CHANG
International Business School Suzhou, Xi'an Jiaotong Liverpool University, Suzhou, China.
ic.victor.chang@gmail.com

Abstract
A system which represents knowledge is normally referred to as a knowledge based system (KBS). This article focuses on surveying publications related to knowledge base modelling and manipulation technologies, between the years 2000-2015. A total of 185 articles excluding the subject descriptive articles which are mentioned in the introductory parts, were evaluated in this survey. The main aim of this study is to identify different knowledge base modelling and manipulation techniques based on 4 categories; 1) linguistic knowledge base; 2) expert knowledge base; 3) ontology and 4) cognitive knowledge base. This led to the proposition of 8 research questions, which focused on the different categories of knowledge base modelling technologies, their underlying theories, knowledge representation technique, knowledge acquisition technique, challenges, applications, development tools and development languages. A part of the findings from this survey is the high dependence of linguistic knowledge base, expert knowledge base and ontology on volatile expert knowledge. A promising technique for knowledge-based business management and other knowledge related applications is also discussed.

Keywords: knowledge acquisition; knowledge-based business; linguistic knowledge base; ontology; expert knowledge base; cognitive knowledge base.

1. Introduction
This article surveys several journal articles, conference papers, serials and books on the implementation and challenges of various knowledge modelling and manipulation technologies. It classifies these technologies according to their development theories and structure, resulting to four categories; the linguistic knowledge bases (Collin F Baker, 2014; Fellbaum, 1998; Speer & Havasi, 2012), expert knowledge bases (Driankov, Hellendoorn, & Reinfrank, 2013; Kerr-Wilson & Pedrycz, 2016; Kung & Su, 2007), ontology (Khan, 2009; Fensel, 2004; David Sánchez, 2010; Studer etal., 1998; Van Heijst et al., 1997) and most recently the cognitive knowledge base (Wang,
Human knowledge is categorized at the levels of data, information, knowledge and intelligence. These categories are the fundamental cognitive objects in the human brain and cognitive systems (Wang, 2015). The development of computers that display an intelligent behaviour has been the foundation of Artificial Intelligence (AI).

A system which represents knowledge is normally referred to as a knowledge based system (KBS). The most important component of any knowledge based system is the knowledge base. Based on the characteristics of knowledge, Dignum & vab de Riet (1991) defined a knowledge base as “a set of statements that describe the knowledge about the truths of the actual world plus a set of constraints that describe statements that must be true in all possible worlds and statements that ought to be true in all possible worlds” (pp. 4).

In the past, KBS development has been viewed as a transfer of human knowledge into the implemented knowledge base (Wielinga et al., 1992). This view was based on the theory that the required knowledge already exists and only needs to be collected and implemented. Usually, the required knowledge is acquired by interviewing an expert, and implemented in the form of production rules. However, this approach did not support a suitable representation of different knowledge types (Studer et al., 1998). The existence of different types of knowledge and the absence of satisfactory justifications of the rules makes the maintenance process difficult and time consuming. Thus, this approach was only feasible in the development of small scale prototypes, thereby initiating a paradigm shift from the transfer approach to the modelling approach (Ramirez & Valdes, 2012). The modelling approach is not intended on simulating the entire cognitive process of an expert, but to create a model which offer similar results in problem solving.

The main objective of this research is to survey existing knowledge base modelling and manipulation technologies, with the aim of identifying different knowledge representation, implementation and acquisition techniques. This survey could provide novice researchers with a tool to select appropriate knowledge representation and implementation techniques, while providing experts with a broader view to introduce novel techniques. Previous researchers have conducted surveys on knowledge based systems, for example, Liao (2003) did an analysis on knowledge management technologies and applications, Plant & Gamble, (2003) discussed the important research in knowledge-based system life cycles and development, while Sahin et al., (2012) reviewed the current approaches and applications of hybrid expert systems. Our current review follows similar methodologies as the previous researchers. However, we focus on a broader aspect of knowledge modelling and manipulation technologies. To the best of our knowledge, this is the first study which provides a broad view on the various categories of knowledge modelling and manipulation technologies, their implementation and challenges. We suggest 8 research questions as shown in table 1. These questions are primarily centred around knowledge base modelling and implementations.

<table>
<thead>
<tr>
<th>S/N</th>
<th>Research Question</th>
<th>Motivation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>What are the common approaches to knowledge modelling and manipulation?</td>
<td>Identify the various categories of knowledge modelling and manipulation technologies</td>
</tr>
<tr>
<td>2</td>
<td>What are the theories supporting each knowledge modelling approach?</td>
<td>Identify the theoretical basis for the implementation of each knowledge base modelling approach</td>
</tr>
<tr>
<td>3</td>
<td>How is knowledge represented in each modelling approach?</td>
<td>Identify the various ways knowledge is represented in each modelling approach.</td>
</tr>
<tr>
<td>4</td>
<td>Which knowledge acquisition technique is commonly applied in each knowledge base modelling approach?</td>
<td>Identify the common knowledge acquisition techniques used in each approach.</td>
</tr>
<tr>
<td>5</td>
<td>What are the common applications of each knowledge modelling approach?</td>
<td>Identify the types of applications for each knowledge modelling approach.</td>
</tr>
</tbody>
</table>
6  What are the implementation tools frequently used in each knowledge modelling approach?  Identify the various implementation tools used in each knowledge modelling approaches.

7  What are the common challenges faced in knowledge modelling?  Identify the challenges of knowledge modelling technologies

8  What are the programming language commonly used in each knowledge modelling approach?  Identify the various programming languages used in each knowledge modelling approaches.

The structure of the article is as follows: First, the outline of the survey process is discussed in section 2. In section 3, we provide detail review of the knowledge base modelling and manipulation technologies. We give a detailed analysis of the results in section 4. Section 5 discusses the need for knowledge based business management. The directions for the further is highlighted in section 6 and the limitations of the survey in section 7. Finally, we present our findings and conclusions in section 8.

2.0  Survey Process

The articles included in our survey were extracted from 4 main digital database of academic journal articles. These digital libraries include Scopus, Web of Science, IEEE Xplore and ACM. These libraries were selected based on their impact evaluation and wide coverage of peer-reviewed journals in multiple academic disciplines. The inclusion of relevant articles in this survey was decided in 5 steps as shown in figure 1, using EndNote Desktop application. First, the digital libraries were searched based on keywords corresponding to the categories of knowledge base modelling technologies and the articles were restricted to publication between the year 2000 and 2015. Then, duplicated articles were removed. Articles published in both journals and conference proceedings, with the same title published by same authors, in the same year are regarded as duplicates. In this situation the article published in journal is selected, neglecting the one published in conference proceeding. Thirdly, the full texts of the remaining articles were searched online. This resulted to a reduced number of articles as seen in the third oval shape in figure 1. The fourth step involved relevance sorting, which required searching the full text based on some combination of keywords relevant to each category. Lastly, the remaining articles were read to identify information related to the different categories: knowledge base representation, implementation, acquisition technique, application, implementation tools and limitations.
This process resulted to 185 articles, excluding the subject descriptive articles which are mentioned in the introductory parts. The analysed articles consisted of journal articles, conference proceedings, books and serials. They were examined based on the publication years, availability and relevance to the research domain.

3.0 Classification of Knowledge Base Modelling Techniques

Based on the fundamental theories of knowledge base modelling and manipulation, knowledge base technology can be categorized into four groups: 1) the linguistic knowledge bases (Baker, 2014; Fellbaum, 1998; Speer & Havasi, 2012); 2) expert knowledge bases (Driankov et al., 2013; Kerr-Wilson & Pedrycz, 2016; Kung & Su, 2007); 3) ontology (Fensel, 2004; Sánchez, 2010; Studer et al., 1998; Van Heijst et al., 1997) and most recently 4) the cognitive knowledge base (Wang, 2015b). The various categories and types of knowledge base modelling
approaches are shown in figure 2. In this section, we briefly explore each of these categories, with the aim of identifying and comparing the techniques and methodology used in representing and manipulating knowledge.

3.1 Linguistic Knowledge Base

![Diagram of Knowledge Models]

Figure 2 Classification of Knowledge Modelling and Manipulation Technologies

Language is a means through which human express thoughts, therefore to model human cognitive processes a lexicon is required. However, linguistic theories which attempt to model human grammar rely on data which is often not well documented (Fellbaum, 1998). Originally, linguists have divided the human knowledge of grammar into phonology, morphology, syntax, semantics and the lexicon. These aspects of grammar which are directed by clear rules allow computational linguistics to design parsers, that identify constituents of phrases and sentences, assigning a phrase structure to them. But, due to the large size of lexicon, there is a huge challenge in capturing the structure and properties of lexicon in a manner which reflects the actual human lexical knowledge (Fellbaum, 1998). Thus, making this an ideal research area for computational language modelling. Typical linguistic knowledge bases are lexical databases such as FrameNet, WordNet and ConceptNet (Wang, 2014). In the preceding sections, we will briefly describe this three lexical databases.

3.1.1 FrameNet

FrameNet (FN) was developed in an effort to build a lexicon of English that is comprehensible by both human and machine, using the theory of frame semantics and backed by means of an annotated corpus of lexical items (Baker, 2014; Fillmore et al., 2003; Lakhff & Laskri, 2015). The knowledge base structure of FN is defined as a relation between frames at various levels of generality. Frames are “script-like conceptual structure that describes a particular type of situation, object, or event along with its participants and props” (Ruppenhofer et al., 2006). Knowledge in FN is represented as frames and an annotated corpus (Baker, 2012; Wandmacher et al., 2011).
Frames represent generalizations over groups of words which illustrate equivalent situations, similar set of roles and related syntactic behaviour (Martínez-Santiago et al., 2015; O’Hara & Wiebe, 2009). In the theory of frame semantics, the roles or common situations which describes a frame are called frame elements (Das et al., 2014; Pimentel et al., 2012). The association between a word form and its meaning is referred to as a lexical unit (Zhang et al., 2015). Word form is a lemma and a given part of speech, while the meaning is represented by a semantic frame and a definition (Baker, 2014). The procedure for creating lexical entries in FrameNet is well explained by (Boas, 2005).

Knowledge acquisition in FN is a manual process of grouping words with semantic overlap according to some explicit criteria to form frames and sentence annotation based on multiple annotation layers (Kettenrová et al., 2008). The annotation process involves FN lexicographers to declare each word in a sentence as a target, then select a frame related to the target, get a set of annotation layers and appropriate frame element tags and the annotate the relevant constituents (Ruppenhofer et al., 2006).

FN version 1.3, is a freely available lexical database which contains a wealth of semantic knowledge of about 1161 Semantic Frames, covering more than 12,600 lexical units, documented with nearly 200,000 manual annotations (Lakhiff & Laskri, 2015). FN is developed based on the eXtensible Markup Language (XML) (Martínez-Santiago et al., 2015). Raw sentences are annotated and converted to XML, using a Java GUI client, and then inserted into tables in a MySQL database (Baker, 2012). FN corpus can be assessed using the Natural Language Toolkit (NLTK) in Python (Garrette & Klein, 2009). The data in FN has been used to develop automatic semantic role labellers (Croce & Basili, 2011; Erk & Pado, 2006; Giannone, 2013; Gildea & Jurafsky, 2002; Padó & Lapata, 2009) and frame-semantic parsers (FSP) (Das et al., 2014). Other end-user applications for FN includes Question answering (QA) (Ofoghi et al., 2008a, 2008b; Sinha, 2008) and information extraction (IE) (Mohit & Narayanan, 2003; Scaiano & Inkpen, 2009). Discourse interpretation (DI) can be supported using existing knowledge in FN (Irmer, 2013).

However, since FN is built on edited text, applying an automatic semantic role labeller (SRL) trained on its current data could give poor results for text such as Twitter feeds (Baker, 2014). Currently, there is no unified annotation model across FNs (Gružitis & Dannells, 2015). Compared to WordNet, FrameNet has a rather limited lexicon (Baker, 2012). Text coherence cannot be handled by FN because it works strictly within the sentence, this makes it impossible to link arguments across sentences (Ruppenhofer et al., 2010). Another limitation of FN is its inability to represent negation and conditional sentences. A major challenge faced now by FN is finding a more cost-effective way to expand, while preserving the richness and accuracy of its annotation. Another challenge is finding ways to combine the frame semantic work currently being carried out around the world, to generate a truly multi-lingual FN (Gružitis & Dannells, 2015).

3.1.2 WordNet

WordNet (WN) is another large scale electronic lexical database conceived based on human semantic organization, where words and their meanings are related to one another via semantic and lexical similarities (Fellbaum et al., 2006). WN is a “proposal for a more effective combination of traditional lexicographic information and modern high-speed computation” (Miller et al., 1990). Unlike FN, this model was constructed based on word association norms, where given a lexical stimulus such as a noun, verb or adjective, responses often remain in specific semantic relations such as synonyms (similar), antonyms (opposite), hyponyms (subordinate) / hypernyms (superior) and meronyms (part) / holonyms (whole) (Fadaee et al., 2013; Fellbaum et al., 2006; Lindén & Niemi, 2014; Marrafa, 2002). For example, “bird” can elicit its hypernym “animal” or its meronym “feather”. Such human associated responses revile the way mental lexicon is organized. Knowledge in WordNet is represented as a semantic network formed by synsets (cognitive synonyms) (Lee, Huh, & McNiel, 2008). A set of synonyms that serve as identifying definitions of lexicalized concepts are referred to as synsets (Maziarz et al., 2013; Montejo-Raez et al., 2014). Similar to FN, the lexical units in WN tend to be derived from their connection with other members of the same synset as well as from their lexical-semantic association with other synsets (Marrafa, 2002; Saif et al., 2015; Wei et al., 2015). The entries in WN are strictly separated based on syntactic categories of Noun, Verb, Adjective and Adverbs only (Dragoni et al., 2015; Uddin et al., 2013).
In WN, lexicon is constructed essentially through automated acquisition or hand-crafted. The manual process is slow, expensive and cumbersome, however it allows one to create entries that will be useful to certain applications. The Lexico-syntactic Pattern Extraction (LSPE) method is one of the automated approaches used by lexicographers in developing domain-dependent knowledge base (Fellbaum, 1998). The technique does not require a knowledge base or complex interpretation modules to suggest new WordNet relations. But, the results provided by LSPE are not comprehensive.

Similar to FN, WN is a freely available lexical database. WordNet 3.0 comprise of 155,287 unique strings with 117,659 synsets and 206,941 word-sense pairs. WN has been converted to Resource Definition Framework (RDF) and Web Ontology Language (OWL), in order to improve interoperability between software applications that use WN (Van Assem et al., 2006).

Using a W3C standard language SPARQL Protocol and RDF Query Language (SPARQL), WN data in RDF/OWL format can be queried (Brengu et al., 2015; Prud’Hommeaux & Seaborne, 2008). Typical queries can pose on WN RDF/OWL once it is loaded in tools such as SWI Prolog’s Semantic Web Library (Wielemaker et al., 2008) or Sesame (Broekstra et al., 2003). Some WN-based word similarity algorithms have been implemented in a Perl package called WordNet::Similarity (Pedersen et al., 2004), Python package called NLTK (Bird, 2006) and Align, Disambiguate and Walk (ADW) in Java (Filevhr & Navigli, 2015).

There are various projects which have utilized WN, for example, 1) BabelNet, a multilingual lexical database in which WordNet is combined with Wikipedia to perform multilingual word sense disambiguation and semantic relatedness (Navigli, 2013); 2) EuroWordNet consists of wordnets from several European languages resulting in a multi-lingual database used for multi-lingual information retrieval (Vossen, 1997) and 3) MultiWordNet, an Italian WordNet which is strictly aligned to English WordNet, while retaining its ability to represent true lexical peculiarity between languages (Planta et al., 2002).

WN is commonly used to improve text clustering (TC) (Wei et al., 2015). In Word Sense Disambiguation (WSD), WN combined with other techniques is been applied as a knowledge base approach to solve ambiguous meaning of words in Information Retrieval (IR) (Kilinc & Alpkocak, 2011; Sachdeva et al., 2014; Tyar & Win, 2015). Nevertheless, WN poses some limitations due to its connotation inconsistencies and exceedingly fine granularity of senses (Erjavec & Fišer, 2006; Prakash et al., 2007).

3.1.3 ConceptNet

ConceptNet (CN) is a large-scale common-sense knowledge base which describes human knowledge and how it is expressed as a semantic graph (Agarwal et al., 2015; Liu & Singh, 2004; Speer & Havasi, 2012). While lexical databases such as WN is developed for lexical categorization and word-similarity determination and FN is optimized to describe a particular type of situation and event, CN is used for making practical context-based inferences (Schuller & Knaup, 2011). The main goal of developing CN is to capture common-sense knowledge that describe the real world (Hsu et al., 2006; Wu & Tsai, 2014). While FN represent knowledge in form of frames and WN in form of a semantic network of synsets, CN represents knowledge in form of graph, where the nodes include higher-order compound concepts composed of action verbs (Bicocchi et al., 2011). It extends WN’s repository of semantic relations from triplet of synonyms to twenty semantic relations (Liu & Singh, 2004). Unlike WN, the knowledge in CN is more informal, defeasible and practically valued for example, HasEffect (‘eating chilli’, ‘stomach ache’) (Hsu et al., 2008; Noah et al., 2010). CN conveys concepts, that are words and phrases which could be extracted from natural language text, along with statements of the ways in which these concepts relate to each other (López et al., 2011; Majewski & Szymański, 2008; Speer & Havasi, 2012).

Knowledge in CN is acquired automatically from the English sentences in the Open Mind Common Sense (OMCS) corpus. It then applies an additional set of procedures to optimize the connectivity of the semantic network (Liu & Singh, 2004). These processes include; extraction, normalization and relaxation. As at 2012, CN contains 12.5 million edges, representing about 8.7 million statements connecting 3.9 million concepts (Speer & Havasi, 2012). Access to CN data is provided by a web API using JavaScript Object Notation (JSON) textual data format (Baydinet al., 2015). Similar to WN, CN was encoded in RDF/OWL language to make it accessible and integrated with Semantic Web applications (Grassi & Piazza, 2011).
CN tool-kit has been used in developing several applications, such as: 1) LifeNet, a probabilistic graphical model of everyday first-person human experience (Singh & Williams, 2003); 2) Emotus Ponens, a textual affect-sensing system that leverages common-sense to classify text using six basic emotion categories (Liu et al., 2003) and 3) GOOSE, a goal oriented search engine for novice users (Liu et al., 2002). CN has been used to improve feature selection in machine learning approaches (Agarwal et al., 2015). The process of selecting meaningful navigational paths in knowledge discovery (KD) can be improved by using CN as a common-sense KB (Freitas et al., 2015). CN is used as a framework to transmit sentiment values for building sentiment dictionaries (Tsai et al., 2013). Latent Dirichlet Allocation (LDA) has been combined with CN to measure similarities between terms in twitter event classification (Vo et al., 2015). However, the coverage of CN’s knowledge is shallow as compared to human knowledge.

3.2 Expert Knowledge Base

Expert knowledge base consists of useful domain knowledge for problem solving. This knowledge is represented as a set of rules in the expert knowledge base. Rules are viewed as a relatively easy technique to represent knowledge. Any rule consists of two parts: the IF part called the antecedent and the THEN part called the consequent. Practically, a rule can have multiple antecedents joined by conjunction (AND) or disjunction (OR) or a combination of both. The antecedent of a rule incorporates two parts; a linguistic object and its value. The object and its value are linked by an operator. The operator identifies the linguistic object and assigns the value. Rules can represent relations, recommendations, directives, strategies and heuristics (Negnevitsky, 2005). Expert knowledge bases are categorized as logical rule-based (LRS) and fuzzy rule-based (FRS) systems. Details of each rule-based knowledge base are discussed in the following sections.

3.2.1 Logical Rule-Based System

Knowledge is represented as binary logic in a logical rule-based system (Park, 2003; Rattanasawad et al., 2013; Selva et al., 2014). In this system, if the antecedent is true, then the consequent is also true (Negnevitsky, 2005). Mostly, in a LRS, knowledge is acquired manually from domain experts through interviews, where experts communicate their knowledge using questionnaires (Connaghan et al., 2013; Dimitroula et al., 2001; Selva & Crawley, 2012). However, knowledge in forms of rules can be acquired automatically, such as RUBRIC which constructs rules from thesauri (Minkoo et al., 2000) and semi-automated like KnowRob, which automatically acquires information from different knowledge sources with the aid of human for correcting mistakes and aligning imported knowledge sources (Tenorth & Beetz, 2013).

Logical rule-based systems have been applied in a number of areas such as, hole maching (Park, 2003), fluorescein angiography (Dimitroula et al., 2001), bridge crane control (Capella et al., 2003), estimating infant immunization coverage (Kowalski & Burton, 2012), Lungs abnormality detection (Lahouar et al., 2012), text extraction system (Ramakrishnan et al., 2012), decision support (Selva & Crawley, 2012) and power grid dispatch and control (H. Zhang & Zou, 2012). Prolog is the programming language used mostly for knowledge representation in logical rule-based systems, as seen in WUENIC (Kowalski & Burton, 2012), sports coaching (Connaghan et al., 2013), KnowRob (Tenorth & Beetz, 2013) and online poker agent (Teofilo et al., 2014). Other development tools used are CLIPS, for the implementation of FUNAGES (Dimitroula et al., 2001).

However, the process of knowledge acquisition is similar to ontologies where it is highly dependent on volatile domain expert knowledge, which is sometime complete and incomplete, precise and imprecise or certain and uncertain (Liu et al., 2013). In addition, capturing an expert’s knowledge entirely is difficult as most of their knowledge is hidden in their skills (Selva & Crawley, 2012).

3.2.2 Fuzzy Rule-Based System

Fuzzy sets are used in representing knowledge in a fuzzy rule-based system (Cordón, 2011). Unlike rule-based systems, in fuzzy rule-based system, if the rule antecedent is true, then the consequent could be partially true. It provides an efficient knowledge representation method for systems that involve continues variables (Banerjee et al., 2001). Fuzzy logic is developed as a method to express and apply human knowledge in a form that reflects an
Knowledge acquisition in fuzzy systems can either be from human experts or data-driven (Bombardier et al., 2007; Zajaczkowski & Verma, 2012; Zhang & Mahfouf, 2011). The human expert approach lends itself to a manual design of fuzzy models based on existing knowledge retrieved from an expert through interviews and open questions (Fay, 2000). On the other hand, the data-driven method identify the structure and parameters of fuzzy models from datasets using various methods such as feature space mapping and C4.5 classification tree (Duch et al., 2001), fuzzy rule learning algorithm (Fernandez et al., 2010), fuzzy scheduler (Prado, et al., 2010), swarm intelligence approach (R. Prado et al., 2010; Zhang & Mahfouf, 2011), differential evolution learning (Prado et al., 2011) and genetic algorithm (Nurnberger, 2004; Ozyer et al., 2007).

Fuzzy rules have been used to represent knowledge in several domains such as railway operation control systems (Fay, 2000), time series prediction (Paul & Kumar, 2002) Hepatitis medical diagnosis, classification system (FRBCSs) (Fernandez et al., 2010), bioinformatics (Zhou et al., 2012), grid computing (Prado et al., 2010), economic analysis of RFID orders (Ustundag et al., 2010), prediction of mechanical properties of alloy steel (Zhang & Mahfouf, 2011), bioinformatics (Nurnberger, 2004; Zhou et al., 2012), eco-system management (Adriaenssens et al., 2004), finance (Boyacioglu & Avci, 2010), to study HIV evolution in infected individuals (Jafelice et al., 2009) and robotics (Bai et al., 2005).

Fuzzy production rules in expert systems can be represented using Fuzzy Petri net (Fay, 2000; Liu et al., 2013). Java programming language has also been used to develop a software tool which supports the design and computation of recurrent fuzzy systems (Nurnberger, 2004). Natural language information analysis method and object role modelling have been applied in creating symbolic fuzzy models representing customer knowledge in a defect recognition system (Bombardier et al., 2007). Definition of membership functions and values captured from human experts, can also be represented using MATLAB fuzzy logic toolbox simulator (Celikyilmaz & Turksen, 2008; Guimaraes & Lapa, 2007; Jafelice et al., 2009). One of the challenges of fuzz rule-based system development is the lengthy process of rule elicitation and every procedure is specific to each application (Banerjee et al., 2001). (Pandey & Mishra, 2009).

### 3.3 Ontology

Ontology, a branch of metaphysics dealing with the nature of being, is the third approach to knowledge modelling and manipulation presented in this survey. While linguistics knowledge models represent knowledge by lexical and semantic relationships and expert knowledge models by logical and fuzzy rules, ontologies represent knowledge as a taxonomy of concepts with their attributes, values and relations (Studer et al., 1998). The main aim of ontologies is to provide a platform which facilitates the sharing and reuse of knowledge between groups in a computational form (Apisakmontri, Nantajeewarawat, Ikeda, & Buranarach, 2016). As Studer et al. (1998) defines ontology as “a formal, explicit specification of a shared conceptualization.” “Conceptualization” refers to an identification of relevant concepts of an abstract model of some phenomenon in the world. While “Explicit” means the concepts identified and the constrains on their use are explicitly defined (Kimble, et al., 2016). Ontologies are made of at least three elements: 1) classes (domain concepts); 2) relations (distinctive sorts of parallel relationship between concepts); and 3) instances (real world phenomenon) (David Sánchez, 2010). Most researchers agree on classifying ontologies based on different generality levels and conceptualization (Borst, 1997; Maalel et al., 2012; Ramos, 2015; Studer et al., 1998; Van Heijst et al., 1997; Zhang et al., 2011). This resulted to the four categories of ontologies discussed in this survey, which are application ontology, domain ontology, generic ontology and representation ontology.

#### 3.3.1 Application Ontologies

Application ontologies, capture all the necessary characterizations required to model knowledge for a particular domain. However, application ontologies have method and task specific extensions (Jin Tan et al., 2005; Savonnet et al., 2015; Simperl, 2009; Van Heijst et al., 1997). An application ontology describes the relationship between concepts based on specific tasks (Liu et al., 2010). This can be seen from its application in courseware management for teachers (Jin et al., 2005), indexing (Assali et al., 2007), OntoNeuroBase which is used as a
common semantic reference for aligning heterogeneous data in neuroimaging (Temal et al., 2008), smart homes (Chen et al., 2009), comparing UML elements type in class diagrams (Robles et al., 2012) and intrusion detection system (Meneses et al., 2015).

Knowledge acquisition in application ontologies involve eliciting reasoning mechanisms used by an expert to do a task or solve a particular problem. Semi-automated techniques have been used to elicit knowledge from documents, technical reports and expert interviews, combined with an inferential modelling technique which supports the knowledge engineer in identifying different knowledge types (Chan, 2004b). However, Legaz-Garcia et al. (2012) developed an application ontology automatically, reusing different ontologies to describe genomic sequence and genetic disorders (Legaz-Garcia et al., 2012).

In application ontology development, Protégé is the most commonly used tool for representing concepts and their relations (Assali et al., 2007; Jin Tan et al., 2005; G. Liu et al., 2010). Other tools used are Goal-Oriented Application Ontology Development Technique “GAODT” (Santos et al., 2013), Knowledge Modelling System (KMS) (Chan, 2004a) and Model Driven Architecture (MDA) (Santoso et al., 2011). The common programming language used in application ontology development is the Web Ontology Language (OWL) (Durbha et al. 2009; Ye et al., 2011). However, Resource Definition Framework (RDF) and OWL have been combined to convert cause-effect relationships of a concept while developing application ontologies (Ebrahimipour & Yacout, 2015). The major challenge of application ontologies is its reusability (Van Heijst et al., 1997). There is an essential need for developing tools which support application ontology reuse processes and the need for a task approach to application ontology reuse (Simperl, 2009).

### 3.3.2 Domain Ontologies

*Domain ontologies* represents conceptualizations that are valid only in a specific domain for example education (Labidi & Sergio, 2000; Mesarić & Dukić, 2007), health care (Batet et al., 2011; Castilho et al., 2008), manufacturing (El-Diraby & Osman, 2011) and environmental studies (Ceccaroni et al., 2004; Mercantini & Faucher, 2012). The main aim of domain ontologies is to eradicate the conceptual and terminological misperception amongst members of a specific community, who need to share documents and information (Navigli & Velardi, 2004).

Several techniques have been proposed for knowledge acquisition in domain ontology construction. The *CommonKADS* methodology was used to formalize the knowledge acquisition process, where the main concepts and characteristics of the domain where extracted manually from experts (teachers) using a questionnaire (Labidi & Sergio, 2000). However, *OntoLearn* system automatically extracts domain terminologies from texts and filters them using Natural Language Processing (NLP) and statistical techniques (Navigli & Velardi, 2004). Soo-Yeon et al. (2004), proposed a semi-automatic method for constructing domain ontology. First a corpus was developed from web documents in the pharmaceutical domain, then concepts were extracted using NLP, afterwards hierarchical structures representing relationships between terminologies were formed. A fully automated approach for domain ontology construction was taken by (Chen et al., 2008). Using Single Value Decomposition (SVD), Adaptive Resonance Theory Network (ART) and Boolean operation, meaningful terms are extracted from domain-related web pages to automatically construct domain ontology. Another automated approach was taken by (Sánchez & Moreno, 2008). Through unsupervised methodology, non-taxonomically related concepts and relationships were extracted from a web corpus.

Domain ontology has been applied in student modelling and knowledge domain construction in collaborative learning system (Labidi & Sergio, 2000). Other applications include information retrieval (Assali et al., 2007), geological hazards (Liu et al., 2010) and data mining (Mansingh et al., 2011). More specifically, domain dependant ontologies are used to support information exchange processes in advertising networks (Al-Safadi & Al-Abdullatif, 2010). They have also been applied in representing antimicrobial treatment rules and providing feedback while prescribing antibiotic (Bright et al., 2012). Scenario Object Model (SOM), a domain ontology, is used for finding, retrieving, auditing, sharing and analysing recorded information from hazard evaluation procedures (Wu et al., 2013).
Various tools for the construction of domain ontology have been introduced. SymOntos support the construction of ontologies using the OPAL (Object, Process, Actor modelling language) methodology (Missikoff et al., 2003). Unified Modelling Language (UML) has also been used for the design of domain ontology models of multi-agent systems (Wei et al., 2003). Property and Class characteristics from Text for Ontology Enrichment (PACTOLE), is used to develop domain ontology by deriving concept lattice from a structured context, forming a binary table which represents a set of objects with their properties (Bendaoud et al., 2008). GRAONTO offers a graph-based approach for automatically constructing domain ontologies (Hou et al., 2011). Zhang & Ma (2014) proposed a tool that transforms eXtensible Markup Language (XML) to ontologies, making it easier to utilize XML data on the web. Similar to application ontology, Protégé can also be used to develop domain ontology (Semenova & Kureychik, 2015). The most language used in domain ontology development is the OWL (Durba & King, 2004; Ochoa et al., 2013; She & Wang, 2009; Soo-Yeon et al., 2004; Su et al., 2005). However, defining ontologies for serving multiple users and application is hard to create and manage (Mena et al., 2000). In complex domains, the task of constructing ontologies are usually controversial due to different expert points of view (Missikoff et al., 2003). Subsequently, there is a lack of standardized methodology for building domain ontology (Huang & Gao, 2009).

3.3.3 Generic Ontologies

Generic ontologies sometimes referred to as top-level ontologies, are usually valid over various domains (Santoso et al., 2011; Xing et al., 2009; Ye et al., 2011). They define concepts like state, event, process, action, component etc. For example, generic ontologies are used in multi-agent systems, permitting agents to interoperate and cooperate through a common ontology (Su et al., 2003). PRIMA is another generic ontology which contains a set of concepts and relations describing risk and its incorporation in various work processes (Makki et al., 2008). Also generic ontologies are used for creating models for policy-based regulations, which incorporate different domains like Food and Drug Administration (FDA), financial regulation, contracts and individuals conducting business (El Kharbili & Stolarski, 2009). The Open Group Architecture Framework (TOGAF) generic ontology contains knowledge about an enterprise based on four domains of business, data, application and technology (Czarnecki & Orlowski, 2010). CoBra ontology aids the pervasive computing environment where different intelligent agents with specific task exchange knowledge and cooperate to reach their goals (Chen et al., 2003).

Knowledge acquisition in generic ontologies is difficult to fully automate, because it involves information scattered in different systems and services. Thus, it is mostly semi-automatic, involving a human expert combined with statistical, syntactic and semantic techniques (Makki et al., 2008). Kohli et al. (2012) manually extracted information using literature reviews, interviews and surveys from experts to develop the generic slum ontology (GSO), which comprises of concept that refer to the morphology of the environment. EmotionsOnto, a generic ontology used to detect and describe emotional expressions, manually acquires information from users (Gil et al., 2015).

OntoEdit and Protégé were used to develop OBELIX, a generic service ontology which represents a formalization of concepts on service management and marketing (Akkermans et al., 2004). Similar to application ontology, Protégé is the most common tool used in the development of generic ontology (Czarnecki & Orlowski, 2010; Darmoul et al., 2011; Dutra et al., 2010; El-Subaihi et al., 2013; Marwat et al., 2014). The OWL language is also used in generic ontology development, because of its adoptability, decidability and efficiency (Dutra et al., 2010; Marwat et al., 2014). However, generic ontology development is still encountering some challenges in knowledge acquisition and lack of sufficiently validated and generalized development methodology (Marwat et al., 2014).

3.3.4 Representation Ontologies

Representational ontologies are not restricted to a particular domain. This kind of ontologies present entities without declaring exactly what ought to be represented. These type of ontologies capture knowledge in a problem-solving independent way (Studer et al., 1998). The major application of ontologies is in the semantic web, which provides a platform for automatically processing data and information using extensible metadata (Feilmaier & Wölff, 2016). The Mediation Bridge Ontology (MBO) is a representation ontology used to store the correspondences between matching ontologies, enabling interoperability in the semantic web (Khan et al., 2015). Another example of a representation ontology is proposed by (Benslimane et al., 2005) for semantic
interoperability of information sources. It uses semantic mediation to allow definition of sharing system, interrogation, uniform and transparent access to diverse information sources.

Knowledge acquisition in ontologies are mostly done manually. Traditionally, the creation of the content of ontologies such as concepts and relations are done by knowledge engineers or domain experts (Zhou, 2007). However, representation ontologies usually define a set of concept representations of generic and domain ontologies (Maalel et al., 2012). Thus, supporting an automatic knowledge acquisition process. To construct representation ontologies, the description logic and stamping technique are used (Benslimane et al., 2005). The stamping technique allows multiple representation of the same concept, allowing a phenomenon to be used in one or more context. Similar to the other types of ontologies mentioned, the representation ontology lacks standard development methodology.

3.4 Cognitive Knowledge Base

Conventional technologies for knowledge base modelling and manipulation such as linguistic knowledge base, expert knowledge base and ontology are man-made rather than machine built. “The absence of thorough and sufficient operations on acquired knowledge, inflexible for learnt knowledge synergy, and weak transformability among different knowledge bases gave rise to a novel approach, the cognitive knowledge base” (CKB) (Wang, 2015a). Based on the previous studies in cognitive science and neurophysiology (Hampton, 1997; Leone et al., 2006), “the foundations of human knowledge in the long-term memory can be represented by an object-attribute-relation model based on the synaptic structure of human memory, which represents the hierarchical and dynamic neural clusters of knowledge retained in memory as well as the logical model of knowledge bases” (Wang, 2014).

The cognitive knowledge base is a structure that manipulates knowledge as a dynamic concept network like the human knowledge processing (Wang, 2008; Wang et al., 2011). In CKB a concept is a cognitive unit which identifies and models real-world concrete entities and a perceived-world (abstract entity) (Pojman, 2001; Wang, 2015a). The basic unit of knowledge in a CKB is a formal concept represented as an OAR model according to concept algebra (Valipour & Yingxu, 2015; Wang, 2007; Wang, 2015a). While complex knowledge such as a theme are represented as multiple associate concepts, which forms a partial dynamic concept network (DCN).

The CKB structure consist of the logical model, physical model, linguistic knowledge base and knowledge manipulation engine. The logical model of knowledge bases shared by humans and cognitive systems is known as the object-attribute-relation model (Wang, 2014). The logical structure is modelled as a hierarchical network of concepts and themes. The logical knowledge base represents knowledge as Cartesian products of formal concepts. The physical knowledge base implements the memory structures of knowledge as a DCN. The linguistic knowledge base comprises of the initial words as modelled in WordNet and the representation of these words and their relation in form of the OAR model.

The CKB is managed by the knowledge manipulation engine (KME), which handles knowledge acquisition, fusions and retrievals such as concept match, relational concept identification and knowledge differentiation (Wang, 2014). Similar to WordNet, the relations among formal concepts are classified as synonyms, antonyms, hyponyms and hypernyms. The knowledge retrieval from CKB is supported by the concept and theme visualization processes.

The cognitive processes of concept memorization and knowledge fusion, similar to human and cognitive systems learning processes, are used for knowledge acquisition in CKB (Wilson & Keil, 2001). Concept memorization involves acquiring concepts based on the formal concept model (Wang, 2014). While, knowledge fusion mimics the mechanism of human learning and knowledge memorization similar to buffered knowledge in the short-term memory before moved to the long-term memory (Wang, 2014).

Web knowledge discovery engine is developed based on concept algebra (CA), simulating the cognitive knowledge representation in the brain (Hu & Wang, 2007). Cognitive knowledge base has also been used to develop AutoLearner, a machine learning system and a cognitive simulator. It visualizes the interactions between internal knowledge representation as well as the thinking and learning processes (Hu & Wang, 2007b). A text knowledge representation method based on CA is used to automatically extract concepts from text and establish
relations between concepts (text association rule mining) (Ye et al., 2013). To automatically implement information restructuring of web documents, knowledge in informal documents are converted to cognitive knowledge represented by CA and concept networks (Hu & Wang, 2007a). In order to improve the semantic properties of knowledge represented in oil companies, an oilfield enterprise knowledge management system based on concept algebra was proposed (Duan et al., 2013).

The Visual Simulator of Concept Algebra (VSCA) tool developed in the MATLAB environment provides a means to explicitly simulate and visualize mathematical concepts (Lin & Wang, 2014). Concept networks can be visualized with concept graphs using an open source library, Java Universal Network/Graph (JUNG) (Hu & Wang, 2007). A formal knowledge representation system (FKRS) which supports autonomous concept generation based on CA was implemented in Java (Tian et al., 2011). Another tool used for visual knowledge representation of concepts, which enables autonomous learning of concepts by machines was developed using java (Tian & Wang, 2007).

4.0 Results and Evaluation

A total of 185 articles which included 3 books, 22 serials, 50 conference proceedings and 110 journal articles were evaluated in this survey. The publication years of these articles are from 2000 to 2015. We present our results in the following subsections based on the research questions presented in table 1.

4.1 Evaluation of Knowledge Base Modelling Approaches

In this survey, we identified the common approaches to knowledge base modelling and manipulation. figure 3, represents the publication years versus related articles on the types of knowledge base modelling categories used in this survey. The aim is to identify the categories of knowledge modelling and manipulation technologies and detect possible trends.

![Figure 3 Number of knowledge base modelling papers per year in survey](image)

Based on our survey, from the year 2000 to 2015, linguistic knowledge base and ontology are more often used for representing knowledge, as compared to expert and cognitive knowledge bases. Cognitive knowledge base is relatively new, and has the least number of publications from the year 2000 to 2015 in comparison with other
knowledge base technologies. Currently, there are more implementations of linguistic knowledge bases. This could be as a result of the increase in mobile gadgets, resulting in the need for more text analysis and information retrieval.

4.2 Comparison of Knowledge Base Modelling Approaches

Based on the knowledge base modelling approaches identified, a comparison is made according to their underlying theories, knowledge representation technique, knowledge acquisition technique and challenges. A detail of this comparison is presented in Table 2. The main objective is to identify and compare the various ways in which knowledge is acquired and represented by each knowledge base modelling approach. The limitations of each approach are highlighted, and possible solutions are discussed.

<table>
<thead>
<tr>
<th>Knowledge Base Technology</th>
<th>Theory</th>
<th>Structure</th>
<th>Knowledge Representation</th>
<th>Knowledge Acquisition</th>
<th>Limitations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Linguistic Knowledge Base</td>
<td>Frame semantics, common sense knowledge and word association norms</td>
<td>Frames, lexical semantic associations between synsets and graph</td>
<td>Frame elements, semantic network and semantic graph</td>
<td>Mostly manual, with some Automated methods such as LSPE and acquiring English sentences from the OMCS corpus</td>
<td>1. Dependence on volatile expert knowledge. 2. Difficult and expensive to build and expand, while preserving the richness of its annotations. 3. CN has shallow knowledge coverage compared to human knowledge. 4. FN cannot handle text coherence 5. WN has exceedingly fine granularity of senses. 6. WN has connotation inconsistencies. 7. FN cannot link arguments across sentences.</td>
</tr>
<tr>
<td>Expert Knowledge Base</td>
<td>Binary Logic and fuzzy set</td>
<td>Linguistic object, value and operator</td>
<td>IF THEN rules</td>
<td>Commonly a manual process where knowledge is acquired from domain experts who communicate</td>
<td>1. Difficulty in capturing expert knowledge. 2. Brittleness of rules.</td>
</tr>
<tr>
<td>Knowledge Base Technology</td>
<td>Theory</td>
<td>Structure</td>
<td>Knowledge Representation</td>
<td>Knowledge Acquisition</td>
<td>Limitations</td>
</tr>
<tr>
<td>---------------------------</td>
<td>--------</td>
<td>-----------</td>
<td>--------------------------</td>
<td>-----------------------</td>
<td>-------------</td>
</tr>
</tbody>
</table>
| **Ontology**              | Nature of being | Classes, relations and instances | Taxonomy of concepts with their attributes, values and relations | Frequently acquired manually from knowledge engineers or domain experts. Could also be data-driven by extracting data from web documents, dictionaries and database schemas. | 1. Difficulty in capturing expert knowledge  
2. Lack of sufficiently validated and generalized development methodology |
| **Cognitive Knowledge Base** | Concept as the fundamental unit of human knowledge | Consist of the logical model, physical model, linguistic knowledge base and knowledge manipulation engine | Object-attribute-relation (OAR) model based on concept algebra | Fully automated processes of concept memorization and knowledge fusion, similar to human and cognitive systems learning processes | Fewer development tools and implementations in comparison with other knowledge representation technologies |

Based on the structure of the various categories compared, the cognitive knowledge base structure consists of a combination of the structure of the other three categories. It has a logical model, a linguistic knowledge base and an object-attribute-relation which is similar to attribute, values and relations of concepts in an ontology. Knowledge acquisition in a cognitive knowledge base is fully automated, unlike ontologies, expert knowledge base and linguistic knowledge base. For this reason, the cognitive knowledge base can be more suitable for developing applications where manual knowledge elicitation is cumbersome and costly. Unlike the cognitive knowledge base, linguistic knowledge base, expert knowledge base and ontology depend highly on volatile expert knowledge.

### 4.3 Evaluation of Knowledge Base Application Categories

We examined 72 articles related to various applications of knowledge base models. As a result, we identified 14 different application categories which are Natural Language Processing (NLP), Question Answering (QA), Information Extraction (IE) / Information Retrieval (IR), Classification, Knowledge Discovery (KD), Engineering, Health Care, Education, Finance, Environment, Business, Machine Learning, Robotics and Forecasting. In figure 4 the knowledge base applications versus the knowledge base modelling categories are presented. The main aim is to identify suitable applications of each knowledge base modelling approach.
Evidently, the linguistic knowledge base is mostly suitable for natural language processing, information retrieval and extraction. Expert knowledge base are mostly applied in engineering and health care domains. Based on our survey, ontologies are the only modelling approach used in education and environmental fields. While the cognitive knowledge base is used for machine learning.

4.4 Evaluation of Knowledge Base Implementation Tools

In this survey, various knowledge base implementation tools were identified from 32 different articles. The 19 tools identified across the 4 categories of knowledgebase modelling techniques are represented in figure 5. The main purpose is to identify the commonly used development tool by each modelling technique.
Protégé is the most common tool used in ontology development. It is popularly used because of its extensible plug and play development environment, which supports rapid prototyping and application development. Subsequently, it can be exported in a variety of formats such as, RDF, RDFS, OWL and XML schema. While MATLAB is mostly used for expert knowledge base implementation. MATLAB provides a natural way to express rules, its built-in graphics makes it easy to visualize and understand data.

4.5 Evaluation on Programming Language

Several programming languages are used in developing knowledge base models. We identified 9 different languages from 26 related articles in this survey. Figure 6 represents the development languages based on the knowledge base modelling categories. The goal is to identify the various programming languages used by each modelling approach.
A wider range of programming languages are used in linguistic knowledge base development in comparison with the other knowledge base technologies. This reflects the complexity in capturing human lexical knowledge. Most ontologies are developed using RDF/OWL, because it facilitates better machine interpretability of information content.

5.0 Knowledge-Based Business

Currently, organizing knowledge as well as knowledge manipulation activities and tools are of great interest among businesses and scientist. Similar to business process management, knowledge management methods and tools are necessary for optimization, automation, formalization and management of knowledge activities within organisations (Chang, 2015). The next growth in economic business is envisaged to come from knowledge-based businesses. Similar to other applications of knowledge base modelling and manipulation technologies, businesses need to know what the customer requires and need in advance. In order, to accomplish this requirements, businesses have to continually monitor news, economic activities, conduct surveys, acquire feedback from their customers, monitor social media to understand the needs of their customers (Yaqoob et al., 2016). With the introduction of big data, many business enterprises have devised ways and techniques to offer products that meet customer expectations (Chang et al., 2016). All this information has to be organised to provide the necessary knowledge to move the business forward.

Existing business modelling methods are concerned with modelling knowledge manipulation activities and its connection to business processes, not considering knowledge management and its relationship with business management processes. Based on the 4 categories of knowledge modelling and manipulation technologies considered, expert systems and ontologies are mostly applied to business management processes. Generic ontologies provide the means to define state, event, process, action and components involved in business management. For example, TOGAF contains knowledge about business management processes of an enterprise. Similarly, the expert knowledge base has been applied economic analysis product orders (Ustundag et al., 2010). However, all these techniques are faced with crucial challenges of knowledge acquisition. The brittleness of rule and the difficulty in maintaining a large rule-based system makes this technique less effective for capturing the knowledge required to provide customers with the required products. Although ontologies prove to be useful in business management, challenges are encountered in knowledge acquisition. The process of building ontologies
is quite cumbersome. The solution to these limitations is through the application of an automated knowledge acquisition process, which is dynamic and adaptable to customer’s changing needs.

6.0 Future Directions

Cognitive knowledge base, a recent knowledge base modelling and manipulation technique which models data based on the fundamental unit of human knowledge, provides a promising approach to knowledge-based business management and other knowledge related applications. Its advantage can be seen from its structure which involves a logical model, linguistic knowledge base and an object-attribute relationship similar to ontologies. Acquiring knowledge using this technique is fully automated, thereby eliminating this cumbersome process, which is common to other techniques. Since the cognitive knowledge base acquires knowledge in the same way as the human, it provides a means to better emulate the customer needs. Its dynamic nature makes it ideal for accommodating the changing needs of customers.

7.0 Survey Limitations

A literature review focusing on a broader aspect of knowledge modelling and manipulation technologies is a tough task based on the requirement for extensive background knowledge. Even though limited in background knowledge, this article presents a brief survey of knowledge base modelling and manipulation technologies from 2000-2015. Conforming to this duration has led to the exclusion of well-known application of knowledge base technology such as MYCIN (Buchanan, 1984) a rule-based expert system for diagnosis of infectious blood diseases. And the KADs (Wielinga, 1992) a common knowledge acquisition technique. Although, we reviewed 185 articles, we could only have access to subscribed journals, thus possibly omitting relevant articles from unsubscribed journals. In order to reduce complexity, we focused on only 4 digital libraries, hence neglecting databases such as google scholar and SAGE. Non-English publications are also not included in this survey. We suggest that further research should be done to provide a detailed review of each category of knowledge base modelling and manipulation technology presented in this survey.

8.0 Conclusion

This survey focused on publications related to knowledge base modelling and manipulation technologies, between year 2000 and 2015. The main aim of the research is to identify different knowledge base modelling and manipulation techniques based on 4 categories; linguistic knowledge base, expert knowledge base, ontology and cognitive knowledge base. This led to the proposition of 8 research questions. The questions focused on the different categories of knowledge base modelling technologies, their underlying theories, knowledge representation technique, knowledge acquisition technique, challenges, applications, development tools and development languages. Based on the results obtained, the linguistic knowledge base, expert knowledge base and ontology depend mostly on volatile expert knowledge. We observed that the cognitive knowledge base least number of publications from year 2000 to 2015 in comparison with other knowledge base technologies. Linguistic knowledge base technologies are obviously well applied in NLP, IE and IR as seen from the results in figure 4. Expert knowledge base and Ontology cover a wider range of applications as compared to linguistic and cognitive knowledge bases. Protégé is the most common tool used in ontology development, while MATLAB is mostly used for expert knowledge base implementation. RDF/OWL language is frequently used in ontology development. A wider range of programming languages are used in linguistic knowledge base development in comparison with the other knowledge base technologies. Our survey reveals the various underlying theories, knowledge representation technique, knowledge acquisition technique, challenges, applications, development tools and development languages of a broad category of knowledge base technologies. This could provide business owners with techniques to relate knowledge management and business management processes. It also provides novice researchers with a tool to select appropriate knowledge representation and implementation techniques, while providing experts with a broader view for introducing novel techniques.

Funding: This work was supported by the University of Malaya Research Grant [RP040B-15AET, 2015]
References


(including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) (pp. 169-179).


