Penerbit Akademia Baru

Journal of Advanced Research in Computing and Applications

Journal homepage: www.akademiabaru.com/arca.html
ISSN: 2462-1927

Semi-automated Ontology based Question Answering System

Open Access

Khairul Nurmazianna Ismail[1,*]

[1] Faculty of Computer Science, Universiti Teknologi MARA Melaka, Malaysia

ABSTRACT

Question answering system enable users to retrieve exact answer for questions submit using natural language. The demand of this system increases since it able to deliver precise answer instead of list of links. This study proposes ontology-based question answering system. Research consist explanation of question answering system architecture. This question answering system used semi-automatic ontology development (Ontology Learning) approach to develop its ontology.

Keywords:
Question answering system; knowledge base; ontology; online learning

1. Introduction

In the era of WWW, people need to search and get information fast online or offline which make people rely on Question Answering System (QAS). One of the common and traditionally use QAS is Frequently Ask Question site which contains common and straight forward answer. Currently, QAS have emerge as powerful platform using various techniques such as information retrieval, Knowledge Base, Natural Language Processing and Hybrid Based which enable user to retrieve exact answer for questions posed in natural language using either pre-structured database or a collection of natural language documents [1,4]. The demand of QAS increases day by day since it delivers short, precise and question-specific answer [10]. QAS with using knowledge base paradigm are better in Restricted-domain QA system since it ability to focus [12]. Most Restricted-domain is develop from scratch using manual ontology development. Manual ontology development is time consuming and costly. Therefore, this paper propose Semi-Automated Ontology Based Question Answering System. By using the semi-automated ontology development approach, it can contribute to shorten the ontology development time and the searching time [8].

2. Related Works

The importance of ontology in categorizing and structuring domain knowledge is exploited in QAS. Most ontology based QAS categorize the specific domain knowledge into ontology structures and then, a list of questions and answers is composed based on the created ontology. This can be

* *Corresponding author.*
E-mail address: Khairul Nurmazianna Ismail (khairul_nur@melaka.uitm.edu.my)

seen through several QAS developed and researched in previous studies using open domain ontology such as AQUA, QASYO, Pythia and NLQA. In AQUA project, developed by Vargas-Vera and Motta based on the combination of natural language processing, ontology, logic and information retrieval techniques. The system has been tested to answer questions about academic people and organizations. Here, ontology is used to formulate the natural language query in the ontological structures [14]. QASYO which use YAGO ontology as it background knowledge which develop from WebNet dataset [9]. Meanwhile, Pythia contains two parts of ontology mainly ontology-specific and ontology-independent part. It is able to parse constructed complex natural language questions and then can subsequently translate into formal queries with respect to a grammar that has been composed in the ontology. Natural Language Question Answering System (NLQA) use domain ontology which populated dynamically for each document in the collections [13]. However, these QAS use manual development approach for its ontology development. Manual ontology development is costly and time consuming compared to semi-automatic and automatic ontology development [8]. Therefore, this research proposes a semi-automatic ontology which may reduce the cost and speed up the ontology development process.

3. Architecture of Semi-Automated Ontology Based QAS

A semi-automated ontology-based QAS architecture is proposed as the main procedure for designing QAS to retrieve the answer using specific domain-based natural language query composed in ontology. The QAS architecture is shown in Figure 1.

In this part, the QAS architecture component is fully described. It is define by the three following main components which are the question parsing, the ontology building and the ontology matching. These three components are the mechanisms to retrieve relevant answer. Each component of the architecture is fully explained in Figure 1.

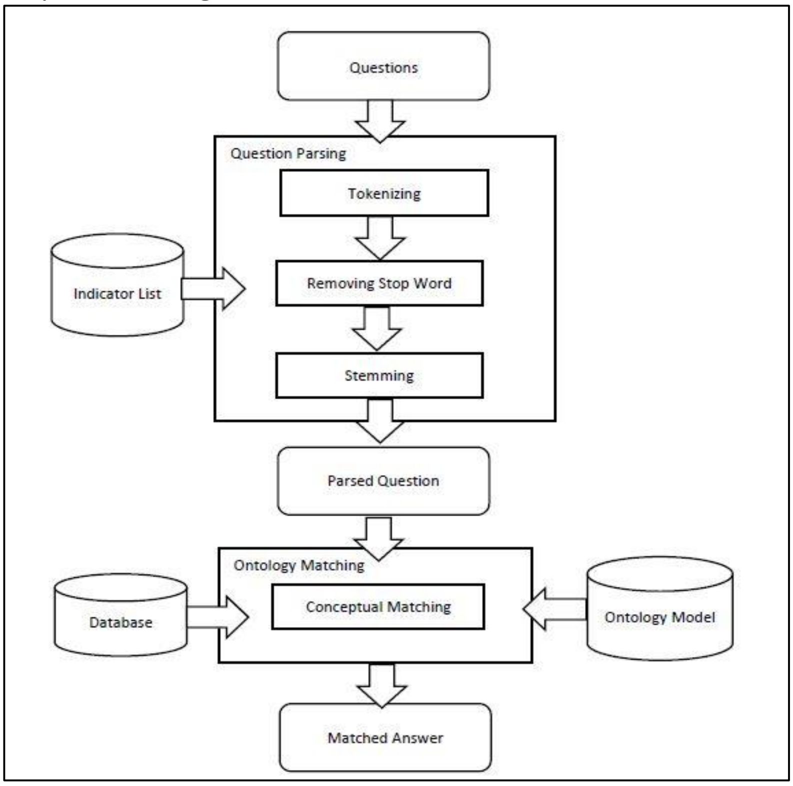

Fig. 1. The Architecture of Semi-Automated Ontology Based QAS

Question Parsing

The first component of this proposed system is the question parsing, in which the natural language questions is solve into logical syntactic form. Factoid question is use in this research which cover what, when, why, where and how. The component has three stages process which are tokenizing, removing stop word and stemming. An indicator list that consists of stop words such as 'the, a, an' is provided for this component to help parsing process. The stop words are deemed irrelevant for searching purposes because they occur frequently in the language. String tokenizing and Porter Stemmer algorithms are used in this component. The reason this component is compulsory is to decrease the high dimensionality problem of processing natural language questions. With this component, any irrelevant words are removed. Parsed question that has been stored as a set of tokens is the outcome result for this component.

By using string tokenizing algorithm, it will tokenize the input question by splitting up the sentence into words and removing whitespace between each words. For the next process, a set of tokenized words is used as input which is removing the stop word. At this phase, the tokenized word are compared with the stop words list. If any tokenized word that is similar with the stop words list is removed from the tokenized words.

Finally, the last process in the question parsing component is stemming. The Porter Stemming algorithm as suggested by Hooper and Paice in this phase [5]. Inside the algorithm lies a suffix rule. A suffix is a letter or group of letters that is added at the end of a word, making a new word. The tokenized word from the previous process has been extracted. If it matches to a suffix rule, then the word is tested with the conditions within the rule. The suffix will be removed once the condition is passed and accepted. The word is passed for testing with other suffix rules for the next step.

Ontology Building using Ontology Learning

This phase has been developed in our previous research. The resulted ontology is use as the knowledge base in currently develop QAS. The ontology development consists five phases. The first phase is to review the ontology development process. The purpose is to identify the suitable method that can involve human intervention in constructing the ontology. Secondly, review the ontology learning tools which will help speed up the process of ontology development. The first and second phase has been discussed in this research paper [7]. Third phase of this research is to extract important concepts to construct the ontology. Then, the graphical representation of the ontology will be presented using ontology editor to ease domain expert to evaluate them. Finally, the ontology evaluation process is in place whereby any modification from domain expert will take note and discuss whether it relevant to the domain ontology and this research (Ismail et al., 2018). Figure 2 shows the framework of this research in modeling the instructional material using ontology. Based on Figure 1, there are five phases involves in this research. However, the aim of this paper is to review ontology development process and ontology learning tools for the purpose of constructing ontology as an instructional material to be used in class.

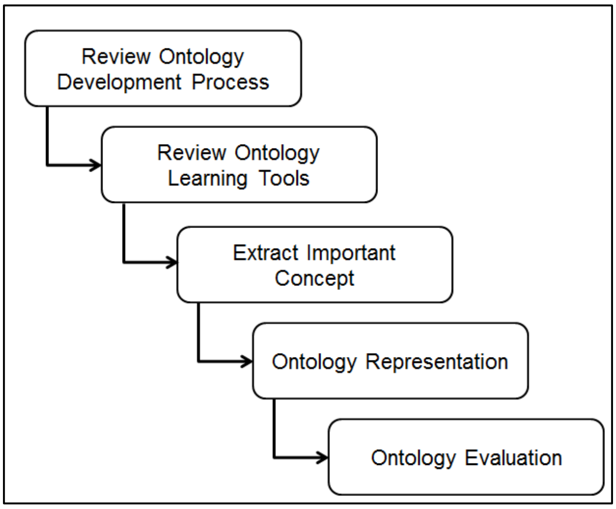

Fig. 2. Ontology development Phases
Source : (Ismail *et al*., [6], Ismail *et al*., [7])

Ontology Matching

The third component of the architecture is ontology that matches the detection of matched responses based on parsed queries and ontology model through the matching process. TF-IDF is a short for term frequency–inverse document frequency, is a numerical statistic that is planned to show how important a word is to a document in a collection or corpus [11]. The value of tf–idf is increases proportionally to the how much times a number of times a word visible in the document and is balanced by the number of documents in the corpus that contain the word, which helps to adjust for the fact that some words appear more frequently in general. Tf–idf is one of the most widely used in term-weighting schemes today. 83% of text-based recommender systems in digital libraries use tf–idf [2]. The parsed questions which in the form of word tokenized will be match with concept of domain ontology. The value of tf-idf for each token and concept is calculated. The concept that has the highest tf-idf value will notified within the domain ontology. Relation *hasDescription* of the concept will be extracted and represented as the answer of the question.

4. Conclusion

This paper described the phases involved development of semi-automated ontology within a Question Answering System (QAS). This QAS consist three phases namely question parsing, ontology building and ontology matching. Question parsing focus on converting natural language query into logical syntactic form which the produce parsed question. Then, the parse question will be the input to the second phase and will be match with domain ontology. The domain ontology has been developed and discuss in previous work. The semi-automated ontology development approach has been used to develop the domain ontology. Meanwhile statistical approach specifically TF-IDF will be used for ontology matching process.

References

[1] Ansari, Ahlam, Moonish Maknojia, and Altamash Shaikh. "Intelligent question answering system based on artificial neural network." In *2016 IEEE International Conference on Engineering and Technology (ICETECH)*, pp. 758-763. IEEE, 2016.

[2] Beel, Joeran, Bela Gipp, Stefan Langer, and Corinna Breitinger. "paper recommender systems: a literature

survey." *International Journal on Digital Libraries* 17, no. 4 (2016): 305-338.

[3] Saad, Saidah, Naomie Salim, Hakim Zainal, and Zurina Muda. "A process for building domain ontology: An experience in developing Solat ontology." In *Proceedings of the 2011 International Conference on Electrical Engineering and Informatics*, pp. 1-5. IEEE, 2011.

[4] Chali, Yllias, Sadid A. Hasan, and Shafiq R. Joty. "Improving graph-based random walks for complex question answering using syntactic, shallow semantic and extended string subsequence kernels." *Information Processing & Management* 47, no. 6 (2011): 843-855.

[5] Hooper, Rob, and Chris Paice. "The Lancaster stemming algorithm." *University of Lancester* (2005).

[6] Ismail, Khairul Nurmazianna, Fadzlin Ahmadon, and Fadilah Ezlina Shahbudin. "Instructional material development using ontology learning." *Social and Management Research Journal* 15, no. 2 (2018): 36-46.

[7] Ismail, Khairul Nurmazianna, Fadilah Ezlina Shahbudin, and Fadzlin Ahmadon. "Modeling instructional material using ontology." In *2017 International Conference on Research and Innovation in Information Systems (ICRIIS)*, pp. 1-5. IEEE, 2017.

[8] Maedche, Alexander, and Steffen Staab. "Ontology learning for the semantic web." *IEEE Intelligent systems* 16, no. 2 (2001): 72-79.

[9] Moussa, Abdullah M., and Rehab F. Abdel-Kader. "Qasyo: A question answering system for yago ontology." *International Journal of Database Theory and Application* 4, no. 2 (2011): 99-112.

[10] Pudaruth, Sameerchand, Kajal Boodhoo, and Lushika Goolbudun. "An intelligent question answering system for ict." In *2016 International Conference on Electrical, Electronics, and Optimization Techniques (ICEEOT)*, pp. 2895-2899. IEEE, 2016.

[11] Rajaraman, Anand, and Jeffrey David Ullman. *Mining of massive datasets*. Cambridge University Press, 2011.

[12] Soares, Marco Antonio Calijorne, and Fernando Silva Parreiras. "A literature review on question answering techniques, paradigms and systems." *Journal of King Saud University-Computer and Information Sciences* (2018).

[13] Unger, Christina, and Philipp Cimiano. "Pythia: Compositional meaning construction for ontology-based question answering on the semantic web." In *International conference on application of natural language to information systems*, pp. 153-160. Springer, Berlin, Heidelberg, 2011.

[14] Vargas-Vera, Maria, and Enrico Motta. "AQUA–ontology-based question answering system." In *Mexican International Conference on Artificial Intelligence*, pp. 468-477. Springer, Berlin, Heidelberg, 2004.