10th International Science, Social science, Engineering and Energy Conference
20th-23rd November, 2019, Rajamangala University of Technology Isan Sakon Nakhon Campus, Thailand

I-SEEC 2019

http://i-seec2019.rmuti.ac.th

A Thai Ontology-Based Framework for Diabetes Question and Answers

Ketsara Phetkrachang[1, a] and Nichnan Kittiphattanabowon[2, b]

[1]Faculty of Engineering, Rajamangala University of Technology Srivijaya

[2]School of Informaties, Walailak University

[a]ketsara.p@rmutsv.ac.th, [b]knichcha@wu.ac.th

Abstract

This paper proposed a Thai ontology-based question answering system for the caring of diabetic patients. Diabetes is a serious health problem in Thailand and in order to prevent complications due to diabetes, routine health checkups are needed. In addition, diabetes patients should consult doctors for diagnosis and treatment of medical conditions. Not only the number of diabetic patients is growing, the demographic of new patients is also expanding to Thai population causing major public health concerns in Thailand. To alleviate the workload of the healthcare staff as well as to improve patient care, a self-care system will be a possible solution. Using such a system, patients will be able to get quicker and informative answers prior to seeking medical consultations. One of the useful and essential resources is the Practice Guidelines for Diabetes (CPGs) from The Clinical Practice Guidelines (CPGs) in which there are approximately suggestions on 1,600 related to diabetes. In this study, 100 suggestions were used for prototype development and preliminary experimentation. Twenty frequently asked questions were used for testing purposes. The system makes use of ontology in classifying the questions and establish the semantically mapping concepts. Weights are then used to rank the concepts in order to deduce the most appropriate answer. There are five main processes involved: (1) data pre-processing, (2) ontology construction, (3) answer weighting, (4) answer ranking, and (5) result visualizing. The preliminary experimental results show that the proposed system gave relevant answers and also capable to expand the question in order to get good answers.

Keywords: *Ontology, Question and Answers, Diabetes*

1. Introduction

Paper concerning diagnosis, management, and treatment for Diabetes Mellitus. The CPGs offer to clinical staff a set of recommendations that are focused on the diagnosis and treatment of specific illnesses. Several researchers have already put effort towards ontology development diabetes care. An example of such work is the intelligent ontological agent for diabetic food recommendations proposed by Lee et al. on the development of ontology for

food recommendations [2]. Sivilai et al., [3] reported an ontology-driven personalized food and nutrition planning system developed for elderlies. They also reported the development and design of a personalized nutrition and food planning system employing a food-oriented ontology and an expert system to provide explanations for older people. To support knowledge management in chronic disease healthcare, a semantic web framework was developed by Buranarach *et al.* [4]. To provide a healthcare knowledge management framework for chronic disease care management. Chalortham *et al.* [5]. presented the ontology for a Type II diabetes mellitus clinical support system. Also, the ontology-based clinical support system together with the ontology development process was provided in the study of Chalortham et al. To support chronic disease healthcare, Buranarach et al. designed an alert service using an ontology-based framework for the development of a clinical reminder system for the diabetes patients [6]. In [7], Buranarach et al. reported a personalized Service Framework used for the development of service systems to support diabetes patient self-management. The framework was applied to the development of two service systems for the coordination of related data, knowledge and interaction in order to provide personalized services for the patients. Nevertheless, all the previously reported systems are in English and they are not applicable for the majority of the population in Thailand. In addition, the unique characteristics of Thai posed additional challenges. An Aqua log system was also examined in the research by Lopez et al. Semantic markup was carried out by using General Architecture for Text Engineering (GATE) NLP platform, string metric algorithms and ontology-based similarity services for relations and classes. According to Ou et al's study, an automatic question pattern generation method was presented for QA system based on an ontology with the use of textual entailment. In this method, predictive questions were identified in a particular domain. These predictive questions were generated based on a domain ontology. With respect to the study by Ferrandez et al. For example, Wang *et al.* [8] discussed a system called PortAble Natural language in Terfaceto Ontologies (PANTO). A generic natural language queries were accepted by this system, and then SPARQL queries were achieved as an output [9]. [10], it suggested a semantic search approach by applying query modification to the given input query. This program was used to retrieve information from a formal ontology by inserting input queries in a natural language format. The result of Damljanovic et al's research [11] revealed that there were syntactic parsing and ontologies containing knowledge in an encoded form in natural language interfaces to ontologies named Feedback Refinement and Extended Vocabulary Aggregation (FREYA). Finally, Athenikos & Han [12] studied the current state of biomedical QA system in agreement with semantic knowledge-based QA approaches. According to their analysis, QA process is classified as semantic-based, inference-based and logic-based methods.

In this study, a Thai ontology-based framework for a question answering system for diabetes care was developed. The purpose of the question-answering system is to allow users to ask questions in natural language and in their own words. The development of the question-

answering services system is based on the CPGs in diabetes with emphasis on query analysis and annotation based on similar words. The rest of the paper is organized as follows the background of study is reviewed in the next section the design of proposed framework is then presented. Results and Discussion are shown subsequently. Finally, the conclusions are summarized.

2.Background of the study

2.1 Clinical Practice Guidelines

The Clinical Practice Guidelines (CPGs) [1] for diabetes document was developed by domain experts based on medical and scientific evidence. They are the authoritative resources containing documents and structured recommendations to guide clinical professionals in making suggestions concerning about specific health circumstances. The main objective of CPGs is to provide clinical staff with a set of advice focused on the diagnosis and prognosis assistance, and treatment of specific health problems [13]. The goal is to heighten the medical attention to be provided to the patients. The CPGs also provides important information to support the patient and his/her family about understanding of the treatment, and CPGs is an important tool to improve medical care quality. This requires the development of a formal system to handle and process the information in CPGs and the ontology approach was therefore adopted in this work.

.

2.2 Question Answering System

The Question Answering system in the computer science discipline belongs to the fields of informational retrieval and natural language processing (NLP), and related to the design and building of automatic answering question systems [14]. In this type of systems, questions are entered by humans in the form of natural language. The system aims to allow users to ask questions in natural language using their own terminology. When a natural language question is input into this system, a corresponding answer is then expected to be returned as the output. Ideally, the system is expected to "understand" a user's question and the user should also be satisfied with the output [15].

2.3 Ontology

Ontology involves the conceptualization and organization of collected attributes or terms within a specific domain. Moreover, the attributes or terms are expected to be related between

each other in different dimensions. To build an ontology, one has to scope and define the domain that the system is based on [16]. In terms of computer science, an ontology consists of controlled vocabulary describing objects and the relationships between them in a formal way. Ontology provides a sound basis for sharing domain knowledge between humans and computer programs, or, between computer programs. Concepts (or classes), individuals (or instances), properties, relationships and constraints are generally described in the ontology. Logical formalization of ontology language ensures semantic interpretation, i.e. Inference, by computer programs. Ontology is an essential instrument toward realization of the Semantic Web vision [17].

3. The Design of Framework

The proposed ontology-based question-answering system for diabetes care consists of three main parts: (1) Document Pre-processing, (2) Ontology Construction, and (3) Question-Answering Processing, as shown in Figure 1. Each part can be described as follows.

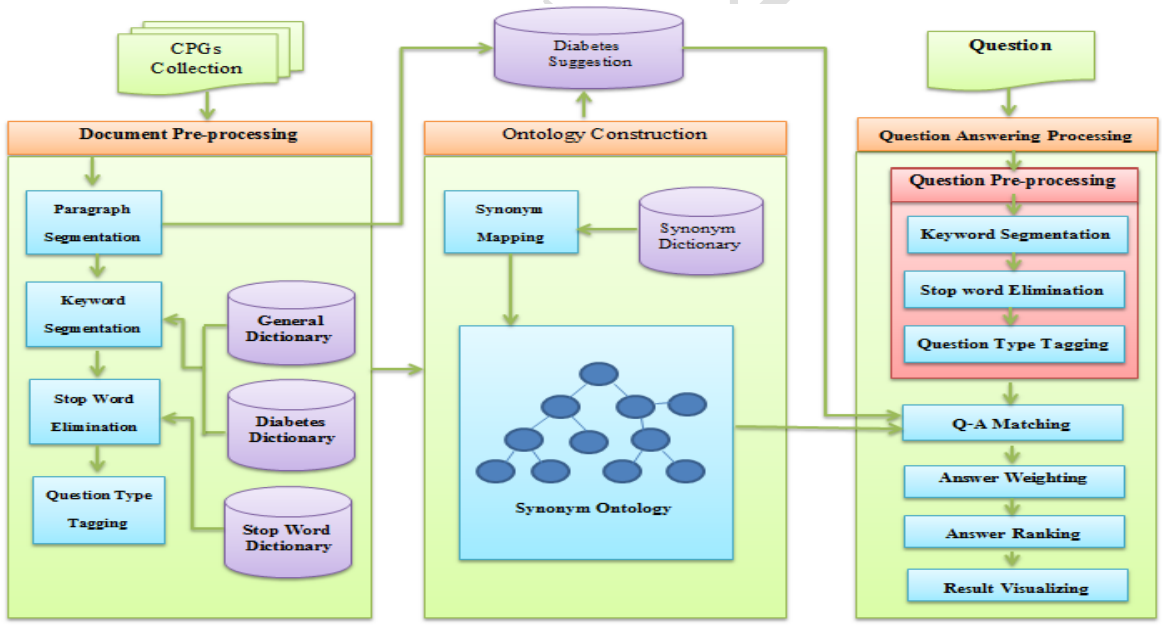

Figure 1 The Design of Framework.

3.1 Document Pre-processing

Part I: Document Pre-processing, consists of paragraph segmentation, keyword segmentation, stop word elimination, and question-type tagging. The paragraph segmentation is used to divide the CPGs documents into paragraphs. Each paragraph is considered as an answer.

To separate sentences into significant words, a keyword segmentation process is then put into the segmented paragraphs. Both general dictionary and diabetic dictionary are applied in the process. Stop words are eliminated from the paragraphs by a stop word dictionary since they are meaningless. An example of the process is illustrated in Table 1.

Table 1 An Example of keyword segmentation and stop word elimination processes.

Paragraph (Suggestion)	Keyword Segmentation and Stop word Elimination
<PAR>ผู้ป่วยเบาหวานขณะตั้งครรภ์ ที่ไม่สามารถถควบคุมระดับน้ำตาลในเลือดด้วยการปรับพฤติกรรมเป็นเบาหวานจากดับอ่อนถูกทำลาย เช่น ตับอ่อนอักเสบเรื้อรัง ถูกตัดตับอ่อน การรักษาเบาหวานต้องอาศัยการฉีดอินซูลิน</PAR>	<PAR>ผู้ป่วย|เบาหวาน|ขณะตั้งครรภ์ |~~ที่~~|~~ไม่~~|สามารถ|ควบคุม|ระดับน้ำตาล|ในเลือด|~~ด้วย~~~~การ~~|ปรับ|พฤติกรรม|~~เป็น~~|เบาหวาน|~~จาก~~|ตับอ่อน|ถูกทำลาย| ~~เช่น~~ |ตับอ่อน|อักเสบ|เรื้อรัง| ถูกตัด|ตับอ่อน| ~~การ~~|รักษา|เบาหวาน|~~ต้องอาศัย~~| ~~การ~~|ฉีด|อินซูลิน</PAR> (Stop words: ที่ ไม่ ด้วย การ เป็น จาก เช่น การ ต้อง อาศัย การ)

The last process in the document pre-processing part is a question-type tagging. The question types, there are 5W1H, including <WHO>, <WHAT>, <WHEN>, <WHERE>, <WHY> and <HOW> tags (labels). The <WHO> tag is who-question which is a person or a group of person such as "ผู้ป่วยเบาหวาน","หญิงตั้งครรภ์". The <WHAT> tag provides a treatment for diabetes, for example, "รักษา", "ป้องกัน", "ดูแล". The <WHERE> tag tells about organs involved in diabetes complications or pain areas, such as "ในเลือด","ตับอ่อน". The <WHEN> tag specifies the time or any period of time that symptoms are expressed for example "ขณะตั้งครรภ์", "ระหว่างตั้งครรภ์". The <HOW> tag shows the symptoms of diabetes, such as, "กระหายน้ำ", "อ่อนเพลีย". The 5W1H tags are adopted for grouping the keywords. They help reduce an ambiguity of any keywords among different question-types.

3.2 The Ontology Construction

The second part is the ontology construction. The ontology is built from the keyword of 5W1H tag. Figure 2 illustrates an example of classes and their Six tags (<WHO>, <WHAT>, <WHEN>, WHERE>, <WHY>, and <HOW>) are defined as classes. Their keywords will be assigned as subclasses, for example, a subclass of <WHO> can be "เด็ก", and "ผู้สูงอายุ". In the synonym mapping process, synonyms will be put to their leaf nodes, e.g., "คนเป็นเบาหวาน" is the synonym of "ผู้ป่วยเบาหวาน". The synonyms are mapped by a prepared synonym dictionary.

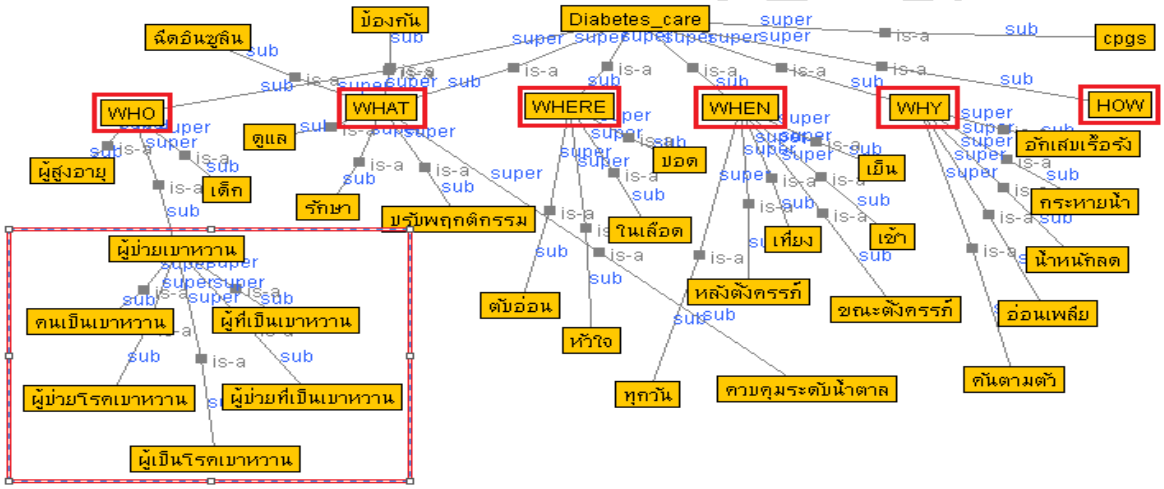

Figure 2 An Example of the ontology construction.

3.3 The Question Answering processing

The final part is the question-answering processing, consisting of a question pre-processing, QA matching, answer weighting, answer ranking and result visualizing. The question pre-processing is the process which prepares the format of a question before being processed. The pre-processing stage is similar to the first step, document pre-processing, except that there is no paragraph segmentation process, so that the question will be segmented and tagged in the following step. The QA matching stage maps the keywords between the question and the ontology which has the same tag. The synonyms of each keyword will be retrieved from the ontology. For every suggestion, they are then matched. As the answers will be ranked according to a weighting scheme, the matched keywords are weighted by TFIDF.

The suggestions are then ordered by the sum of their keyword's weights in the answer ranking process. The top-5 suggestions will be selected to be the answers for result visualizing.

4. Results and discussion

In this paper, a proposed ontology-based framework of a Thai Question and Answer for diabetes care is presented. The system is able to choose answers from the corpus based on questions from the users. Synonym words were used to solve the problems of different Thai sentences but similar meanings. Regarding the information for the efficiency evaluation of the system, twenty frequently asked questions and answers in diabetes care were collected to answer to meet user's needs. The most essential thing of this question answering systems is the understanding of user's questions with an objective to meet the user's needs. It could be seen that the evolution of research results was undertaken only by TF-IDF weighing so that the results tended to provide the correct answers according to the algorithm. For future research, the database will be expanded and other weighting methods will be used.

5. Conclusions

A Thai ontology-based framework for question answering system of diabetes care was proposed and presented in this paper. Three main parts were introduced: (1) Document pre-processing (2) ontology construction, and (3) question answer processing. Ontology of 5W1H tags was built to decrease the ambiguity of their keywords. Weighting term is a significant factor to rank the suggestions processed by the question answering matching process. Finally, the best suggestion from ranked answers was displayed to the patients.

Acknowledgement

The authors acknowledge the provision of the *Clinical Practice Guidelines* of the Endocrine Society of Thailand for the experimental dataset. This work was also supported by Rajamangala University of Technology Srivijaya.

References

[1] Endocrine Society of Thailand. Clinical Practice Guideline for Diabetes Mellitus 2014, 2014.

[2] CS Lee, MH Wang, HC Li and WH Chen. Intelligent ontological agent for diabetic food recommendation. *In*: Proceedings of the IEEE International Conference on World Congress on Computational Intelligence Fuzzy Systems, Hong Kong, 2008, p. 1803-10.

[3] S Sivilai, C Snae and M Brueckner. Ontology-driven personalized food and nutrition planning system for the elderly. *In*: Proceedings of the 2nd International Conference in Business Management and Information Sciences, Phitsanulok, Thailand, 2012.

[4] M Buranarach, T Supnithi, N Chalortham, V Khunthong, P Varasai and A Kawtrakul. A semantic web framework to support knowledge management in chronic disease healthcare. *Metadata Semantic Res.* 2009; **46**, 164-70.

[5] N Chalortham, M Buranarach and T Supnithi. Ontology development for type II diabetes mellitus clinical support system. *In*: Proceedings of the 4th International Conference on Knowledge Information and Creativity Support Systems, Seoul, Korea, 2009.

[6] M Buranarach, N Chalortham, P Chatvorawit, Y Thein and T Supnithi. An ontology-based framework for development of clinical reminder system to support chronic disease healthcare. *In*: Proceedings of the 4th International Symposium on Biomedical Engineering, Thailand, 2009.

[7] M Buranarach, N Thatphithakkul, A Kawtrakul, S Wongrochananan, N Kulnawan and W Jiamjarasrangsi. Development of service systems to support diabetes patient self-management using a personalized service framework. *In*: 2011 Annual SRII Global Conference, USA, 2011, p. 363-70.

[8] C Wang, M Xiong, Q Zhou and Y Yu. Panto: A portable natural language interface to ontologies. *In*: Proceedings of the 4th European conference on The Semantic Web: Research and Applications, Austria, 2007, p. 473-87.

[9] V Lopez, V Uren, E Motta and M Pasin. AquaLog: An ontology-driven question answering system for organizational semantic intranets. *Web Semantics* 2007; **5**, 72-105.

[10] O Ferrández, R Izquierdo, S Ferrández and JL Vicedo. Addressing ontology-based question answering with collections of user queries. *Inform. Process. Manag.* 2009; **45**, 175-88.

[11] D Damljanovic, J Petrak, M Lupu, H Cunningham, M Carlsson, G Engstrom and B Andersson. Random indexing for finding similar nodes within large RDF graphs. *In*: Proceedings of the 8th Extended Semantic Web Conference, Greece, 2012, p. 156-71.

[12] SJ Athenikos and H Han. Biomedical question answering: A survey. *Comput. Meth. Programs Biomed.* 2010; **99**, 1-24.

[13] F Pech-May, I Lopez-Arevalo and VJ Sosa-Sosa. *Information Extraction Approach for Clinical Practice Guidelines Representation in a Medical Decision Support System*. InTech Open Access Publisher, Croatia, 2011.

[14] B Webber and N Webb. *Question Answering. In*: A Clark, C Fox and S Lappin (eds.). The Handbook of Computational Linguistics and Natural Language Processing, Wiley-Blackwell, USA, 2010, p. 630-54.

[15] O Kolomiyets and MF Moens. A survey on question answering technology from an information retrieval perspective. *Inform. Sci.* 2011; **181**, 5412-34.

[16] NF Noy and DL Mcguinness. *Ontology Development 101: A Guide to Creating Your First Ontology*. Stanford Knowledge Systems Laboratory Technical Report KSL_01_05 and Stanford Medical informatics Technical Report SMI-2001-0880, USA, 2001.

[17] R Mizoguchi, E Sunagawa, K Kozaki and Y Kitamura. The model of roles within an ontology development tool: Hozo. *Appl. Ontology* 2007; **2**, 159-79.

Ketsara Phetkrachang received a bachelor and master degree in Management of Information of technology from Ramkhamhaeng University, Thailand in 1995 and Walailak University, Thailand in 2003, respectively, and doctoral degree in Management of Information technology, Walailak University in 2019 .Her research interests are Information Retrieval, Information Engineering, Question answering systems and Expert System.

Nichnan Kittiphatttanabawon received a bachelor and master degree in Computer Science from Rangsit University in 1993 and Prince of Songkhla University in 1999, respectively, and doctoral degree in Technology)International Program(from Sirindhorn International Institute of Technology, Thammasat University in 2012 .Currently, she is a lecturer in information technology and management of information technology, school of informatics, Walailak university .Her current research interests include text mining, data mining, document relation discovery, and predictive analytics.