Comparison of Question Answering Systems

Tripti Dodiya and Sonal Jain

Abstract. Current Information retrieval systems like Google are based on keywords wherein the result is in the form of list of documents. The number of retrieved documents is large. The user searches these documents one by one to find the correct answer. Sometimes the correct or relevant answer to the searched keywords is difficult to find. Studies indicate that an average user seeking an answer to the question searches very few documents. Also, as the search is tedious it demotivates the user and he/she gets tired if the documents do not contain the content which they are searching for. Question-answering systems (QA Systems) stand as a new alternative for Information Retrieval Systems. This survey has been done as part of doctoral research work on "Medical QA systems". The paper aims to survey some open and restricted domain QA systems. The surveyed QA systems though found to be useful to obtain information showed some limitations in various aspects which should resolved for the user satisfaction.

Keywords: Information retrieval systems, Question Answering system, Open QA systems, Closed QA systems.

1 Introduction

Current search engines are based on keywords wherein the result is in the form of list of documents. The number of retrieved documents is large. For instance: querying about obesity results in more than 186,000,000 documents. The user searches these documents one by one to find the correct answer. Sometimes the correct or relevant answer is difficult to find. Studies indicate that an average user seeking an answer to the question searches very few documents. Also, as the search is tedious, it demotivates the user and gets tired if the documents do not contain the content which they are searching for.

Tripti Dodiya
GLS Institute of Computer Applications, Ahmedabad
e-mail: triptidodiya@glsica.org

Sonal Jain
GLS Institute of Computer Technology, Ahmedabad
e-mail: sonal@glsict.org

A. Abraham and S.M. Thampi (Eds.): Intelligent Informatics, AISC 182, pp. 99–107.
springerlink.com

QA systems, unlike information retrieval (IR) systems, can automatically analyze a large number of documents and generate precise answers to questions posed by users. These systems employ Information Extraction (IE) and Natural Language Processing (NLP) techniques to provide relevant answers. QA systems are regarded as the next step beyond search engines [8].

While the research on automated QA in the field of Artificial Intelligence (AI) dates back to 1960s, more research activities involving QA within the IR/IE community have gained momentum by the campaigns like TREC evaluation which started in 1999 [10]. Since then techniques have been developed for generating answers for three types of questions supported by TREC evaluations, namely, factoid questions, list questions, and definitional questions.

2 Literature Survey

QA systems are classified in two main parts [8]: (a) open domain QA system (b) restricted domain QA system.

Open domain QA systems deal with questions about nearly everything and can only rely on general ontology and world knowledge [8]. Alternatively, unlimited types of questions are accepted in open domain question answering system.

Restricted domain QA systems deal with questions under a specific domain (for example, biomedicine or weather forecasting) and can be seen as an easier task because NLP systems can exploit domain-specific knowledge frequently formalized in ontology. Alternatively, limited types of questions (or questions related to a particular domain) are accepted in restricted domain system.

Some of the open domain QA systems and restricted domain QA systems are surveyed as part of the research work on "Medical QA Systems".

2.1 START(SynTactic Analysis Using ReversibleTransformations)

START is the world's first Web-based open QA system developed by Boris Katz and associates in December 1993 [3]. Currently, the system answers questions about places, movies, people, dictionary definitions, and more. START parses incoming questions, matches the queries created from the parse trees against its knowledge base (KB) and presents the appropriate information segments to the user. It uses a technique "natural language annotation" which employs natural language sentences and phrases "annotations" as descriptions of content that are associated with information segments at various granularities. An information segment is retrieved when its annotation matches an input question. This technique allows START to handle variety of questions from different domains [3].

START system was tested using some sample questions related to various domain, and most of them were answered. Resluts were precise and few of them were supported with images. Figure 1 shows the question and the result displayed along with the source. However, dissatisfaction was observed in some cases. For instance, output for question "*procedure for kidney stone removal*" was unsuccessful as shown in Figure 2.

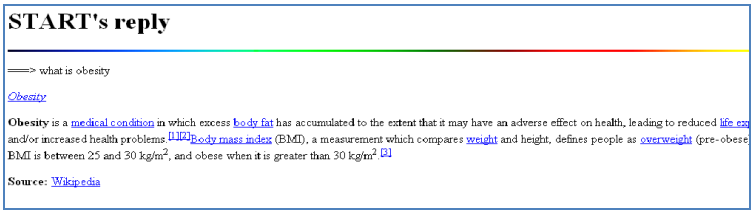

Fig. 1 START QA result along with source

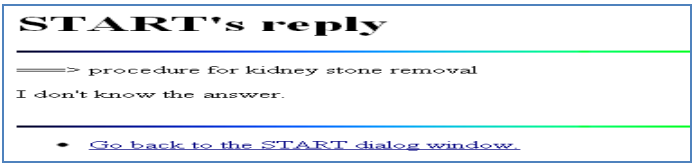

Fig. 2 START QA result for restricted domain question

Limitations

The survey results concluded that START answered to almost many questions pertaining to different domains, but when questions are more related to restricted domain, many were unsuccessful.

START can be accessed at http://start.csail.mit.edu/

2.2 Aqualog

Venessa lopez et al. in [11,12] discusses Aqualog as a portable QA system that takes queries in natural language, an ontology as input, and returns answers drawn from knowledge bases (KBs), which instantiate the input ontology with domain-specific information. The techniques used in Aqualog are:

- It makes use of the GATE NLP platform in linguistic component
- String metrics algorithms
- WordNet (open domain ontology)
- Novel ontology-based similarity services for relations and classes, to make sense of user queries with respect to the target knowledge base.

AquaLog is coupled with a portable and contextualized learning mechanism to obtain domain-dependent knowledge by creating a lexicon [11,12]. It is portable as its architecture is completely independent from specific ontology's and knowledge representation systems. It is also portable with respect to knowledge representation, because it uses a modular architecture based on a plug-in mechanism to access information about an ontology, using an OKBC-like protocol. AquaLog

uses KMi ontology and is coupled with the platform WebOnto , a web server which contains the knowledge models or ontologies in an OCML format.

Limitations

- AquaLog does not handle temporal reasoning. The results are not proper for questions formed with words like: yesterday, last year etc.
- It does not handle queries which require similarity or ranking reasoning. For instance: "*what are the top/most successful researchers*?".
- It does not handle genitives. For example: What's, Project's etc.

2.3 MedQA (Askhermes)

MedQA developed by Lee et al. [7] is a biomedical QA system. It is developed to cater to the needs of practicing physicians. The system generates short text answers from MEDLINE collection and the Web. The current system implementation deals with definitional questions, for example: "*What is X?*". Figure 3 shows sample question "kidney stone removal" with output extracted from Google, PubMed and OneLook along with time taken to search the result.

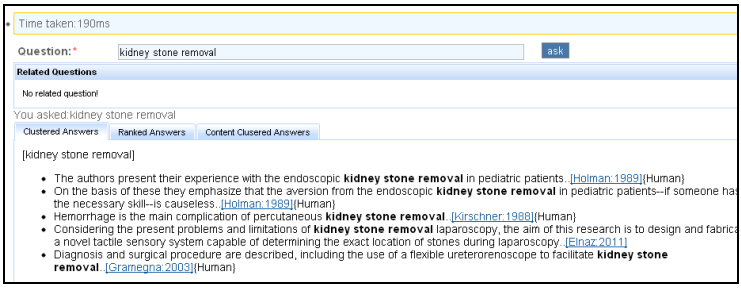

Fig. 3 MedQA results with the posted question

Limitations

Minsuk Lee et al. in [2,7] concluded that MedQA out-performed three online information systems: Google, OneLook, and PubMed in two important efficiency criteria; a) time spent and b) number of actions taken for a physician to identify a definition. However, some limitations observed are:

- Capacity is limited due to its ability to answer only definitional question.
- Another important issue is speed. As the QA systems needs to process large documents and incorporates many computational intensive components it consumes more time.
- It does not capture semantic information which plays an important role for answer extraction and summarization.

MedQA can be accessed at www. askhermes.org/MedQA/

2.4 HonQA

HonQA developed by HON (Health On the Net) foundation is a multilingual biomedical system useful for both patients and health professionals [13]. It is restricted domain and has multilingual support in English, French and Italian languages. Figure 4 shows the advanced search option with selection of the language and source for the search. By default the search is done in the database of certified websites.

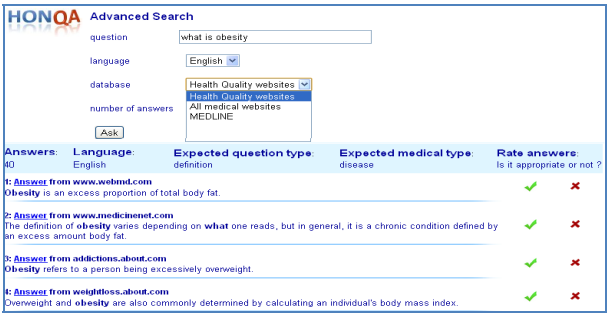

Fig. 4 Search results of HonQA

The results displays multiple definitions, source, question type and medical type. The user is also asked to rate the answer for further reference.

Limitations

The survey results shows that HonQA results are short definitional answers, whereas the users expect more details pertaining to the query. Time taken to extract the answers is high.

HonQA can be accessed at www.hon.ch/QA/

2.5 QuALiM

Michael kaisser et al. in [6] describes the demo of QuALiM open domain QA system. The results are supplemented with relevant passages. In [6] Wikipedia is used as search engine. QuALiM uses linguistic methods to analyse the questions and candidate sentences in order to locate the exact answers [5].

Limitations

Michael kaisser et al. in [6] concluded with some issues faced regarding the delay caused due to search engines API's, post processing delay and results fetched from Wikipedia that needs to be resolved.

3 QA Systems in Other Languages

QA Systems generally have been developed in English as majority of documents available on the internet are in English. However, there are examples of QA systems that have been developed in foreign languages like Arabic, Spanish etc.

Table 1 QA systems in other languages

Name	Language	Details
AQUASYS [9]	Arabic	Fact based questions
QARAB [1]	Arabic	Extracts answers from Arabic newspapers
GeoVAQA [4]	Spanish	Voice Activated Geographical Question Answering System

Table 2 Comparative study of open and restricted domain QA systems

System observed [paper ref.] / Features	Start [3]	Aqualog [11,12]	MedQA [2,7]	HonQA [13]	Qualim [5,6]
Open/Restricted domain	Open	Restricted	Restricted	Restricted	Open
Language used for Implementation	Common LISP	Java	Perl	Not specified	Java
Wordnet support	Yes	Yes	No	Not specified	Yes
Ontology support	Yes	Yes	No	Not specified	No
Ontology portability	Not required	Yes	No	No	Not required
Supports multilingual documents	Yes	No	No	Yes	No
Language used to display result	English	English	English	English, Italian, French	English

4 Evaluation and Results

For the evaluation, sample questions were generated for open and restricted domain QA systems. The medical QA systems were evaluated by doctors and their satisfaction level with 5 point likert scale was recorded. Restricted domains were tested with MedQA, HonQA and START while for open domain START was used. Some sample questions were not answered which we considered as dissatisfactory for the result. Table 3 lists some sample questions.

Table 3 Sample questions input to QA systems

Open Domain	Restricted Domain
Who is the president of India	What is length of large intestine
When was Gandhiji born	What is the function of pancreas
Who invented electricity	What is normal blood pressure in adult
Which is fastest flying bird	Which are the reproductive organs of human body
Which country has the largest army	How to calculate BMI
Who is world's fastest runner	How many bones are there in human body
How many planets are there in the milky way	When was penicillin invented

The evaluation results in case of restricted and open domain systems are given below.

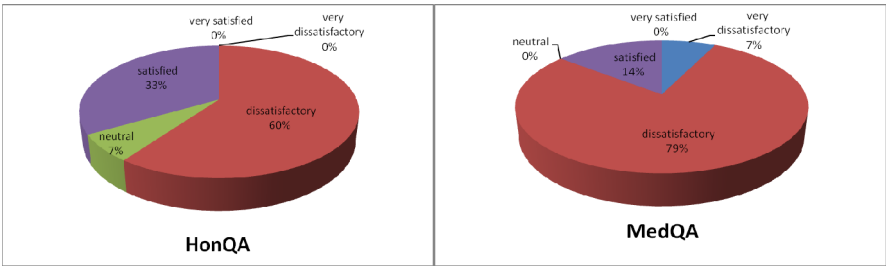

Fig. 5 Results of HonQA and MedQA for restricted domain sample questions

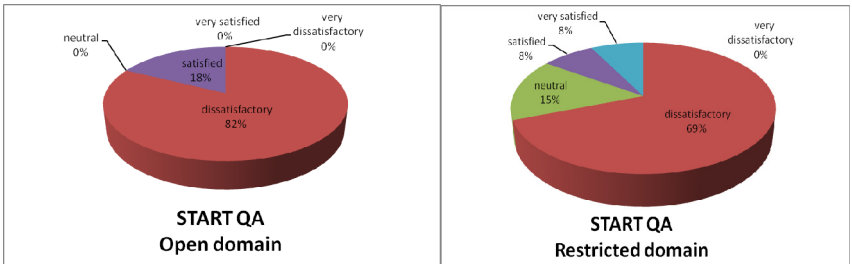

Fig. 6 Results of START QA for restricted and open domain sample questions

Looking at the above results, we can conclude that the user satisfaction level in case of restricted and open domain QA systems is less. As per the oral inputs given by the users, the dissatisfaction was also because of the format and the content provided in the results.

5 Conclusion

QA systems represent the next step in information access technology. By delivering precise answers they can more effectively fulfill users' information needs. Corresponding to the growth of information on the web, there is a growing need for QA systems that can help users better utilize the ever-accumulating information. The surveyed QA systems though found to be useful to obtain information, showed some limitations in various aspects which can be resolved for the user satisfaction. Also, continued research toward development of more sophisticated techniques for processing NL text, utilizing semantic knowledge, and incorporating logic and reasoning mechanisms, will lead to more useful QA systems.

6 Future Scope

In the survey, many of the limitations discussed needs to be overcome. The results should be formatted to be more understandable to the user. Speed is an important issue. Obviously, the higher is the speed, the greater is the user satisfaction. It makes a challenge for QA systems to deliver optimal response times.

References

[1] Hammo, B., Abu-Salem, H., Lytinen, S., Evens, M.: QARAB: A Question Answering System to Support the Arabic Language. In: ACL 2002 Workshop on Computational Approaches to Semitic Languages, Philadelphia, PA, pp. 55–65 (July 2002)

[2] Yu, H., Lee, M., Kaufman, D., Ely, J., Osheroff, J.A., Hripcsak, G., Cimino, J.: Development, implementation, and a cognitive evaluation of a definitional question answering system for physicians. Journal of Biomedical Informatics 40, 236–251 (2007)

[3] http://start.csail.mit.edu/start-system.html

[4] Luque, J., Ferrés, D., Hernando, J., Mariño, J.B., Rodríguez, H.: Geovaqa: A voice activated geographical question answering system

[5] Kaisser, M., Becker, T.: Question Answering by Searching Large Corpora with Linguistic Methods. In: The Proceedings of the 2004 Edition of the Text Retrieval Conference, TREC 2004 (2004)

[6] Kaisser, M.: The QuALiM question answering demo: supplementing answers with paragraphs drawn from Wikipedia. In: Proceedings of the 46th Annual Meeting of the Association for Computational Linguistics on Human Language Technologies: Demo Session, Columbus, Ohio, June 16, pp. 32–35 (2008)

[7] Lee, M., Cimino, J., Zhu, H.R., Sable, C., Shanker, V., Ely, J., Yu, H.: Beyond Information Retrieval-Medical Question Answering. In: AMIA Annu. Symp. Proc., pp. 469–473 (2006)

[8] Kangavari, M.R., Ghandchi, S., Golpour, M.: Information Retrieval: Improving Question Answering Systems by Query Reformulation and Answer Validation. World Academy of Science, Engineering and Technology 48, 303–310 (2008)

[9] Bekhti, S., Rehman, A., Al-Harbi, M., Saba, T.: AQUASYS: An Arabic Question-Answering system based on extensive question analysis and answer relevance scoring. International Journal of Academic Research 3(4), 45–54 (2011)
[10] Athenikosa, S.J., Hanb, H.: Biomedical question answering: A survey. Computer Methods and Programs in Biomedicine 99, 1–24 (2010)
[11] Lopez, V., Motta, E.: Aqualog: An ontology-portable question answering system for the semantic web. In: Proceedings of the International Conference on Natural Language for Information Systems, NLDB, pp. 89–102 (2004)
[12] Lopez, V., Uren, V., Motta, E., Pasin, M.: AquaLog: An ontology-driven question answering system for organizational semantic intranets. Web Semantics: Science, Services and Agents on the World Wide Web 5, 72–105 (2007)
[13] http://www.hon.ch