International Journal of Engineering and Technical Research (IJETR)
ISSN: 2321-0869, Volume-3, Issue-2, February 2015

A survey of medical question answering systems

Samrudhi Sharma, Huda Patanwala, Manthan Shah, Khushali Deulkar

***Abstract*— A Question Answering system is an intelligent information retrieval system that retrieves precise answers for natural language queries in a short span of time. Such systems are particularly useful in the field of biomedicine where medical literature continues to increase. In this paper, a brief overview and comparison of four biomedical systems is presented. The systems AskHermes, MedQA, HONQA, and MiPACQ are compared over features such as corpus, architecture, user interface, etc.**

***Index Terms*— Question Answering, Information Retrieval, Information Extraction, Corpus, Comparison, Architecture, User Interface, AskHermes, HonQA, MedQA, MiPACQ.**

I. INTRODUCTION

Health care professionals often have to refer to medical literature and documents while seeking answers for medical queries. Medical databases such as Thompson Micromedex, PubMed or search engines such as Google are powerful resources of upto date medical knowledge. However, the existing documentation is large and makes it difficult for professionals to retrieve answers quickly in a clinical setting. The problem with search engines and informative retrieval engines are that these systems return a list of documents rather than answers. Instead, health care professionals can use question answering systems to retrieve short sentences or paragraphs in response to medical queries. Such systems have the biggest advantage of generating answers and providing hints in a few seconds

Question Answering (QA) systems can be defined as the task whereby an automated machine (such as a computer) answers arbitrary questions formulated in natural language. [1] In general such a system analyzes a large amount of information in a short amount of time to give answers. Incorporating QA systems with Natural Language Processing (NLP) give a machine the ability to better understand a doctor's query and fetch precise answers. Analyzing clinical questions is one of the biggest hurdles a system must face. This is because the questions can range from factoid questions to extremely complex, verbose questions.

In this paper, we will review and compare four biomedical question answering systems 1) AskHermes 2) MedQA 3) HONQA 4) MiPACQ. Our objective is to explore the various design and implementation decisions undertaken while building such complex systems.

1) AskHermes: The objective of this system is to perform robust semantic analysis on complex clinical questions and output question focused extractive summaries as answers. [2] Queries are allowed in natural language without much formulation. Moreover, the system can handle complex questions by using a structured domain-specific ontology.

Manuscript received February 11, 2015.

2) MedQA: The goal of this system is to automatically generate paragraph-level text for definitional questions (i.e. "What is X?"). [3]

3) HONQA: The application developed by HON is available in both English and French. In addition, the user is able to choose the domain of research i.e. Websites accredited by HON or all the websites, regardless of accreditation. By default the research is based on the accredited sites as the quality of the available information is paramount, far more so than the quantity. [4]

II. PROCEDURE FOR PAPER SUBMISSION

A. Corpus

1) AskHermes: MEDLINE abstracts, PubMed Central full-text articles, eMedicine documents, clinical guidelines, and Wikipedia.

2) MedQA: Web documents (using Google: Definition) and MEDLINE abstracts.

3) HONQA: Two corpora were used: the first was composed of 136 questions in English and the second, 140 questions in French. In both cases, the questions were collected from Frequently Asked Questions (FAQ) found in forums specialized in health as well as discussions on the Internet.

4) MiPACQ: Medpedia, 4654 questions from the Clinques corpus.[6]

B. General Architecture

1) AskHermes:

1. *Data sources and preprocessing*: It can be sub divided into two sections: Data collection, and Pre-processing for retaining semantic content. In the data collection phase medical literature and articles are indexed.
2. *Question analysis*: Questions are classified into 12 general topics to facilitate information retrieval. Support Vector Machines (SVMs) showed the best results for question classification. Since a question can be assigned to multiple topics, the team developed a binary classifier (Yes or No) for each of the 12 topics. For keyword identification, they formulated it as sequence labeling problem using conditional random fields (CRFs) model. In addition to basic lexical features (e.g. unigram, bigram) and syntactic features (e.g. parts-of-speech), the lexical tool MMTx, an implementation of MetaMap was incorporated.
3. *Document Retrieval*: The latest version of the probabilistic relevance model BM25 is empirically tuned for retrieval. [7]
4. *Passage Retrieval*: Instead of using naturally occurring paragraph or a fixed window, AskHermes dynamically generates passage boundaries.
5. *Summarization and answer presentation*: This module is sub divided into two sections:

➔ Topical clustering, ranking and hierarchical answer presentation

➔ Redundancy removal based on longest common substring

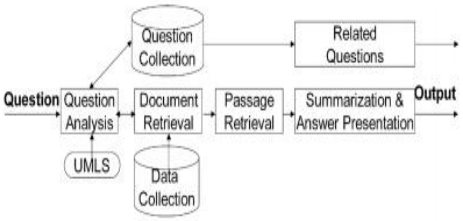
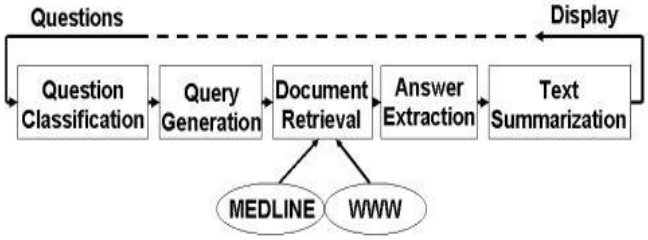
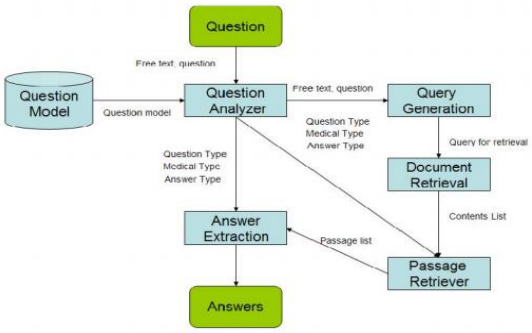


Fig. 1.
The AskHERMES architecture.

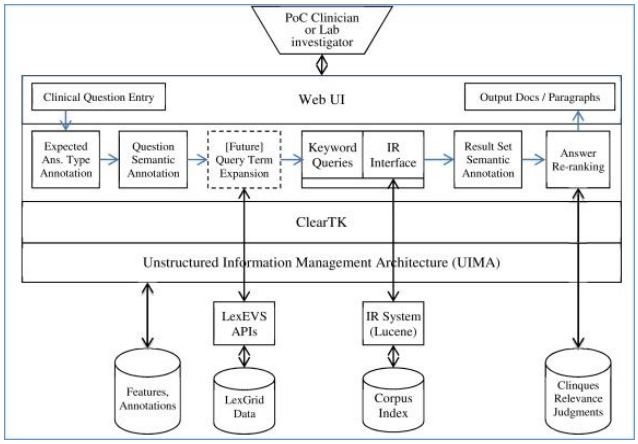
2) MedQA:
1. *Question classification:* This module automatically classifies a question posed by a physician into a question type for which a specific answer strategy will be developed.
2. *Query Generalization:* This module analyzes the question to extract noun phrases as the query terms.
3. *Document Retrieval:* Query terms are applied to retrieve documents from either the Web documents or the locally-indexed MEDLINE collection.
4. *Answer Extraction:* Automatically identifies the sentences that provide answers to questions.
5. *Text Summarization:* Removes the redundant sentences and condenses the sentences and condenses the sentences into a coherent summary. The user is presented the summary of the answer.



3) HONQA: In a first step classifying questions automatically is preferred, with no information about their categories. For this task we used the software Rapid Miner which is an environment for machine learning. It consisted of automatically detecting clusters only from the corpora according to the characteristics of the questions. Moreover, according to the algorithm used, it is possible to fix or remove the number of classes of exits. Initially, the questions were presented in the untreated form (called unigram form). Several algorithms of clusterisation were tested (K-Means, EM, DBSCAN, Kménoids, and Kernel-KMeans). Following this, questions were proposed with bigrams of words or co-occurrences. The bigrams of words allow identifying the expressions composed. The co-occurrences associate the distant words within a sentence belonging to the same linguistic unit. It is considered that there is co-occurrence when the presence of a word in a text gives an indication to the presence of another word. The interest to apply such work to the set of questions is to locate the expressions, associations of words, which are recurring on the entire corpus. Non supervised classification did not emphasize new classes of questions. However, we could observe that the software creates clusters according to the type of interrogation term. For instance, all questions beginning by "What is" were in the same cluster. Consequently, this work was not useless and has prepared the step of conception by providing distinctive structures of sentences and useful expressions.



4) MiPACQ: Questions are submitted by a clinician or investigator using the web-based user interface. The question is then processed by the annotation pipeline which adds semantic annotations. MiPACQ then interfaces with the information retrieval (IR) system which uses term overlap to retrieve candidate answer paragraphs. Candidate answer paragraphs are annotated using the annotation pipeline. Finally, the paragraphs are re-ordered by the answer re-ranking system based on the semantic annotations, and the results are presented to the user.



System implementation:
1) AskHermes: The AskHERMES system is built on the J2EE framework, in which JBoss is used for the application server and the JBoss Seam for building the user interface. JBoss has built-in EJB (Enterprise JavaBeans) caching and a reuse mechanismthat enables heavy load accessing. We also built a round-robin load-balancer in the front web server to distribute the accessing load among six backend servers. The six servers are running Linux/Solaris operating systems in which AskHERMES is deployed. Currently, AskHERMES system response time averages 20 s.

2) MedQA: MedQA is implemented with Perl as the core platform and is running on a Macintosh PowerPC (dual 2 Ghz CPU, 2 GB of physical memory, Mac OS X server 10.4.2). The distributions of time spent among different components are 6 seconds for *Document Retrieval*, 6 seconds for *Answer Extraction*, and 6 seconds for *Text Summarization*.

Availability and user interface:

1) AskHermes: http://www.AskHERMES.org

2) MedQA: http://askhermes.org/MedQA/

3) HONQA: http://services.hon.ch/cgi-bin/QA10/qa.pl

3) MiPACQ: The system is not available online.

III. CONCLUSION

Thus, we have presented a detailed and in-depth analysis of the various question answering systems in the table below.

	AskHermes	MedQA	HONQA	MiPACQ
Web Address	http://www.AskHERMES.org	http://askhermes.org/MedQA/	http://services.hon.ch/cgi-bin/QA10/qa.pl	-
Corpus	MEDLINE abstracts, eMedicine, clinical guidelines, PubMed Central, and Wikipedia	Web documents (using Google: Definition), MEDLINE abstracts	HON certified websites	Medpedia, 4654 questions from the Clinques corpus
Answers	Multiple sentence passages	Paragraph-level answers	Sentence	Ranked paragraphs
Language	English	English	English/ French	English
Interface	Simple	Very Simple	Very Simple	-
QA	Yes	Yes	Yes	Yes

REFERENCES

[1] Mollá, D. and Vicedo, J. (2007). Question Answering in Restricted Domains: An Overview. Computational Linguistics, 33(1), pp.41-61.
[2] Cao, Y., Liu, F., Simpson, P., Antieau, L., Bennett, A., Cimino, J., Ely, J. and Yu, H. (2011). AskHERMES: An online question answering system for complex clinical questions. Journal of Biomedical Informatics, 44(2), pp.277-288.
[3] Lee M, Cimino J, Zhu HR, Sable C, Shanker V, Ely J, et al. Beyond information retrieval--medical question answering. AMIA Annual Symposium Proceedings. 2006:469–473.
[4] Supervised approach to recognize question type in a QA system for health. Stud Health Technol Inform. 2008; 136(136)):407–412.
[5] Cairns BL, Nielsen RD, Masanz JJ, Martin JH, Palmer MS, et al. (2011) The MiPACQ clinical question answering system. AMIA Annu Symp Proc 2011: 171–180.
[6] Ely J, Osheroff J, Chambliss M, Ebell M, Rosenbaum M. Answering Physicians' Clinical Questions: Obstacles and Potential Solutions. J Am Med Inform Assoc. 2005 Mar-Apr; 12(2):217–224.
[7] Robertson, S., Zaragoza, H. and Taylor, M. (2015). Simple BM25 extension to multiple weighted fields. CIKM '04 Proceedings of the thirteenth ACM international conference on Information and knowledge management, pp.42-49.