Project Title

Medical Question Answering System: Revisited

Project Members

John Patrick Amata

Immanuel Espiritu

Jayvee Febrer

I. Introduction

1.1 Background of the Problem

The medical profession requires a great deal of memorization and information retrieval of medical texts. Doctors and physicians often need to refer to medical documents and most get theirs from the internet. Due to their profession, time is of great essence. A study showed that it took an average of more than 30 minutes for a healthcare provider to search for answers from PubMed.

The problem lies on the cause of information overloading; a simple query immediately presents hundreds of articles. Question answering systems provide answers for questions in a natural language form. They accomplish this task by analysing the text and providing a summarised answer from a database of articles and documents. An early feasibility study on Medical Question Answering systems showed that In QA as in IR, one cannot search for every kind of information on the internet, since they are not equally represented. Due to the promising value that Medical Q&A systems bring, our aim is to study the modern trends of information retrieval and question answering models and their application to the Medical field.

1.2 Statement of the Problem

What are the impact of modern advancements in QA in making a Medical Question Answering system?

1.3 Objectives

Our objective is to experiment and study applicable trends in Question Answering and its application to a Medical Question Answering system.

1.4 Significance

Our study will provide a modern model(s) of a question answering system for researchers and developers to follow in creating their own Medical Question Answering system. The promising value that Medical Q&A systems bring in regards to the summarising texts and answering quick manner, addresses the problem of information overload that medicine practitioners encounter.

II. Scope and Limitations

Our study will only examine and experiment with Factoid Question Answering systems. For the context of our corpus, we shall gather 100 questions about Diabetes. With these serving as our usage context, we shall compare the results between two methods of Question-Answering: Traditional Information Retrieval and Neural Networks.

III. Related Literature

Question Answering: From a very general perspective QA can be defined as an automatic process capable of understanding questions formulated in a natural language such as English and responding exactly with the requested information.

(Handbook of Natural Language Processing, Second Edition (Chapman & Hall))

New Trends in Automatic Question Answering

**Computational Linguistics**

|  |  |
| --- | --- |
| Maximum Entropy Model | The basic idea of the maximum entropy model is, that the probability distribution which best represents the current state of knowledge is the one with the largest entropy. It is used to build up models of many different sources with limited information. |
| Decision Tree Learning | Decision tree learning is a method used in many domains of knowledge discovery, pattern recognition and data mining. Decision trees are hierarchical trees which try to predict an output variable for a given input. Each leave represents an attribute. Each path is a conjunction of the attributes which are on the path. |
| Artificial Neural Networks | Artificial neural networks are mathematical models based on the biological neuron structure of the brain. The neurons, also called nodes, together with the weighted connections are the basic components. The advantages of artificial neural networks are the possibility of enhancing the processing speed by parallelization, adaptation of knowledge, robustness and implementation in low power applications |

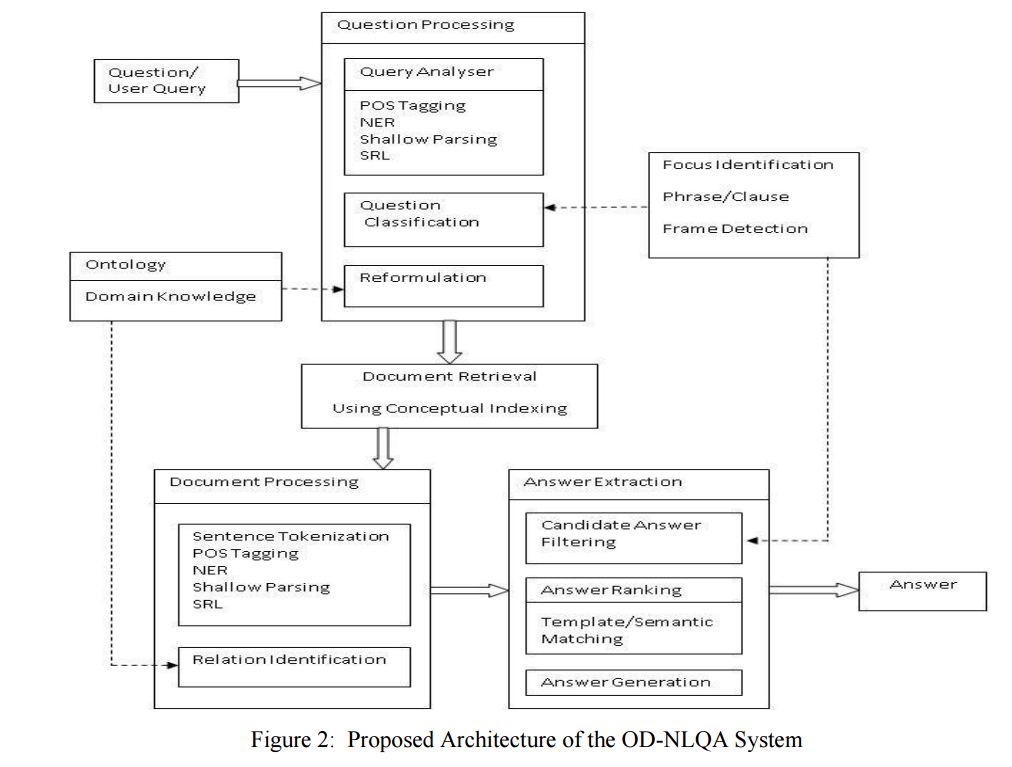
**Information Extraction**

|  |  |
| --- | --- |
| Entity extraction | It identifies and classifies all phrases in a free text which refer to objects of semantic classes like names, nouns, pronouns etc. In addition all object mentions are linked together which refer to the same entity. |
| Relation extraction | The relations between entities are identified. A relation is always represented by two entities and can be described in many languages. |
| Event extraction | It is also a common application in order to derive specific knowledge from a text. Event extraction identifies events of particular types and the corresponding arguments. A type of an event would be for example ”car crash” or ”natural disaster”. |

<http://www.iicm.tugraz.at/0x811bc82b_0x0011c036>

Architecture of an Ontology-Based Domain-Specific Natural Language Question Answering System

The proposed architecture of an ontology-based domain-specific NLQA system is depicted in Figure 2. The model integrates key components such as Natural Language Processing techniques; Conceptual Indexing based Retrieval Mechanism, and Ontology Processing.

****