Vision and Scope Document

for

Medical Question Answer System

Version 1.0 approved

Prepared by : Espiritu, Immanuel  
Febrer, Jayvee  
Amata, John Patrick

SOCIT DEPARTMENT

10/12/16

Table of Contents

Table of Contents ii

Revision History ii

1. Business Requirements 1

1.1. Background 1

1.2. Business Opportunity 1

1.3. Business Objectives and Success Criteria 1

1.4. Customer or Market Needs 1

1.5. Business Risks 1

2. Vision of the Solution 2

2.1. Vision Statement 2

2.2. Major Features 2

2.3. Assumptions and Dependencies 2

3. Scope and Limitations 2

3.1. Scope of Initial Release 2

3.2. Scope of Subsequent Releases 2

3.3. Limitations and Exclusions 3

4. Business Context 3

4.1. Stakeholder Profiles 3

4.2. Project Priorities 4

4.3. Operating Environment 4

Revision History

|  |  |  |  |
| --- | --- | --- | --- |
| **Name** | **Date** | **Reason For Changes** | **Version** |
| Amata, Espiritu, Febrer | 10/13/2016 | Initial Draft | 1.0 |
|  |  |  |  |

# Business Requirements

## Background

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.

## Business Opportunity

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.

## Business Objectives and Success Criteria

There is currently no established business objective as our QA System -- at its core – is a research project. As our research progresses and directions change, objectives will be added. At the very least, our only success criteria is to build a running QA System that is able to answer simple factoid Medical Questions.

## Customer or Market Needs

The promising values that Medical Q&A systems bring is the quick retrieval of information from a set of documents as queried from a natural language form.

The most commonly used for these needs are information retrieval systems. These information retrieval systems work by returning the list of documents according to user’s query. It is typical for the results to be a voluminous number of text/documents that it frequently makes the doctor review these texts after his clinical hours.

There is a need for a system that can process documents and answers in response to (clinical) questions that are poised by not just these physicians, but also to those doing research work.

**1.5 Business Risks**

The timing of our research is vulnerable to the current studies being done that can affect our whole research. By this we need to take on a regular research reviews for our project to be versatile and reliable at a certain time frame. Existing modern and sophisticated search engines is also a problem we are facing. Our study will be compared to those search engines and of course the line of thinking would be, “why create a simpler system if there is a more sophisticated existing system”. But the main goal of our study is to provide simpler framework that can be used as a basis of starting up a new study that will help the innovation or invention of new ideas in the future. Sophisticated or hybrid systems can be hard to study in the first part of understanding this field of study and that is the one of the main purpose of our study to provide an aid to that problem.

# Vision of the Solution

## Vision Statement

Medical Question Answering System is a software that is designed to aid researchers, physicians, and developers with their line of work. Our software shall address the concerns of information dump frequently experienced by medical researchers and the lack of model frameworks in the Question Answering side of NLP.

Time permitting, Medical Question Answering System will support text summarisation, grammar recognition, word misspelling recognition, online integration, and the use of delimiters.

## Major Features

1. Question Classification
2. Information Extraction Components
3. Lexical Parsers
4. Text Selection
5. Answer Construction
6. Modifiable Corpus
7. Trainable Data

## Assumptions and Dependencies

We are dependent on using the following:

* Stanford CoreNLP
* Apache OpenNLP
* IntelliJ Ultimate IDEA

The system must be compatible to run with low end PCs.

# Scope and Limitations

## Scope of Initial Release

The Initial Release of our product is slated for December 18, 2016. The initial release shall only cover the most important functions and features to serve as a proof of concept. It will be used to gather validated learning and insights for the continued development towards the final product. It shall also serve as a form of market analysis in the field of QA.

This Release (1.0) will include:

* Question Classification: The first stage, this determines the type of answer that shall be provided.
* Text Selection: This module typically uses typically uses several components that apply increasingly complex NLP techniques on a gradually reduced amount of text. The focus of this stage is to identify the documents or paragraphs in the document set that are likely to contain the answer, and a filter preselects small text fragments that contain strings of the same type as the expected answer.
* Answer Construction: This module expands on previous stage -- text selection -- by looking for further clues in the text to determine if the candidate answer can indeed answer the question.

## Scope of Subsequent Releases

Release 1.1 will include:

* Neural Network Integration for Question Classification
* Basic Intelligent Expert System
* Expanded Corpus for Expert System

## Limitations and Exclusions

We shall limit our scope onto examining and experimenting with Factoid Question Answering systems.

# Business Context

## Stakeholder Profiles

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Stakeholder** | **Major Value** | **Attitudes** | **Major Interests** | **Constraints** |
| Physicians | Integration with existing software and their clinical work | Desire a product that is intuitive, maintainable, and customizable | Richer feature set than competitors, cheaper, intuitive, and reliable. | Close-Domain QA |
| Researchers | Consider project as an opportunity to apply academic and expertise | Receptive as resource for project success | Flexible and has all the modules of a complete question answering system | Must run on low-end workstations |
| Developers | Saves time for development as they use the product as a framework | Resistant unless product is keystroke-compatible with current system | Product is well documented, complete, and developed using the appropriate software engineering principles | Must run on low-end workstations |

## Project Priorities

|  |  |  |  |
| --- | --- | --- | --- |
| **Dimension** | **Driver (state objective)** | **Constraint (state limits)** | **Degree of Freedom (state allowable range)** |
| Schedule | Release 1.0 to be available by December 18, 2016 |  | The release date for the next releases are subjected to change from the feedback received on the 1.0 release. |
| Features |  |  | High priority features must be included in release 1.0. These include Question Classification and Answer Construction. |
| Adaptability |  | The created application must be able to expand its corpus of questions and answers. |  |
| Quality |  |  | 90-95% of user acceptance tests must pass for release 1.0, 95-98% for release 1.1 |
| Staff |  | Team size is 3 developers |  |
| Cost |  |  | Dependent on the scale of the project |