Spring 2011

Medical Analysis Question and Answering Application for Internet Enabled Mobile Devices

Loc Nguyen
San Jose State University

Follow this and additional works at: http://scholarworks.sjsu.edu/etd_projects

Part of the Databases and Information Systems Commons

Recommended Citation
http://scholarworks.sjsu.edu/etd_projects/190

This Master's Project is brought to you for free and open access by the Master's Theses and Graduate Research at SJSU ScholarWorks. It has been accepted for inclusion in Master's Projects by an authorized administrator of SJSU ScholarWorks. For more information, please contact scholarworks@sjsu.edu.
Medical Analysis Question and Answering Application for Internet Enabled Mobile Devices

A Master's Project
Presented to
The Faculty of Computer Science
San Jose State University

In Partial Fulfillment
Of the Requirements for the Degree
Master of Science

by
Loc Nguyen
Spring 2011

© 2011
Loc Nguyen
ALL RIGHTS RESERVED
ACKNOWLEDGEMENT

I would like to thank Dr. Tseng for helping, and guiding me throughout the semester in my master’s project. Dr. Tseng’s support has been an important factor that drives my productivity throughout the project. I also would like to thank Dr. Teoh, and Dr. Crunk for being part of my committee and supporting the concept.
ABSTRACT

Mobile devices such as smart phones, the iPhone, and the iPad have become more popular in recent years. With access to the Internet through cellular or WIFI networks, these mobile devices can make use of the great source of information available on the Internet. Unlike a desktop or laptop computer, an Internet enabled mobile device is designed to be carried around and available to the owner almost instantly at any moment of the day. Despite having such great advantage and potential, searching for information with a mobile device remains a difficult task. Mobile device users have to juggle between different search engines through the web browser to seek useful information. The small screen of most mobile devices makes it very difficult to interact with the search engines using the device's web browser. This project focuses on the development of a mobile application that attempts to make information searching easier for mobile device users. In addition, the application is designed to answer search queries that are related to the medical field. Search query in the medical field requires more selective results and that the results come from trusted
sources before it can be used. The application addresses this problem by introducing a database of community knowledge designed to store credible medically related articles that could help answer medical specific search queries. This would help minimize the need to look for answers from the Internet, which often could contain unreliable information. In addition to answering search queries, the application also allows user to store personal medical records and conveniently share that information as needed using email.
TABLE OF CONTENTS

1. Introduction.................................................9
2. Concepts and Methods........................................13
   2.1. Information and Its Representation .....................13
   2.2. Information on the Internet .........................17
   2.3. Reliability of Internet Sources and Community Knowledge Proposal .............................................21
   2.4. Improving the Sharing of Community Knowledge ........24
   2.5. Personal Knowledge .....................................25
3. Application Overview.........................................30
4. Application Workflow: .......................................31
5. Implementation...............................................35
   5.1. Application Screenshots .................................35
      5.1.1. Homescreen ........................................35
      5.1.2. Search Keyword Suggestion .......................36
      5.1.3. Search Results from the Internet ...............37
      5.1.4. View a Search Result ...............................38
      5.1.5. Save a Search Result into Personal Knowledge ....39
      5.1.6. Save Personal Medical Records ....................41
      5.1.7. Send Email of Search Results .....................42
      5.1.8. Question Suggestion Wizard .......................43
5.1.9. Select Type of Returned Results ........44
5.2. Titanium Appcelerator ......................44
5.3. Server Side Scripts and Web Services ........45
  5.3.1. Server Side Scripts ...................45
  5.3.2. Web Services ..........................46
5.4. Technical Specifications ....................48
5.5. Software Methodology .......................49
5.6. Testing ..................................49
  5.6.1. Verify Result Type .....................50
  5.6.2. Verify Search Internet ................50
  5.6.3. Verify Search Community ...............50
  5.6.4. Verify Search Personal ................51
  5.6.5. Verify Send Result By Email ..........51
  5.6.6. Verify Saving Knowledge ...............51
  5.6.7. Verify Question Suggestion Wizard ......51
6. Advantages of the Application ...............52
7. Limitations and Possible Improvements .......55
8. Project Schedule/ Deliverables ...............57
  8.1. Schedule ..................................57
  8.2. Deliverables .............................58
9. Conclusion ...................................59
10. References ................................60
### TABLE OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Figure 1</td>
<td>Application Overview</td>
<td>30</td>
</tr>
<tr>
<td>Figure 2</td>
<td>Cyclical methodology</td>
<td>49</td>
</tr>
</tbody>
</table>
1. Introduction

One of the most successful applications of the Internet is the search engine. A search engine allows users to search through the massive information available on the Internet. With the growing world wide popularity of Internet enabled mobile devices such as smart phones and tablets, mobile device users looking for information from the Internet are challenged by the relatively small screen size, which greatly limits the ability to interact with the search engine and navigate through the search results (Luca, 2005). On the other hand, with the advantage of being instantly available at anytime and anywhere, Internet enabled mobile devices could become the first and only tool that many people use to find information from the Internet. Despite having numerous researches focusing on advancing search engine technologies, there have been limited research on the user interface of search engine for mobile devices. In contrast to searching from a computer where a long list of search results might be sufficient, searching on a mobile device, because of its compact size, would require greater attention to the user interface as well as the search experience as a whole. One way to improve the
mobile search experience might be to change the way information searching is done on a mobile device. Instead of getting thousands of search results from numerous sources all over the Internet for each query, searches from mobile device should be filtered for more reliable sources that would provide more accurate answers. In addition, because a mobile device is generally considered a personal device, i.e. only the owner of the device uses it, search results that have helped answer user question should be saved on the mobile device for future searches.

Medical analysis question answering application for Internet enabled mobile devices is an application that focuses on an intuitive and flexible user interface to help mobile device users quickly find medically related information from reliable sources as well as from the Internet. Using the application, the user will provide search input as either keywords or a natural language question (English). The application processes the question and will attempt to answer the question using information in a personal knowledge database first. If no answer is found, the application will attempt to answer the question using information stored in a community knowledge database, which contains information from reliable sources. Lastly, if no answer is found, the application will submit the
search input to several well-known search engines such as Google, Wolfram Alpha, and Trueknowledge. As the application receives search results from these search engines, it will perform post result processing and combine the results into a format that is suitable for display on a mobile device. The problem with information from the Internet in general is the reliability of the sources. It is important for medically related information to be verified with reliable sources such as personal physician and medical experts before the patient should trust it. The proposed application addresses this issue by allowing users to enter pieces of information they found from the searches into a database and tag them so that the information can be retrieved and verified later with more reliable sources. By tagging search results, the users can also take personal notes of the information found. In addition, as the user tags information found from the community knowledge, the user specified tag is applied back to the corresponding record stored in the community knowledge. The additional tags created by the user will be used as keywords for the corresponding information in the community knowledge hence making it searchable for everyone. In addition to storing knowledge from the search results, the users of the proposed application could also
store and keep track of their personal medical information and have the ability to send personal medical information such as health recording and monitoring data to their personal physicians without having to pay a visit to the doctor’s office.

With the rising cost of healthcare, it is important that everyone should be more in touch with his/her personal medical conditions. With the help of technologies such as Internet enabled devices and the proposed application, patients can keep themselves knowledgeable about the diseases/illness/conditions that they might have while maintaining a personal health record that is most relevant to them. The proposed application can help reduce the cost of healthcare because it would help patients monitor their health records and have quick contact with their primary care physician by sending their records directly to the doctor without waiting for the next visit. In addition, a well-informed, well-educated patient will help save lives and reduce the cost of healthcare for the society in the long run.
2. Concepts and Methods

2.1. Information and Its Representation

Information is an essential part in the process of building knowledge. According to Wikipedia, knowledge is a collection of facts, information, and/or skills acquired through experience or education or (more generally) the theoretical or practical understanding of a subject. (Knowledge, 2011) The preservation of knowledge by sharing and passing on information is important to the evolution of human beings. In the digital age, the Internet offers new ways of communication, and with that new ways of sharing information. Paper textbooks are converted into digital text; paper pictures, diagrams are now created in digital image; classroom lectures are video recorded in digital format. The advantage of storing information in digital format is that it can be produced, transported, searched, and shared quickly at a fraction of the cost of hard copies. The Internet offers a perfect environment for people around the world to create and share these types of digital information. This project focuses on making the sharing of these types of digital information easier on Internet enabled mobile devices and promotes the seeking
and learning of knowledge. As the rest of the document will discuss methods to improve the process of searching, and learning information as well as building knowledge from the information found, I find that having a consistent format that best describes a unit of information will help in the process of creating, searching, and sharing information. The unit must reflect the most common format of information on the Internet, i.e. text, image, and video, as well as other attributes that are related to the information. Consequently, I select a term called "information record" to refer to information as a unit of record with content and attributes that belong to the information.

The following table describes the attributes of an information record:
<table>
<thead>
<tr>
<th><strong>Table 1 Format of an Information Record</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Topic</strong></td>
</tr>
<tr>
<td><strong>Title</strong></td>
</tr>
<tr>
<td><strong>Text content</strong></td>
</tr>
<tr>
<td><strong>Image URL</strong></td>
</tr>
<tr>
<td><strong>Video URL</strong></td>
</tr>
<tr>
<td><strong>Source</strong></td>
</tr>
<tr>
<td><strong>Source URL</strong></td>
</tr>
</tbody>
</table>
Tag is used to describe how the information is related to topic. Tags are created by people who have read or created the information.

Storing text with formatting is a complicated problem. Even though text formatting could play an important role in improving user experience on a mobile question and answering application, it is not a focus of the application. For the purpose of simplicity, minimal formatting is used for information represented as text. The minimal formatting includes plain text with no format except for a special mark up that signals a separation of content. A document usually contains several paragraphs where each paragraph contains several sentences. A paragraph could refer to a subtopic that holds a meaning for itself despite its relative relationship with the topic of the document. A similar statement is true for some individual sentences or groups of sentences within a paragraph. In this project, the occurrence of such a paragraph, group of sentences, or individual sentence, having its own meaning, is considered as separation of content and can be annotated in the plain text by a special
mark up. The special mark up helps the application break down the information quickly to allow the user to take actions on the pieces of information that matter the most to him or her (see 2.5).

For storage conservation and simplicity, the binary content of the image or video file is not stored with the information record. The public Internet URL of the image or video is available instead.

Not all information records will contain a value for each attribute listed above. For example, text-based information will have text content but might not have a value for image URL or video URL. In addition, depending on where the information record comes from and how it has been used within the application, some attributes such as topics or tags might or might not have any significance to the application. Later sections will discuss in more detail when and how these attributes are used.

2.2. Information on the Internet

The project focuses on medical information available to mobile device through the Internet hence it is important to discuss the availability of such information presently on the Internet. The Internet, also called the Web, is a collection of pages of information interconnected by web links. The web links, which are listed on the pages, allow
reader to browse a portion or the entire collection. Each page in the collection generally contains information about a subject. The information is usually presented by the following formats: text, image, or video. In addition to exploring the information through the web links, many users rely on the help of search engines. There are several types of search engines but the most popular are crawler-based search engine e.g. Google Search Engine (Web Search Engine, 2011). In a typical crawler-based search engine implementation, a web-crawler, a computer program that is used to index web pages, is used to visit web pages and their web links to extract meta data and index the web pages based on the meta data found. Meta data are keywords within a web page that best describe the content of the page. Depending on search engine implementation, meta data could either be extracted from keywords specified by the page owner or from keywords found in the title and headings of the document on the page. One of the disadvantages of using a crawler-based search engine to search for information is that the meta data extracted from the page only represents the perspective of the person that created the meta information. In addition, some information that is presented in the document, which could be of interest to someone else, might not be represented by the meta data
because it might not be the main point of the document. On the other hand, crawler-based approach is fast and reasonably effective. For example, the Google search engine returns search results in a fraction of a second; and, even though many results are not relevant, users can change their search keywords and quickly get another set of results until they are satisfied with an answer or simply give up. In this project, a crawler-based search engine is selected as one of the search methods to help answer queries that require information retrieval from the Internet. The crawler-based method is chosen because its advantages could compliment other search methods in finding information that is most relevant to the search queries. The application implemented for this project will use crawler-based search engines that have a publicly available application programming interface (API). Since the application focuses on answering medically related questions, a customized crawler-based search engine is used in which the search engine will only return results from a configurable list of medically related websites.

Despite having been a popular and quite successful solution for the Internet search problem, a crawler-based search engine's simple approach does not satisfy many researchers. In the recent years, there have been several
attempts to promote search engines that can take a complete formal question and return more specific answer rather than a list of web pages. These new search engines are informally classified as answer engines. Some of the answer engines that have gained public notice are Wolfram Alpha and Trueknowledge. The answer engines store information as facts and attempt to answer user's question using those facts (Trueknowledge Technology, 2011). The problem of taking a natural language question and computing a single answer using a database of facts is a complicated problem with many sub-problems including the following: question classification, question processing, context detection, answer extraction, answer formulation, etc (Answer Engine, 2011). Unfortunately, the companies that implemented Wolfram Alpha and Trueknowledge answer engines do not share the details of their technologies; therefore, it is not possible to fully understand how these search engines answer a search query. Through personal testing, I notice that these search engines are only specialized in some subjects and thus would often only be able to answer correctly questions related to those subjects. For example, Wolfram Alpha is specialized in science subjects whereas Trueknowledge can answer more general questions and less scientific ones. In my opinion, the answer engine
is a relatively new technology thus still requires many years of research before it can replace crawler-based search engines. On the other hand, because answer engine works out its answers on facts, I believe that its results might be more reliable than those of crawler-based search engine. For my application, I choose to make use of both crawler-base search engines and answer engines. However, instead of focusing on improving search results from these search engines, I will focus on other aspects to help improve user interactions with the search results. On the other hand, because it is hard to control the contents returned by these search engines while medically related question requires answers that come from reliable sources, I will focus on developing an alternative source of information that can help answer medically related questions. The alternative source for searching is discussed in more detail in section 2.3.

2.3. Reliability of Internet Sources and Community Knowledge Proposal

Many people believe that information from the Internet should be used with caution because a large portion of information on the Internet comes from unreliable sources. This belief has merit since just about anyone can publish information on the Internet. Except for reputable websites
where contents are carefully edited and reviewed before appearing on the sites, many other informational websites such as web blogs, discussion groups, and forums contain information that is posted by individual contributors whose contents are seldom verified for correctness. To address this problem the application implements a scheme that directs users to search results that come from more reliable source before Internet results are used. Such a reliable source is referred to as community knowledge. Community knowledge is a collection of information created by a trusted community. Information from the community knowledge is shared among members of the community who have the same interests. Information retrieved from community knowledge is considered more reliable than that of the Internet because the individual or group of individuals that created the community knowledge is trusted by the community as having more knowledge about the topics that the community is most interested in. The concept of community knowledge is especially useful for applications answering medically related questions because such queries require answers that are based on reliable sources. For example, if the users of the mobile application are medical staff of a hospital, the hospital could create a community knowledge that contains medically related information that
might be useful to the staff. Because it is a mobile application, the medical staff could quickly search for the information using their mobile device while the hospital management could feel more at ease that search queries are answered using community knowledge first.

In the mobile application for medically related questions and answer, information of community knowledge is represented by the format described in 2.1. The application chooses to implement community knowledge with a set of medically related articles extracted from Wikipedia. The content from Wikipedia is chosen because most of its contents are free from copyrights yet the information it contains is widely accepted by the general public. Even though the content of each Wikipedia page is created by individual contributors on the Internet, the standards used by Wikipedia require that information be referenced with a reliable source. As information is manually extracted from Wikipedia pages, information records are created following the format described in 2.1, beside attributes that hold the content of the information, topic and tags are carefully selected because they determine if the information record matches with a search query. Once the information record is created, it can be retrieved with a
search query that contains topic and relating keywords that match the information record's topic and tags.

2.4. Improving the Sharing of Community Knowledge

As discussed in section 2.3, topic and tags are important elements that determine if an information record in the community knowledge is included in the search results. This section discusses ways of improving the user experience in the search for information in the community knowledge.

When information in community knowledge is created, because different people might have different interpretations given the same content, the topic and tags used by the content creator might not be a preferred choice of some users. This could result in user having to modify the search query several times to find a match. This problem is addressed in this project by allowing the user, after having found the information he or she is looking for, to specify a tag that the user would consider as most related to the information. The user will specify this tag when he or she chooses to save the information into personal knowledge, a process that will be discussed in details in section 2.5. The user specified tag would be used in conjunction with the tag chosen by the content creator in subsequent search queries by everybody. Because
user specified tags might be biased or incorrect, the application could put a higher weight on the tags created by the content creator and also provide a way for the content creator to disapprove any tag that is not actually related to the content. Another way to lower the effect of incorrect or biased tags is to favor answers that have higher number of people that tag it using the same tag.

Information represented as text, image, or video can have multiple sections that might contain different useful facts for the users. The tags that the content owner created might not convey every message delivered in an information record. Using user specified tags as described above, where the user can add additional tags for the information record, could help improve information retrieval while limiting the amount of work required by the content owner. As more users specify tags for a given information record, the probability of the users using the same tags is higher. The application can then add more weight to popular user specified tags to help rank the search results.

2.5. Personal Knowledge

Currently, when users search for information from the Internet using a computer web browser, neither the search engine nor the web browser offers any intuitive solution
that allows the users to save the search result on the computer for later use. Although storing the answer found from a search session could be useful for future reference, there might be reasons that such an option is not popular in the current practice. One of the reasons lies in the way popular search engines work. For instance, crawler-based search engines, such as Google Search, do not return a direct answer to a query. In fact, the search engine returns a list of web pages that might contain the answer. In addition, because the Internet is constantly changing, saving links to pages that might contain the answer is not practical; therefore, crawler-based search engines do not promote the saving of search results. On the other hand, in the mobile application for answering medically related queries, the advantages of saving search results into personal knowledge for future reference are clearer and the implementation of such a feature is more practical. As discussed in section 2.4, allowing the user to tag a specific information record in the community knowledge could benefit the community as a whole; however, making the users volunteer to add tags to an information record could be a challenging problem. A more practical approach would be to promote the benefits of saving search results into personal knowledge with personal tagging such as offline
and personalized search capabilities. For example, in the proposed mobile application, when the user finds the answer to his or her search query, the user is given an option to save one or more of the search results into a personal knowledge database stored on the mobile device. To save the information record the user needs to specify at least one tag that describes how he or she interprets the information being saved. If the information being saved is based on an information record from the community knowledge, the user specified tag is also posted back to the information record in the community knowledge, hence helping the community as a whole. Because the user chooses the tag, it is more likely that when the user is looking for the same information again, he or she would key in the same keywords and thus would be able to retrieve the same information record. In addition, since the information is stored locally on the mobile device, the user can expect to get his or her answer even when the mobile device does not have Internet access. Given these benefits, it is more likely that the user will take advantage of the personal knowledge feature while at the same time contribute to improving the searching of the community knowledge.

New information records could be created for personal knowledge from search results that come from personal
knowledge, community knowledge, or the Internet. An information record that is represented as plain text could contain more than one fact that is worth taking note of; therefore, the application has a unique tool to help users quickly extract different parts of the plain text and turn them into new information records for personalized note taking. Besides taking note about an information record, users can use the note-taking feature as a by-product to store personal medical records. For example, in the mobile application, users can store readings of their blood pressure with each reading being recorded as one information record in the personal knowledge database. User can create different types of records such as blood pressure, weight, sugar level, etc, ... Users also have the option to search for these records and conveniently export the records to an email and send them to a doctor for review.

In section 2, I have explained the concepts behind the design of the mobile application for answering medically related questions. I have also discussed at a high level the methods used to implement the application. By design, the application promotes information searching and the process of learning and turning that information into personal knowledge. The design highlights the importance
of community knowledge over information retrieved from the Internet. The application encourages users to save information found from their search results into personal knowledge for future reference while taking advantage of the users' personal tags to help improve the search results of everyone in the community.
3. **Application Overview**

![Diagram of Application Overview](image)

*Figure 1 Application Overview*
4. **Application Workflow:**

The following is a typical workflow of the application:

1. User enters a question from the mobile device's user interface
2. The question is submitted against a personal knowledge database, which resides on the mobile phone. If an answer is found, the result is returned immediately to user.
3. If there is no answer found in the personal knowledge, the question is submitted to a processing server.
4. The server next checks the question against the community knowledge. If an answer is found, it is returned to the mobile device to be displayed for the user.
5. If no answer is found from community knowledge, the question is sent to the Internet through several search and answer engines such as Google Search, Trueknowledge, and Wolfram Alpha. These engines will always return some results. The results are returned to the mobile phone for display.
The application also provides the following functionalities to the user:

1. Suggested keywords: as the user types a search query, the application instantly provides a list of suggested keyword to help the user complete the query. This is helpful when the user has to type long technical terminologies or when the user does not remember the exact spelling of a medical term.

2. Choice of different result types: There are three types of results returned by the application. The user can choose from text only, image only, video only, or everything. This option can be changed in the profile page.

3. A question suggestion wizard: When the user is not sure how to ask a question, the question suggestion wizard is useful as it will ask the user several questions and will suggest some question for the user to use.

4. When the user successfully finds the answer to his or her query, the user is provided an option to save the answer into his or her personal knowledge. When saving the information, the user is required to enter at least one tag to specify the user's own interpretation of the information being saved. In
addition to saving the information into the personal knowledge, the application will also sends the original question, the tag, and a reference of the community knowledge (if applicable) to the server. The server will store this information into the database to be used in the question suggestion wizard. While saving an information record, the user can choose to further break down the information being saved. This way, given an article from the community knowledge, the user can zoom in to only a portion of the article and save only that portion into personal knowledge.

5. Normally, the user can save search results for the purpose of supporting future searches; however, using the same save feature, user could turn the saving record into personal health records. A button named "my" is used to turn the information record into a personal health record with a date stamp. For example, the user can store records of blood pressure measurements daily as well as notes of other measurements. The user can use the search query to obtain all personal health record by starting their search query with the word "my" plus the keyword being the type of record. For example, a query of "What is
my blood pressure" will return all records of user's blood pressure measurements.

6. Finally, with any search results the user can send a cumulative email of all search results to anyone for further evaluation. The user can use this functionality to send search results or medical records to his or her doctor for review.
5. **Implementation:**

5.1. **Application Screenshots**

The following screenshots illustrate a typical flow of usage of the mobile application:

5.1.1. **Homescreen**
5.1.2. Search Keyword Suggestion

The user can click on an entry in the table to have the word completed in the search bar.
5.1.3. Search Results from the Internet
5.1.4. View a Search Result

Title: What Is Cancer?

Mar 19, 2010 ... Cancer is the general name for a group of more than 100 diseases in which cells in a part of the body begin to grow out of control.

See Google
5.1.5. Save a Search Result into Personal Knowledge
Knowledge Breakdown

Mar 19, 2010

Cancer is the general name for...

Cancer is the general name for a group of more than 100 diseases in which cells in a part of the body begin to grow out of control.

Ok  Cancel

Topic

cancer,

Title  What Is Cancer?

Knowledge

Cancer is the general name for a group of more than 100 diseases in which cells in a part of the body begin to grow out of control.

Break  Reset  Clear

Tags

Definition  my

Add to my knowledge  Cancel

Adding knowledge for topic cancer,

Topic

cancer,

Title  What Is Cancer?

Knowledge

Knowledge has been added to your personal record.

Alert

Add to my knowledge  Close

From Community Knowledge

Definition of Cancer

Email  QnA
5.1.6. Save Personal Medical Records
5.1.7.  Send Email of Search Results

To:

Cc/Bcc:

Subject:

Here is some information I'd like to share with you.

Record #1:

Title: Definition of Cancer
Content: cancer (medical term; malignant neoplasm) is a class of diseases in which a group of cells display uncontrolled growth, invasion that intrudes upon and destroys adjacent tissues, and sometimes metastasis, or spreading to other
5.1.8. Question Suggestion Wizard
5.1.9. Select Type of Returned Results

The selection can be made from the Profile tab.

5.2. Titanium Appcelerator

The mobile application is implemented using a third-party tool called Titanium Appcelerator. The tool allows the coding to be done in Javascript programming language then built into platform specific binaries for different platform deployments. In order to write a mobile application for the iOS platform, which is the Operating System for Apple mobile devices such as iPhone, iPod touch, and iPad, developers have to use Objective-C programming language. On the other hand, in order to write mobile
application for the Android platform, developers have to use the Java programming language. Given this scenario, developers targeting both platforms have to implement their application logic twice, once in each of the programming languages. Titanium Appcelerator solves this problem by creating a middleware Javascript API that allows developer to write mobile application in the Javascript language while still targeting both the iOS and Android platforms. Titanium Appcelerator chooses the Javascript programming language perhaps because it is the most popular programming language and is considered easier to learn than Java and Objective-C.

5.3. Server Side Scripts and Web Services

5.3.1. Server Side Scripts

The implementation of keyword suggestion is done using a server side PHP script that reads from a database of common English keywords combined with a database of common medically related keywords. The community knowledge is stored in the server side in a MySQL database where its format is described in 2.1. All data returned from the server side to the mobile device are encoded using the JSON data format.
5.3.2. Web Services

The server is also responsible for querying the Internet when a particular search query cannot be answered using the personal knowledge or community knowledge. There are several web services APIs used to help answer the search query:

5.3.2.1. Google Custom Search API

The Google Custom Search API is an interface to the Google crawler-based search engine. The API returns text results for a given search query. This search API can also be customized to include results from a particular website or list of websites. The API is a RESTful web service where the query can be made with a formatted URL and the data is returned in several choices of formats. In this project, for the matter of simplicity, JSON format was chosen.

This is an example of a custom search query using HTTP Get request:

https://www.googleapis.com/customsearch/v1?key=DEVELOPER-KEY&cx=CUSTOM-SEARCH-ID&q=Cancer&alt=json

DEVELOPER-KEY is a unique key that is associated with the host name of the server where the request is initiated.

CUSTOM-SEARCH-ID is a unique key that is associated with the custom search profile.
5.3.2.2. **Google Image Search API**

Google Image Search API is an interface to Google Image Search. For a given search query, the API returns a list of images that match the search query. The results include URLs of the images but not the actual image files. Using the URLs, the mobile application will download the image file on demand to display on the mobile device. Similar to Google Custom Search API, the data is returned in JSON format.

5.3.2.3. **Youtube API**

Youtube API is an interface to search for Youtube video from the Youtube.com website. The results include URLs of the videos but not the actual video streams. Using the URLs, the mobile application will start the video stream on demand as the user chooses to play the video from the mobile device. The Youtube API can be accessed using a set of PHP libraries developed by the developers of Youtube. The PHP libraries require a part of the PHP Zend Framework.

5.3.2.4. **Trueknowledge Search API**

Trueknowledge Search API is an interface to search the Trueknowledge.com website. Trueknowledge.com is an answer engine rather than a typical crawler-based search engine.
where search query can be submitted as natural language query. The engine includes facts that are used to answer the natural language query. Trueknowledge Search API is generally good for asking general questions. The API can be accessed using a set of PHP libraries developed by Trueknowledge developer. The PHP libraries do not require any special library.

5.3.2.5. *Wolfram Alpha Search API*

Similar to Trueknowledge Search API, the Wolfram Alpha Search API is another answer engine. The Wolfram Alpha focuses on answering question regarding science subjects such as Math and Chemistry.

5.4. **Technical Specifications**

Open Source Library and Application: Titanium Appcelerator

Languages: PHP 5.x, JavaScript

Database: MySQL 5.x

Web Server: Apache 2.x.x

5.5. Software Methodology

In a Cyclical methodology, the software application is started at the analysis phase then follows through the sequence of phases as demonstrated in figure 9. When there is a change to the application such as a feature change, the cycle restarts at analysis phase to make sure that the application logic, design, implementation, and integrity are not compromised by the change. The production of this application follows the Cyclical software development methodology.

5.6. Testing

The application has been tested with various queries in general medical field. Since it is difficult to find a
volunteer tester who is a professional in the medical field, the application has only been tested by general testers whose search queries are considered general medically related questions. A small set of articles from Wikipedia on medical diseases were chosen and broken down into smaller sections to be used as medical community knowledge for logic evaluation purposes only. The following specific test areas were performed:

5.6.1. **Verify Result Type**

There are three types of results: text, image, and video. User can choose to have one of the result types or all. Verify that the user's selection is correctly reflected in the search results.

5.6.2. **Verify Search Internet**

Enter a query that cannot be answered by personal or community knowledge. Make sure the Internet returns with some result. Internet should always return some results.

5.6.3. **Verify Search Community**

Enter a query that can be answered by community knowledge. Make sure the results are relevant to the search query.
5.6.4. **Verify Search Personal**

Enter a query that can be answered by personal knowledge. Make sure the results are relevant to the search query.

5.6.5. **Verify Send Result By Email**

For personal, community, or Internet result, make sure that clicking the email button will display the email dialog with correct result content in the email. Make sure the email is correctly sent to intended party.

5.6.6. **Verify Saving Knowledge**

For personal, community, or Internet result, make sure user can save the result to personal knowledge. Verify that if the user does not specify a topic or tag then the entry is not recorded and error message is shown. Verify that the information break down button works correctly as it breaks down the text by paragraph and/or sentence.

5.6.7. **Verify Question Suggestion Wizard**

Activate the wizard and make sure that the user has to specify at least a topic before some suggested questions are shown. Verify that when the user selects a suggested question, the application immediately processes the question.
6. Advantages of the Application

Information retrieval is not a new problem. On a computer, users can already search for most information by using various search engines. With the advantage of being mobile and always on, a mobile device can be a great tool for information searching. Even though there are a number of applications for searching and information retrieval on a mobile device, there has not been a mobile application that focuses on the searching, learning, and personalizing the information as the application discussed in this paper. Most of the available mobile applications are mobile version of the equivalent web version. For example, Google Search mobile application is a copy of the Google Search website; Wolfram Alpha mobile application is a copy of the Wolfram Alpha web version. Medically related application such as WebMD for mobile provides a subset of the information available on the WebMD website with no additional feature. There are also many other medically related mobile applications of different sizes with different features on various medically related topics; however, there is no mobile application that allows user to search for information from reliable sources like community
knowledge while at the same time allows searches from the Internet. There is also no mobile application that focuses on helping the users keep track of the information that they searched. Some of the advantages of the implemented mobile applications over other available mobile applications are:

- The application would benefit mobile device users who have to use the mobile web browser to search multiple search engines. The search results for Internet search in the application are combined and displayed in a format easy for mobile device user. In addition, some search engine can be customized to limit the searches from a set of reliable websites.

- The application allows user to search for information from a more trusted source: the community knowledge. The community knowledge could be set up by an organization that the mobile user is associated with. It could contain specialized knowledge that best serves the interests of mobile users in that community. Community knowledge search results having precedence over Internet search results will help produce more related results.

- The mobile device user can store search results into his or her personal knowledge. The user can express
his or her own understanding of the knowledge by the use of personal tagging rather than relying on the keyword chosen by the content provider. User can also keep record of their personal medical record and share those records as needed with their medical advisors.
7. Limitations and Possible Improvements

The effectiveness of the application relies on the accuracy of the information that it is provided.

For search results from the Internet, while some search engines such as Wolfram Alpha and Trueknowledge work best when they are given a search query in the form of natural language question, other crawler-based search engines such as Google work best when the search query includes the search keywords that the user is looking for. By combining the search results from different types of search engines, the application takes advantage of the benefits from all search engines. On the other hand, the combined list of results could be overwhelming for the user, hence having a method to rank the search results would help the user save time looking at several search results before finding the best one.

For search results from community knowledge, the accuracy of the result relies on the accuracy of the tags that are associated with the community result. There are two ways for tags to be added to a community result: when the information is added, with tags, into the community knowledge by the community administrator and when search
users choose to save the information, with a tag, into personal knowledge and the information came from community knowledge. Considering that the community administrator is the owner of the community content, this combined tagging approach incorporates the view of the content owner with those of the users. The drawback of this approach is that users could pollute the tagging system by adding irrelevant tags to information. On the other hand, if we have a larger community of responsible users, incorrect tags would not affect the decision of the system because it would still pick the tags that many users have voted on.

A community is often more effective when it is small. As a community grows, it is harder to manage, as it is harder to define the common interests among a larger group of users. Depending on the type of community, the optimal community size could differ. A similar statement could be made of community knowledge. As the community knowledge grows to include more information on various topics, subdividing the community knowledge to its optimal size might be necessary to ensure its search effectiveness. It is then necessary to implement the mobile application to support hierarchy of community knowledge databases and easy switching between the databases.
## 8. Project Schedule/ Deliverables

### 8.1. Schedule

<table>
<thead>
<tr>
<th>Date</th>
<th>Work Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Week 1-4</td>
<td>Design application's architecture and workflow. Determine core features of the application</td>
</tr>
</tbody>
</table>
| Week 5-13 | Implemented the application with core features  
  a) Developed the keyword suggestion script. Built a database containing keywords for the suggestion.  
  b) Implemented server side scripts to use the Image Search API and Video Search API. Other search engine APIs were already done in previous works (CS297). All search APIs are combined into a single web service.  
  c) Developed the database for community knowledge and the server side script to make the data available as web service.  
  d) Developed the application for mobile device. The application provides user with an interface to enter search query and |
examine search results returned by the web services.
e) Developed the functionality to allow user to add search results into personal knowledge. Created a database on user's mobile device and added logics to allow searches from personal knowledge.
f) Implemented all other minor features that help with the workflow of the application.

| Week 13-14 | Test, debug, and optimize the application |
| Week 15    | Prepare project report and project defense |
| Week 16    | Present the application and project defense |

8.2. **Deliverables**

1. An application for a mobile device that allows user to search for information from personal knowledge, community knowledge, and the Internet. User can save search results and share them with others by email.

2. CS298 Final Project Report
9. Conclusion

This paper presents a solution for mobile device users who have the need to search for information from their own knowledge, from a group of users with similar interests, or from the Internet. The implementation application focuses on search queries that are medically related; however, the concepts described in this paper can be applied to applications that are built to answer general questions or questions from fields other than medical. To make the application work with search queries in other fields, community knowledge of such fields need to be built and the user can toggle between the fields to get the answers most relevant to their search field.
10. References


