Chapter 6

Software Requirement Specification

6.1 Introduction

OncoSys is developed to make the Oncologists work easy by automating the manual method of capturing patient information during their visit. Also organizing and indexing the scattered knowledge available among a group of Oncologists in a hospital. The software should be designed to capture patient therapy information from Oncologists and demographic information from clerks of the hospital. It should also include documents uploading, search engine with crawler which parses and indexes contents of these offline documents or online web documents. Crawler should be authenticated before indexing the contents of any website. This software also provides customization of crawler by specifying the number of levels of scanning. Crawler should display necessary user friendly messages while it is in progress. OncoSys should also provide case base wherein various statistical data of patients over a period of time should be available to Oncologists. It enables Oncologists to update their medication
and treatment policy. These statistical data should be in printable form. The software should be able to handle the most complex tasks in friendly manner.

6.2 Product Features

6.2.1 General Description

OncoSys is a web application to create and manage cancer patient’s health information and knowledge base with the search engine for Oncologists at cancer hospital. It also includes case base to present various statistical reports done on patient database.

6.2.1.1 Purpose

The purpose for developing this type of product is to facilitate Oncologists at cancer hospital.

6.2.1.2 Scope

The scope of OncoSys is within a hospital i.e. it should be accessed through intranet within the hospital. Registered users must be able to login their accounts, after approved by administrator, accessing the OncoSys application and signing with their username and password anytime.

6.2.1.3 Abbreviation

OncoSys - System for Oncology

6.2.1.4 Overview

As the OncoSys is able to have a user interface, it should have menus to navigate across the site. A profile form has to be provided to store each
individuals profile and load them when the user signs in to the application. It should provide the oncologists the upload tool to upload the contents and allow adding URL’s to scan and index for search engine. As an administrator, it should allow them in monitoring user activities. As a developer, it should make user interface which is user friendly.

6.2.2 Overall Description

6.2.2.1 Product Perspective

OncoSys will be integrated with Hospital Information System (Cancer Hospital) thereby making it a Decision Support System for Oncologists of cancer hospital. OncoSys should be able to provide a basic and easy way to collect data and search their content. It should be able to capture the patient’s visit to hospital and retrieve their past history if required. It should have a knowledge base and facility to write articles with formatting tools available to all the Oncologists.

6.2.2.2 Product Functions

The following are the functions of the OncoSys:

Patient Information  OncoSys will provide web based forms to capture patient health information from the therapies given. These forms will be processed and served as the history of the patients treatments over a period of time.
Table 6.1: Patient Information

**Knowledge Base:** The system will provide UI for Oncologists to upload and unload documents like doc, pdf, ppt etc. Indexing these documents will be done automatically. It will be provided with features to send and receive message and write custom articles or notes.

<table>
<thead>
<tr>
<th>INPUT</th>
<th>Title, Description, FilePath</th>
</tr>
</thead>
<tbody>
<tr>
<td>PROCESSING</td>
<td>Data is stored in FilesIn table and file is uploaded to local server</td>
</tr>
<tr>
<td>OUTPUT</td>
<td>Added Successfully</td>
</tr>
</tbody>
</table>

Table 6.2: Knowledge Base

**Case Base:** The system will provide statistical data from Hospital Information System in the form of report. It will provide options to print and download them as image. User provides the two dates i.e. from and to period, to view these statistics.

<table>
<thead>
<tr>
<th>INPUT</th>
<th>FromDate, ToDate</th>
</tr>
</thead>
<tbody>
<tr>
<td>PROCESSING</td>
<td>Retrieve statistics from HIS</td>
</tr>
<tr>
<td>OUTPUT</td>
<td>Graphical Charts</td>
</tr>
</tbody>
</table>

Table 6.3: Case Base

**Web Crawling:** The system will provide a web crawler. It can create an index of supplied URL’s. A base URL is provided as a starting point. After it
starts it gets the text from each document it reaches and also the hyperlinks if it is a webpage. These hyperlinks can be put in a queue after verifying if they point in the same domain. These links will be provided to the crawler to visit them. The crawler should not be limited to web documents only i.e. it should read the text and index pdf, doc, ppt and other popular files.

<table>
<thead>
<tr>
<th>INPUT</th>
<th>URL, Depth Level</th>
</tr>
</thead>
<tbody>
<tr>
<td>PROCESSING</td>
<td>Start crawler and index the document found at URL</td>
</tr>
<tr>
<td>OUTPUT</td>
<td>Indexing Completed</td>
</tr>
</tbody>
</table>

Table 6.4: Web Crawling

**Word Extracting:** The system will provide string splitting and replacing. After the web crawler index is the pages, the text of each page has to be extracted from that page. String splitting and replacing method is used to extract the content words after removing symbols and notations.

**Create Index:** The system will provide functions to create an index. Using the extracted words from a string parser, create index for searching. The main attributes are Word, FileID, Position, Count

**Ranking:** The system will provide a ranking function for search results and order them.

**Searching and Ordering:** The system will provide search functionality for users. It will also have support for AND search and searching the offline contents available. The search results from search function are not in order. Ordering function can reorder the results by ranking or user rating.
### Software Requirement Specification

**INPUT** | Keyword, OrderBy, SearchType
---|---
**PROCESSING** | Search keyword from index table and send in order required
**OUTPUT** | Search results in ordered manner

| Table 6.5: Searching And Ordering |

**Sending Message:** The system will provide facility to communicate by providing text messaging feature. So the users of system can exchange information among them.

<table>
<thead>
<tr>
<th>INPUT</th>
<th>Sender Username, Message Content</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>PROCESSING</strong></td>
<td>Store in database</td>
</tr>
<tr>
<td><strong>OUTPUT</strong></td>
<td>Message Sent</td>
</tr>
</tbody>
</table>

| Table 6.6: Sending Message |

**Notes:** Each Oncologist can write notes or articles and store them in the database. These pages will also be indexed. These pages will be dynamically generated on request to view. It includes various formatting options for article.

<table>
<thead>
<tr>
<th>INPUT</th>
<th>Title, Description, Content</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>PROCESSING</strong></td>
<td>Store in database</td>
</tr>
<tr>
<td><strong>OUTPUT</strong></td>
<td>Notes added</td>
</tr>
</tbody>
</table>

| Table 6.7: Notes |

**Administrator tools:** The system will provide a set of tools for administrator to manage users, notes, messages and files indexed. Apart from these features it should also support registration, login, changing account settings.
6.2.2.3 User Characteristics

The target users are administrators and Oncologists. Each oncologist can have only his/her username and password after registration which has to be approved by the administrator. This prevents unauthorized access and hacking the product. The administrator is responsible for monitoring all database activities and also answers the user queries. Oncologists add and search contents available online and offline. They need not know the working of the crawler so UI is provided to provide seed for indexing.

6.2.2.4 General Constraints

The following constraints existed to limit the project development:

**Human Resources**- As this project is carried out at an individual; it is difficult to build a search engine which competes with Google, Yahoo, MSN and other famous search engines.

**Hardware** - The search time is very short for famous search engines. Most of them can output search result within 0.5s. They are capable of such great performance due to their high computing processors.

**Time for researching** - The success of search engine also results from the algorithms used, hardware’s used etc. This project’s development time is about half a year and therefore the time for discussion and implementation is less.

6.2.2.5 Assumptions and Dependencies

OncoSys should work even if the local server is not connected to internet and it is assumed that the search keywords includes only English alphabets.
6.2.3 Specific Requirements

6.2.3.1 External Interface Requirements

**User Interface:** The external users are the Oncologists and administrator of the hospital. The Oncologists can access their account to upload or unload contents, search a query, write articles, send messages to other Oncologists etc. The administrators have an account to monitor documents or files uploaded, managing users etc. Web forms are provided as UI to interact with users.

**Hardware Interface:** The interface for accessing the OncoSys is personal computers of Oncologists and of the administrator. The PCs may be laptop or desktop connected to local server. Apart from these no other hardware is required.

**Software Interface:** MS SQL 2005, a database server to store the information. Web Server i.e. IIS, to serve the user requests from their browsers like Mozilla Firefox, Internet Explorer etc. End users operating system can be any version of Windows, Linux or any other which supports TCP/IP protocols.

**Communication Interface:** The communication interfaces is a local area network connected to local hosting server

6.2.3.2 Performance Requirements

The PCs used must be at least Pentium 4 machines so that they can give optimum performance of the product.
6.2.3.3 Design Constraints

The constraints at the designing time are that the Oncologist’s needs at hospital may change over a time so the design should be modular and hence updateable.

6.2.3.4 Attributes

The following are the attributes of OncoSys

- Files uploaded should be easily downloadable
- It should facilitate updating the account of Oncologists

6.2.3.5 Security Requirements

OncoSys should allow only registered and approved users to use the functionalities of system. User privacy should be maintained. Users should not be allowed to use their account whenever they are not approved or disabled.

6.2.3.6 Other Requirements

As the knowledge is increased, search engine should still search and display results faster removing duplicate links.

6.3 Process Model

This application development is a mixture of XP and TDD. Extreme Programming (XP) is a software development methodology which is intended to improve software quality and responsiveness to changing customer requirements, as a type of agile software development. Other elements of extreme programming include: programming in pairs or doing extensive code review,
unit testing of all codes, avoiding programming of features until they are actually needed. This method is used to obtain simplicity and clarity in code, and when we expect changes in the customer’s requirements as time passes, it makes better understanding of the problem, and frequent communication is kept with the customer. It does not advocate death march work schedules, but instead working at a sustainable pace.

Test-driven development (TDD) is a software development technique that relies on the repetition of a very short development cycle: first the developer writes a failing automated test case that defines a desired improvement or new function, then produces code to pass that test and finally refactors the new code to acceptable standards.

**Design Phase**  The design is developed based on SRS and general features of search engine. In this phase, sub processes of a particular process are highlighted. The impact of a change in a processes/subprocesses is explored. None of the process / subprocess possess the expected functionality at this stage. A component should be independently deployable and should encapsulate its constituent features. The components should be maintained with clear specifications of what it provides and the inputs it requires.

**Implementation and Testing**  Pseudo coding phase will be introduced after the design phase. This pseudo code can be sent to the client, reviewed by him and his comments can be incorporated in the pseudo coding stage itself. All the expected functionality of the process/sub-process is incorporated into the prototype. As the impact analysis of every process/sub-process is done, those features, which are highly interrelated, are implemented at the outset and later those features whose dependence on other features is less are developed. Thus the development is modular. Developer level testing has to
be done for each process before exposing the feature to the actual prototype. Since the customer doesn’t have any idea about the features or advantage of search engine, this application has to be built as a standalone application where most of the search engine features are incorporated by looking into the problem.

6.4 Development Approach

Top Down Development approach is followed to build this software. According to this approach we have to consider OncoSys as a whole which is divided into three subsystems

- Patient Information
- Knowledge Base
- Case Base
- Search Engine

The Search engine involves several sub processes. It can also be used independently with any other system. It includes following sub process

- Crawler
- Robot Checker
- Text Extractor
- Ranking the keywords
- Indexing
- Searching
• Ordering the search results

Once the crawler is started by providing base url; Robot Checking, Text Extracting, Ranking and Indexing of keywords are done automatically. Once they are indexed Searching can be done on this indexing and they can be ordered per requirement.

6.5 Duration

Duration: 1.5 Years