Development of Web Portal  
For

Patent Records

Software Requirements Specification

13-01-2017

* Tapish Pratap Singh (150001035)
* Shivam Tayal (150001034)
* Manish Chavare (150001008)
* Shivam Parashar (150001033

Prepared for

CS 258 Software Engineering

Spring 2017

**Table of Contents**

1. Introduction i

1.1 Purpose i

1.2 Scope i

1.3 Overview i

2. General Description ii

2.1 Product Perspective ii

2.2 Product Functions ii

2.3 User Characteristics ii

2.4 General Constraints ii

2.5 Assumptions and Dependencies ii

3. Specific Requirements ii

3.1 External Interface Requirements iii

3.1.1 User Interfaces iii

3.1.2 Hardware Interfaces iii

3.1.3 Software Interfaces iii

3.1.4 Communications Interfaces iii

3.2 Functional Requirements iii

3.2.1 Extraction of information from PDF files iii

3.2.2 Building database from extracted information iii

3.2.3 Creation of excel file from database iii

3.2.4 Development of web-portal with various sorting options iii

3.3 Classes / Objects ii

3.4 Non-Functional Requirements iii

3.4.1 Performance iv

3.4.2 Reliability and Availability iv

3.4.3 Security iv

3.4.4 Maintainability iv

3.4.5 Portability iv

3.5 Design Constraints iv

3.6 Logical Database Requirements iv

# 1. Introduction

This document is a Software Requirement Specification (SRS) for the **Web Portal for Patent Records**. This is the initial draft for the SRS and it will be used for the extensions.

The aim of this project is to extract the information from Patents records which are in PDF format and then write the same in a .xlsx format file and in a SQL database. Under this project

we will build a web portal which will allow to search for data in the SQL database using various filters and download the information in a excel file format.

## 1.1 Purpose

The purpose of this document is to specify complete description about out project. Through this document, the workload needed for development, validation and verification will ease. To be specific, this document is going to describe functionality, external interfaces, performance, attributes and the design constraints of the system which is going to be developed. Therefore, intended reader groups for this software requirement specification are client and users.

## 1.2 Scope

The project is intended to :

*(1)* Reduce the efforts of searching information in patent records.

*(2)* Provide patent record information in excel format for further uses.

*(3)*  Provide a user-friendly web-portal for searching and downloading required information.

## 1.3 Overview

We are going to focus on describing our project in terms of project perspective, project features, user characteristics and dependencies on the following section of this document. Next, we will address specific requirements of the project, which will enclose requirements of the system, performance requirements, and other requirements.

# 2. General Description

This section gives background information about specific requirements of our project to be developed in brief. Although we will not describe every requirement in detail, this section will describe the factors that affect the final product.

## 2.1 Product Perspective

This software product is eventually intended for the research work. Product will be deployed to web portal and all users of the product will access the product through web portal. Web portal will be main user interface where users can operate all the provided functionality. However, this web portal will be only a part of our project. The main conversion and extraction will be done off-line on the main system for building up the database. Web portal will only be the interface for the user data and the execution of provided functionalities.

## 2.2 Product Functions

This project will extract the information from Patents records which are in PDF format and then write the same in a .xlsx format file and in a SQL database. There will be a web portal which will allow to search for data in the SQL database using various filters and download the information in a .xlsx file format.

## 2.3 User Characteristics

The user is expected to be able to use basic functionalities of .xlsx files. The user is expected to be Internet literate. The user is expected to be Windows literate and to be able to use button, pull-down menus, and similar tools.

## 2.4 General Constraints

The extraction of information from PDF files is format dependent and will work only on some specific formats, so if format of patent record is changed, conversion may not work properly.

The tools thats are used for conversion of PDF files to HTML files during extraction process are open-source tools compatible with Linux based operating systems(as of now).

## 2.5 Assumptions and Dependencies

* This project uses some open-source tools such as PDFTK(<https://www.pdflabs.com/docs/pdftk-cli-examples/>) and PDF2TXT([http://www.unixuser.org/~euske/python/pdfminer/index.html#pdf2txt](http://www.unixuser.org/~euske/python/pdfminer/index.html" \l "pdf2txt)).
* The extraction part of project requires tools that are compatible with Linux based operating system, so this functionality of project will work on Linux-based operating system(as of now).
* The user must have a web browser(Google-chrome, Mozilla-firefox etc) installed on the system to access the web-portal.

# 3. Specific Requirements

## 3.1 External Interface Requirements

### 3.1.1 User Interfaces

The user interface for the project shall be compatible to any browser such as Internet Explorer, Mozilla or Netscape Navigator by which user can access the web-portal. The user interface shall be implemented using any software package like HTML, CSS,Python and PHP etc.

### 3.1.2 Hardware Interfaces

No additional hardware is required.

### 3.1.3 Software Interfaces

The software will use PDF2TXT and PDFTK (Linux based tools) for conversion and extraction of information. This software will use various python modules like

* XlsxWritter for writing information to xlsx files
* Url2lib for reading HTML files.
* Commands for controlling terminal.
* Sqlite3 module for managing database.

### 3.1.4 Communications Interfaces

The user must have access to Internet for accessing web-portal.

## 3.2 Functional Requirements

### 3.2.1 Extraction of information from PDF files

### 3.2.2 Building database from extracted information

### 3.2.3 Creation of excel file from database

### 3.2.4 Development of web-portal with various sorting options

## 3.3 Classes / Objects

Nothing(as of now)

## 3.4 Non-Functional Requirements

### 3.4.1 Performance

No requirements(As of now)

### 3.4.2 Reliability and Availability

The database will be available both on-line and off-line and backup will also be created and stored on GITHUB and Google-drive.

### 3.4.3 Security

No security requirement (as of now).

### 3.4.4 Maintainability

The project is dependent on pattern of information in pdf files, so if new pattern is found in future, required changes must be implemented.

### 3.4.5 Portability

The web-portal can be accessed from any machine having Internet connectivity and web browser .

## 3.5 Design Constraints

No design constraints(As of now).

## 3.6 Logical Database Requirements

No Requirements(As of now)