

**ÇANKAYA UNIVERSITY FACULTY OF ENGINEERING
COMPUTER ENGINEERING DEPARTMENT**

**Project Report
Version 1**

**CENG 407 Innovative System
Design and Development I**

201901 PATENT COMPARISON SYSTEM

Özlem Simay CELKAN

201511009

Neval Zeynep GÜNDEŞLİ

201511027

Öznur ÖZTÜRK

201511045

Mert KIZILDAĞ

201411032

Advisor: Dr. Öğr. Üyesi Murat SARAN

Table of Contents

Abstract	4
Özet.....	5
1.Introduction	5
1.1 Motivation	5
1.2 Company Background	5
1.3 Problem Statement and Solution Statement.....	6
2. Literature Search.....	6
2.1 Introduction	6
2.2 What is Patent and How to Get a Patent?	8
2.3 What should be done before the Application?	8
2.3.1 Processes	8
2.3.2 Fees (as of November 2019)	8
2.4 Patent Searching	9
2.4.1 USPTO Patent Database	9
2.4.2 Espacenet.....	9
2.4.3 CIPO	9
2.4.4 Google Patents.....	9
2.4.5 Free Patents Online	10
2.5 Natural Language Processing	10
2.6 Rapid Keyword Extraction (RAKE).....	10
2.6.1 What is Rapid Keyword Extraction	10
2.6.2 How does Keyword Extraction work?	10
2.7 Database	10
2.8 Mapping Patent Classifications	11
2.8.1 Methods	11
2.8.2 Data	11
2.8.3 Distance Measures.....	11
2.8.4 Clustering	12
2.8.6 Portfolios	12
3. Summary	12
3.1 Technology Used.....	12
4. Software Requirements Specification	13
4.1 Introduction	13

4.1.1 Purpose	13
4.1.2 Scope of Project.....	13
4.1.3 Overview of the Document	13
4.2 Overall Description	14
4.2.1 Product Perspective	14
4.2.2 Patent Characteristic	14
4.3 Requirements Specification.....	14
4.3.1 External Interface Requirements	14
4.4 Functional Requirements.....	15
4.4.1 Profile Management Use Case	15
4.4.2 Options Menu Use Case for Setting.....	16
4.5 Non-functional Requirements	17
4.5.1 Performance Requirement	17
4.6 Software System Attributes	17
4.6.1 Portability	17
4.6.2 Performance	18
4.6.3 Usability	18
4.6.4 Adaptability	18
4.6.5 Scalability	18
5. Software Design Document(SDD)	18
5.1 Introduction	18
5.1.1 Purpose	18
5.1.2 Scope	19
5.1.3 Overview of Document	19
5.1.4 Motivation	19
5.1.5 Design Approach	19
5.2 Architecture Design.....	22
5.2.1 Profile Management	22
5.2.2 Options Menu	23
5.2.3 Class Diagram	23
5.2.4 Activity Diagram.....	24
5.2.5 Database Diagram.....	25
5.3 Use Case Realization.....	25
5.3.1 Brief Description of Figure	25
5.3.1.1 Environment Design	25
5.4 Environment	26

5.4.1 Modelling Environment.....	26
6. Survey Result	31
7. Conclusion	34
8. Acknowledgement	35
9. References	35

Figure Of Table

Figure 1 Patent Procedure	7
Figure 2 Formulas.....	11
Figure 3 Profile Management Use Case.....	16
Figure 4 Options Menu Use Case	17
Figure 5 Design Approach.....	21
Figure 6 Class Diagram	23
Figure 7 Class Diagram	24
Figure 8 Database Diagram	25
Figure 9 Home Page	26
Figure 10 Comparison Page	27
Figure 11 Login & Register Page	28
Figure 12 Create Report Patent Page Design	29
Figure 13 Search Patent Page Design	30
Figure 14 Survey Result Question 3.....	31
Figure 15 Survey Result Question 4.....	31
Figure 16 Survey Result Question 5.....	32
Figure 17 Survey Result Question 6.....	32
Figure 18 Survey Result Question 7.....	33
Figure 19 Survey Result Question 8.....	33
Figure 20 Survey Result Question 10.....	34

Abstract

This project aims to increase the quality of the patent search process in terms of time and accuracy in the researches conducted in the Turkish Patent Institute. This project also aims to enhance the limited comparison features of existing patent comparison systems by developing an alternative, more advanced patent comparison system. In this project, we first investigate the extensive Patent Comparison literature to see the current status in the patent comparison field.

Özet

Patent, bir buluş için buluş sahibine devlet tarafından verilir ayrıca sahibinin izni olmadan başkalarının buluşu üretmesini, kullanmasını veya satmasını belirli bir süre boyunca engelleme hakkı verir. Yani her bir patent kendine has yenilikler içermelidir. Bir buluş başka bir buluşu taklit etmemeli veya ona belli bir orandan fazla benzememelidir. Bu projede yapılmak istenen karşılaştırma sisteminde, patent sahibi kişi yayınlanmış patent içerikleriyle kendi patent içeriği arasında bir kıyaslama yapabilecek ve bir benzerlik oranı bulabilecektir. Böylece kendi patent içeriğini korumuş olacak ve olabilecek risklere karşı hakkını korumuş olacaktır.

1.Introduction

The aim of this project, to ensure that the patent owner finds a similarity rate between his own patent and other patents registered in the system.

1.1 Motivation

As 4th-grade computer engineering students, we did much research for our graduation project. One of the main reasons we chose the project was a project that people could easily use and make their lives easier. Since we believe that we can use web-based programming and search algorithms effectively, we decided to develop our project in these aspects. Our biggest dream as a team is to develop and bring this project to better places after graduation. Forcing our limits and seeing what we can do and preparing ourselves for the future is our greatest motivation.

1.2 Company Background

ONUR; Was established in 1980 and has been serving the defense industry and public institutions of our country for more than 35 years. In the beginning years, ONUR started to gain experience in the defense industry by cooperating with consortia and in the following years, it has developed its integration ability by working with foreign companies in various projects. ONUR is one of the few companies capable of working as the main contractor providing the needs of the Turkish Armed Forces, the Defense Industry Presidency and the State Airports Authority, especially in the field of air traffic control systems. There are 5 main areas of activity: Voice Communication Systems, Air Traffic Control Systems, Network Supported Talent and Communication Systems, Cyber Security Systems and Meteorological

Systems.ONUR aims to be one of the major actors in the world by focusing on export opportunities in the fields of civil aviation and defense industry.[1]

1.3 Problem Statement and Solution Statement

There are too many patent data in the patent databases. Therefore, it takes time to search among the available patents and find similarity rates. It is necessary to make another grouping among the existing patents other than the IPC value to shorten the search time. Since the criteria chosen to find similarity rates vary from person to person, the similarity rate to be found may not carry a 100 percent accuracy rate. To overcome this problem, we need to improve the search algorithm with future feedbacks.

2. Literature Search

This literature contains information that we think is necessary to understand the progress of the project. We gave information about what to do before and after the patent application. We have listed the most preferred sites from patent search sites. We have provided information about how to search these sites and about their content. You can also find useful articles about patent classification in our literature research.

2.1 Introduction

The invention or invention is the result of a human effort of a product that has not been previously found. That is, each invention has its unique features. The most convenient way for an individual to protect his invention is to provide patent protection for the invention. For an invention to be patentable, the inventor must apply to a patent office and wait for a specified period. Those who want to patent applications in Turkey should apply to the Turkish Patent Institute. The procedure for patenting an invention is shown in Figure 1.

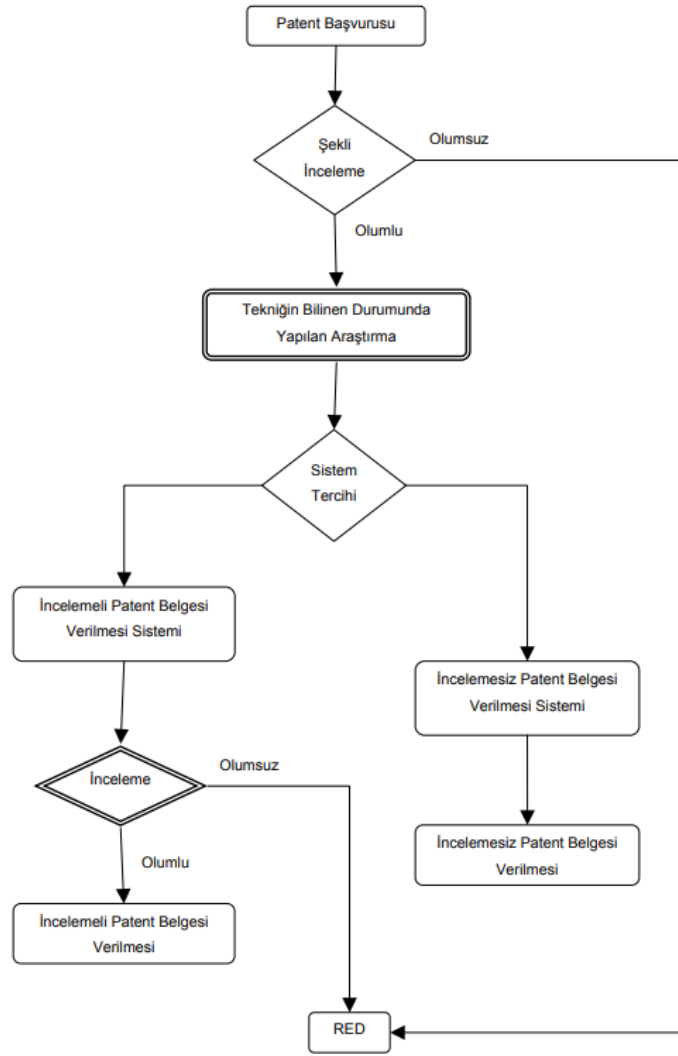


Figure 1 Patent Procedure

There are IPC values that indicate what the patent contents are. In other words, the IPC value indicates which class the patent belongs to. The aim of our project is to find out which patent belongs to which class and to compare whether there is another patent with the same content. Patent contents need to be cleared from meaningless words (and, or, and etc.) to find more meaningful results. We use one of the natural language processing libraries to clean our data. We thought that our search algorithms would work slowly because there are too many patent contents available in our data. Therefore, we wanted to make a classification other than IPC value to get faster results. In making this classification, we use phrases from patent contents. We use rapid keyword extraction to prepare these phrases. It can be used in multiple different methods for the classification process. Clustering is an example of this process.

2.2 What is Patent and How to Get a Patent?

A patent is a right granted to an inventor by the federal government that permits the inventor to exclude others from making, selling, or using the invention for a while. patent is the guarantee of the rights of the person who invented an invention or a new product. To put it more clearly; the person making or inventing the invention is entitled in the fields of sale, reproduction and marketing of the invention or product, and no one other than the owner of the product may sell a product or a similar product. In short, all rights on the product are granted to the owner and secured by the patent. The patent system is designed to encourage inventions that are unique and useful to society. Congress was given the power to grant patents in the Constitution, and federal statutes and rules govern patents. It is necessary to understand the steps that constitute the process to answer the question of how to obtain a patent. The formal institutions of the process of obtaining a patent to pay fees and fill in application forms and writing technical specifications and official institutions and review relevant articles from conveying to it are to respond to the report. We can examine the process of obtaining patents in 3 main titles. These are financial, technical, and bureaucratic.

2.3 What should be done before the Application?

First, an expert patent attorney does preliminary research on the product, which is claimed to be an invention. If the preliminary research assistant decides it is an invention, he or she receives detailed documentation and pictures of the product. The patent attorney duly writes the patent description and sends it to the inventor for examination. The inventor makes suggestions on this description or makes necessary corrections/additions/subtractions/modifications and sends them to the patent attorney after the patent description is completed with the inventor.

2.3.1 Processes

- MAKE AN APPLICATION
- RESEARCH REQUEST
- MAKING THE FIRST REVIEW REQUEST
- MAKING THE SECOND REVIEW REQUEST
- MAKING THE THIRD REVIEW REQUEST

2.3.2 Fees (as of November 2019)

- Application fee: 45 TL
- Research fee: 590 TL
- Examination Fee: 355 TL
- Patent Certificate Fee: 350 TL

2.4 Patent Searching

The sites mentioned below allow you to search for patents using specific phrases. On the TPE website, in the comparison system provided to users, it is possible to search for published patents according to many different criteria (the title of the invention, applicant, IPC class...). As can be seen, no application allows finding similarity between patents.

2.4.1 USPTO Patent Database

USPTO has its patent database. This resource is free to use. USPTO offers the full text of U.S. patents and patent applications from 1790 until the present.

You can search by between 1790 to 1975 [2]:

- date
- classification number
- patent number

You can search by from 1976 until the present:

- patent number
- date of issue
- technical features in the patent
- keyword
- name of inventors
- name of the company
- date of application

2.4.2 Espacenet

The official database of the European Patent Office (EPO). EPO is the largest free patent database with more than 90 million patents. You can download the full-text part of patents [1].

2.4.3 CIPO

The Canadian Intellectual Property Office (CIPO) checks the Canadian patent database. In CIPO you can search the contents of patents that have been available in French or English since the 1870s. Patent searches can be made using the following.

- keywords
- number of patent document
- Date of patent

2.4.4 Google Patents

Google Patents contains more than 8 million full-text patents. Google Patents work with the same logic as the Google search engine. Google Patents, users USPTO, EPO, etc. Is a worldwide patent search engine that allows you to search for patents. Google Patents makes several suggestions to give better results. Furthermore, the easy-to-use interface of Google Patents and the quick download of patent documents in PDF format are the biggest advantages.[6]

2.4.5 Free Patents Online

Sign up for an account at [Free Patents Online](#) and get access to:

- U.S. patents and applications
- European patents
- WIPO documents

You can only make quick searches for US patents,. An advanced search includes foreign patents. You can save and download patents that you find [3].

2.5 Natural Language Processing

The first thing we can do with the Natural Language library is to recognize which language a text is written in. It shows what languages our text is like in different percentages. One of the most important features of the library is to separate the text into words. Later, it allows grouping by using some classes in the system. [8]

2.6 Rapid Keyword Extraction (RAKE)

2.6.1 What is Rapid Keyword Extraction

Keyword extraction (also known as keyword detection or keyword analysis) is a text analysis technique used to automatically find the most important phrases and expressions in a text. It helps to summarize the contents of a text and to recognize the main headings. Keyword extraction allows you to find the most important words from large data sets within seconds.

2.6.2 How does Keyword Extraction work?

Keyword extraction is a text analysis technique that simplifies the task of finding the most important words and phrases in the text. One of the advantages of keyword extraction models is that they are easy to install and implement. There are different techniques you can use to perform automatic keyword extraction. From simple statistical approaches to finding keywords by counting word frequency, to more advanced machine learning approaches that allow you to build more complex models.

2.7 Database

The existing data set to be used in this project is quite large. In addition, new data will be added to the data set periodically. Therefore, we wanted to use a database that can work more efficiently and faster with the large data set. Based on our research and the advice of our consultant, we focused on two different databases. These;

- PostgreSQL
- MongoDB

Both databases had similar and active properties. However, as a result of our internet research, we found that mongoDB provided the most effective and fast search. That's how we decided to use MongoDB in our database.

2.8 Mapping Patent Classifications

There are many patented products and inventions in the world. To facilitate access to patented products, we should group similar products. There are many forms of construction for patent comparison. To facilitate this benchmarking and comparison process, we can classify patents in-house. In this article, patent classification and portfolio analysis are explained. This article also revises the parameter selection of patents.[2]

2.8.1 Methods

There are many methods for patent classification. These are data, distance measurements, clustering and portfolios. We will explain these methods briefly.

2.8.2 Data

Each of the patents contains different or similar data sets. Therefore, we can include more than one class when classifying a patent. In this way, we can easily access data sets and patent track.

2.8.3 Distance Measures

In this project, two different equations were created to find similarity measures from 654 different data set classes. One was used to calculate the similarity of the Jaccard and one of the Cosine. These formulas.

$$Jaccard = \frac{\sum_{i=1}^n x_i y_i}{\sum_{i=1}^n x_i^2 + \sum_{i=1}^n y_i^2 - \sum_{i=1}^n x_i y_i} \quad (1)$$

$$Cosine = \frac{\sum_{i=1}^n x_i y_i}{\sqrt{\sum_{i=1}^n x_i^2} \sqrt{\sum_{i=1}^n y_i^2}} \quad (2)$$

Figure 2 Formulas

The Jaccard matrix allows the specified patents to be counted in binary and numerical values are used to calculate cosine values. The cosine is a measure of proximity in the vector space and thus provides the measure of distance.

2.8.4 Clustering

Both the Jaccard and the cosine matrices were decomposed using the routine `decomp. exe.` classification is made through top-down hierarchical clustering. Jaccard and cosine values generated based on the 2-mode matrix of patents versus classes can be very small since these matrices are sparse. During the decomposition, for example, VOSviewer saves files with fewer decimals than the cosine values so that these values are rounded to zero. This inflates the modularity and generates small clusters.

In order to avoid this effect, we multiplied all values by 1000. For the global map, this makes no difference in the case of cosine values and hardly any difference for a map based on Jaccard values; but the larger values improve the decomposition because the additional zeros otherwise increase the number of small clusters. However, the larger values can no longer be used for distance measurement because they can be larger than one. For this reason, we use the cosine between aggregated citations among classes as the proximity measure in the routines about portfolio management.

2.8.6 Portfolios

Overlays in maps can be used for portfolio management by science-policy makers and R & D management. These maps have different characteristics and colors, making it easier for us to cluster and classify. In addition, difference maps can be used to compare the portfolios of different analysis units.

3. Summary

3.1 Technology Used

To design a web interface in our project; html, bootstrap technologies will be used. Our application will be designed as a single page. Our website will also be compatible for mobile phones. For database we will use MongoDB which is good in terms of performance in big data. We will use the Python programming language to write our algorithm and make it the main purpose of our program. We will write this language using the Anaconda IDE.

4. Software Requirements Specification

4.1 Introduction

4.1.1 Purpose

The purpose of this document is to provide information and details about the project to be conducted on patent comparison. Patent comparison will be developed in our project to find similarities of different products with multiple features and to prevent the repeated use of the same names. The most important goal here is that companies will be able to follow up on their own without having to pay extra fees for patent follow-up.

4.1.2 Scope of Project

There is currently no program to compare the patent. Our goal is to make the patent classification based on the title and content. Patent comparison relates to comparing a patented product to a product intended to be produced. Patent Analysis and Comparison is indispensable for preventing the negative material - spiritual consequences that may be encountered later and for correct R & D structuring. In our project, we aim to program with Java or Python. Once our data set is provided by our company, our first job will be to process this data into the database. In doing so, we will explore several classification and clustering algorithms. If we want to give a few examples; k-means, Bayes Classifier ex. Thanks to the structure we have established, we aim to carry out our transactions quickly and accurately.

4.1.3 Overview of the Document

The second part of the document describes functionalities of the Patent Comparison System. Since we use a large data set in this project, we decided to choose a NoSQL database. NoSQL databases use a variety of data models to access and manage data, including documents, charts, key-values, in-memory, and search. Such databases are, specially optimized for applications requiring large data volumes, low latency, and flexible data models. We decided that we need an e dedicated server to store the our data set. Java and python will be used as software languages. We are continuing our research on software language.

4.2 Overall Description

4.2.1 Product Perspective

Patent Comparison System is a web project aiming to compare patents and give a certain ratio and reporting. The Patent Comparison System project will be implemented using technologies such as machine learning, language processing and search algorithms.

4.2.1.1 Development Methodology

Successful projects are managed well. Therefore, it is very important that we choose the methodology that will work best in our project. While developing the project, we have decided to use Scrum which is an agile software development methodology. Scrum; is one of the project management methodologies and it is used to manage complex software processes. In performing this management, it split the whole and follows a method based on repetition. It provides that the target is achieved through regular feedback and planning. It has a structure that is flexible for needs and open to innovations. Communication and teamwork are very important. The one most advantage of scrum is that reviewing each sprint before moving to another that testing is conducted throughout the process, so permits teams to change the scope of the project at whatever point.[5]

4.2.2 Patent Characteristic

4.2.2.1 Registration

We will use the database in the registration system of this project. We will create a database with the data sets. In order to search the system, we need to categorize the data in the database. This way, we will be able to easily access the data in our registration system and compare the similarity rate.

4.3 Requirements Specification

4.3.1 External Interface Requirements

4.3.1.1 User interfaces

Our software will be able to work actively on all web-based platforms.
What you can do in the user interface:

- Register to the system,
- Login to the system,
- Search for patents,
- Update Patent database,
- Create Patent report,
- Create Simple Patent report (for Non-members)

Unlike the user, the administrator will be able to make the features listed below.

- Test and train the system,
- Can comment on files uploaded by the user,
- Add data to database,
- Will be able to access the information uploaded by the user,
- Can data statistics in uploaded files

4.3.1.2 Hardware interfaces

- Enough free space for database size.

4.3.1.3 Software interfaces

- We will develop a web-based software interface during the testing phase of this project.

4.3.1.4 Communications interfaces

- There are no external communications interface requirements.

4.4 Functional Requirements

4.4.1 Profile Management Use Case

Use Case:

- Register
- Login
- Sign Up
- Validation
- Exit

Diagram:

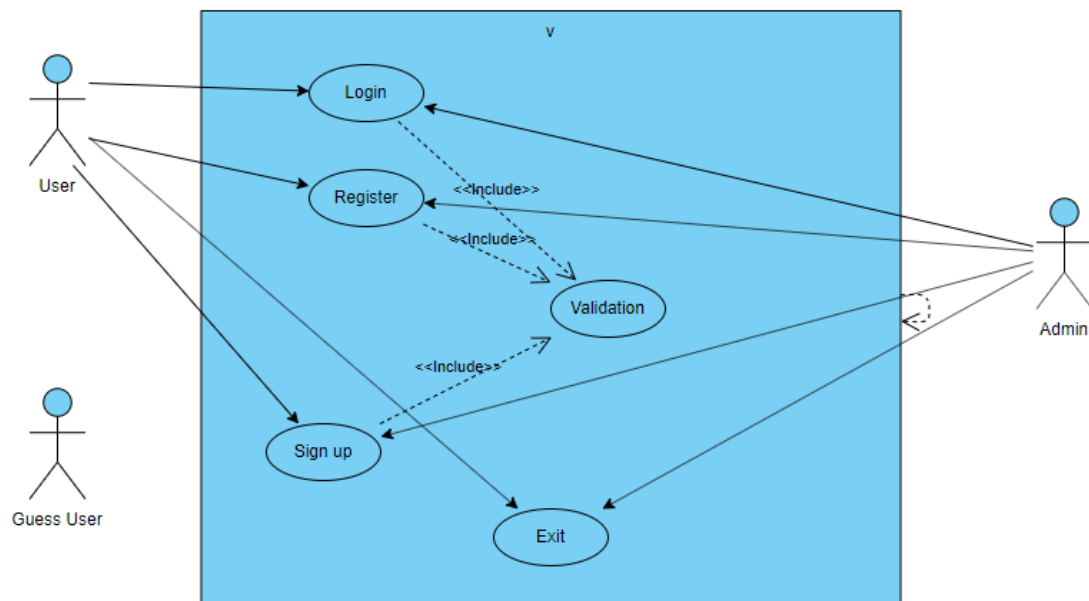


Figure 3 Profile Management Use Case

4.4.2 Options Menu Use Case for Setting

Comparison Mode Use Case

Use Case:

- Login
- Enter Patent Name
- Search Patent
- Create Patent Report
- Create Simple Patent Report

Diagram:

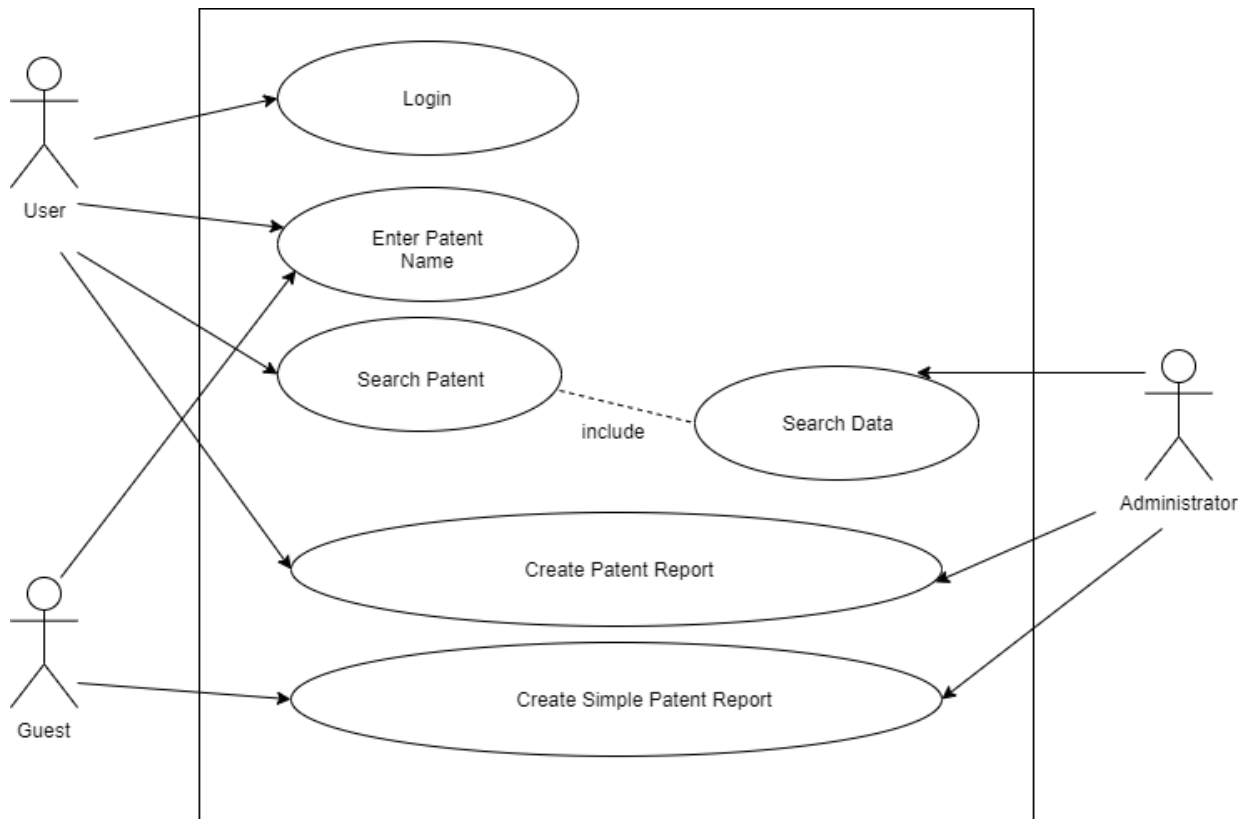


Figure 4 Options Menu Use Case

4.5 Non-functional Requirements

4.5.1 Performance Requirement

The large size of our data set will significantly delay the user finding patent similarity rate. So we're going to need a quick search algorithm in this project. In order to give the user similarity rate in a short time, we need a quick search algorithm in this Project. We also need to create word classes. According to these classes, it will be easier to find out which category the patent belongs to.

4.6 Software System Attributes

4.6.1 Portability

- Patent Comparison System will be developed as a web application by using Java or Python programming language. So, our project will be available at all platforms like Mobile Phones, Tablets and PCs having an Internet browser application.
- The system is a web based project and it not require different hardware for users.
- All computers and phones will be available to use our system.

4.6.2 Performance

- Patent name search should be fast.
- The user should not be able to obtain patent names previously obtained.
- The data should be classified well in the database.

4.6.3 Usability

- If the participant does not fill the places that need to be filled, he/she will receive an error message.

4.6.4 Adaptability

- Patents must be in English language.

4.6.5 Scalability

- Many participant uses the system at a time, there is scalability requirement.
- The system may generate a similarity ratio in line with the criteria we set.

5. Software Design Document(SDD)

5.1 Introduction

5.1.1 Purpose

The purpose of this Software Design Document, "a program capable of Contents Benchmarking Benchmarking Patent" is entitled to provide the details of the project. Target audience, people who want to own and have all invention. The system will create opportunities to compare Patent contents. The program will serve to show us the similarity rate is another patent of the patent. The aim of the Patent Comparison System project is to design the "Content Comparative Patent System" in order to find the similarities of the contents, both innovative and different from other comparison programs, taking into account the currently used Patent comparison programs. The system will operate in two main modes. The first is "User" mode and the other is "Guest" mode. In User mode, the user will be able to compare the patent he / she wants to compare in detail and see the similarity ratio. It can also generate a patent report or search for patents. In Guest mode, the user will be able to perform limited operations on the system without being a member. In order to provide a better comprehension, this SDD includes various diagrams such as UML diagram of the project, activity diagram and block diagram.

5.1.2 Scope

This document provides a complete description of the design of the Patent Comparison System Project. Procedures for creating, programming, and designing the system environment through Java or Python programming languages will be performed. It is an object-oriented language in two languages. Java's use ranges from embedded devices to mobile phones, from corporate servers to supercomputers. We decided to use Clustering Algorithms for the operation of the system (K-Means, Neural Networks ex.) We will design our system in English. When the patents are published in Turkish patent institute, our system cannot read it because it is in pdf format. That's why we will convert those documents and save them in our database. In finding similarity, we will set limits for ourselves and act within that framework. First of all, after the completion of these parts, we intend to do the patent valuation, which is the main purpose of our project.

5.1.3 Overview of Document

The remaining chapters and their contents are listed below.

Section 2 is the Architectural Design which describes the project development phase. Also it contains class diagram of the system and architecture design of the simulation which describes actors, exceptions, basic sequences, priorities, preconditions and postconditions. Additionally, this section includes activity diagram of scenario generator.

Section 3 is Use Case Realization. In this section, a block diagram according to the SRS document is shown.

Section 4 is related to Environment. In this section, we have shown the sample frames of environment from the prototype and have described scenario.

5.1.4 Motivation

As 4th grade computer engineering students, we have done many researches for our graduation project. One of the main reasons we chose the project was that it was a project that people could use easily and make their lives easier. As we think that we can use web based programming and search algorithms effectively, we decided to develop our project in these aspects. Our most important source of motivation is effective in our project.

5.1.5 Design Approach

In this project, we chose SCRUM to carry out our projects. SCRUM is a step-by-step software development model that enables you to build a plan successfully. Agility is just one of the features that we emphasize in this methodology. When we consider the enormous dimensions of modern software projects, scrum methodology is a method that is in the foreground in most projects in order to facilitate this complex structure and reduce the future difficulties of planning the project at the beginning. Scrum methodology approaches this problem with three principles and tries to reduce the general problems that may occur in the project.

- Agile Development
- Scrum Methodology
- Agile Process in the Project

Transparency: Ensures that processes and improvements made during project operation can be recorded instantly and presented to other developers and customers to be followed by all project managers.

Inspection: The desired products and functions in the project are delivered within a specified time period (usually in large projects, this process runs for 2 weeks) and can be evaluated before the end of the project. This is an important opportunity to prevent any mistakes at the beginning.

Adaptation: In projects, requests can vary continuously. At the beginning of the project, if it is done according to the requirements, it will affect both the producer and the customer in terms of time and budget. In Scrum technology, re-evaluation is made after each delivery and the process continues according to the adapted requests. Thus, improvements can be made more easily through the raw project. The ultimate goal in Scrum technology is to produce the desired product in the shortest possible time and with the highest quality. Keeping these difficult factors together is another important factor that increases the likelihood of being the most appropriate application development method for teamwork. Scrum process model is formed on three different roles. These; The product is divided into the owner, software development team and Scrum master. The software development team is also divided into three different roles. These are; a developer, tester and designer. In the category of external roles or side roles, there are customers, businesses, users and marketing.

Owner: User requests and status suggestions are shared with the software development team in detail. Sends requests to the team within 2 weeks. We call these weeks Sprint. **Software Development Team:** Shares requirements details, requested and obtained data, through which timetable the project will be performed, result data with the product owner.

Scrum Manager: The person who acts as a bridge between Software Development Team and Product Owner. It is concerned with keeping the product at a certain standard level, dealing with negative situations that may occur during the scrum process, and removing obstacles to the team. Works with the Software Development Team but is not subject to the team. During the meetings, the desired product is developed within the framework of the project. Figure 1 represents four sprints of the project. The product functions specified during the planning period are delivered to Sprint. The Software Development Team's work is shaped according to the Task Board and progresses according to the user requests specified by the Product Owner.[2]

SPRINT 1	SPRINT 2	SPRINT 3
NEEDS FOR SEARCH ALGORITHM	GUI DESIGN	DATABASE DESIGN
1. Data Editing	1.Login Panel	1.Create Table
1.1 Converting data from pdf to appropriate format	2.Register Panel	2.Deciding relationships between tables
1.2 Create clustering set	3. Options Panel	3.Add data to database
1.3 Create a training set	3.1 Guest	
2. Create Word Tree	3.2 User	
2.1 Create a tree for similar meaningful words	4. Search Panel	
2.2 Create a tree for synonyms	5. Report Panel	
2.3 Creating trees for prepositions and conjunctions	5.1 Similarity Rate	
3. Testing	5.2 Other Patents	

Figure 5 Design Approach

5.1.5.1 Scrum Board with Tasks

This section summarizes our overall responsibilities in this project. It consists of three parts. These; to-do list is a list that briefly summarizes what has been done and what has been done.

To Do

- A survey will be prepared to be given to employees at TPE
- Presentation report will be prepared
- Prepare word trees
- Design draft of website will be created.
- Logo design will be made for our project.
- Data will be processed into the database.
- Text will be prepared for the presentation.
- the data set will be cleared of meaningless words.

In Progress

- Website is being designed.
- Trying on logo.

Done

- Literature Review.
- SRS
- SDD.
- Project Web Page.
- Research and Collection of Information for Software Requirements Specification.
- Project Work Plan (Will be prepared as Github wiki page).
- Meeting with the Company and Detailing the Project.
- Opening a Github account and Creating a Project Group.
- Survey was prepared.

5.2 Architecture Design

5.2.1 Profile Management

Summary: It is a system used by the user and the administrator. The user can enter, save and update personal information and exit the system, compare patents, view reports and search for patents. In addition, an administrator can delete an account, approve participant accounts, and add a new administrator.

Prerequisite: The user must turn on the system.

Basic array:

- If an account is not available, the user must register.
- The user must login to the system by entering the username and password.
- The user can update their personal information by selecting the update button from the user menu.
- User can search through patent search button.
- User can compare patent using patent comparison button
- The user can view their reports.
- The user can browse the site without being a member.
- The administrator can delete a user account by selecting the delete button from the admin menu.
- The administrator can confirm a recently registered user account by selecting confirm from the admin menu.
- The administrator can add a new administrator by selecting the Add new administrator button from the admin menu.
- The user can exit the system by selecting the exit button.

Exception: Database connection may fail.

Message Conditions: None

5.2.2 Options Menu

Summary: The user can perform transactions on the system with or without a member. The user can see the similarity percentage of his searches with other patents.(Kullanıcı sistem üzerinde üye olarak ya da üye olmadan da işlem yapabilir. Kullanıcı yaptığı aramaların diğer patentlerle benzerlik yüzdelerini görebilir.)

Actor: User

Prerequisite: The user must be able to select the search from the options button.

Basic Sequence:

- The user can search and become a member.
- The user can select the continue button without being a member.
- The user can view the search results.

Exception: None

5.2.3 Class Diagram

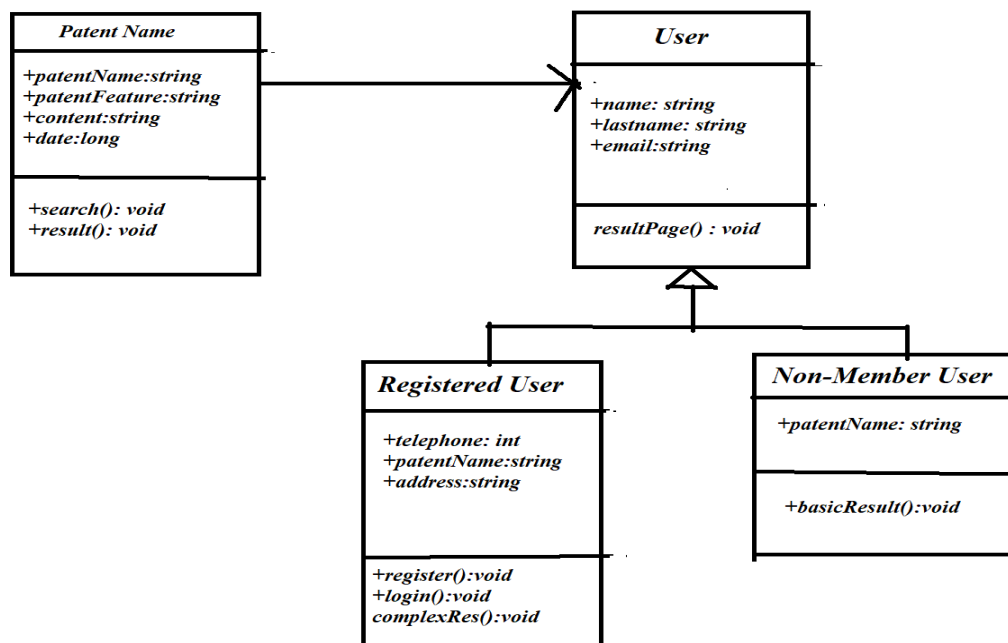


Figure 6 Class Diagram

5.2.4 Activity Diagram

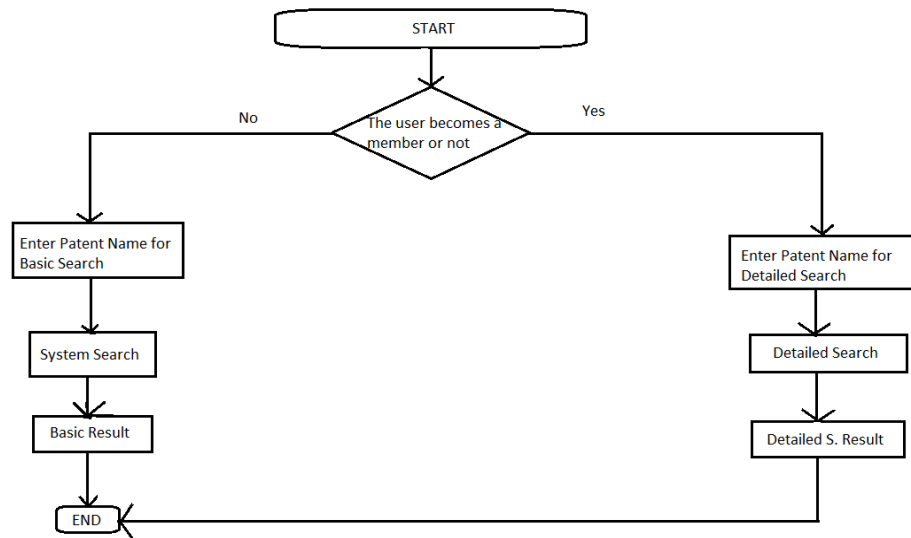


Figure 7 Class Diagram

This diagram shows how users will use it. The user can continue with or without a member. The user enters the search for the patent name he wishes to investigate. System searches according to similarity ratio. Then these rates are listed.

5.2.5 Database Diagram

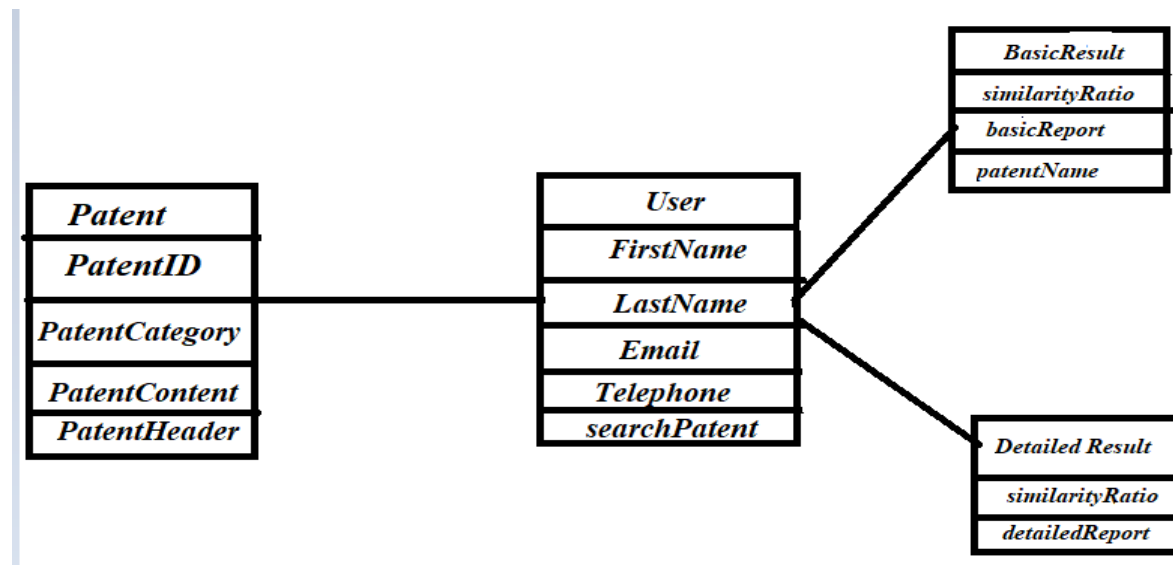


Figure 8 Database Diagram

5.3 Use Case Realization

5.3.1 Brief Description of Figure

The components of the Patent Comparison System are shown in Figure X. All designed parts of the system are shown in the block diagram in the figure. The system has five main components with its own subsystems.

5.3.1.1 Environment Design

Environment Design is responsible for managing environment which the user interacts with objects.

5.4 Environment

5.4.1 Modelling Environment

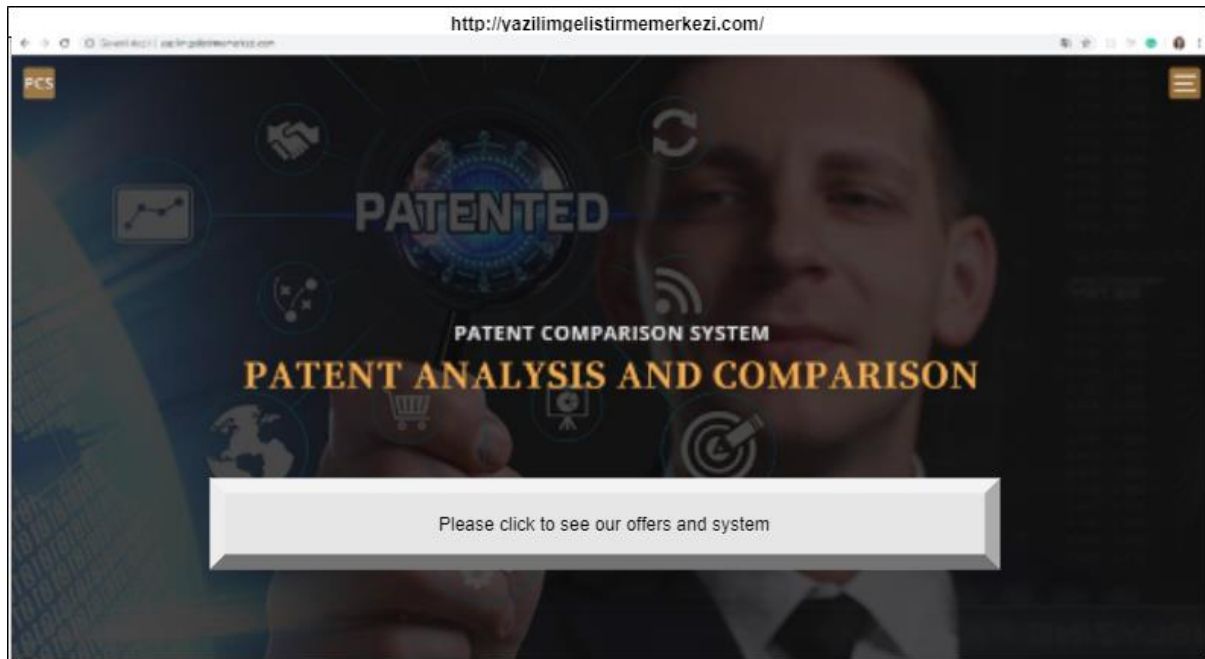


Figure 9 Home Page

We have chosen to include titles on the homepage that may be useful to inform the user about the patent. The button at the bottom of the page directs the user to the page for patent comparison

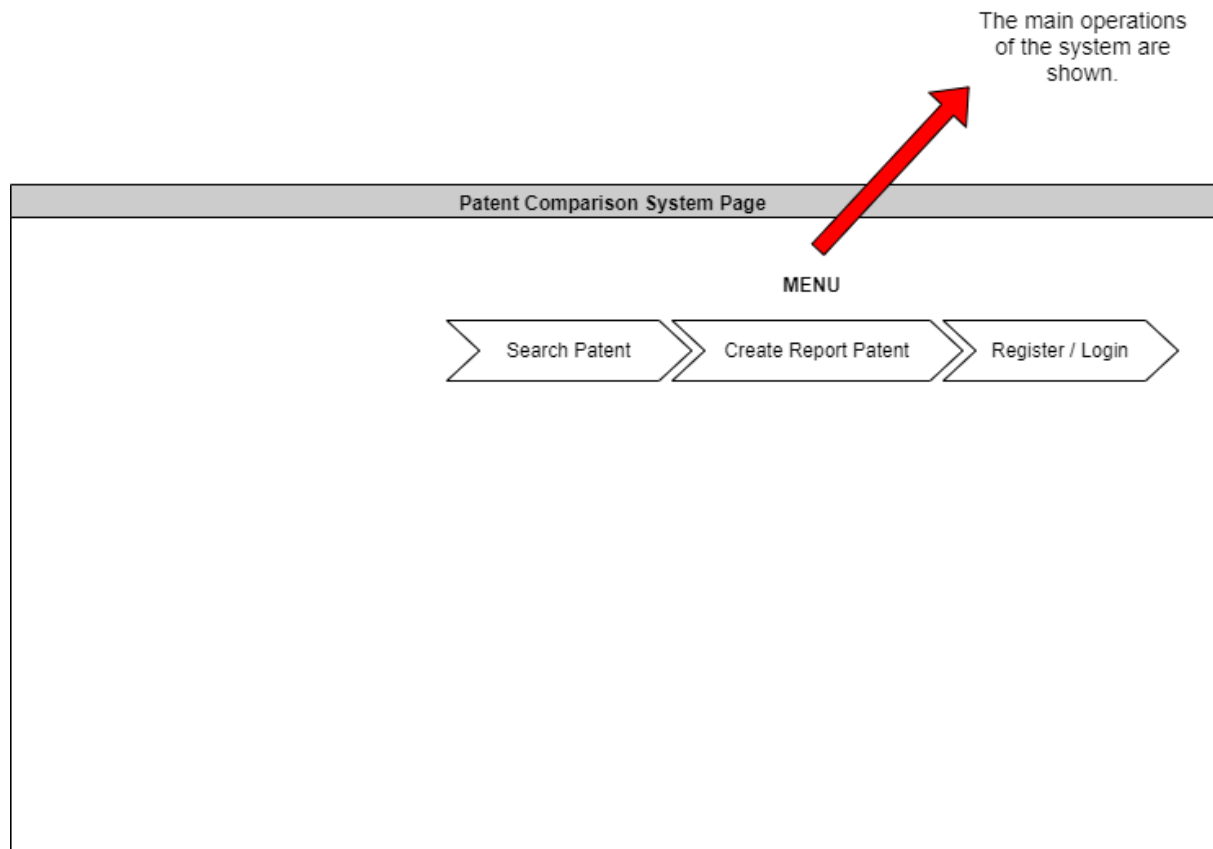


Figure 10 Comparison Page

We wanted to show Search Patent, Create Report Patent and Register / Login pages as the main features of our program in the menu bar. We also chose to include a page that describes the patent process and links that may be useful.

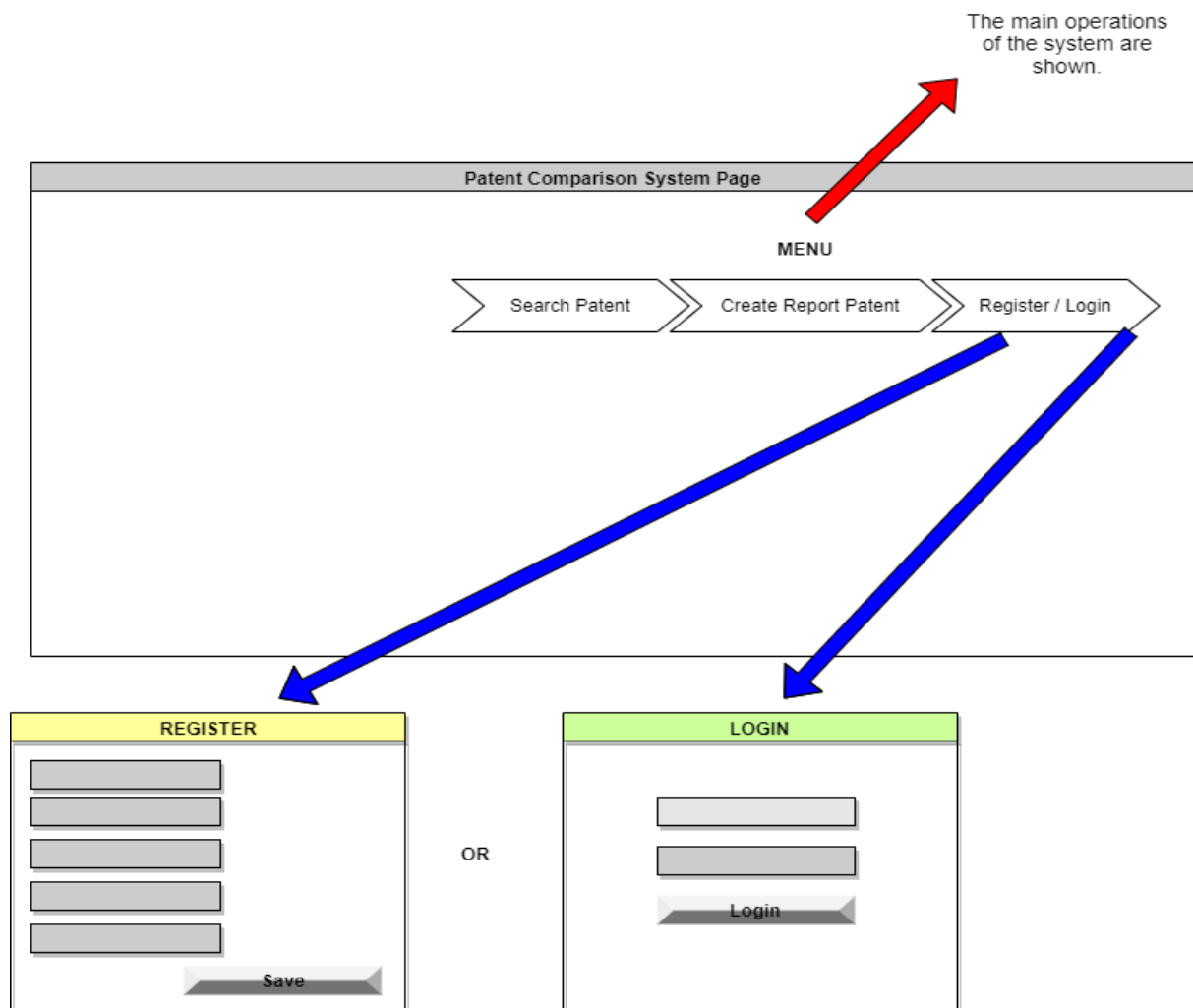


Figure 11 Login & Register Page

The user has 2 options to use the application. You can become a member or continue as a guest. Members have more privileges than guest users.

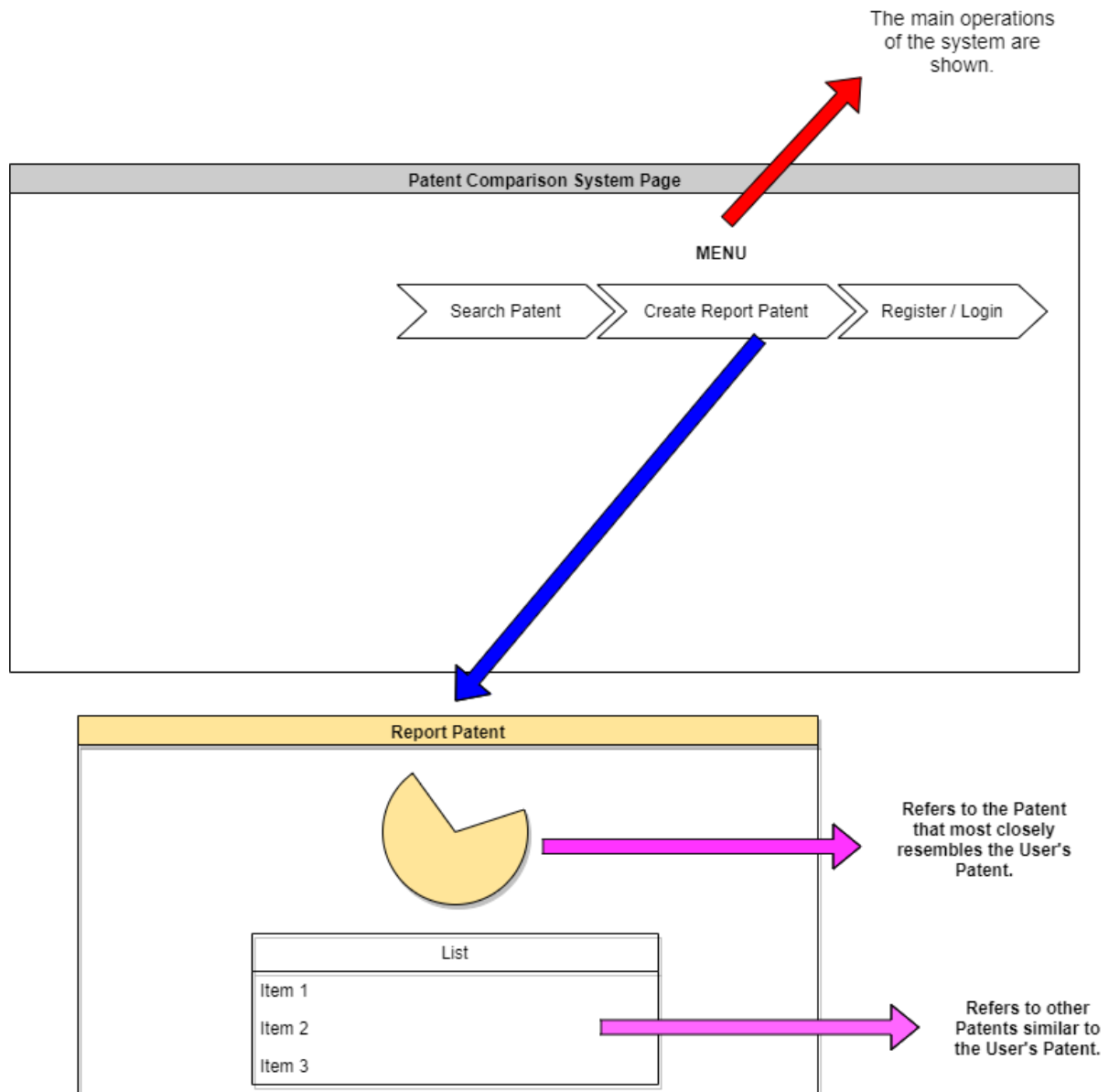


Figure 12 Create Report Patent Page Design

In this section, we will see the similarity rates of the patents we have compared. At the same time, the system will report to the user about similarities.

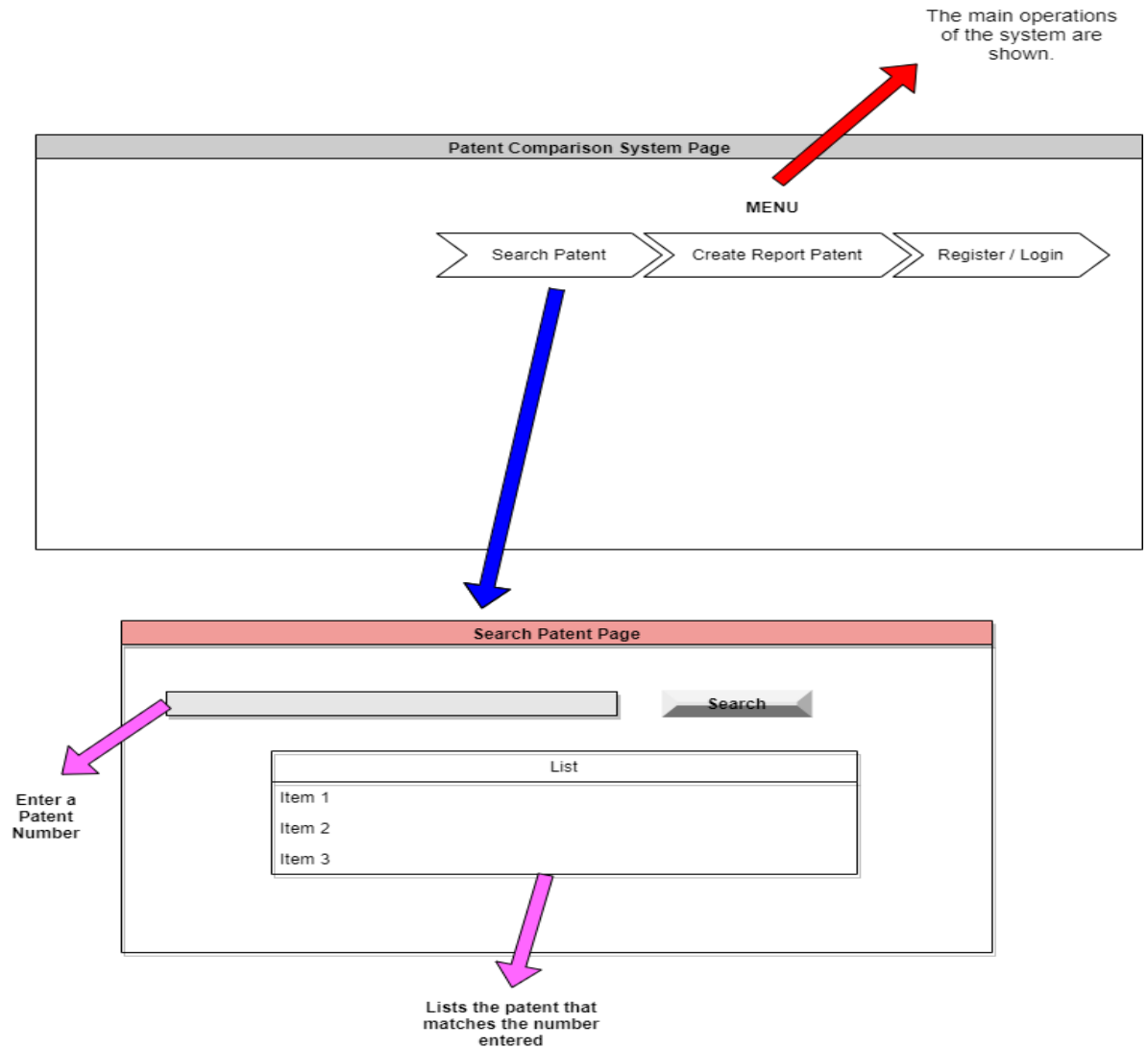


Figure 13 Search Patent Page Design

If the user wants to search for a patent, the system starts searching by obtaining the patent number and a few required information from the user. And the patents that match this information are listed.

6. Survey Result

In case of some results of our survey:

3 Patent benzerlik oranı hangi aralıkta olursa sizin için kritik sayılır?

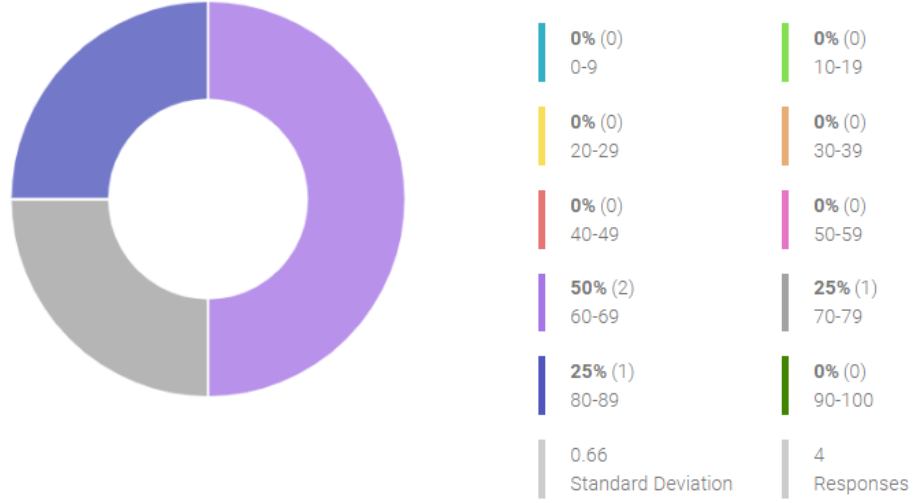


Figure 14 Survey Result Question 3

4 Uygulamamızda kıyaslanan patentlerden benzerlik oranı hangi yüzdelik dilimde olan diğer patentleri görmek istersiniz?

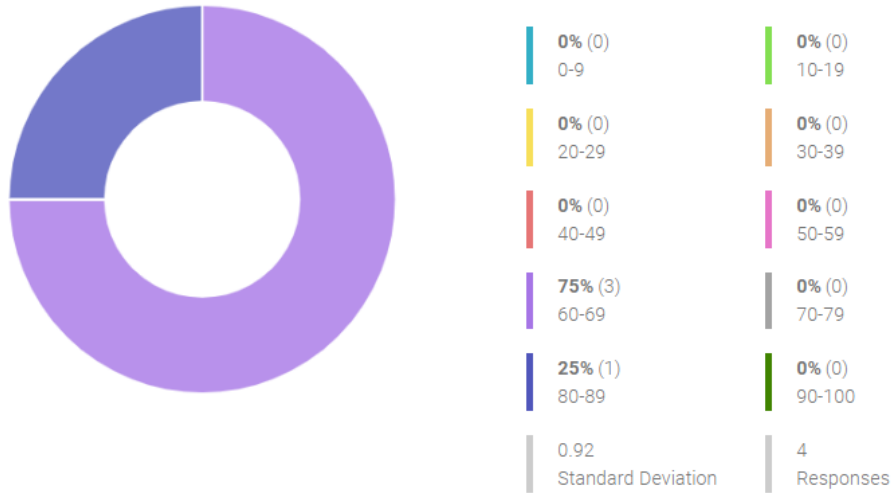


Figure 15 Survey Result Question 4

5 Şu anda bir patent arama/karşılaştırma sistemi kullanıyor musunuz?

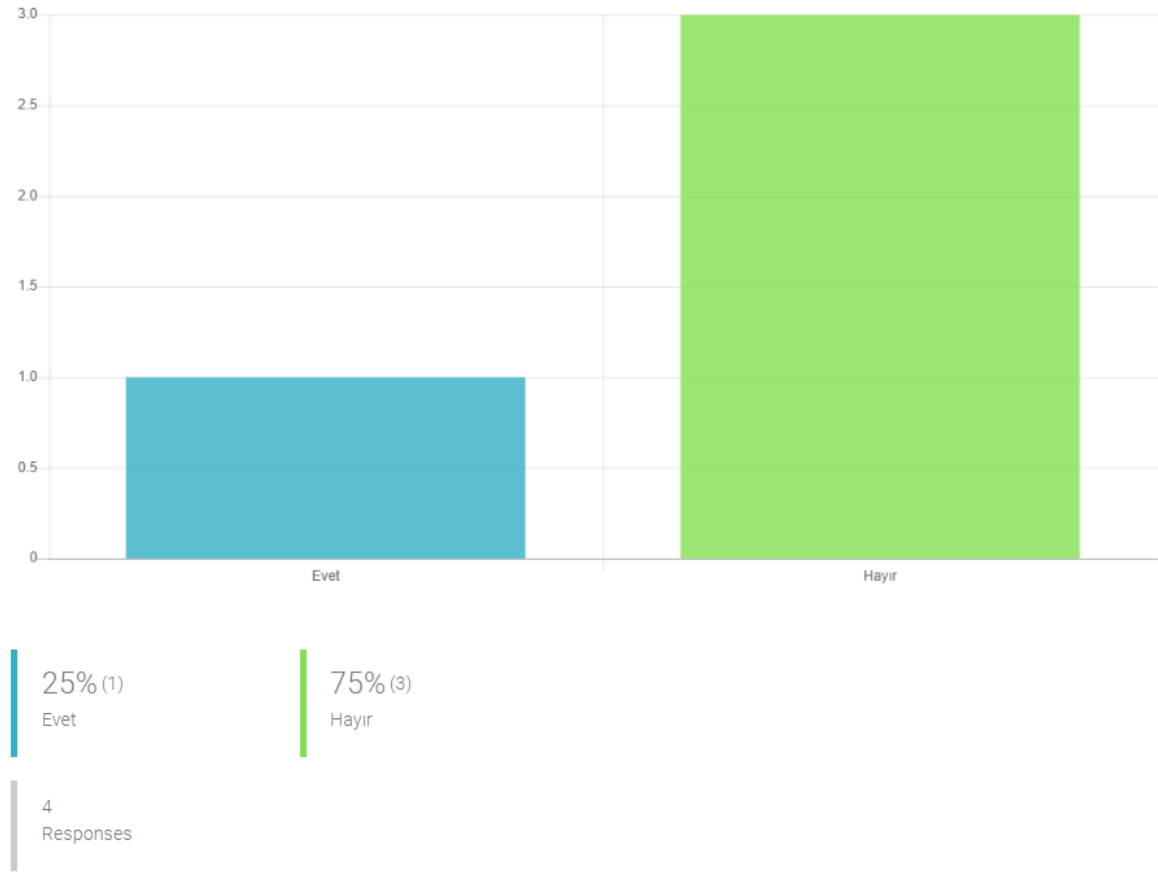


Figure 16 Survey Result Question 5

6 Cevabınız evetse hangi sistemi kullanıyorsunuz?

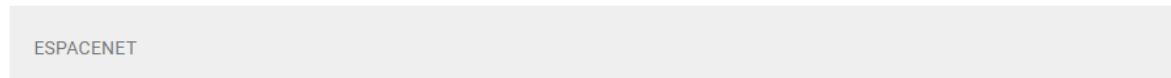


Figure 17 Survey Result Question 6

7 Patent karşılaştırma uygulamasını kullanmak için ücret öder miydiniz?

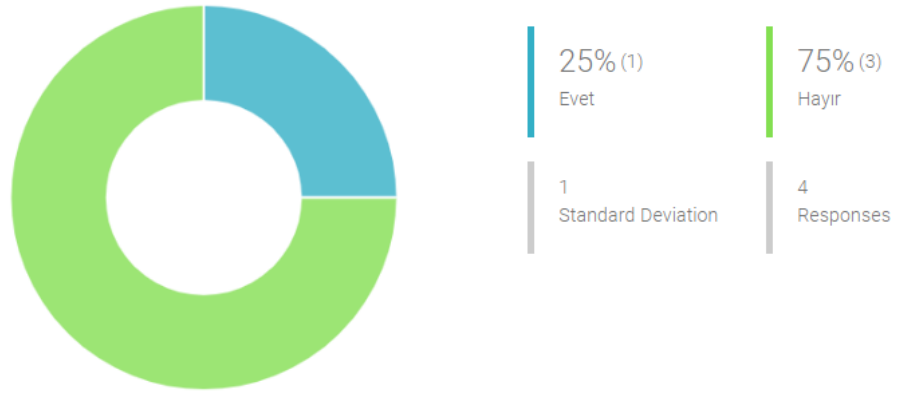


Figure 18 Survey Result Question 7

9 Şu anki mevcut patent arama/karşılaştırma sisteminizi değiştirmek ister misiniz?

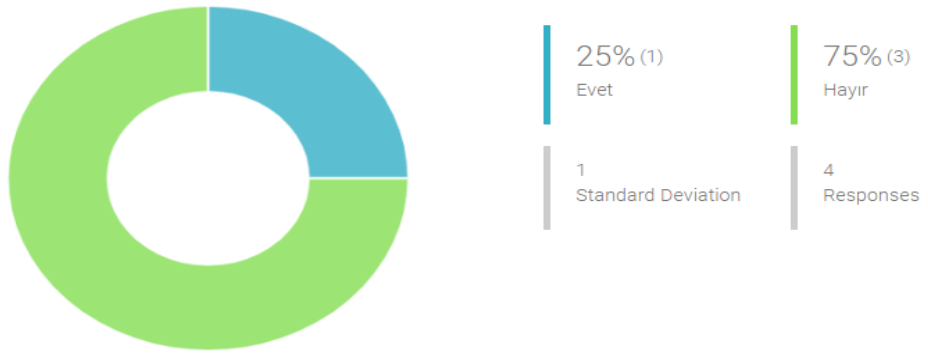


Figure 19 Survey Result Question 8

10 Patent karşılaştırma uygulamasında hangi özelliklerin olmasını istersiniz?

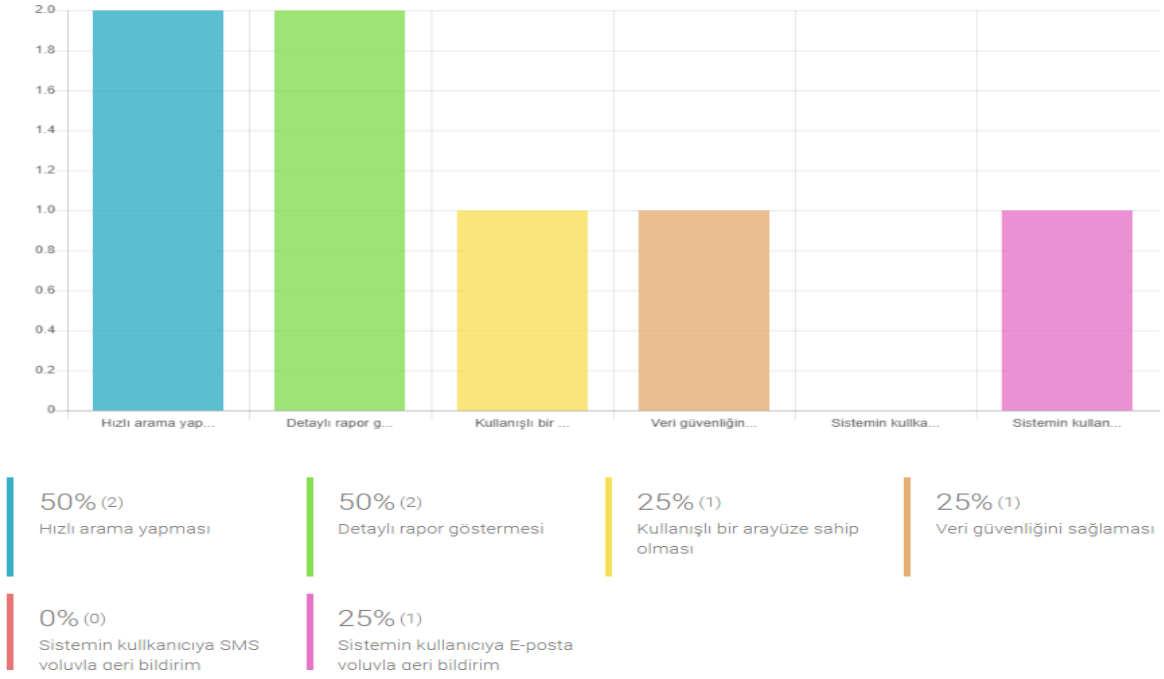


Figure 20 Survey Result Question 10

We prepared a survey to develop our project for the right purposes. The purpose of this survey is to determine what kind of needs people have to use our web-based application and to proceed in that way. Individuals who are willing to obtain patents and who are in the process of getting patents answered our survey.

In response to the responses of the respondents, we gained a new perspective for our project. We got answers to our questions such as how to make our project more effective and useful.

7. Conclusion

In this project that we are currently working on, some specific points can not be handled as a whole instead they should be dealt in parts. The most important thing among these parts is how much a patent is similar to related patents. The first thing that we will do is correctly classify the patents according to their IPC numbers. After the classification, the titles and content of the patents will be compared with each other. The basis will be critical similarity level that we have defined after the results obtained from a survey that targets patent holders. A report containing similarity levels of the patents will be presented to the user. In today's world researches for patents are still carried out by humans. This project aims to decrease the required manpower and time that is allocated to this business.

8. Acknowledgement

We would like to thank our consultant Murat Saran for helping us to design and implement this project. We believe that we will be more successful in this project with the advices of our advisor. We would also like to thank our other instructors who assisted and advised in this process. Finally, we would like to thank Onur Engineering for believing in us and for giving us this project.

9. References

-
- [1] Onur Mühendislik, 2019. [Online] Available at: <https://www.onur.net/tr/kurumsal/tarih%C3%A7e-1/>
- [2] Wang Chong-feng "The quality difference analysis between environmental patents and non-environmental patents in China", 2011 International Conference on Management Science & Engineering 18th Annual Conference Proceedings. [Online] Available at: <https://ieeexplore.ieee.org/document/6070135>
- [3] Loet Leydesdorff; Dieter Franz Kogler Bowen Yan "Mapping patent classifications: portfolio and statistical analysis, and the comparison of strengths and weaknesses", September 2017. [Online] Available at: https://www.researchgate.net/publication/314080370_Mapping_Patent_Classifications_Portfolio_and_Statistical_Analysis_and_the_Comparison_of_Strengths_and_Weaknesses
-
- [4] Patent Database Comparison: Everything You Need to Know. [Online] Available at: <https://www.upcounsel.com/patent-database-comparison>
- [5] Development Methodology [Online] Available at: <https://medium.com/@secilcor/scrum-nedi%CC%87r-6a4326951dd8>
- [6] Development Methodology [Online] Available at: <https://www.synopsys.com/blogs/software-security/top-4-software-development-methodologies/> (Development Meth.)
- [7] D. Kukolj; Z. Tekic; Lj. Nikolic; Z. Panjkov; M. Pokric; M. Drazic; M. Vitas; D. Nemet "Comparison of Algorithms for Patent Documents Clusterization ", 2012 Proceedings of the 35th International Convention MIPRO, 2012. [Online] Available at: <https://www.sciencedirect.com/science/article/pii/S1877705814002379>
-
- [8] Nadin Tamer ,Natural Language: Swift ile Doğal Dil İşleme. [Online] Available at: <https://medium.com/turkishkit/natural-language-swift-ile-do%C4%9Fal-dil-i%CC%87%C5%9Fleme-ca316b8e218d>
- [9] Google Patents: The global patent search engine [Online] Available at: https://www.researchgate.net/publication/280301154_Google_Patents_The_global_patent_search_engine