MOROCCAN PORTAL
Capstone Design

Nariman Agharbi

May 2020

Supervised By

Dr. V. Cavalli Sforza
MOROCCAN PORTAL

Capstone Report

Student Statement:
I affirm that I had applied ethics to the design process and in the selection of the final proposed design. Also, I avow that I have held the safety of the public to be paramount and have addressed this in the presented design wherever may be applicable.

________________________
Nariman Agharbi

Approved by the Supervisor

________________________
Dr. V. Cavalli Sforza
ACKNOWLEDGEMENTS

My sincere gratitude to Dr. Cavalli Sforza for accepting my project proposal and for providing me with the help needed throughout my capstone project. I am honored to have you as my supervisor because you backed me up at each stage of this project. It allowed me to be at peace. I knew your dedication to your work would motivate me to work even harder.

I consider this capstone experience to be a special and riveting opportunity. Thanks to Dr. Cavalli Sforza and Al Akhawayn University, I had the privilege to explore the software engineering field, and scrutiny the Search Engine Optimization field. I had the chance to apply my existing skills and sharpen others.

Furthermore, my journey at AUI played a great part of finding out what I am and what I am passionate about. Thanks to Al Akhawayn University, I got to learn and grow as a person and as an engineer.

Finally, I would like to express my utmost gratitude to my parents who empowered me to believe in myself and the path I have chosen for myself. Also, I would like to thank my siblings and friends who supported me emotionally with their unconditional love.
# TABLE OF CONTENTS

1. Introduction  
2. Feasibility Study  
   2.1. Managerial  
   2.2. Operational  
   2.3. Financial  
3. Software Specification  
   3.1. Functional Requirements  
      3.1.1. User Interaction  
      3.1.2. Content Management  
      3.1.3. Optional Features  
   3.2. Nonfunctional Requirements  
      3.2.1. Product Requirements  
      3.2.2. Organizational Requirements  
      3.2.3. External Requirements  
4. Software Architecture  
   4.1. Use Case Diagram  
   4.2. Sequence Diagram  
   4.3. Class Diagram  
5. Technology Enablers  
6. Software Design and Implementation  
   6.1. The Crawling and Scraping Process  
   6.2. Sending Documents to Elasticsearch  
   6.3. Response  
   6.4. The Creation of the Website  
7. STEEPLE Analysis  
   7.1. Social  
   7.2. Technological  
   7.3. Economic  
   7.4. Environmental  
   7.5. Political
7.6. Legal 22
7.7. Ethical 22

8. Challenges 23
9. Future Work 24
10. Conclusion 25
11. References 26
**LIST OF FIGURES**

<table>
<thead>
<tr>
<th>Figure Reference</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.2.1</td>
<td>Nonfunctional Requirements</td>
<td>6</td>
</tr>
<tr>
<td>4.1</td>
<td>Creately</td>
<td>9</td>
</tr>
<tr>
<td>4.1.1</td>
<td>Use Case Diagram</td>
<td>10</td>
</tr>
<tr>
<td>4.2.1</td>
<td>Sequence Diagram</td>
<td>11</td>
</tr>
<tr>
<td>4.3.1</td>
<td>Class Diagram</td>
<td>12</td>
</tr>
<tr>
<td>6.1.1</td>
<td>Crawling Mohammed V Website</td>
<td>16</td>
</tr>
<tr>
<td>6.1.2</td>
<td>Parsing and Paginating Mohamed V Website</td>
<td>16</td>
</tr>
<tr>
<td>6.2.1</td>
<td>Import Elasticsearch Library</td>
<td>17</td>
</tr>
<tr>
<td>6.2.2</td>
<td>Instantiation</td>
<td>17</td>
</tr>
<tr>
<td>6.2.3</td>
<td>Creation of indices</td>
<td>17</td>
</tr>
<tr>
<td>6.2.4</td>
<td>Mapping</td>
<td>18</td>
</tr>
<tr>
<td>6.3.1</td>
<td>Index Search results</td>
<td>18</td>
</tr>
<tr>
<td>6.4.1</td>
<td>Home Page</td>
<td>19</td>
</tr>
<tr>
<td>6.4.2</td>
<td>HTML PAGE</td>
<td>19</td>
</tr>
<tr>
<td>6.4.3</td>
<td>CSS Home Page</td>
<td>20</td>
</tr>
<tr>
<td>6.4.4</td>
<td>About Page</td>
<td>20</td>
</tr>
</tbody>
</table>
LIST OF TABLES

Table 3.1.1.1: User Interaction 3
Table 3.1.2.1: Content Management 4
Table 3.1.3.1: Optional Features 5
Table 3.2.1.1: Product Requirements 7
Table 3.2.2.1: Organization Requirements 7
Table 3.2.3.1: External Requirements 8
ABSTRACT

Many Moroccan students face the anxiety of starting the search for universities to apply to and understanding the several steps needed for the application. This project aims at minimizing the frustration of students during the period of searching for a suitable program. This project is an attempt to solve the issue of unifying all academic information for Moroccan users on a single platform. The solution is in the form of a website that makes use of an optimized search engine to allow these students to access this information by fewer clicks. The solution presented in this report is a website application that provides Search Engine Optimization (SEO) services for Moroccan students wishing to access information about universities in Morocco in an effective and efficient way.

Keyword: search engine, search engine optimization, data collection, web application, track, scraping, crawling, security
1 INTRODUCTION

The heart of this capstone project is Search Engine Optimization (SEO) for a platform to reduce the amount of time that high school students lose to reach information for applying to universities.

The development of this software was an idea I came across when my sister complained about the frustration of not finding a single platform that eases access to information about Moroccan universities. I, too, personally underwent this experience, which led me to struggle to fulfill what is expected from me and finding the next step of studies to pursue after high school. This is what motivated me to solve this issue for Moroccan high school students. This project is an attempt to alleviate the adversities faced by students by facilitating the process of browsing, comparing, and applying to the programs that fit the paths they wish to take.

This project consists of two main parts: the search engine, and the web application. The first part is about optimizing the search engine of the website to extract the necessary information for the students, and the second part is about the presentation of the search results in a manner that would be organized and appealing to the students.

I believe that this platform will be the learning portal that helps them to discover and explore the different education and career options for after high school. The Moroccan Portal is a tool that will serves the students to view and access all possible resources at the comfort of their own home.
2 FEASIBILITY STUDY

2.1 MANAGERIAL
The software is to be produced respectively following the software process: Software specification, design and implementation, validation, and software evolution.
The software process model that has been used in the production of this final software product is the incremental model. I mean to emphasize the development of increments; based on modified requirements.

2.2 OPERATIONAL
The software application is an attempt to ascertain the needs and the requirements necessary for Moroccan students or anyone who want to study in Morocco are met. The students’ needs of available information are directing us to address the questions related to collecting data about universities, analyzing and presenting them to the users. For that reason, the solution is harnessed through many resources to ease the process of implementation, as I explore some of the responsibilities of data scientists in handling data.

2.3 FINANCIAL
This project is an online website application software that makes use of free sources and tools to create it. It is an initial prototype or a sample of the requirements the product has to meet. For that reason, I am only using free software tools and sources for this capstone project.
3 SOFTWARE SPECIFICATION

3.1 FUNCTIONAL REQUIREMENTS

This project is fusing data science concepts into the creation of a product that fits the needs of the Moroccans’ students. This software product is going to be implemented as a search engine that is to be optimized accordingly \[2\].

In this section, I present a description of the functionalities and services of the Moroccan portal search engine. I brainstormed and came up with the following functional requirements based on my experience as a high school student and on my sister’s. The following requirements must be valid, complete, consistent, realistic, and verifiable to some extent to achieve certain criteria.

3.1.1 User interaction

Table 3.1.1.1 is a description of the interaction flow functionalities that underpin the tasks to be performed and accomplished.

<table>
<thead>
<tr>
<th>Functional Requirement ID</th>
<th>User Requirement Statement</th>
<th>System Requirement Statement</th>
<th>Essential/Desirable Criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>The student shall be able to access and navigate through all the webpages</td>
<td>The system shall be able to crawl and discover new webpages</td>
<td>Essential</td>
</tr>
<tr>
<td>1.2</td>
<td>The student shall be able to use the search engine without the need to log into their personal account</td>
<td>The system shall allow students or any users to use the search engine</td>
<td>Essential</td>
</tr>
<tr>
<td>1.3</td>
<td>The user shall be able to select an option from the listed searching results</td>
<td>The system shall provide as many relevant options to select from by the users</td>
<td>Essential</td>
</tr>
</tbody>
</table>
### 3.1.2 Content Management

Table 3.1.2.1 evinces the functional requirements needed for the application to perform internally:

<table>
<thead>
<tr>
<th>Functional Requirement ID</th>
<th>User Requirement Statement</th>
<th>System Requirement Statement</th>
<th>Essential/Desirable Criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>The user shall be able to view the latest results from the search engine</td>
<td>The system shall index all the newly discovered URLs</td>
<td>Essential</td>
</tr>
<tr>
<td>2.2</td>
<td>The user shall get relevant search results to fulfill their query</td>
<td>The system shall be presented with a relevant set of quality searching results</td>
<td>Essential</td>
</tr>
<tr>
<td>2.3</td>
<td>The user shall be able to view only options that are located in Morocco</td>
<td>The system shall be location-dependent so the search queries are searching results available in only Morocco</td>
<td>Essential</td>
</tr>
<tr>
<td>2.4</td>
<td>The users shall be able to get searching results relevant to the keywords they entered in the search engine</td>
<td>The system shall crawl a list of website for the required keywords relevant to the users</td>
<td>Essential</td>
</tr>
<tr>
<td>2.5</td>
<td>The user shall get their search result delivered in a structured and clean format</td>
<td>The system shall provide the users structured results</td>
<td>Essential</td>
</tr>
</tbody>
</table>
3.1.3 **Optional Features**

Table 3.1.3.1 displays additional features that could be included in the web application for their added value.

<table>
<thead>
<tr>
<th>Functional Requirement ID</th>
<th>User Requirement Statement</th>
<th>System Requirement Statement</th>
<th>Essential/ Desirable Criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1</td>
<td>The user should be able to create a personal account</td>
<td>The system should let the users create their own accounts</td>
<td>Desirable</td>
</tr>
<tr>
<td>3.2</td>
<td>The user should be able to modify their personal information available in their personal account</td>
<td>The system should allow the users to modify and save the modification of their personal account</td>
<td>Desirable</td>
</tr>
<tr>
<td>3.3</td>
<td>The user with an account should be able to bookmark their favorite pages of the searching results</td>
<td>The system should allow the users who are registered to have a bookmark list which presents a list of favorite programs to pick from</td>
<td>Desirable</td>
</tr>
<tr>
<td>3.4</td>
<td>The users should be able to filter their searching options according to the tuition fees, degree types and duration</td>
<td>The system should provide filtering options to sort the results of the search query</td>
<td>Desirable</td>
</tr>
</tbody>
</table>
3.2 NONFUNCTIONAL REQUIREMENTS

In order for the software product to come into existence, it is important to define the constraints that could be imposed on it. For that reason, in this section, I am representing a set of standards I set to describe the quality attributes of the Moroccan Portal application. The following figure is a representation of the nonfunctional requirements that must be taken into consideration when developing any software product. The next subsections address and specify each of the nonfunctional requirements as in Figure 3.2.1.

![Nonfunctional Requirements Diagram]

**Figure 3.2.1 Nonfunctional Requirements**

3.2.1 Product Requirements

The product requirements are basically specifications that constrain the behavior of the search engine product as shown in Table 3.2.1.1.
Table 3.2.1 Product Requirements

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Usability</td>
<td>The student shall be able to go through the webpages due to the user-friendly web design</td>
<td>The program should provide an easy and intuitive interface</td>
<td>Desirable</td>
</tr>
<tr>
<td>Reliability</td>
<td>The student shall be able to get their search results consistently with the least failure possible</td>
<td>The system should be reliable such that the website does not crash easily</td>
<td>Desirable</td>
</tr>
<tr>
<td>Portability</td>
<td>The user shall be able to run the software on different platforms like Microsoft Windows and macOS</td>
<td>The system shall be viewed and run by users from multiple software platforms like Microsoft Windows and macOS</td>
<td>Essential</td>
</tr>
<tr>
<td>Efficiency</td>
<td>The user shall get fast and quality response to their search query</td>
<td>The system shall require a maximum of 10 seconds to perform the searching task. The system should not overwhelm the users with many results</td>
<td>Essential</td>
</tr>
</tbody>
</table>

3.2.2 Organizational Requirements

The organizational requirements are organizational policies and procedures that constrain the production of the project as shown in Table 3.2.2.1.

Table 3.2.2.1 Organizational Requirements

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Delivery</td>
<td>The student shall be able to successfully use the search engine</td>
<td>The system shall be successfully designed, developed, and tested</td>
<td>Essential</td>
</tr>
<tr>
<td>Implementation</td>
<td>The student should be able to read and understand the search engine results</td>
<td>The system shall be implemented using the Python programming language.</td>
<td>Desirable</td>
</tr>
<tr>
<td>Standards</td>
<td>The system shall follow the IEEE format based on the Software and System Engineering Standards</td>
<td>The system shall follow the IEEE format based on the Software and System Engineering Standards</td>
<td>Essential</td>
</tr>
</tbody>
</table>
3.2.3 External Requirements

The external requirements are external factors that put constrains on the development process of the software product as shown in Table 3.2.3.1.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Interoperability</td>
<td>The user shall be able to use this software on any device any time</td>
<td>The system shall be compatible with the user’s device</td>
<td>Essential</td>
</tr>
<tr>
<td>Ethical</td>
<td>The student shall be able to use the search engine solely for the purpose of browsing academic programs offered in Morocco</td>
<td>The system shall not display any misleading and false information to harm their best interests</td>
<td>Essential</td>
</tr>
<tr>
<td>Legislative</td>
<td>The software product shall operate within the legal jurisdiction in Morocco</td>
<td>The software product shall operate within the legal jurisdiction in Morocco</td>
<td>Essential</td>
</tr>
</tbody>
</table>
4 SOFTWARE ARCHITECTURE

The Moroccan Portal system is created using the Docker approach. It is an open source platform that eases the development process and the deployment of software. The Defense-in-depth model, adapted by the Docker, relies on securing the production and the development of software by shielding the containers with multiple defenses for the attacker to breach. I choose to make use of the Docker platform due to its numerous advantages, as follow [3]:

- Speed: The Docker development ecosystem benefits from its containers. That is, the latter is fast-paced which helps to run a variety of kits that may be required to develop a certain application.
- Scalability: Docker is highly scalable due to the easy movements it provides for its containers from one environment to another which eases the adjustments for developers. Thus, APIs can work with different systems.
- Security: Due to the nature of Docker containers, Docker daemon sockets keep applications isolated from one another by letting each container to own a network stack. Thus, when running the Docker daemon, the attacker has to overcome many levels of obstacles before harming the system.

Before starting developing the Moroccan Portal Website, it is necessary to build a good design. For that reason, I used visual software called Createley (Figure 4.1) to convey the conceptual design phase through UML drawings portraying the primary and secondary functional requirements.

Figure 4.1 Creately
4.1 USE CASE DIAGRAM

The Use Case diagram (Figure 4.1.1) is primarily used to specify the expected behaviors of the system and the actors.

Figure 4.1.1 Use Case Diagram
4.2 SEQUENCE DIAGRAM

This type of UML diagram (Figure 4.2.1) was used to depict the sequential order of the interaction happening between the main objects of the system.

Figure 4.2.1 Sequence Diagram
4.3 CLASS DIAGRAM

Figure 4.3.1 is a static UML diagram that describes the interactions that happen between the classes of the Moroccan Portal system.

Figure 4.3.1 Class Diagram
5 TECHNOLOGY ENABLERS

- **Docker**: It is a tool that simplifies and empowers application development for developers. A Docker file is a text file that contains commands to assemble an image. It makes use of isolated and lightweight containers to manage various application deployments. It is basically a list of essential parts of the image created by a software developer\[^4,5\].

- **Docker Compose**: It is the tool that unifies multiple container Docker applications. The Docker Compose is basically a file that can run various services related to one application in an isolated environment. This file can be based on a layered foundation of other Docker files to bundle different containers for creating a service\[^6,7\].

- **ELK stack**: It is a collection of 3 open-source products which are **Elasticsearch**, **Logstash**, and **Kibana**. I choose this tool because the Elasticsearch engine is a full-text engine based on Apache Lucene. Logstash is a source tool that handles streams of log data from different sources. And to perfectly match these tools, Kibana comes as a browser user interface that visualizes the documents stored within the indices of Elasticsearch.

- **LXML**: It is a Python Library that is used for web scraping. LXML deals with parsing XML and HTML files to extract specific information. An example would be retrieving certain attributes from an existing tag name of a specific element like root tag (html). LXML is used not only to parse web elements (tags) but also strings in between a tag, and traversing particular element’s parents or siblings\[^9\].

- **Flask**: It is a micro web application framework that allows building a web application in Python. I choose to use Flask because of its simplicity in mapping HTTP requests to Python since I wanted to create a basic search engine application\[^10\].

- **Scrapy**: It is a framework that helps in building web scrapers easily through extracting useful data from CSS selectors and XPath expressions. In my project, I am using the
Anaconda platform because it provides with useful packages for data science in Python.

- Pycharm: It is an integrated development environment that helps in developing a web application using Python. This IDE supports the frontend and the backend development of web applications.\(^{[11]}\)

- Python: It is a high-level programming language that is used throughout this project. Due to its easy syntax and the large collection of libraries that come handy for web scraping.

- HTML: Hypertext Markup Language is responsible on managing and defining the content of an application.

- CSS: Cascading Style Sheets is responsible on styling and visualizing the contents of the webpage layouts from colors to fonts.
6 SOFTWARE DESIGN AND IMPLEMENTATION

6.1 THE CRAWLING AND SCRAPING PROCESS

In order for any search engine to work, web crawlers need to launch their spiders to fetch, analyze, and store contents of the web pages from the internet into organized indices. Thus, a search engine bot allows its users to retrieve relevant information from the Internet. The following steps depict the meticulous steps to obtain search results of the Moroccan Portal search engine:

- Collecting Phase: The Scrapy framework provides spiders in the form of defined classes that describe the strategies used by crawlers to collect raw data, the type of the data to target such as URL lists, sitemaps, and web crawls. When the user enters a search query into the engine, the search engine generates an initial request to crawl the URLs in order to assemble the URL lists into the relevant index table depending on the content of a certain web page as shown in Figure 6.1.1. The key factor of any indexing process is based on finding the pattern between the collections of users’ queries.

- Parsing Phase: After crawling the raw data, we parse the web page response (Figure 6.1.2) to the initial crawling request using selectors like lxml to extract data items of the web pages from the HTML source. The XPath expressions serve as a mechanism in Scrapy to extract element attributes and organize them under classes of type Item.

- Storing Phase: Also called the piping phase, each Item Pipeline component starts processing the various items that were selected by the selectors. The main purpose of this process is to decide the validity of the scraped data before it can pass through the pipeline and be stored in a database.

- Presenting Phase: It is about undergoing a process in which the information stored in the database is converted to a user-friendly visual context of the URLs. The latter
must be relevant to the user’s query in the form of a simple website with a description about its content.

```python
import scrapy
from scrapy.http import Request
import csv
import os
from selectorlib import Extractor
import re

# Crawling Mohammed V University website

class CrawlerSpider(scrapy.Spider):
    name = 'crawler'
    allowed_domains = ['http://www.um5.ac.ma/um5/']
    start_urls = ['http://www.um5.ac.ma/um5/']
    extractor = Extractor.from_yaml_file(os.path.join(os.path.dirname(__file__), '..resources/search_results.yml'))
    max_pages = 5

    def start_requests(self):
        with open(os.path.join(os.path.dirname(__file__), '..resources/keywords.csv')) as search_keywords:
            for keyword in csv.DictReader(search_keywords):
                search_text = keyword['Ingénierie']
                url = 'http://www.um5.ac.ma/um5/search?searchText={}&viewType=3'.format(search_text)
                yield scrapy.Request(url, callback=self.parse, meta={'search_text': search_text})

    def parse(self, response):
        data = self.extractor.extract(response.text, base_url=response.url)
        for Ing in data['Ingénierie']:
            yield Ing

        if data['Ingénierie']:
            if '&page=' not in response.url and self.max_pages >= 2:
                yield Request(response.request.url + '&page=2')
            else:
                url = response.request.url
                current = re.findall('page=(\d+), url[0]
                next = int(current) + 1
                url = re.sub('(^.*?&page=)(\d+(.*$)', rf'&g<1>{next}\g<3}>', url)
                if next <= self.max_p:
                    yield Request(url, callback=self.parse)
```

Figure 6.1.1 Crawling Mohammed V Website

Figure 6.1.2 Parsing and Paginating Mohammed V Website
6.2 SENDING DOCUMENTS TO ELASTICSEARCH

After scraping websites using Python and Scrapy, we need to connect our Python code to Elasticsearch server. It is important to connect the Python client to the Elasticsearch JSON to allow the retrieval and the storage of documents within the Elasticsearch database and ingest their data. So, I import the Elasticsearch library to Python (Figure 6.2.1), and instantiate the Elasticsearch class (Figure 6.2.2).

![Figure 6.2.1 Import Elasticsearch Library](image)

```python
from elasticsearch import Elasticsearch

elastic = Elasticsearch("localhost:9200")
print("\nElasticsearch instance:", elastic)
print("Attributes:", dir(elastic))
```

![Figure 6.2.2 Instantiation](image)

At first, I run Elasticsearch on the browser. Then, run Kibana to create two nodes in a cluster in Elasticsearch. A cluster is a collection of nodes that stores important data into the Elasticsearch database. In the Kibana Console, we create Indices within one of the nodes just as in Figure 6.2.3. Then, the Elasticsearch system automatically replicates the data stored within the indices to increase the throughput of them and their availability when the users need them.

![Figure 6.2.3 Creation of indices](image)

Elasticsearch provides us with an inverted index consisting of unique terms that appear in the documents stored within the shards (split pieces of an index) in our cluster. An inverted index has this specialty of storing texts in an efficient and fast structure for full-text searches. The
documents in JSON formats within our index are chosen whenever the terms within the Elasticsearch Database match with the search query entered by the user.

An important step to add is the creation of a Schema or mapping, as it is called in Elasticsearch. The structure of the documents being stored within Elasticsearch is fully focused on the mapping structure defined by its specific fields (Figure 6.2.4). As long as the documents sent from the Python client are conforming to the mapping structure defined beforehand in Elasticsearch, there will not be a problem in recording the data within a new index.

![Figure 6.2.4 Mapping](image)

### 6.3 RESPONSE

Finally, the final step comes, which consists of querying the records of the documents or data at hand. The search criteria would be the keywords entered by the user in the searching box, and any indices that fall into the categories matching the fields that were instantiated within the shards (parts of documents stored within indices) of the cluster in Elasticsearch (Figure 6.3.1). Then, we can pass the search criteria and the matching indices to the Python Flask application to showcase the collection of files and elements of URLs, and to check the results.

![Figure 6.3.1 Index Search results](image)
6.4 THE CREATION OF THE WEBSITE

The Moroccan Portal website was created using the brackets frontend framework where I created two simple pages. The first page shows the Home page where the users can enter their search query to get their results.

![Figure 6.4.1 Home Page](image)

It was created using the following HTML and CSS code:

```html
1. <DOCTYPE html>
2. <html>
3. <head>
4.   <title>SEARCH FOR A UNIVERSITY</title>
5.   <link rel="stylesheet" type="text/css" href="css/style.css">
6. </head>
7. <body>
8.   <header>
9.     <div class="main">
10.    <div class="logo">
11.      <img src="logo.png">
12.    </div>
13. </header>
14. <ul>
15.   <li class="active"><a href="#">Home</a></li>
16.   <li><a href="about/index.html">About</a></li>
17. </ul>
18. <div class="title">
19.  SEARCH FOR A UNIVERSITY
20. </div>
21. <form method="POST">
22.   <input type="text" name="SEARCH" placeholder="Search for the Major">
23.   <button type="Submit">Search</button>
24. </form>
25. </header>
26. </body>
27. </html>
```

![Figure 6.4.2 HTML PAGE](image)
Figure 6.4.3 CSS Home page

And the second one is the **About** page so that users can get an idea about what are the services that are provided by the website.

Figure 6.4.4 About Page
7  STEEPLE ANALYSIS

7.1  SOCIAL

This project mainly targets the social aspect of the hardships faced by students in searching for an institution where to continue their studies after high school. This application will alleviate the hassles that students go through to find universities in Morocco. The main points that we are addressing are two:

1. Nowadays, everyone relies on the Internet to interconnect with all type of services through mere clicks. Hence, we are trying to use this as a way to fulfill the wishes of students in exploring offers in Morocco easily and efficiently without having to be present physically at universities to apply to them.

2. Concerning the anxiety of comparing the vexation and frustration of finding information about universities, we have created this unified platform to ease the access to all academic information needed to apply to universities.

7.2  TECHNOLOGICAL

This software is a web application that can be viewed and accessed from any laptop and phone. Therefore, there is no need for prior knowledge to use this platform because it is a user-friendly website solely serving Moroccan students or anyone who is interested in studying in Morocco.

7.3  ECONOMIC

This software project won’t have an effect on the economy since it is intended to be a free web application. It is open source software that will be available online for all Moroccan students wanting to pursue more studies after getting their high school degrees.
7.4 ENVIRONMENTAL

No bad repercussions come to mind. If anything, the software will benefit for the environment since it will cost less energy and time for students to know about universities. In general, this project does not have a negative impact on the environment.

7.5 POLITICAL

The administrative procedures are simplified to Moroccan students for the purpose of helping them to unravel the next step they want to take. Thus, this project does not get engaged into politics.

7.6 LEGAL

The project does not violate any law or regulations. It abides by the laws in the Moroccan constitution because all the resources used for developing this software are open source. Moreover, all the collected information from universities sites is public which does not violate any legal copyrights.

7.7 ETHICAL

This application cannot be used for malicious intent because the reason it was developed was only to provide Moroccan students with the universities’ academic information. In other words, this application can’t harm anyone or could be used in a manipulative way.

Besides that, it is my duty as a software developer to follow the code of ethics and make sure that this application won’t harm the society.
8 CHALLENGES

During the semester, I encountered unexpected challenges. That is, I was not just diving into a new part of computer science, but I had my laptop crashing a few times by the end of February. At the time, I was trying to comprehend what tools I needed to get right into the development phase. I could not distinguish what caused my computer to crash, but I thought it would be fine to borrow a desktop computer on campus at AUI so I can work on my project. Little did I know, the coronavirus outbreak would force the country to lockdown during the spring break. For that reason, I had to work with what I had. I did not succeed in fixing my laptop, so I decided to use the desktop computer, which I shared with my other siblings.

Splitting the time between us to attend classes decreased the time I could spend to work on my project. I studied during nights, which led to a decrease in overall performance and productivity.

I tried to learn how to use Elasticsearch and Scrapy framework. I know the basic concepts of how Elasticsearch searches, and add new indices to its database, and how I can scrape a few pages of a website. However, I haven't had the chance to use the full potential of Kibana to analyze the data being stored in Elasticsearch to find the indices patterns. It would have helped me to create an efficient search algorithm and mapping structures of documents stored in the indices.

I stand at a place where I know now basic things on how to use them, but I have not yet got to grasp the full potentials of each of these development tools.
9  FUTURE WORK

The Moroccan Portal project was meant to be a prototype model. I believe that the future project should be focused on adding interesting features like returning search results that are sorted by date relevance. Another research area that would be interesting and beneficial for this project is cross-language information retrieval. It deals with retrieving information written in Arabic and/or Spanish language from the user’s search query. At times, the user may submit an Arabic query but wait for the search results to be written in French or English. Although it is challenging, it would be a great feature for the web search engine if it supports multiple languages since many Moroccan students are multilingual [8].

Another great feature that could be an added value to this project would be filtering settings such as search results that are based on the location of the students or the desired location to move to, their financial budgets, and their type of degree. These advanced filtering settings can direct students to find the universities that are near where they live, in the targeted location.

I feel that there are more features that could be added, especially, to ensure the safety and the security of students from getting biased results. As for now, this research aims to redirect high school students to the web pages that display public and trustworthy information shared by the universities itself or trustworthy sites.
The focal point of my capstone project was to ease the frustration experienced by students planning for their next step of education. Thus, I designed a web search engine that can guide high school students to access trusted public resources provided by universities.

In the course of my project, I went through many difficulties in developing the search engine since I did not have any prior experience. Yet, I am glad I choose to work on this new project which allowed me to acquire new skills.

This project allowed me to learn and explore more about search engines. I intended to use my project to learn how search engine functions, and make use of the knowledge acquired to create a web search engine prototype. It is clear that the web application needs to be enhanced before deploying it. However, it is a good starting point that opens for further considerations over the website’s design and the quality of the search results.
11 REFERENCES


