University of Babylon

College of Information Technology
Software Department

Search engine

Research introduced by
Dev.Sakina Muhammed
Dev.Hussein Qassem
Dev.Hussein Mahdi

Supervisor by
Dr. Ahmed Selim

Abstracted

We are in a rapid development in the field of programming (databases) as well as the ways in which this data is processed and in order for this data to be displayed to the user on the web interface or mobile phone or any other interface, so we need a search engine that displays the results that we want excellently and closer to the user request, it was facing a problem which is the presentation of results which may sometimes be close to what is required and sometimes far and this is due to several reasons of them (Data request is unclear, the order of data is wrong, ...) So we're working to reduce these problems through the search engine we're creating.

Introduction

In this report we will discuss the search engine we have created which we want to be a fast and accurate search engine in the data presentation. All the search engines we see now such as (Google, Ping, Yahoo , Brava , ...) operate on three axes that make up the search engine (crawling, indexing, ranking) These three actions work together to form a search engine, in this article we will explain another way of working consisting of another structure at work that will be partly better at displaying results for the user The work came from a research letter from Stanford University PhD students Larry Page and Sergey Brin , whose research project was named BackRub.

In this report we will address how the parts of the search engines work and who is above the level in the presentation of results, as well as the problems facing the search engines, we will explain the mechanics of the work of our search engine, as well as what defects appear in it, as well as the future of this project and what will be added to it other technologies

Background

The search engine, a software system that works to inquire data from previously stored data (may be a relational or non-relational database) and displays the results requested by the user, this is the work of the search engine simply but how it works basically and what depends, as mentioned earlier it depends on three basic parts which are the essence of its components (creeper, Indexing, ranking) . The creeper , which is a software bot created and which starts on the Internet between HTML pages in order to draw results from them and send them to data base and received from it indexing, his movements between pages through the code, <a> which is a expression of link refers to the next page, but the work of the indexer It puts the received values into a group with its label definition to be queried at a later time by the user, After the indexer's work comes in the work of the Ranking , which is algorithms that take these totals and arrange them according to the degree of convergence of each other as well as the degree of convergence of the user request In public search engines , the data is displayed according to the degree to which it is close to the user's query, Here comes the work of the project where the project does not contain any of these components the reason is simple because the data that we want to search for is present and knowledge and integrated all that needs is to connect the user to this data in an effective way to display this data, how did this !!? The data we have is granted by companies that we have indicated that they have prevented reptiles from search engines from accessing them, so they offer them free of charge to developers and we as developers of a search engine have withdrawn this data but more than one source to be large and diverse and thus achieved effective access to data without accessing sites to search for results or information .
We also try to provide more than one source as well as several ways to enter data we have completed the search by text but we try to reach the integration of search by sound and image . The search engine also gives data by user, so if the user has recorded his personal data in the system, the results in the display are more different from the visiting person .

Problem description and statement

The problems that caused this project to be carried out are a group of problems, but which are:

1. Results are displayed for the most visited data but these statements may be false .

2. Data that is displayed does not suit the user because the system deals with the most data soon (if the user wants to search for a specific book and be "my child" we will notice that the search engines return us a song (video) while the search was for a book and this is due to a reason in arranging and indexing the data because it deals with the most visited and famous on the Internet .

3. If certain data are searched, for example, for example, the search product is locations for robberies, images or videos of the work of the parameters, but it should have displayed the parameters we need, this is an example of a model of data.

4. Also, do not forget the problem of fake links that sometimes lead to the breach of user data because they are not secured .

the problem we are referring to may worsen in the future because most companies with large databases have forced reptiles owned by search engines not to access this data and this causes the results to be reduced or the results may be far from what we want.The problem is also shared by the user because the values that are sought are the focus of determining the result because the query should be understood because we are dealing with a and it does not understand what the user is trying to reach in this query .

We are in the process of reaching solutions that reduce this problem, the problem lies not in the presentation of results but also the problem may affect the work of the server that works on the operation of this search engine for example if a user entered a query for example mobile so the results are a website about a company while the user wanted to know about what is this device use by humans and its history.

We need here in these cases to introduce the "natural language processing" algorithms that we want to add to the project, as well as expand the receipt of the query through different formats such as text, sound and image.

Problems facing the project

And according to what we have come to know, it is possible to build such a project. It is true that it faces a problem in the number of resources that must be provided to the user as well as entering into competition with major companies. We will also have trouble enabling other human-speaking natural languages like (Arabic, Chinese, Spanish, Russian, ...) and this too will create a problem with the voice search feature (because the voice recognition algorithms) have not yet reached the high level that We can deal with most users in their different languages

Differences between the project and search engines in :-

terms of technology

We use a ready database which has been processed and this facilitates work as well as gives the best results from the rest of the engines, and this is contrary to the work of search engines which provides for the presence of a creeper collecting data and then it is indexed and then arranged, but the disadvantage of this technology that our data base should be updated with the latest results so that the data is modern and not old data and this is one of the differences we face with the rest of the search engines

terms of work, time and speed

In terms of work, search engines have high-level work because there are 3 parts in its work and this is unlike our search engine, which is done in pulling results only from the database, in terms of time our search engine is faster because its work is to pull data unlike other search engines but also depends on the range that the database gives us from processing on its data, In terms of speed, our search engine is faster because it consists of simple parts unlike

other search engines, but there may be a difference in the speed at which data is displayed.

terms of its algorithms

We try to add the best algorithms in the field of voice recognition, image, natural language processing and word processing to be competitive with other search engines.

Conclusion

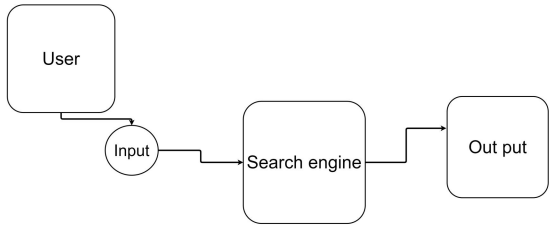
In conclusion, we seek to build a high-results search engine through the use of highly rated technologies and algorithms, we are not in the process of competition but we seek to reach the best user experience in the use of search engines as well, we so far as we indicated earlier the search has been completed by text but we still have a lot of hard work, We hope to reach this ease of research and presentation of results if it may not be in the future and with the development of technology and algorithms there will be architectures in the research as well as all this applies under the concept of search engine optimization

Schemes and design

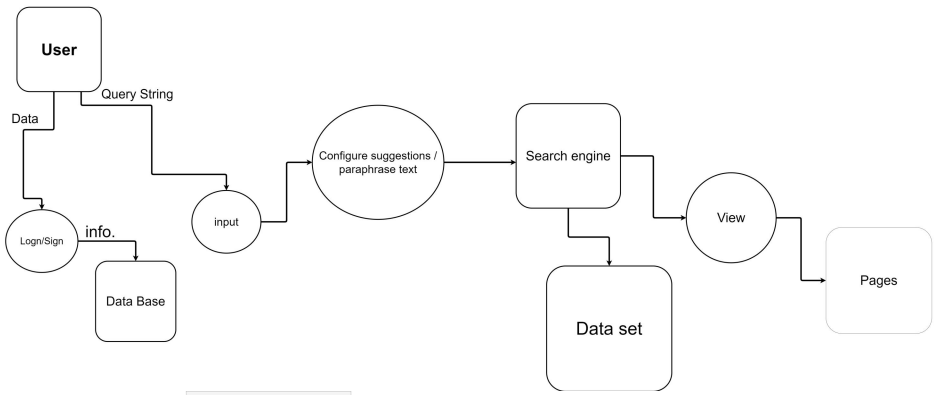
Data function Diagram

Level 0

Data FunctionsDigram (DFD)

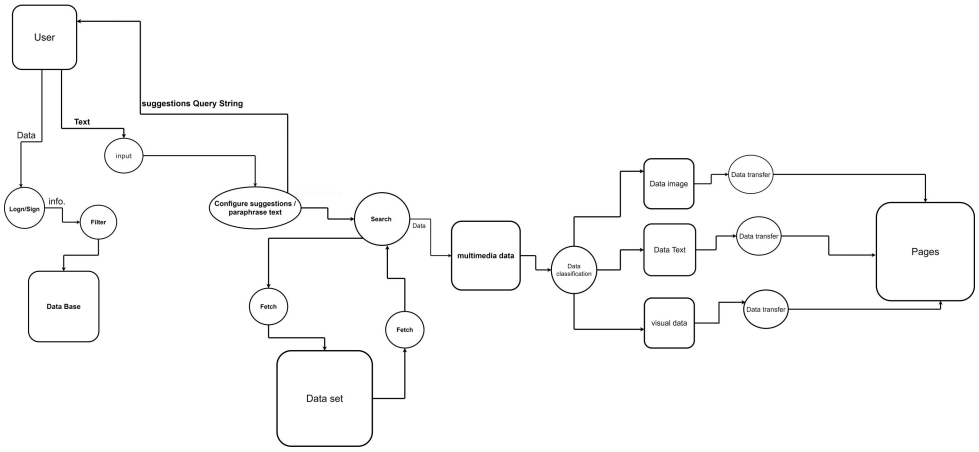


Level 1

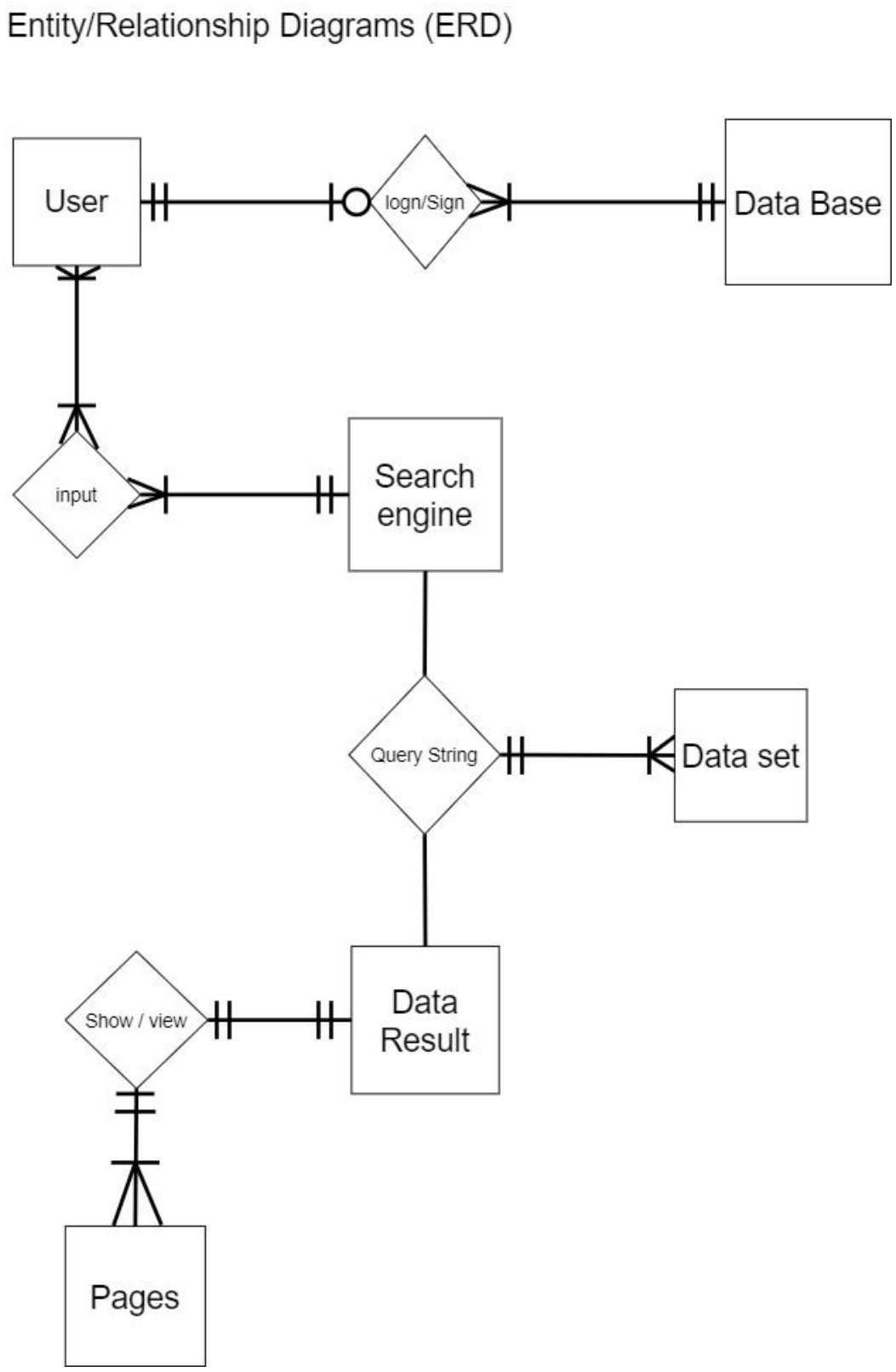


Data Functions Diagram (DFD)

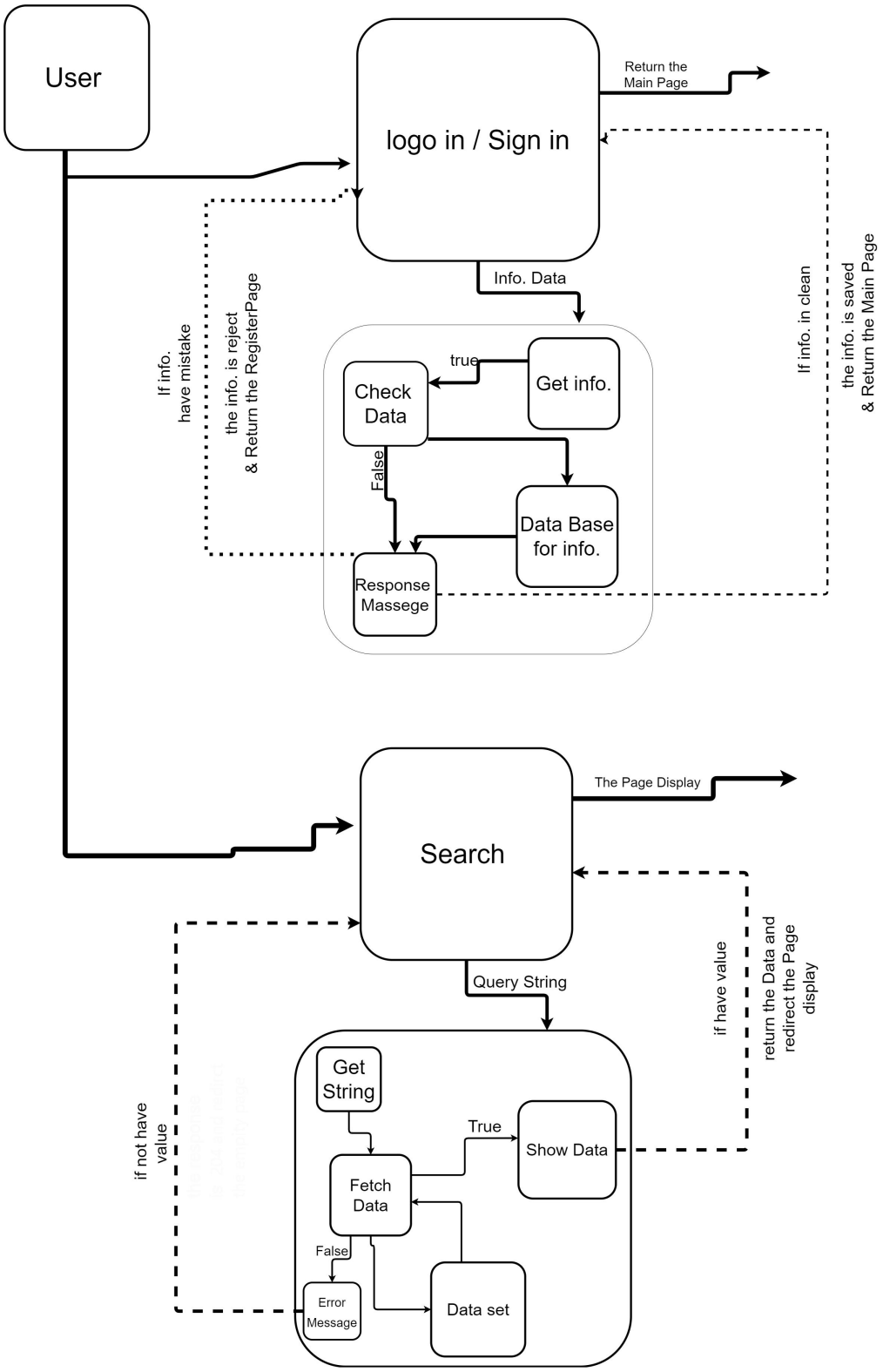
Level 2



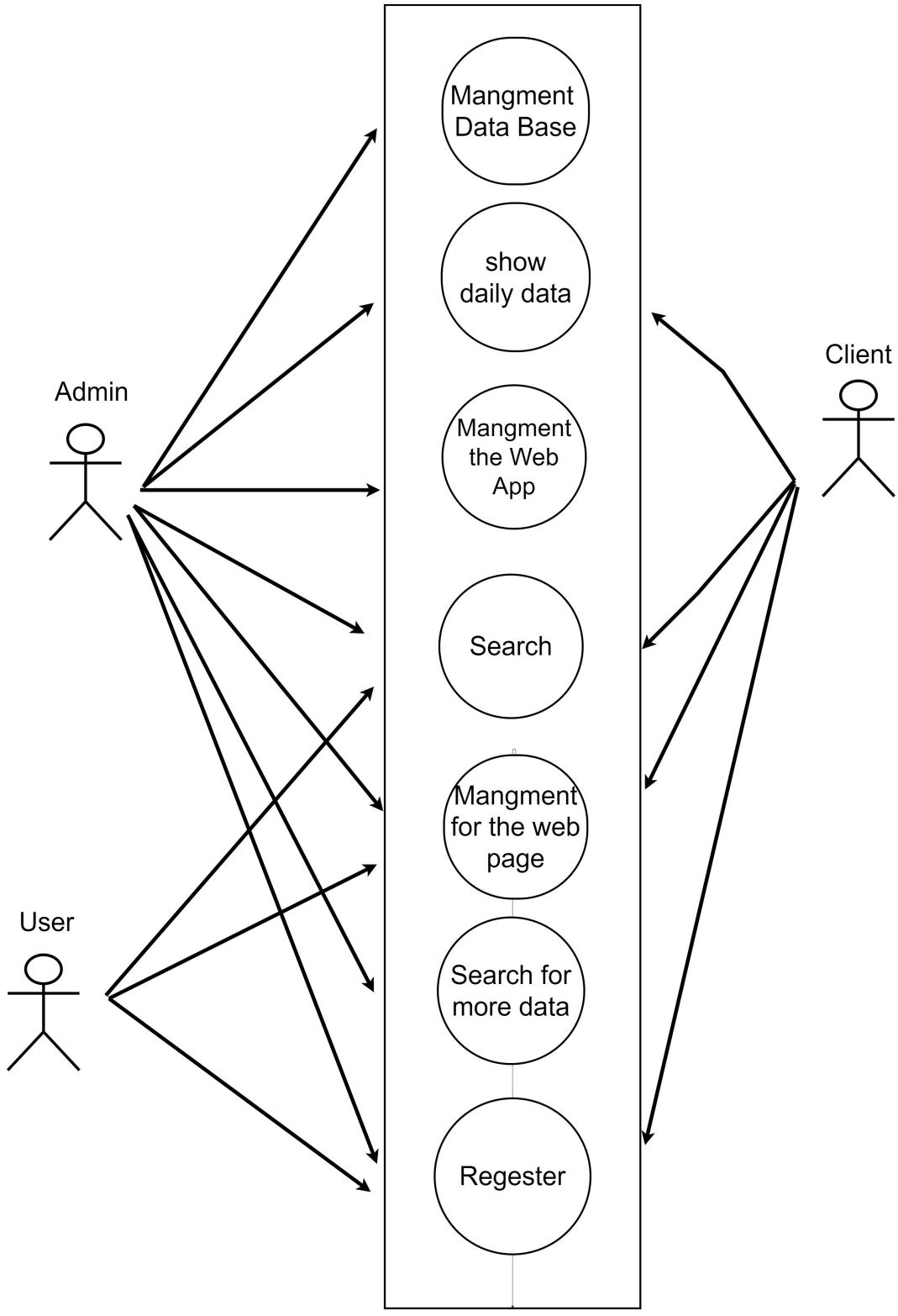
Entity/Relationship Diagrams (ERD)



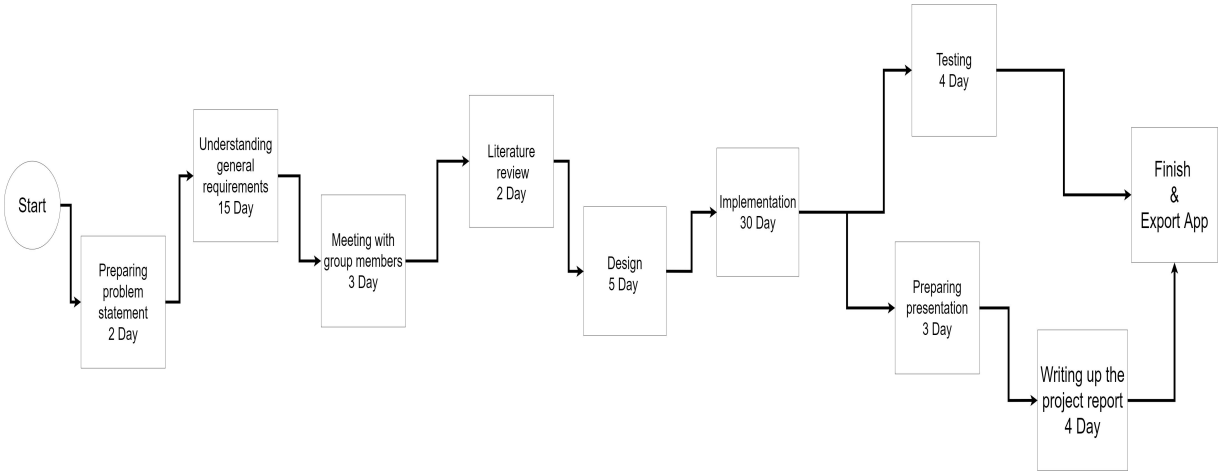
State-Transition Diagram



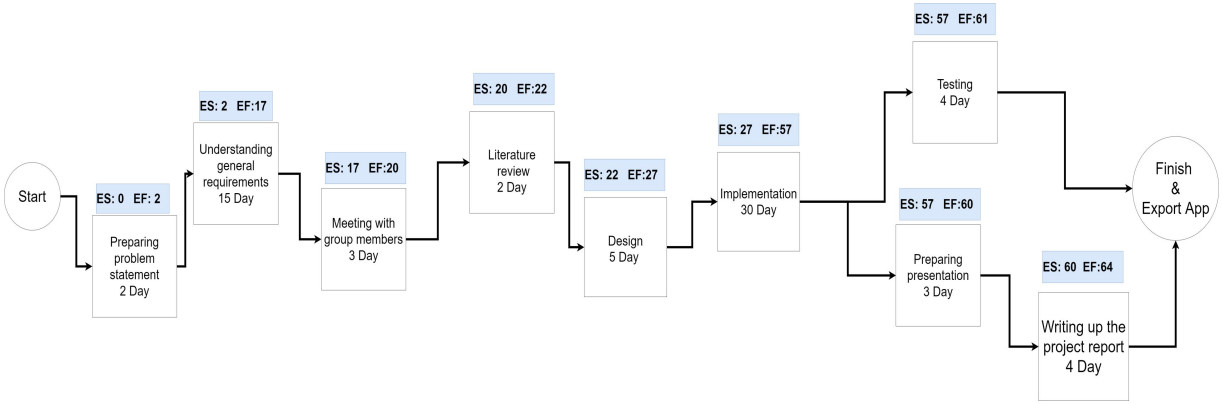
User case


Mangment Data Base
show daily data
Mangment the Web App
Search
Mangment for the web page
Search for more data
Regester
Admin
User
Client

Activity networks of the MIS problem

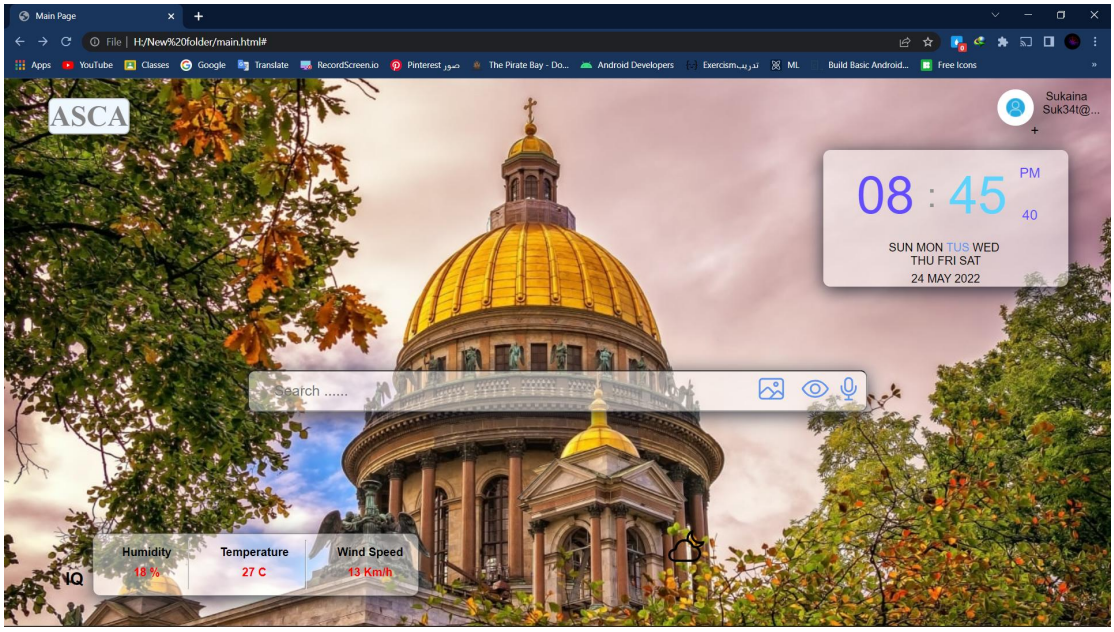


Earliest start and finish dates

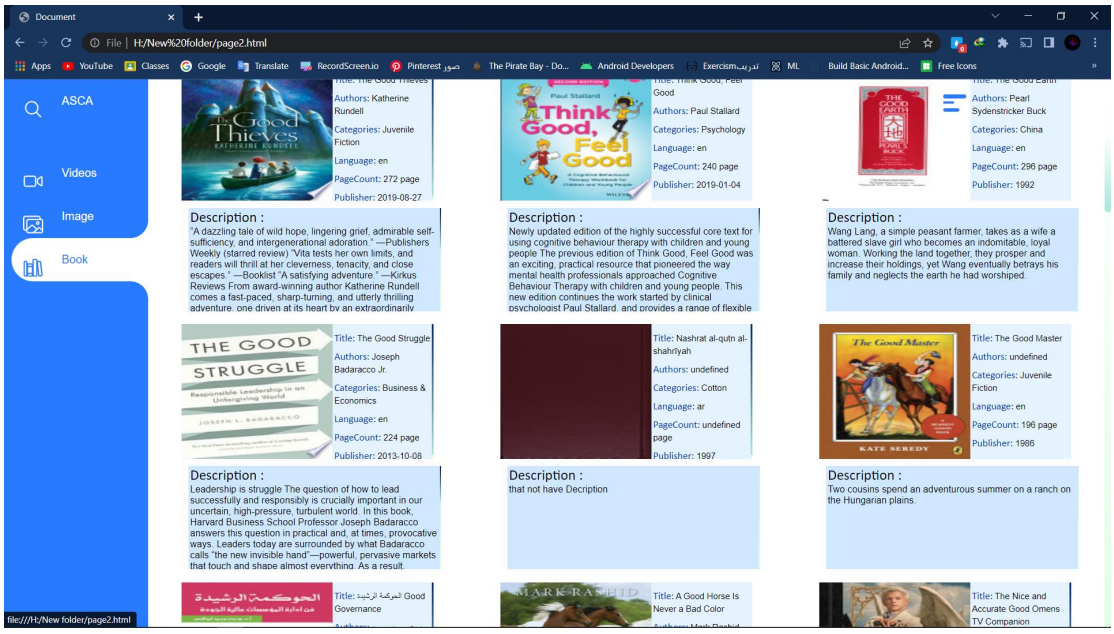


Page view

Main page



Book page



Video page

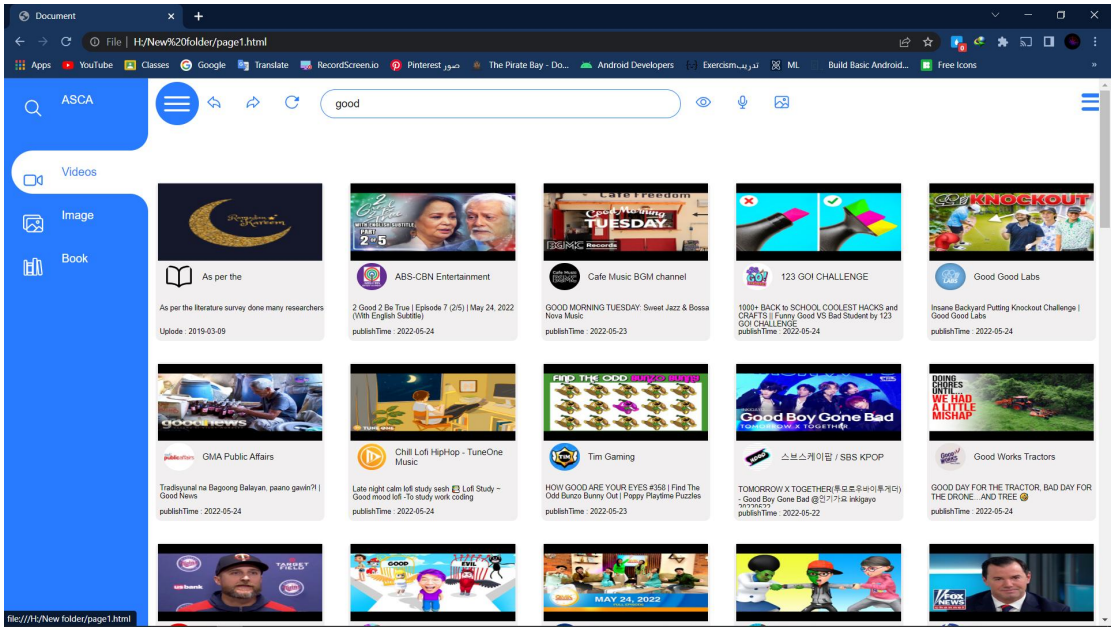
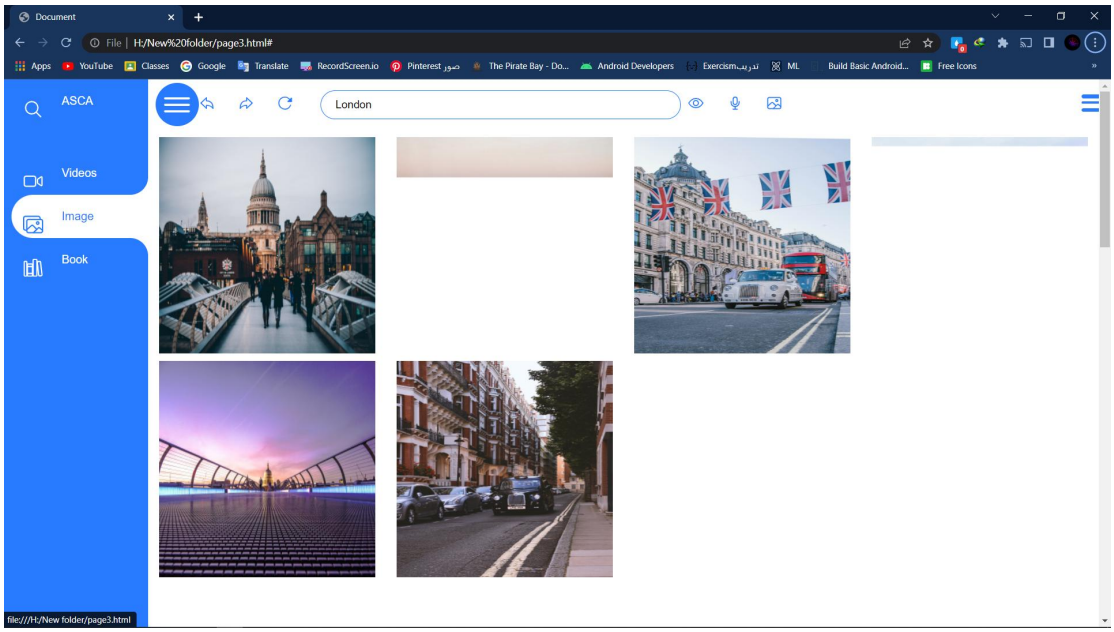
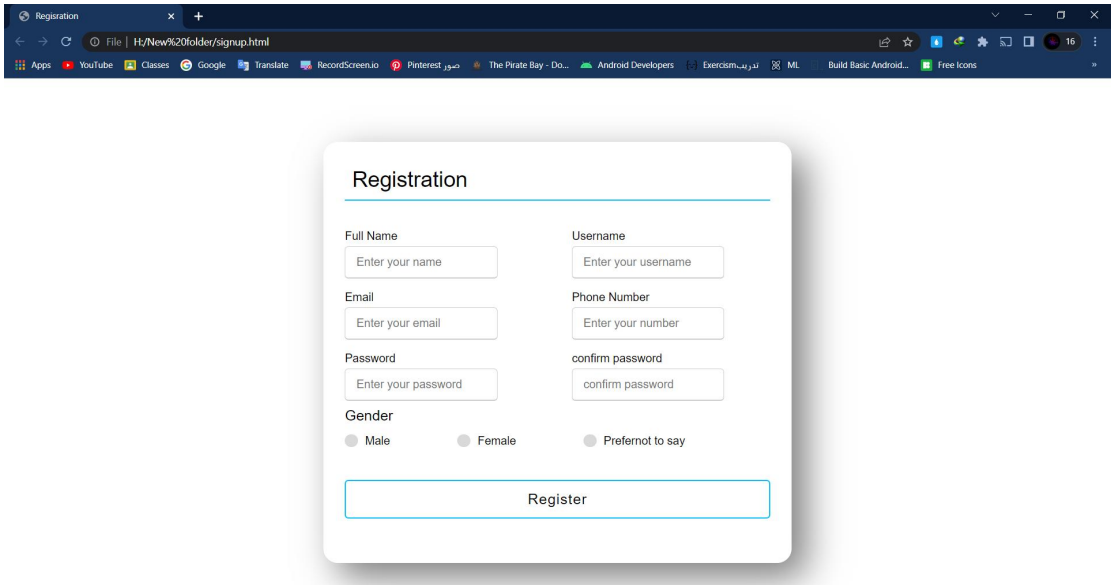
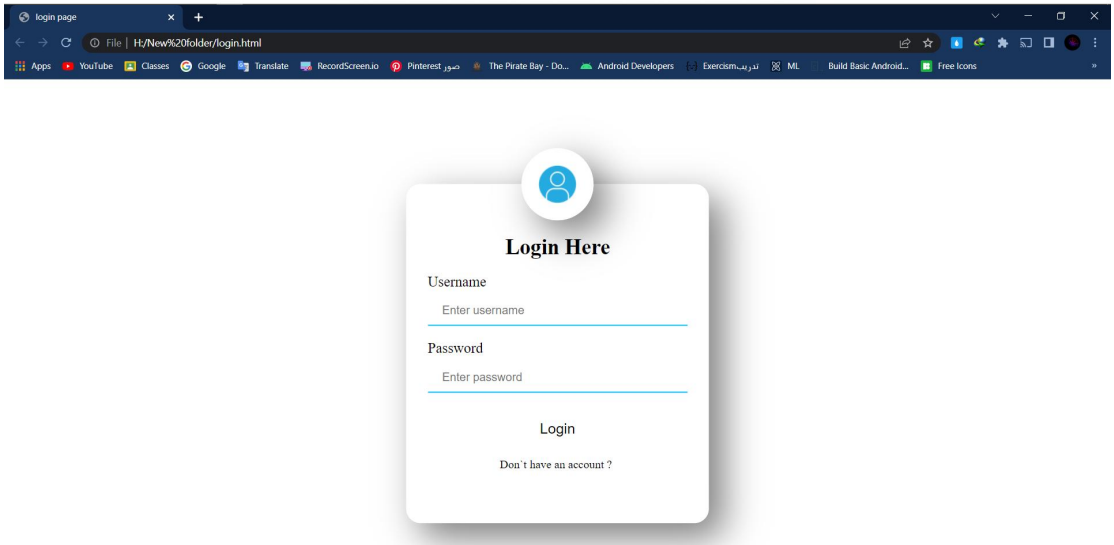


Image page



Login & sign


Regisration
File | H:/New%20folder/signup.html
Registration
Full Name
Enter your name
Username
Enter your username
Email
Enter your email
Phone Number
Enter your number
Password
Enter your password
confirm password
confirm password
Gender
Male
Female
Prefernot to say
Register


login page
File | H:/New%20folder/login.html
Login Here
Username
Enter username
Password
Enter password
Login
Don't have an account ?

Project Report

No	Task	Duration	start	End
1	Preparing problem statement	2 day	05/03/2022	07/03/2022
2	‘ Understanding general requirements	15 day	08/03/2022	22/03/2022
3	Meeting with group members	3 day	27/3/2022	29/3/2022
4	Literature review	2 day	30/3/2022	02/04/2022
5	Design	5 day	03/04/2022	08/04/2022
5	Implementation	30 days	09/04/2022	10/05/2022
6	Testing	4 day	11/05/2022	15/05/2022
7	Preparing presentation	3 day	16/05/2022	19/05/2022
8	Writing up the project report	5 day	20/05/2022	25/05/2022

Project Diealy : Hussein Mahdi

No	Date	From	To	Description
1	Collectproject requirements	15/03/2022 12:00 am	15/03/2022 01:30 pm	The opinion of the users and what they need in a project similar too urswas taken,as well as all the functional and non-hospital requirements were determined,* Bearing in mind that it is subject to change
2	Accountsignin/login interface	12/03/2022 02:00 pm	17/03/2022 12:52 am	Simple interfaces indealing,through which the use rcanregister an account so that he can follow up and get more features
3	A questionnaire	8/03/2022 12:15pm	8/03/2022 1:15pm	We write.We did a simple survey with the students.Through this survey ,we asked some questions to them.What are the most prominent problems you face using Google?
4	finish work from DFD , SDT , ERD , SRS	29/03/2022 9:30 pm	29/03/2022 1:15 am	finish work from One of the main functions One of the non-mainfunctions Domain requirements are Entity/Relationship Diagrams (ERD)State-Trans ition Diagram (STD) Data Flow Diagrams (DFD)
5	project design page	12/04/2022 9:30 pm	12/0/2022 1:30am	Design as the structure of the project and from how many interfaces it consists and details of each interface and its content with the help of the team and I took the opinion of each member

6	project design view page	26/04/2022 11:00 Am	26/04/2022 12:30 Pm	Data display interface design (the structure adopted in data display)
7	project design view page video	26/04/2022 11:30 pm	26/04/2022 3:30am	An interactive interface has been worked out to display video search results, taking into account the user experience
8	Critical Path Method& Activity networks	26/04/2022 10:00 pm	26/04/2022 11:30 pm	A visual layout of the project plan was made, taking into account the time factor and the number of days
9	Completed API Photo Filter	10/5/2022 11:50 pm	11/5/2022 2:00 am	It pulls data for 30 photos from unsplash website
10	The article filter and word translation have been completed	6/5/2022 9:00 pm	10/5/2022 11:00 pm	This filter provides the ability to search within books or search for books as well as a translation of vocabulary
11	Preparing the data base for the project	22/5/2022 8:00 pm	22/5/2022 11:00pm	Work was done to build Database for the purpose of storing user information as well as session time on the site
12	search engine test	24/5/2022 5:30pm	24/5/2022 9:00 pm	The project has been tested in terms of work and consistency of pages among them
13	Prepare research project	23/5/2022 3:00pm	23/5/2022 9:00pm	The research for the project was written from a plan, an idea, requirements, tools, work and theories