

**D.K.T.E. Society's Textile and Engineering Institute,  
Ichalkaranji.**

**(An Autonomous Institute)**

**DEPARTMENT OF INFORMATION TECHNOLOGY**

**2020-2021**



**Project report on  
Snippet Visualization System**

**Under the guidance of  
Prof. S. C. Sagare**

**Group Members**

**RollNo.**

<b>1. Saloni P. Bamane</b>	<b>17UIT11003XX</b>
<b>2. Prajakta K. Bugade</b>	<b>17UIT11006XX</b>
<b>3. Aditi R. Wankhede</b>	<b>17UIT71017XX</b>
<b>4. Kaushik S. Talange</b>	<b>19UIT653XX</b>
<b>5. Harshal S. Shelar</b>	<b>16UIT52050XX</b>
<b>6. Sohel R. Zari</b>	<b>16UIT12058XX</b>

**D.K.T.E. Society's Textile and Engineering Institute,  
Ichalkaranji.**

**(An Autonomous Institute)**

Department of Information Technology

**CERTIFICATE**

This is to certify that

<b>Group Members</b>	<b>RollNo.</b>
<b>1. Saloni P. Bamane</b>	<b>17UIT11003XX</b>
<b>2. Prajakta K. Bugade</b>	<b>17UIT11006XX</b>
<b>3. Aditi R. Wankhede</b>	<b>17UIT71017XX</b>
<b>4. Kaushik S. Talange</b>	<b>19UIT653XX</b>
<b>5. Harshal S. Shelar</b>	<b>16UIT52050XX</b>
<b>6. Sohel R. Zari</b>	<b>16UIT12058XX</b>

have successfully completed the mini project work, entitled,

**Snippet Visualization System**

in partial fulfillment for the award of degree of Bachelor of Technology in Information Technology. This is the record of their work carried out during academic year 2020-2021.

Date:

Place: Ichalkaranji

Prof. S. C. Sagare  
[Project Guide]

Prof.Dr. D. V. Kodavade  
[Head of Department]

Prof.(Dr.) P. V. Kadole  
[Director]

## **DECLARATION**

We the undersigned students of B.Tech I.T. declare that, the project work report entitled **Snippet Visualization System** written and submitted under the guidance of **Prof. S. C.Sagare** is our original work. The empirical findings in this report are based on the data collected by us. The matter assimilated in this report is not reproduction from any readymade report.

Date:

Place: Ichalkaranji.

Signature

Name

- |                       |              |
|-----------------------|--------------|
| 1. Saloni P. Bamane   | 17UIT11003XX |
| 2. Prajakta K. Bugade | 17UIT11006XX |
| 3. Aditi R. Wankhede  | 17UIT71017XX |
| 4. Kaushik S. Talange | 19UIT653XX   |
| 5. Harshal S. Shelar  | 16UIT52050XX |
| 6. Sohel R. Zari      | 16UIT12058XX |

# INDEX

CONTENTS	PAGE
1. Abstract	5
2. Table of contents	6
3. Introduction	7
4. Scope and Limitations	8
5. Need of the project with motivating example	9
6. Objective	10
7. Literature Overview	11
8. System Requirement	12
9. System Designs	13
10. Module	20
11. Testing and Integration	24
12. Installation Guide and User Manual	25
13. Cost Estimation	26
14. Ethics	27
15. Technology Used	28
16. References	29

## ABSTRACT

We propose a visualization technique to display the results of web queries aimed at overcoming such limitations. It combines the neighborhood preservation capability of multidimensional projections with the familiar snippet-based representation by employing a multidimensional projection to derive two dimensional layouts of the query search results that preserve text similarity relations, or neighborhoods. Similarity is computed by applying the cosine similarity over a “bag-of-words” vector representation of collection built from the snippets.

If the snippets are displayed directly according to the derived layout they will overlap considerably, producing a poor visualization. We overcome this problem by defining an energy functional that considers both the overlapping amongst snippets and the preservation of the neighborhood structure as given in the projected layout. Minimizing this energy functional provides a neighborhood preserving two-dimensional arrangement of the textual snippets with minimum overlap. The resulting visualization conveys both a global view of the query results and visual groupings that reflect related results, as illustrated in several examples shown.

## TABLE OF CONTENTS

SR.NO	POINTS
1	Abstract
2	Introduction
3	Scope and Limitations
4	Need of the project with motivating example
5	Objective
6	Literature Overview
7	System Requirement
8	System Designs
9	Implementation
10	Modules
11	Testing and Integration
12	Installation Guide and user Manual
13	Cost Estimation
14	Ethics
15	Technology Used

## INTRODUCTION

Searching for information on the web is routine task to millions of users. The typical procedure consists in providing textual queries to a search engine, which returns a ranked list of textual snippets each containing a content summary and a link to the referred document (or web page). A ranked list of snippets is quite simple, straightforward to interpret, and it turns out to be effective in focused search tasks that require locating a particular web page or document. None the less ,it also has limitations likely to hamper user experience when exploring and analysing search results in other scenarios. In fact, ranked lists fail to provide an overview of the collection retrieved, making it difficult and time consuming to figure out how documents relate content wise. For example, if a user queries Google's search engine on the keywords "jaguar features", the first page returned includes snippets on at least four distinct subjects, namely, the animal, the car brand, a fan club of old Jaguar cars, and a video game console. Surely users may refine the search, however, if he/she needs a global picture there is no other option but navigating through the pages in the list and manually group the snippets according to their topic.

## Scope and Limitations

This project requires a very simple pre-processing step, but some tricky issues remain. For example, setting appropriate values for Luhn's lower and upper cuts in scenarios where little information is available, as it is the case here, is not straightforward and deserves further investigation. In our examples we typically employed a lower cut of three and no upper cut and removed the query terms from the vector representation. Finally, clustering in visual space will produce good results as long as the projection technique does a good job of preserving the relevant neighborhood. Our choice of the LSP method is justified by its being known to perform quite well in terms of neighborhood preservation, which is confirmed by the results. Nonetheless, if this is not the case, the visualization of groups may be misleading. Users may investigate alternatives playing with the number of clusters while observing the visualization.



## **Need of the project with motivating example**

The standard approach of displaying query results as a linear list of snippets is quite effective for most tasks performed by users of search engines. However, when users carry out an exploratory search on a broad topic or subject, linear lists are not so helpful, demanding additional effort towards gathering and mentally organizing the relevant information. The visualization technique introduced in this article, this project, has been designed to assist users in these exploratory scenarios. As such, it is not intended as a substitute for lists of snippets, but rather as an additional resource to improve user experience in specific situations. Therefore, the proposed visualization system aims at helping users to gain a more comprehensive view of the query results, highlighting related documents and web pages while still retaining, as much as possible, the good properties of the conventional list-based paradigm, namely, the rank information and the summary content provided by the snippets.

## **OBJECTIVE**

- To provide visualization based technique to make the navigation easy and more friendly.
- To provide multidimensional projection which helps in greater visualization by plotting the snippets on 2D layout.
- To calculate similarity among document contents.
- To optimize placement of snippets so as to avoid overlapping.

## LITERATURE OVERVIEW

**Erick Gomez-Nieto [1]** obtained a BSc degree (2009) in Informatics Engineering from San Pablo Catholic University - Arequipa, Peru. He holds a MSc degree (2012) in Computer Science from University of São Paulo - SP, Brazil. He was lecturer/researcher at Computer Science School of San Pablo Catholic University, in 2012/2013. Currently, he is working toward the PhD degree at University of São Paulo. His research interests are in Interactive Visualization and Geometry Processing.

**Frizzi San Roman [2]** received her Informatics Engineering degree from the San Pablo Catholic University, Peru, in 2013, and her MSc degree in Computer Science from the University of São Paulo, Brazil, in 2012. She is currently an Assistant Professor at the Computer Science Department of the San Pablo Catholic University. Her research interests are in Visual Text Analytics, Visual Data Mining and Information Visualization.

**Paulo Pagliosa [3]** received the Ph.D. degree in structural engineering from the Escola de Engenharia de São Carlos at Universidade de São Paulo (1998). He is currently an associate professor of computer science in the Faculdade de Computação at Universidade Federal de Mato Grosso do Sul. His research interests include geometric processing, visualization, physics-based animation and general-purpose computation on graphics hardware.

**Maria Cristina F. de Oliveira [4]** received the BSc in computer science from the University of São Paulo, Brazil, in 1985, and the PhD degree in electronic engineering from the University of Wales, Bangor, in 1990 (now Bangor University). She is currently a Professor at the Computer Science Department of the Instituto de Ciências Matemáticas e de Computação, at the University of São Paulo, Brazil, and has been a visiting scholar at the Institute for Visualization and Perception Research at University of Massachusetts, Lowell, in 2000/2001. Her research interests are in Visual Analytics, Visual Data Mining and Information Visualization. She is currently the head of the Computer Science Department at ICMC, and the chief editor of the Journal of the Brazilian Computer Society, published by Springer. She is a member of the ACM, IEEE and of the Brazilian Computer Society.

# SYSTEM REQUIREMENT

## Software Requirement

- **Language used and technology** : php5.x, java script , AJAX

Visual Studio Code

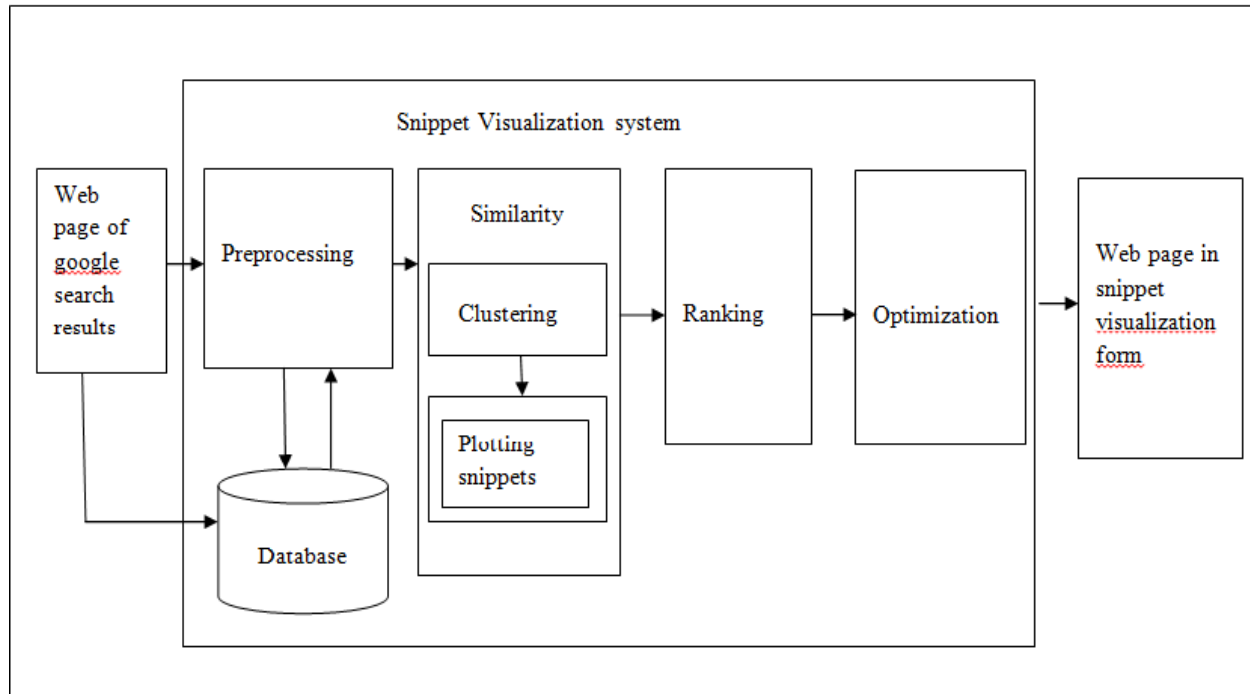
- **Server** : XAMPP
- **Database** : phpMyAdmin

## Hardware Requirements :

- **Processor** : 800MHz Intel Pentium III or equivalent
- **RAM** : 2 GB
- **Disk space** : 750 MB of free disk space
- **Internet connectivity**

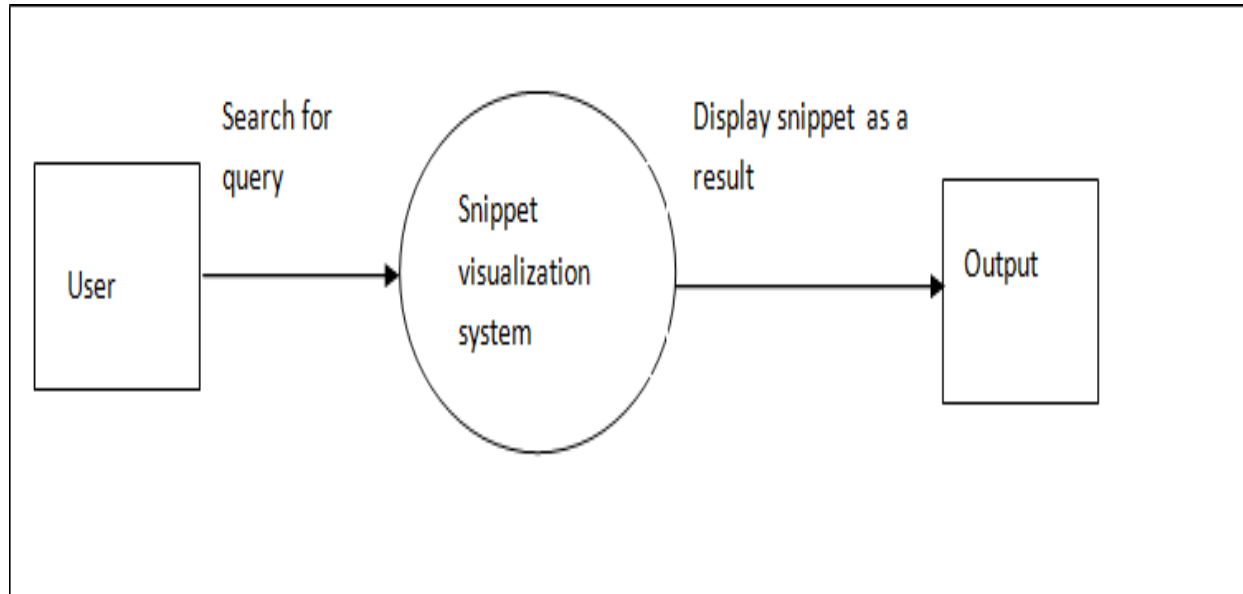
# SYSTEM DESIGNS

Architectural Diagram :-

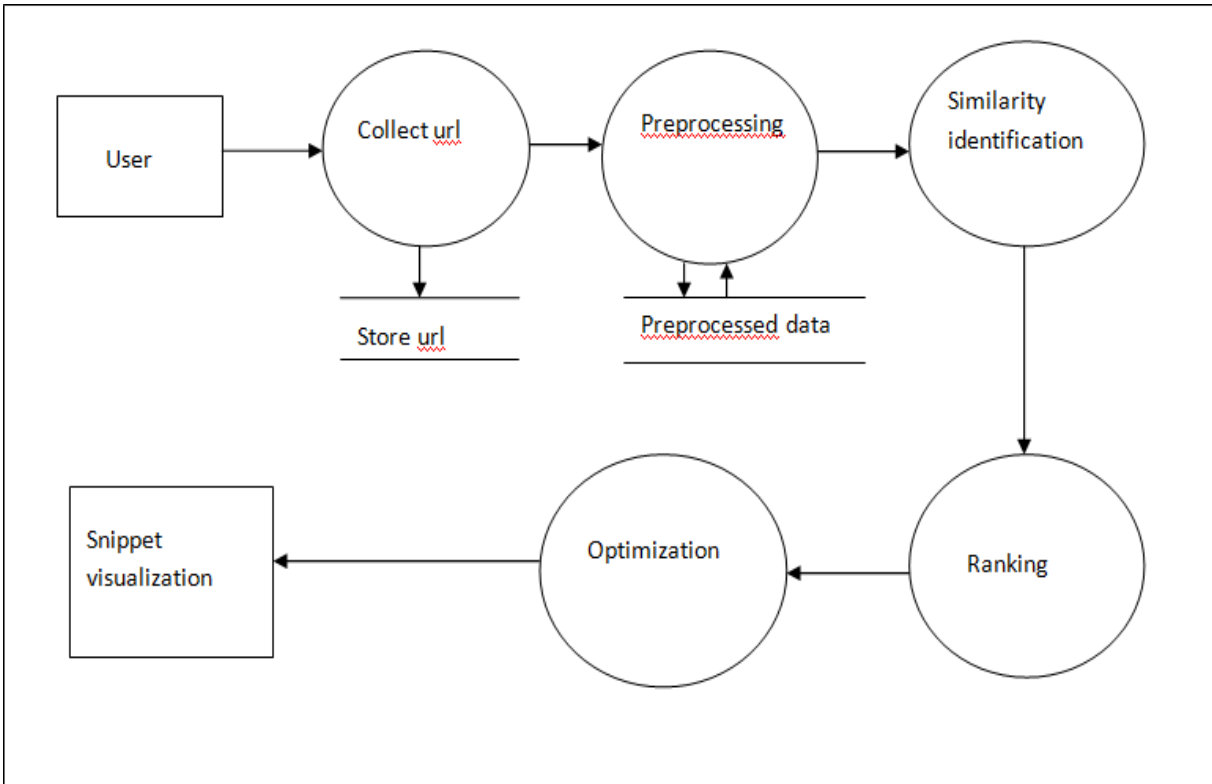


# DATA FLOW DIAGRAM

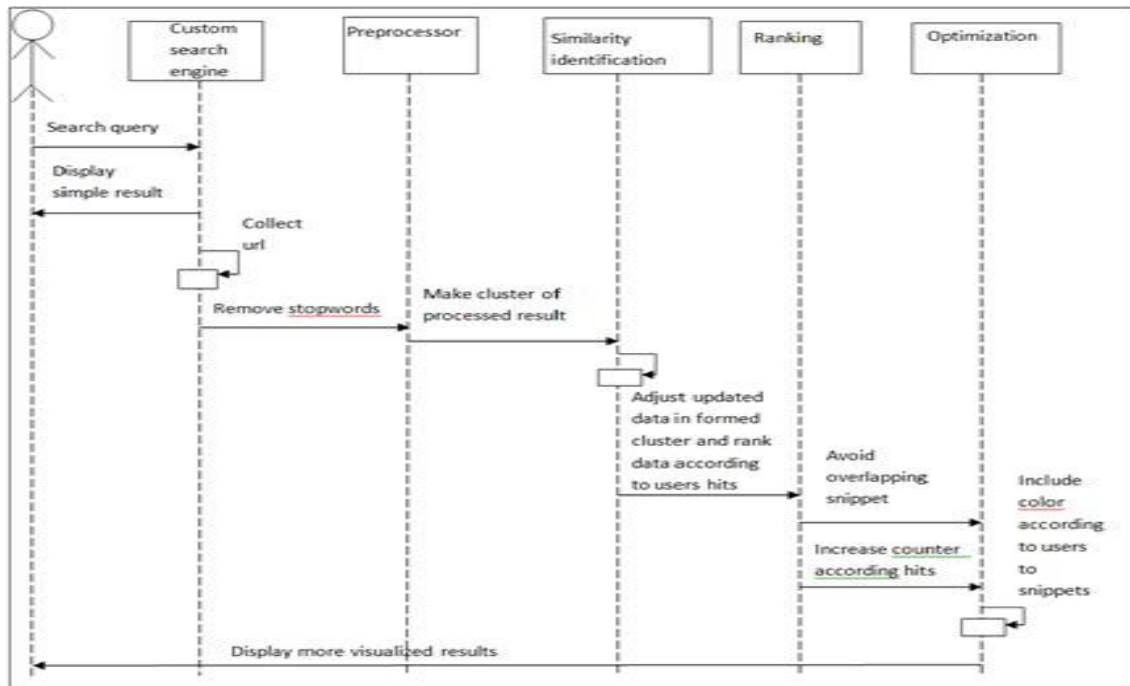
DFD 0:-



## DFD 1:-

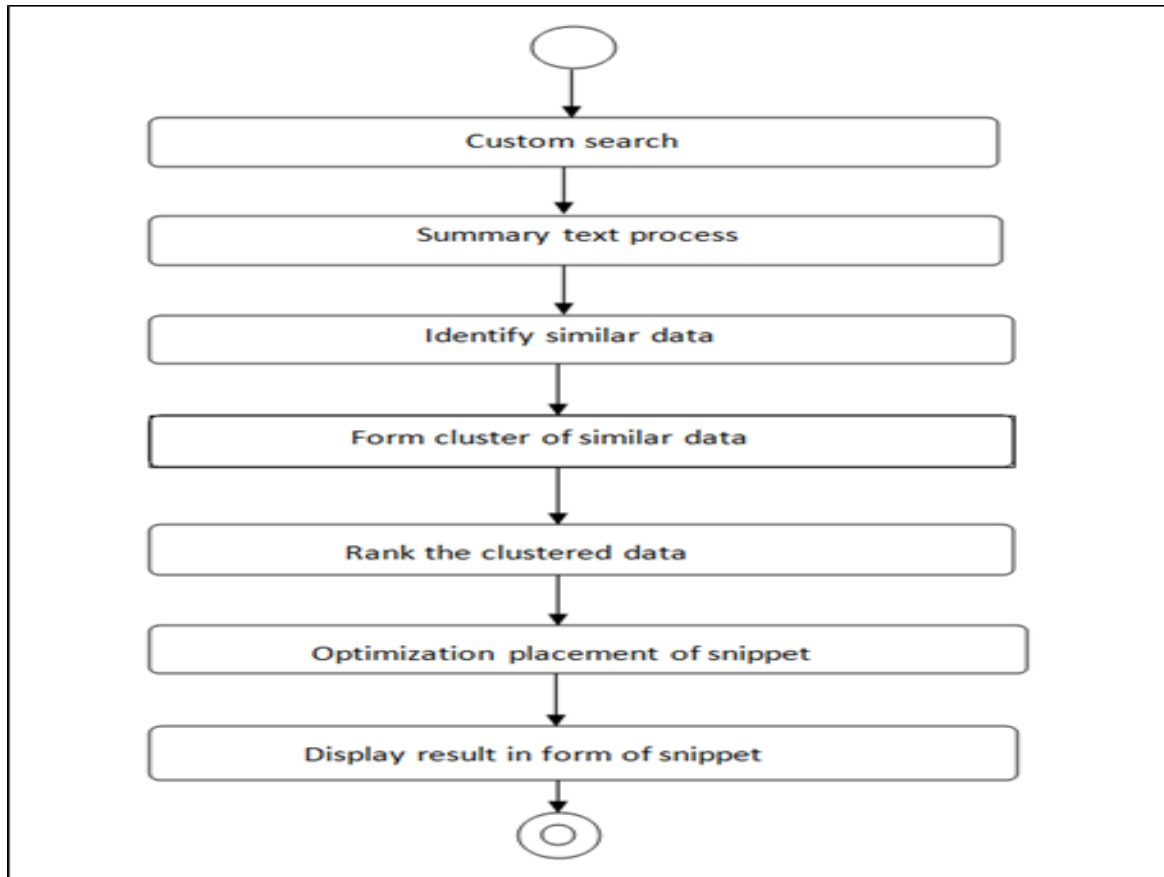


## SEQUENCE DIAGRAM :-

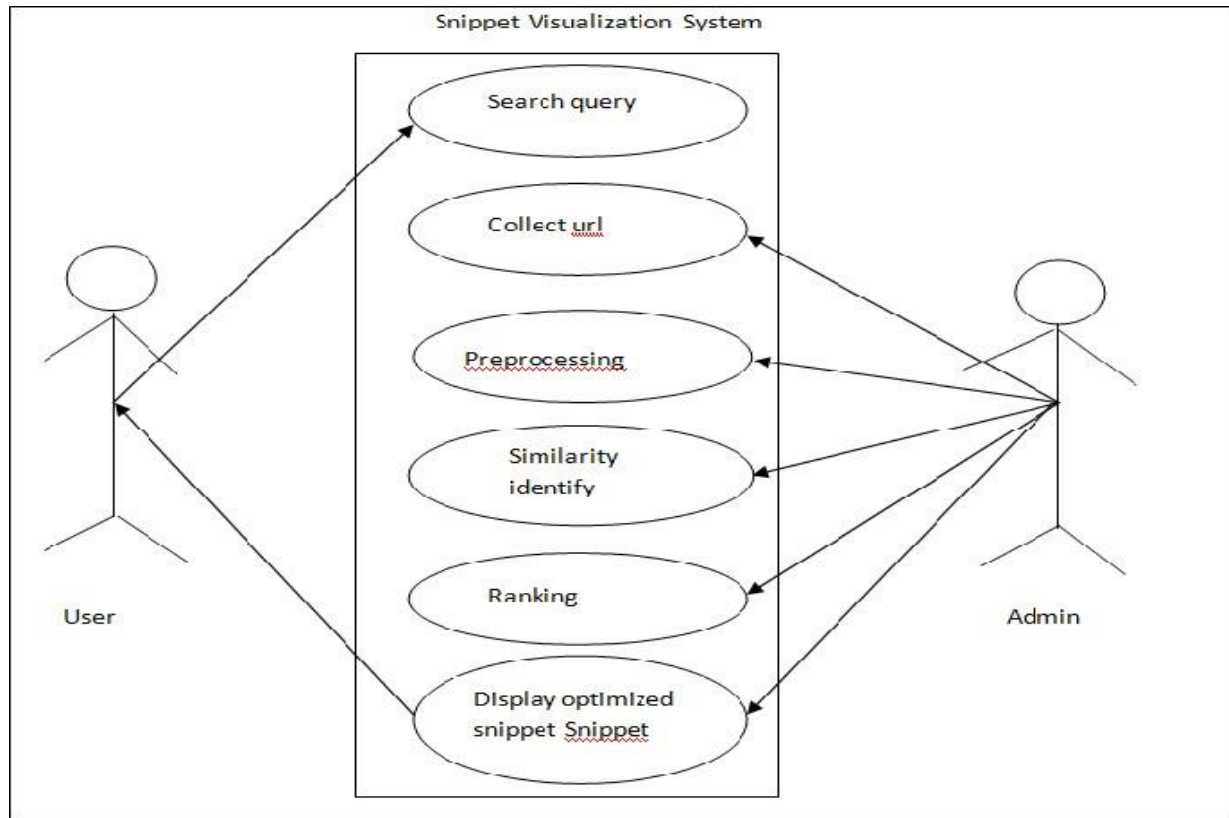




## ACTIVITY DIAGRAM



# USE CASE DIAGRAM



## **IMPLEMETATION**

Once the design is approved technical implementation begins. This often the shortest phase because research and design have been done in advance.

For implementation we have used different web technology ie., php, javascript, ajax and open source database i.e., MySQL

# MODULES

## **1. Collecting information:**

Admin will search for query in the search engine. The URL Will be displayed in the form of snippets by most viewed snippets at the top. Then Admin will save all that queries by clicking on save URL button then all that queries will be saved in the database in the form of text. Also, in the database, the queries will be saved in the specific manner with the respective table and column. After collecting all the URL from the search engine, when user wants some information the resulted queries will be displayed directly which are saved into the database.

## **2. Pre-processing:**

It is the second step where each entry returned from a textual query is processed and its term frequency vector extracted. The resulted queries get extracted from the database and get counted by how many number they are get visited. Here we remove all stop wards (i.e. common words) present in search results. This is necessary to get only exact keywords.

## **3. Similarity Identification:**

Multidimensional projection techniques may be employed to generate visualization that favour the perception of groups of similar documents. Here we have implemented the K Means++ algorithm to plot similar results close to each other. A straight forward way to build layouts where similar documents recovered by a search engine are placed close to each other. This minimization

Process is likely to spoil the neighborhood structure snippets far apart in the final visualization.

## **Ranking:**

In database we have written one field as counter it increases as user clicks on particular results. As users hits a results it automatically increase counter. Those snippet results has highest counter means that user visited this more time than those has less counter. Here, Rectangles height and width are settled to reflect the rank of its corresponding snippets in the retrieved documents list so that better ranked snippets are assigned larger rectangles.

## **5.Optimization:**

Here snippet has colored based on ranking module results as who as highest count has to displayed in darker color than those who have less count. Hence it make easy navigation to user while searching. Here, the high dimensional data is mapped to the visual space and snippets are embedded into rectangles. Optimization is applied to avoid overlap while preserving neighborhood. Smoothness is an important property here as it allows resorting to simple and efficient optimization methods which are mandatory for quick generation of the final visualization.

## TESTING AND INTEGRATION

### Testing Performed

Methods	Time Required
Admin Login	5 sec
Searching Query	10sec
Collecting URL	5 sec
Data Pre-processing	10sec
Similarity Identification	7 sec
Optimization	6 sec

### Output

Collecting URL :- The snippets which are displayed by searching a query get collected in the database by clicking a collect URL button.

Data Pre-processing :- The collected present in the database get pre-processed and get displayed by the number of visits.

Similarity Identification :- The Similar queries get removed in the database if there are some keyword similar to each other.

SrNo	Title	Pre Title	Pre URL	Visite Count
1	Exposure Notifications: Helping fight COVID-19 - Google	Exposure Notifications: Helping fight COVID-19 - Google	https://www.google.com/covid19/exposurenotifications/	0
2	Google	Google	https://www.google.com/	0
3	Apple's 16-inch MacBook Pro is here, and it has a good keyboard ...	Apple's 16-inch MacBook Pro is here, and it has a good keyboard ...	https://www.google.com/.../AFnwnKXmEulHgLL3JdvxLpNVtLxgaXfju24dKIwaNZSoBR8YVaa...	0
4	Google Alerts - Monitor the Web for interesting new content	Google Alerts - Monitor the Web for interesting new content	https://www.google.com/alerts	0
5	Google Podcasts on the App Store	Google Podcasts on the App Store	https://www.google.com/.../AD-FnEzIxeUb4twVuug5zJcFwcVoF3FV0s5kfib203SUNE1TFDYPzQponK6...	0
6	Google Drive - Cloud Storage & File Backup for Photos, Docs & More	Google Drive - Cloud Storage & File Backup for Photos, Docs & More	https://www.google.com/drive/	0
7	Safari User Guide for Mac - Apple Support	Safari User Guide for Mac - Apple Support	https://www.google.com/.../AFIPHzU5AxxVrncWPrMr0qJdJtKZKHfgABinZ9Cj18KeZu3o5ENDrGL...	0
8	Google My Business - Get Your Free Business Listing Stand Out on ...	Google My Business - Get Your Free Business Listing Stand Out on ...	https://www.google.com/business/	0
9	Intelligent Tracking Prevention   WebKit	Intelligent Tracking Prevention   WebKit	https://www.google.com/.../APDk4sN48rCev8UYInfrGpLGiqB3TuU5mpMq4UrxycQ6pLBsxSs5lGm1...	0
10	Google Earth	Google Earth	https://www.google.com/earth/	0
11	Shoes Zone, 114 Banegårdsvej, Glostrup - Google Maps	Shoes Zone 114 Banegårdsvej Glostrup - Google Maps	https://www.google.com/.../search/?...Shoes%20Zone%2C%20114%20Banegårdsvej%2C%20Glostrup...	0
12	Google Webmasters - Resources to Track Website & Search ...	Google Webmasters - Resources to Track Website & Search ...	https://www.google.com/webmasters/	0
13	Liberty Shoes Ltd. (526596 : BOM) Stock Price & News - Google ...	Liberty Shoes Ltd 526596 BOM Stock Price News - Google	https://www.google.com/finance/quote/526596:BOM	0

**Optimization :-** The queries which are collected in the database get optimized and ranked according to the number of visits to that query.

<input type="text" value="apple"/> <input type="button" value="Search"/>		
<b>Apple unveils Apple Watch Series 5 - Apple</b> 10 Sep 2019 ... Apple today announced Apple Watch Series 5, debuting an Always-On Retina display that never sleeps. <input type="button" value="Visit Page!"/> <input type="button" value="Similar Results!"/>	<b>Apple Inc (AAPL : NASDAQ) Stock Price &amp; News - Google Finance</b> Get the latest Apple Inc (AAPL) real-time quote, historical performance, charts, and other financial information to help you make more informed trading and... <input type="button" value="Visit Page!"/> <input type="button" value="Similar Results!"/>	<b>Jony Ive leaving Apple after nearly 30 years to start new design firm ...</b> 27 Jun 2019 ... Apple's chief design officer Jonathan Ive is departing the company, bringing an end to a career of nearly 30 years spent crafting some of... <input type="button" value="Visit Page!"/> <input type="button" value="Similar Results!"/>
<b>Apple's 16-inch MacBook Pro is here, and it has a good keyboard ...</b> 13 Nov 2019 ... Apple has launched a new 16-inch MacBook Pro today, available for pre-order now and in stores later this week. The starting price of \$2399 is... <input type="button" value="Visit Page!"/> <input type="button" value="Similar Results!"/>	<b>Apple unveils Apple Watch Series 5 - Apple</b> 10 Sep 2019 ... Apple today announced Apple Watch Series 5, debuting an Always-On Retina display that never sleeps. <input type="button" value="Visit Page!"/> <input type="button" value="Similar Results!"/>	<b>Apple Inc (AAPL : NASDAQ) Stock Price &amp; News - Google Finance</b> Get the latest Apple Inc (AAPL) real-time quote, historical performance, charts, and other financial information to help you make more informed trading and... <input type="button" value="Visit Page!"/> <input type="button" value="Similar Results!"/>
<b>Jony Ive leaving Apple after nearly 30 years to start new design firm ...</b> 27 Jun 2019 ... Apple's chief design officer Jonathan Ive is departing the company, bringing an end to a career of nearly 30 years spent crafting some of... <input type="button" value="Visit Page!"/> <input type="button" value="Similar Results!"/>	<b>Apple's 16-inch MacBook Pro is here, and it has a good keyboard ...</b> 13 Nov 2019 ... Apple has launched a new 16-inch MacBook Pro today, available for pre-order now and in stores later this week. The starting price of \$2399 is... <input type="button" value="Visit Page!"/> <input type="button" value="Similar Results!"/>	<b>Apple unveils Apple Watch Series 5 - Apple</b> 10 Sep 2019 ... Apple today announced Apple Watch Series 5, debuting an Always-On Retina display that never sleeps. <input type="button" value="Visit Page!"/> <input type="button" value="Similar Results!"/>

# INSTALLATION GUIDE AND USE MANUAL

## Installation :-

- 1) First Install VS Code IDE.
- 2) Then install the XAMPP Server
- 3) Then use phpMyAdmin database in admin panel of Xampp Server.

### 1) Steps to install VS Code-

- 1) download the Visual Studio Code installer for Windows. Once it is downloaded, run the installer.
- 2) accept the agreement and click on next.
- 3) click on "*create a desktop icon*" so that it can be accessed from desktop and click on Next.
- 4) click on the install button.
- 5) after installation completes, click on the finish button, and the visual studio code will get open.

### 2) Steps to download XAMPP Server

- 1) To download the XAMPP server, visit the "[Apache Friends](#)" website in your web browser.
- 2) Click on "XAMPP for Windows". Then, navigate the downloading location and the file will be automatically downloaded.
- 3) Double-click the downloaded file to launch the XAMPP installer.
- 4) "Setup" window will appear on the screen. Then, click on the "Next" button.
- 5) Select the components that you want to install and click on the "Next" button.
- 6) Choose a folder to install the XAMPP and click on the "Next" button.
- 7) "Ready to Install" window will appear on the screen, then click on the "Next" button.
- 8) Click on the "Finish" button.



### 3) Steps to use phpMyAdmin database —

- 1) Navigate to XAMPP in your system or simply launch it by clicking the XAMPP.
- 2) Click on the "**Start**" button corresponding to **Apache** and **MySQL** modules.
- 3) Now click on the "**Admin**" button corresponding to the **MySQL** module
- 4) Click on the "**Database**" tab. Here you can see the **Create** option.

# COST ESTIMATION

## Project Cost

### a. Hardware Cost:

<u>Hardware</u>	<u>Cost</u>
<u>Computer System</u>	<u>Rs. 46000/-</u>
<u>Internet</u>	<u>Rs.800/-</u>
<u>Light Source</u>	<u>Rs. 300/-</u>
<u>Total</u>	<u>Rs.47100/-</u>

In this project the Cost Estimation based on COCOMO (Constructive Cost Model) the formula for the this Model is follows

Effort = Constant  $\times$  (Size) scale factor  $\times$  Effort Multiplier

Effort in terms of person-months

Constant: 2.45 in 1998 based on Organic Mode –

Size: Estimated Size in KLOC–

Scale Factor: combined process factors

Effort Multiplier (EM): combined effort factors

The basic COCOMO equations take the form

Effort Applied (E) =  $ab(KLOC)^b$  [ man-months]

Development Time (D) =  $cb(Effort Applied)^d$  [months]

People required (P) = Effort Applied / Development Time[count]

Where, KLOC is the estimated number of delivered lines (expressed in thousands) of code for project. The coefficients ab, bb, cb and db are given in the following table.

# ETHICS

## Declaration of Ethics

As A Information Technology Student, I believe it is Unethical To,

1. Surf the internet for personal interest and non-class related purposes during classes
2. Make a copy of software for personal or commercial use
3. Make a copy of software for a friend
4. Loan CDs of software to friends
5. Download pirated software from the internet
6. Distribute pirated software from the internet
7. Buy software with a single user license and then install it on multiple Computers
8. Share a pirated copy of software
9. Install a pirated copy of software

## **Technologies Used**

### **1 .PHP:**

PHP (recursive acronym for php: Hypertext Preprocessor) is a widely-used open source general-purpose scripting language that is especially suited for web development.

### **2.JavaScript:**

JavaScript is a lightweight, interpreted programming language. It is designed for creating network-center applications. JavaScript is very easy to implement because it is integrated with HTML.

### **3.AJAX:**

AJAX stands for Asynchronous JavaScript and XML. AJAX is a new technique for creating better, faster, and more interactive web applications with the help of XML, HTML, CSS, and Java Script.

### **4. phpMyAdmin Database :**

phpMyAdmin database is used to save the url after searching a query and also used to retrieve the collected queries. Each table is connected with each other in the database. This database is used by accessing the Xampp server MySQL Admin tab.

# REFERENCES

## 1. Papers :

1. [1] “Similarity Preserving Snippet-Based Visualization of Web Search Results” ErickGomez-Nieto, Frizzi San Roman, Paulo Pagliosa, Wallace Casaca, Elias S. Helou, Maria
2. Cristina F. de Oliveira, and Luis Gustavo Nonato, IEEE transactions on visualization and computer graphics, vol. 20, no. 3, MARCH 2014.
3. “Visualization on Web Snippets”: A Survey Shefali Kedia<sup>1</sup>, Prof. K. P. Wagh, Dr. P.N. Chatur International Journal of Innovative Research in Science, Engineering and Technology. Volume 6, Special Issue 1, January 2017.

## 2. Websites :

1. <https://www.analyticsvidhya.com/blog/2014/11/text-data-cleaning-steps-python>
2. <https://ctrlq.org/code/20076-google-search-api>
3. <https://searchengineland.com/how-to-use-fetch-as-googlebot-like-seo-samurai-214292>

## 3. Books :

1. The Joy of PHP Programming: A Beginner’s Guide – by Alan Forbes
2. PHP & MySQL Novice to Ninja – by Kevin Yank
3. Java the complete reference book by Herbert Schildt