**Chapter 1**

**Introduction**

**1.1 Overview**

Cognitive Web is an amazing website projected to start in year 2014. It is expected to gain many users since it is one of its kind. All you have to do is set up an account in the website, choose from variety of interests and you are done. Then you can browse through thousands of articles of your favourite interests (movies, sports, art, history, computer, gadgets, technology, nature, rock music’s and many more). For website/blog owners, Cognitive Web is great way to find lot of information about their favourite interest. It also allows person to like, share and comment on page they are visiting. Cognitive Web have so many useful features that you can’t miss. So this is just a basic introduction of Cognitive Web.

Cognitive Web discovers web sites based on your interests, learns what you like and brings you more.

Using a combination of human opinions and machine learning to immediately deliver relevant content, Cognitive Web presents only web sites that have been suggested by other like-minded users. Each time the Cog button is clicked, the user is presented with a high quality web site based on the collective opinions of other like-minded web surfers.

Cognitive Web terms itself as a Discovery Engine. That may sound like a scientific term, but it is in fact a website for you to discover new interesting sites. There is no toolbar to download or any software to install. You just need to go to the site, sign up and click on the Cog button (or any of your favourite tags) to start discovering new sites. If you are not happy with the result, simply click on the Cog again to go to the next site.

**1.2 Problem Definition**

Reach and importance of internet is known to all. In present age, due to evolution of social networks, internet has become important part of our life. But use of internet becomes monotonous after some time. Most of the social networking sites does not provide assistance to increase our knowledge and it is tiresome to search for such websites. Thus a website which will help to increase our knowledge across different genre and is easy to use will be cheered. The proposed website known as “COGNITIVE WEB” will provide easy access to numerous pages across different genres. The proposed website will have large database containing the name, rating, genre of each web page. The website will also employ the system known as opinion mining. Opinion mining, as the name suggests, will help to extract the opinion about different web pages visited through structured data like rating as well as unstructured data like comments. The website will have various options to ask for inputs such as areas of interest. If areas of interest are provided then web pages are sorted in such a manner that web pages which fits given criteria are shown first starting and then rest of the pages. Else pages are shown according to ratings. If feedback is given about a web page then rest of the pages are sorted accordingly.

**1.3 Objective/ scope of the project**

**1.3.1 Consistent & quality content**

Consistency is required for any type of computing systems. It is one of those key components which holds the whole system and its users together. Consistency should be the priority in case of any system. The system should be able to provide unnerved performance always.

Content is another important component in any website. But more importantly, it should be relevant content. The user should not only be satisfied, but also happy with the contents provided by the website, and that’s what we are aiming for.

**1.3.2 Service quality standards and maintenance**

The quality of the system is the most important thing which we aim for. We try to set high standards and also achieve them with optimum quality. Because in the end, it is the quality of service that binds any user with the system.

In our project, Database maintenance will be necessary. With thousands of entries in the database, it becomes mandatory to keep a backup of the database and maintain it.

**1.3.3 Easy navigation**

As Leonardo DaVinci said, “Simplicity is the ultimate sophistication”, we will be trying to keep out user interface as simple as possible for easy navigation between the pages.

**1.3.4 Key objectives**

1. To ensure that the content, visual design and navigation of the website is appealing to prospective and current users;
2. To review website content to ensure currency, relevance and general appeal to users
3. To review and improve the website Information Architecture for easier access to essential information for users;
4. Improve the integration with applications that are currently delivered using the website;
5. Ensure that navigation of the website and visual design is consistent throughout.

**1.3.5 Target Audience**

Our current audience comprises of alpha and beta testers, also few people who have had an intense web surfing experience in the past.

We expect our future audience to be people from all kind of community. As we will be trying to cover all general interests, we expect a wide variety of users from all over the internet. So, in conclusion, out audience should reflect to the general concentration of users all around the web.

**Chapter 2**

**Review of Literature**

Many interesting works exist that focus on extracting the opinions from the customer reviews. Some works focus on performing opinion mining to identify the semantic orientation of a review overall, whereas others focus on identifying and extracting the opinion words that will determine the semantic orientation. This line of work further divides into those who focus on the opinion word identification and semantic orientation, and those who also employ features as an additional tool in representing the semantic orientation of a review. Significant information extraction from Web content has been a major focus for many researchers, where different degrees of information, such as words, text passages, or WebObjects, have been taken into account.

[1] Proposed the system that will not only mine the opinions but will also extract useful information related to the item’s features and use it to rate them as positive, neutral, or negative. This feature based opinion mining will help user to focus on features of product he/she is interested in. Authors proposed a system with two algorithms. One to identify and extract the features that are deemed as the most important and characteristic of each review, and one which takes as input these features, assigns ranks to them and decides the final classification of the review as positive, neutral, or negative. [2] Calculate the semantic orientation of words based on their semantic association with pre-determined positive and negative words. [3] Classifies the opinion words individually and then the polarity of the opinion sentence is calculated by combining the individual opinion word polarity. [4] Determines the orientation of the subjective terms by utilizing the term definitions contained in the glossaries and dictionaries. [5] Propose a way to automatically mine the product features and the opinions by integrating the semi-structured and unstructured review sources. In this approach the mining results of the semi-structured reviews are treated as prior knowledge and used as a base to mine opinion and product features from the unstructured source using clustering based approach. It then finally integrates product features and opinions to form feature-opinion pairs. [6] Provides a systematic literature survey that contains a comprehensive overview of recent research trends, advances and challenges. [7] Defines the OM problem. ‘‘Given a set of evaluative text documents D that contain opinions (or sentiments) about an object, opinion mining aims to extract attributes and components of objects that have been commented in each document dЄD and to determine whether the comments are positive, neutral or negative”. [8] Presented an unsupervised learning algorithm for classifying reviews as recommended (thumbs up) or as not recommended (thumbs down).

**Chapter 3**

**Proposed System**

**3.1 Methodology**

**3.3.1 How it works?**

The key to COGNITIVE WEB is that its users can share their views on interesting Web content. Other users can then view that content, through topic-orientated browsing or purposeful searching, and reciprocate the favour by sharing their own views on interesting finds.

COGNITIVE WEB offers a personalized Web experience. At one time or another, every Internet user experiences difficulties getting the intended match from a query entered in a search engine. Narrowing search results can be a challenge, especially when you're not sure what you're looking for. Let's say you'd like to check out some cool photos taken around the world. Type "cool photos" into a regular search engine and you might get some worthwhile links mixed in with some questionable, unrelated or uninteresting content. Try the same thing with COGNITIVE WEB by selecting you interest as photos, and not only will you receive hits related to quality photography sites, the photos will crosscheck with your as well as others interests. A tool like COGNITIVE WEB helps filter Web content into a manageable, customized experience. And the more you indicate what you like, the better COGNITIVE WEB gets at offering content you'll appreciate.

**3.3.2 Terms**

Cog!: Look to the top of your homepage and you'll see a button that says Cog! If you click that button, you're taken to a random, totally new Web site. However, that Web site isn't as random as you might think. Being directed to this new site is the result of a combination of several inputs entered into an algorithm. Your interests and positively rated pages are factors, as well as other information like what your friends and similarly interested users have viewed and rated positively.

Users: Users are one of the most integral part of COGNITIVE WEB community. Because it is their inputs that we use and suggest Websites for other users with similar interests.

Comments: This is the traditional way of opinion mining used by COGNITIVE WEB. User is allowed to enter his/her views on the website and an algorithm is run through the comments to determine whether it is a positive reply or a negative reply. Then, accordingly the rating is saved I the database and the user’s experience is altered accordingly. However, we realize that commenting may be a tedious part on user’s side, so we came out with a new system to perform opinion mining, which is Thumbs-up and Thumbs-down.

Thumbs-up and Thumbs-down: As you are surfing or enjoying conventional Web browsing, you can use special buttons in your COGNITIVE WEB homepage to rate different Web sites you come across: either a thumbs-up (meaning you liked the site) or a thumbs-down (meaning you didn't). This rating system helps COGNITIVE WEB determine the kind of pages you'll enjoy surfing. The pages you've thumbed-up are added to the Pages Liked section of your profile.

**3.3.3 Recommendation Engine**

COGNITIVE WEB is all about site discovery. User clicks on the “Cog!” button and the recommendation engine would return some random site based on the categories user said he/she was interested in. But then the more user uses it, better sites would be sent on his way. This is because it’s not actually random, but rather sites are served up based on a series of processes that go on within the COGNITIVE WEB Recommendation Engine.

As you can see in the chart above, there are two key parts to the Recommendation Engine. There are pages from the topics you marked that interest you and peer endorsed pages. Peer endorsed pages are ones from users who have similar voting habits (giving a site the thumbs up or thumbs down) as you.

When a site is first visited, it is put through both the Classification Engine and the Clustering Engine as shown above. The Classification Engine filters the page by topic and tags. While the Clustering Engine suggests the pages based on users past experience. The Clustering Engine sorts out the votes a site is getting so it can determine which sites are the quality ones that should be served. This engine is a key component which holds the system together.

This all makes for a system of “quality plus relevance”. And as with any of the recommendation engines, the more data you have, the better it’ll perform.

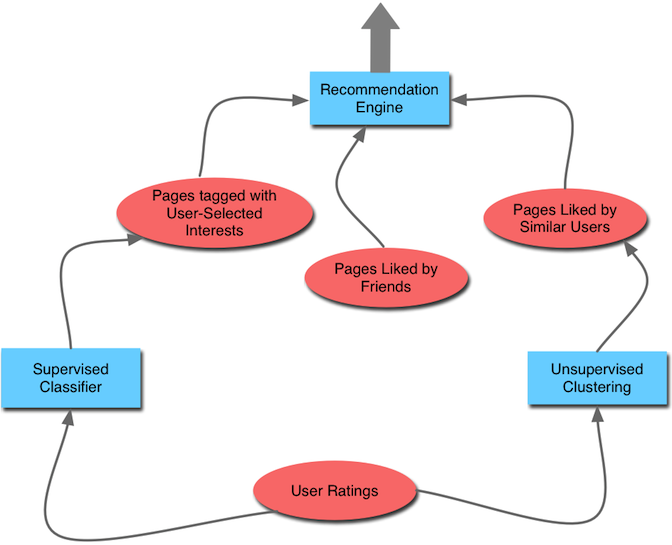


Figure 3.1 Recommendation Engine

**3.2 Algorithm**

**3.2.1 Flow chart**

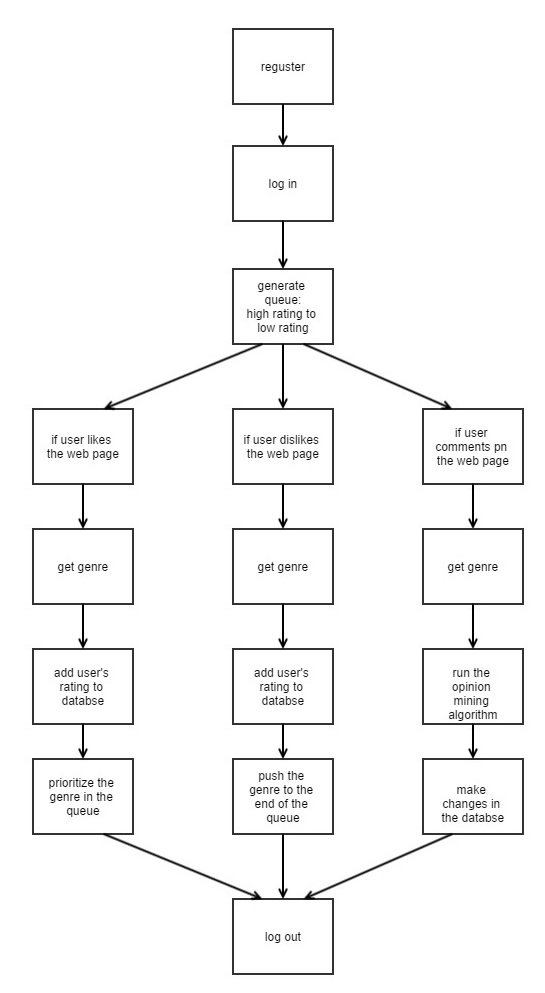
****

Figure 3.2 Flow chart of algorithm

**3.2.2 Pseudo code**

1. User Signs Up

1.1 User Logs In

2. Generate Queue: Priority: High Rated pages to low Rated

2.1: If User likes a Link

2.1.1: G=Get(Genre)

2.1.2: Add User’s rating to the database

2.1.3: Prioritize Genre G in that queue

2.1.4: Display Webpages on click.

2.2: If User dislikes a Link

2.1.1: H=Get(Genre)

2.1.2: Add User’s rating to the database

2.1.3: Push Genre H down to the end of the queue

2.2.4: Display Webpages on click.

2.3: If user comments on a link

2.3.1: J=Get(Genre)

2.3.2: Run Opinion mining algorithm

2.3.3: make changes in the algorithm

2.3.4: Display Webpages on click.

2.4: Else Display Random Webpages.

3. Update and Make copy of the database.

**3.3 System Architecture**

The proposed website will perform opinion mining in structured format as well as unstructured format. The system architecture will be as shown in figure 3.2

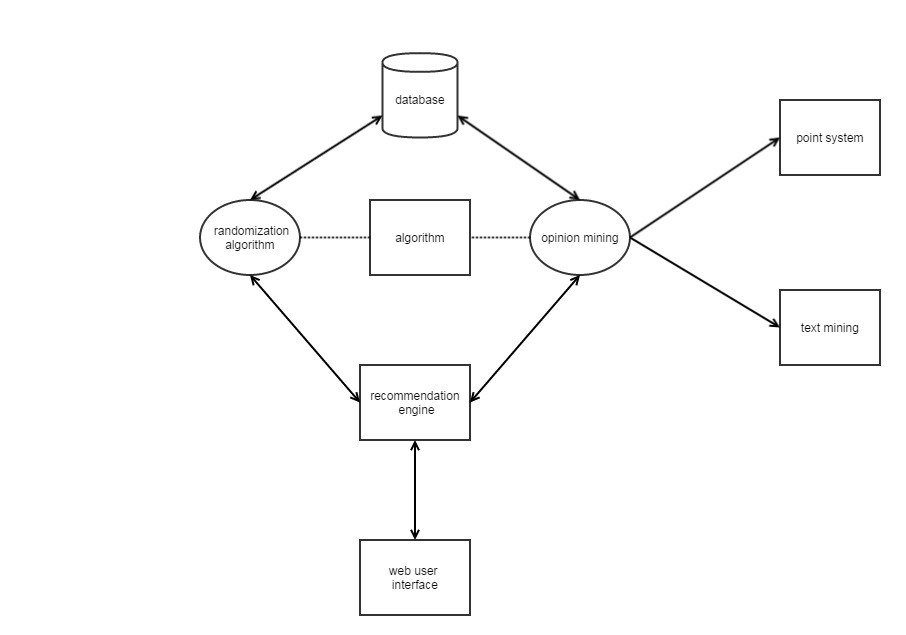


Figure 3.3 System Architecture

Proposed website will have following components:

1. Web user interface
2. Recommendation engine
3. Randomization algorithm
4. Opinion mining
5. Database
6. Web user interface:This component will include a user interface to interact with user. There will be a facility to register and log in. While registering user will provide his/her areas of interest and other necessary information which will be useful for user authentication. User interface will also provide different web pages according to user preference. User can choose to rate the system or comment on the web page or both. After using the web site user can logout of the system.
7. Recommendation engine: Recommendation engine will provide web sites to user either randomly or by the preference given. If the user is new to proposed website and user hasn’t given areas of interest then web pages are shown randomly. If user has given preferences then web sites are shown according to the give preference.
8. Randomization algorithm: Randomization algorithm will contain an algorithm which will randomize the pages shown to the user. If user has given any feedback previously or user has given preferences then algorithm will sort web pages accordingly.
9. Opinion mining: This is the most important component of proposed system. This component will include algorithm for opinion mining. Opinion mining will be of 2 types. First will be point system. Point system will provide buttons to user. One if user likes the page and other one if user doesn’t like the page. Second type of opinion mining will be based on text mining. In this user opinion about particular web page will extracted from unstructured data like comments.
10. Database: This component will hold the information about different web pages shown in the proposed website. This will include name of the web page, link to that web page, genre to which it belongs, user rating. Besides this, each web page will have a unique id to identify them. Database will hold data about users as well such as login information, areas of interest etc.

**3.4 UML Diagrams**

**3.4.1 Use case diagram**

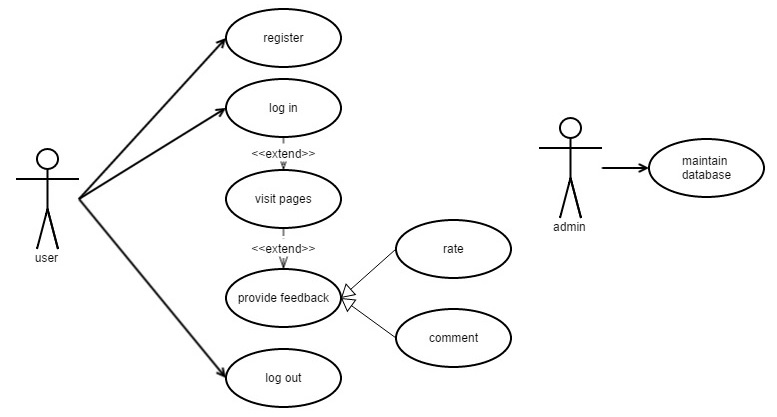


Figure 3.4 Use case diagram

**3.4.2 Sequence Diagram**

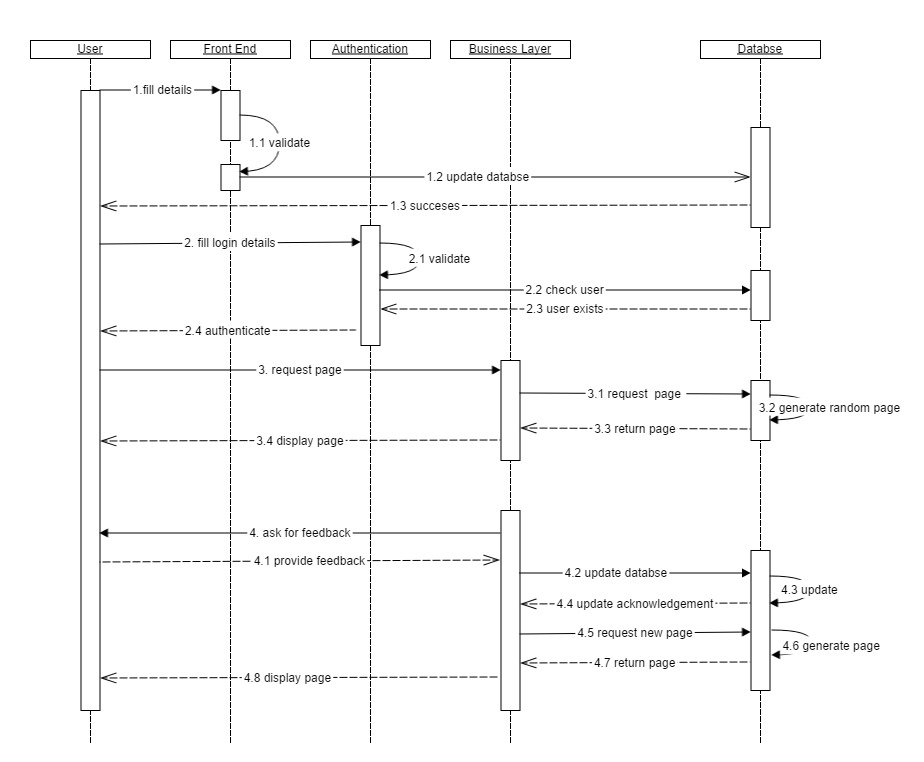
****

Figure 3.5 Sequence Diagram

**3.4.3 Activity Diagram**

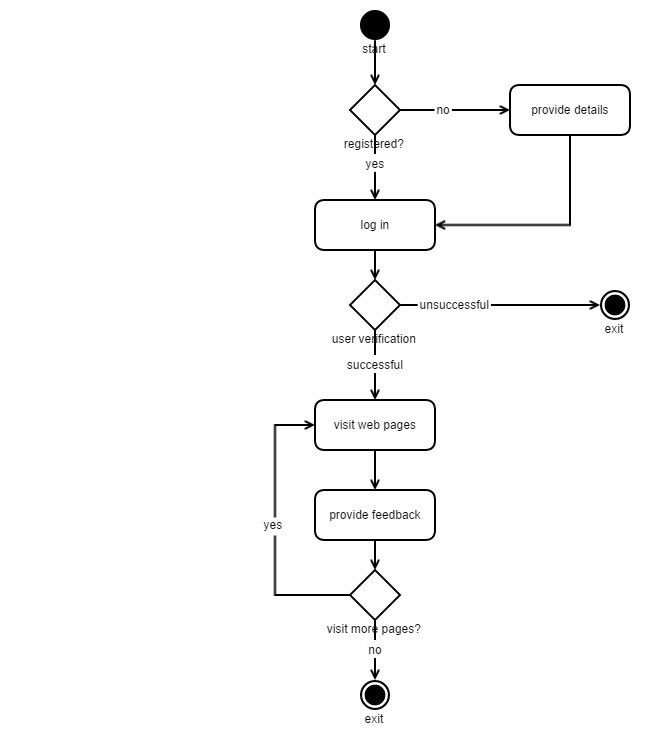


Figure 3.6 Activity Diagram

**3.4.5 State-chart Diagram**

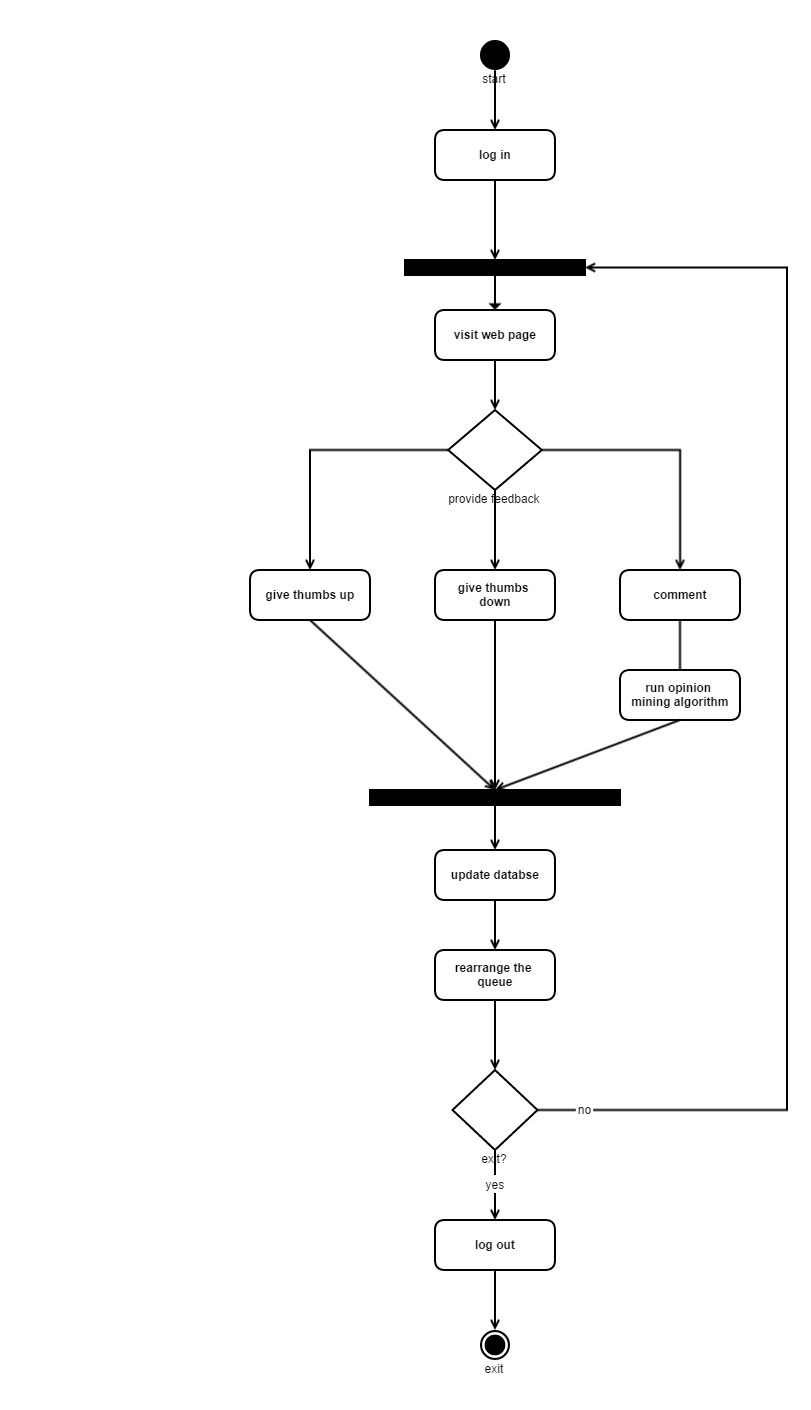
****

Figure 3.7 State-chart Diagram

**3.4.6 Deployment Diagram**

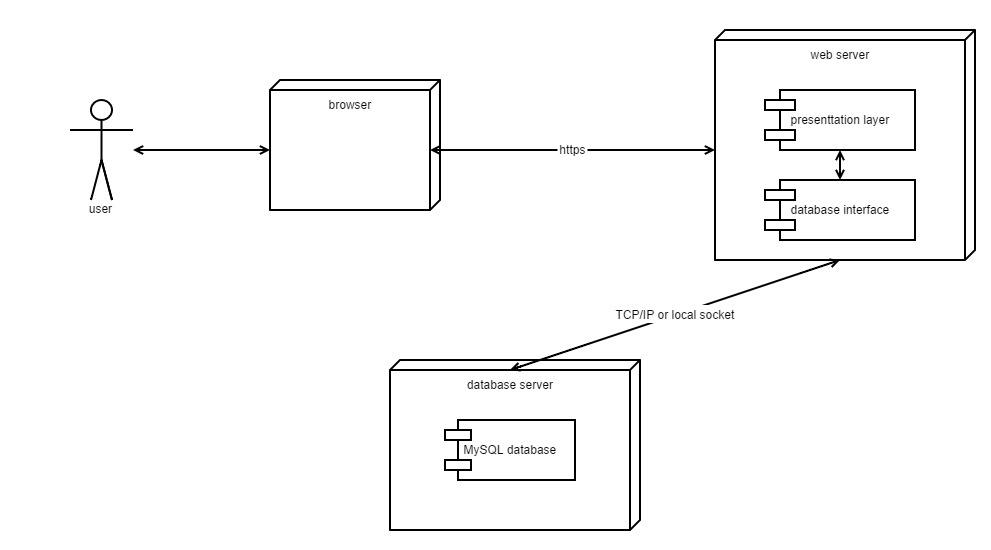


Figure 3.8 Deployment Diagram

**3.5 Software Requirements**

**3.5.1 Internet Connection Requirements**

Best: T1, T3; wired cable; wired DSL; uncongested, strong wireless

Good: congested or weak wireless; satellite; ISDN

Average: dial-up connections, non 24-7 connections

**3.5.2 Minimum Software Requirements**

A version "5"+ browser is required. It should allow "Javascript" or "Active scripting (IE)". It also should check for newer content at every visit and allow requested window to be opened (Opera and some Mozilla class browsers). The following browsers are tested to work with the above settings in effect:

1. On Windows 2000 and XP: Internet Explorer 5.0 and above, Netscape Navigator 6.2 and above, Mozilla 1.4 and above, Mozilla Firefox 0.9.3, and Opear 7.11, All versions of Chrome.
2. On a Mac OS X 10.3 machine: Internet Explorer 5.2.3 and Safari 1.25, All versions of Chrome.
3. On a Linux Redhat 8.0 machine: Mozilla 1.0 and 1.7, and Galeon 1.4, All versions of Chrome.

**3.5.3 Minimum hardware Requirements**

1. 120MHz Power Macintosh or Pentium-class PC
2. 32MB of RAM
3. Monitor displaying Thousands of colors at a resolution of 1024 x 768

**3.6 Database Structure**

**3.6.1 Table Structure for website table**

|  |  |  |  |
| --- | --- | --- | --- |
| Field | Type | Null | Default |
| id | int(11) | Yes | NULL |
| name | varchar(25) | Yes | NULL |
| link | varchar(1000) | Yes | NULL |
| rating | int(3) | Yes | NULL |
| genre | varchar(25) | Yes | NULL |

Table 3.1 Website Table Structure

**3.6.2 Table Structure for comments table**

|  |  |  |  |
| --- | --- | --- | --- |
| Field | Type | Null | Default |
| user | varchar(25) | Yes | NULL |
| link | varchar(100) | Yes | NULL |
| comment | varchar(100) | Yes | NULL |

Table 3.2 Comments Table Structure

**3.6.3 Table Structure for login table**

|  |  |  |  |
| --- | --- | --- | --- |
| Field | Type | Null | Default |
| id | int(11) | Yes | NULL |
| username | varchar(25) | Yes | NULL |
| password | varchar(25) | Yes | NULL |

Table 3.3 Login table structure

**3.7 Screenshots**

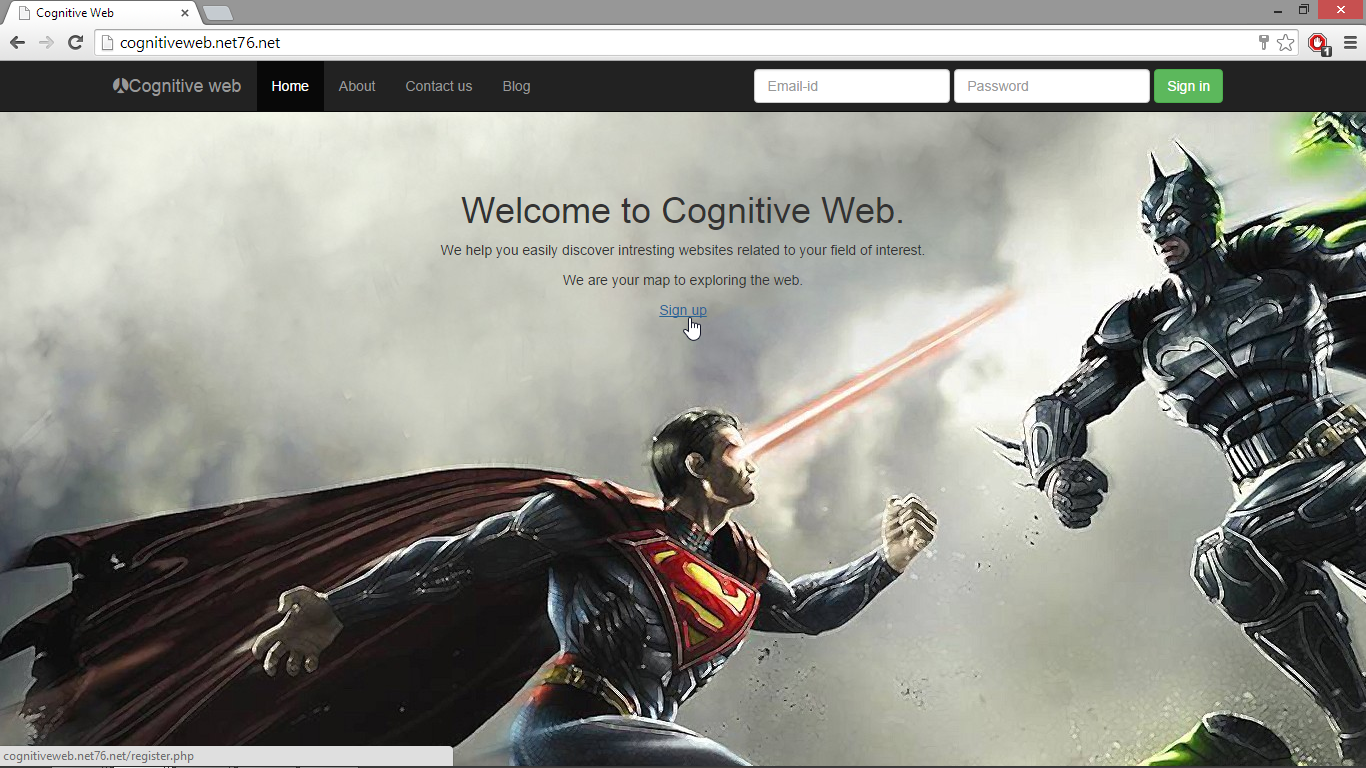


Figure 3.9 Homepage

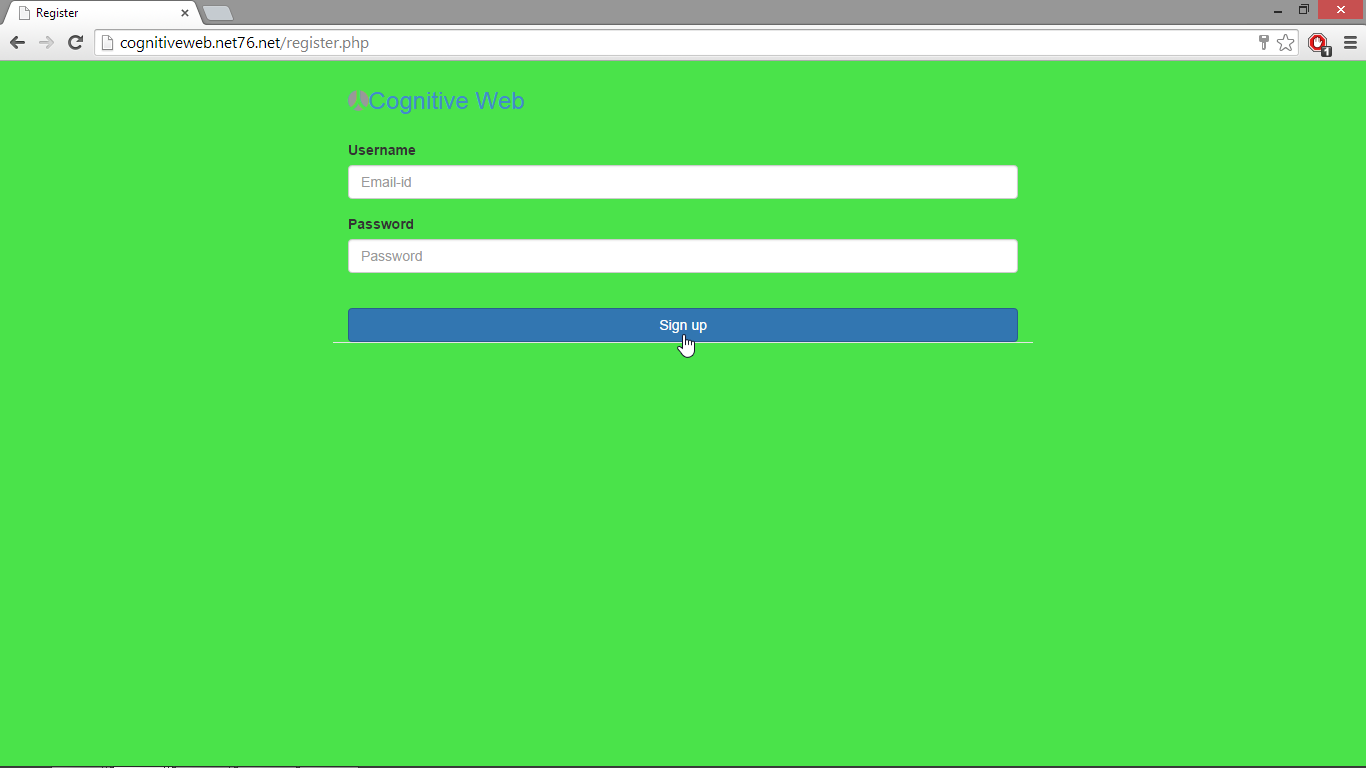


Figure 3.10 Signup page

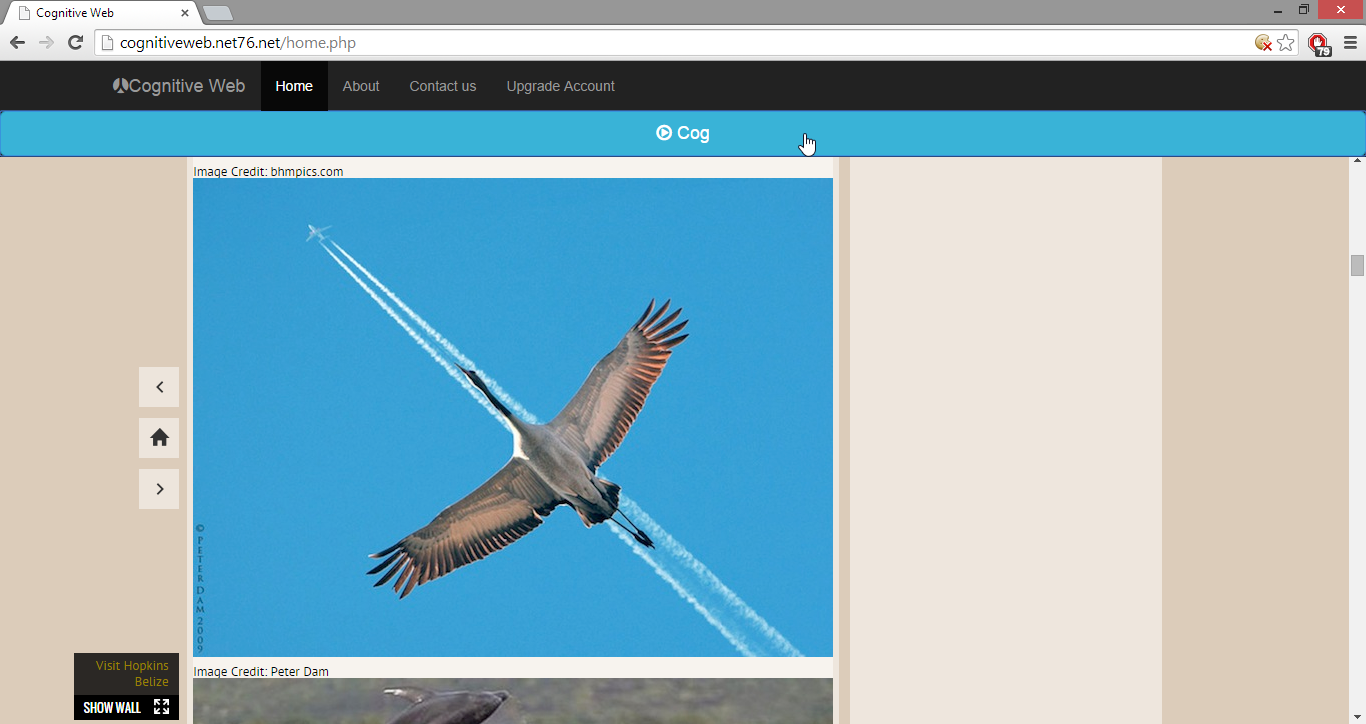


Figure 3.11 Output page 1

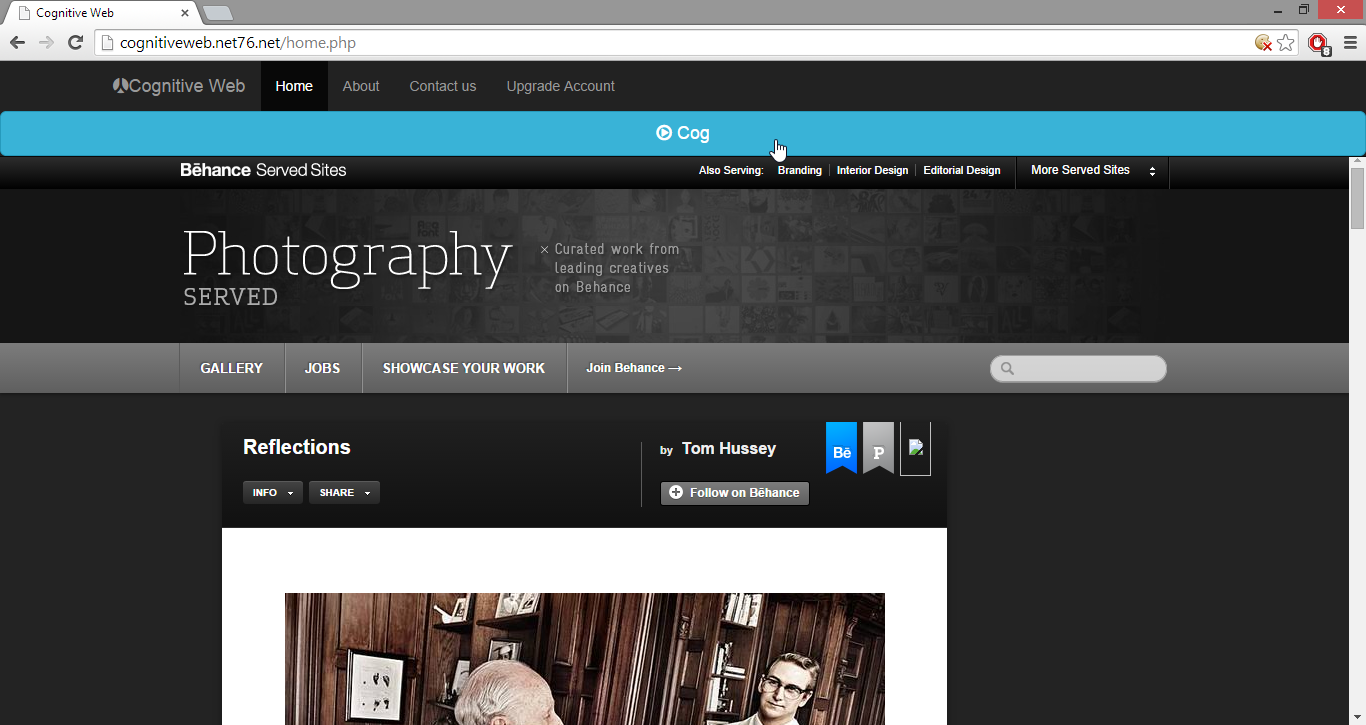


Figure 3.12 output page 2



Figure 3.13 Login Database Table

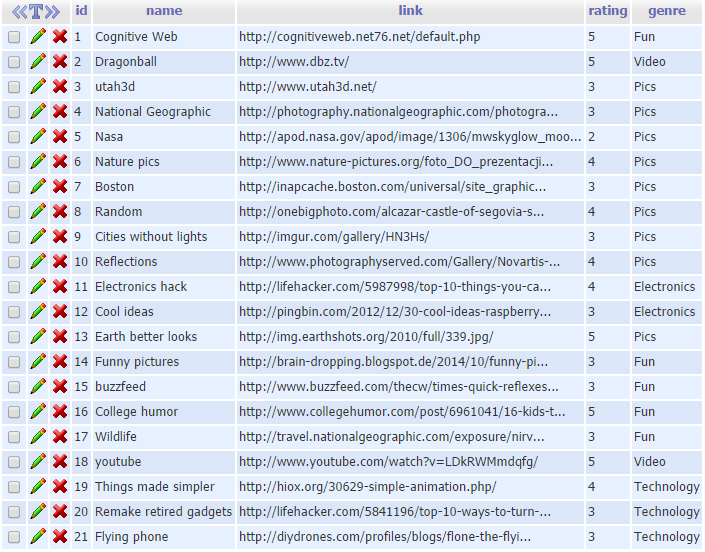


Figure 3.14 Websites Database Table 1

****

Figure 3.15 Websites Database Table 2

**Chapter 4**

**Report on present investigation**

**4.1 StumbleUpon**

Presently there is a website named “StumbleUpon.com” which is the leader in websites recommending engine. It was founded in 2001 and it is currently one of the leading websites on the internet. That is because it is one of a kind. What they basically do is they recommend websites to user from enormous database of websites. It is quite astonishing that a website that sophisticated as theirs has only 75-100 employers. It sets a perfect example for existing start-ups such as ours that it is the idea and conviction that matters, far more than our resources. StumbleUpon has a button named “Stumble” which handles all the page requests from the users. It also allows user to submit their own websites.

**4.2 Pinterest:**

Pinterest is a new competitor of StumbleUpon in content discovery sites. Unlike StumbleUpon, pinterest focuses on images, video, graphics than articles, text. But it have some limitation compare to StumbleUpon. It does not have vast amount of interests like StumbleUpon.

**4.3 Alltop:**

Finding new information from particular website/blog is easy. User can also submit your website to Alltop. All categories are divided in more sub-categories. This depth of categories makes discovery of content easier.

**4.4 Flipora:**

It is one of new content discovery and sharing site you can find. With over 60 topics, flipora provides advanced content discovery to their users. ”Flip” button is used to discover more websites.

**Chapter 5**

**Conclusion and Future scope**

**5.1 Conclusion**

The Internet is just a world passing around notes in a classroom. What are trying is to impart a bit of knowledge in those notes.

The proposed idea is comparatively unique and its scope is wide. Instead of selecting one generic topic, we tried to select a topic which includes many generic ideas.

In this age of media and Internet access, we are much more talkative than ever before. We are planning to use this expressiveness for a good cause, to make people see what they are missing out on.

So, to conclude, the idea is dreamt and the purpose is set, not is the time to pursue that dream.

**5.2 Future Scope**

Our project is still at the prototyping stage. In future, we wish to flawlessly implement all the features from our proposed system with maximum efficiency.

Database expansion: We plan on expanding out database to a minimum of 1500 websites before the final presentation of the project. Such a huge number of websites will really give the users a wide variety of websites.

User reach: We will plan to reach out to as many users as we can. Because, in the end, user interaction plays a key role in our algorithm. We want users to take the front row and carry the show further.

**Chapter 6**

**References**

**Journal Paper,**

[1] Magdalini Eirinaki, Shamita Pisal, Japinder Singh, Feature-based opinion mining and ranking, Journal of Computer and System Sciences (2010), 1-3

[2] P.D. Turney, M.L. Littman, Measuring praise and criticism: Inference of semantic orientation from association, ACM Trans. Inf. Syst. 21 (2003) 315–346.

[4] A. Esuli, F. Sebastiani, Determining the semantic orientation of terms through gloss classification, in: Proceedings of the 14th ACM International Conference on Information and Knowledge Management, CIKM’05, 2005, pp. 617–624.

[5] Q. Miao, Q. Li, R. Dai, An integration strategy for mining product features and opinions, in: Proceeding of the 17th ACM Conference on Information and Knowledge Management, CIKM’08, 2008, pp. 1369–1370

[6] Khairullah Khan, Baharum Baharudin, Aurnagzeb Khan, Ashraf Ullah, Mining opinion components from unstructured reviews, Journal of King Saud University – Computer and Information Sciences (2014), 1-5

[7] B. Liu, M. Hu, J. Cheng, Opinion observer: Analyzing and comparing opinions on the web, in: Proceedings of the 14th International Conference on World Wide Web, WWW’05, 2005, pp. 342–351

**Proceeding Paper,**

[3] S.-M. Kim, E. Hovy, Determining the sentiment of opinions, in: Proceedings of the 20th International Conference on Computational Linguistics, COLING’04, 2004.

[8] Turney Peter, 2002. Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews. Paper presented at the Proceedings of the 40th Annual Meeting on Association for Computational Linguistics, Philadelphia, Pennsylvania, USA

**Websites,**

[9] MG Siegler, There's a method behind StumbleUpon's madness, August 4, 2008, http://venturebeat.com/2008/08/04/theres-a-method-behind-stumbleupons-madness/

[10] "About StumbleUpon.", StumbleUpon.com (28/09/2014),

http://www.stumbleupon.com/about/

[11] “About Pinterest”, pinterest.com (01/10/2014), https://about.pinterest.com/en

[12] “alltop”, alltop.com (10/08/2014), http://alltop.com/

[13] “flipora.com”, flipora.com (09072014), http://login.flipora.com/login.jsp