DEVELOPING SEMANTIC RICH WEB SEARCH ENGINE

Mr. Suneet Joshi , Mr. Ashish Jain , Dr Binod Kumar
suneetjoshi_2000@yahoo.com
Assistant Professor in Medicaps Inst. of Science & Tech, Indore
Associate Professor IET DAVV , Indore
Assoc. Professor in Computer Sc. Dept. at Medi-Caps Institute of Science & Tech. , Indore

Abstract-The amount and variety of information on the World Wide Web (Web) is increasing in an unstructured way. This makes navigation and retrieval of information a difficult and sometimes frustrating process. For instance, Web search engines are becoming less effective as the documents available on the Web proliferate and users' queries return hundreds of links. Users may become lost or frustrated because navigation is unintuitive and semantic meanings are not used to evaluate the relevance of the links, most of which may be unrelated to what the user wants. Gathering useful and interesting information from the Web or discovering knowledge from hypertext data is a problem that may be solved by implementing measures to make Web information understandable by a Web search engine or other types of software [8]. This project discusses about to classify web search data according to the need of web search engine user. Classification leads' the sorting, indexing, combining different results, categorized the output data. More over it provide a help to write query for the search engine.

Keyword: www, web search engine, sorting, indexing, query.

I. INTRODUCTION

A web search engine is designed to search for information on the World Wide Web and FTP servers [1]. The search results are generally presented in a list of results and are often called hits. The information may consist of web pages, images, information and other types of files. Some search engines also mine data available in databases or open directories. Unlike web directories [2], which are maintained by human editors, search engines operate algorithmically or are a mixture of algorithmic and human input.

II. PROBLEM DEFINITION AND PROPOSED SOLUTION

The task is to apply semantic search which seeks to improve search accuracy by understanding searcher intent and the contextual meaning of terms as they appear in the searchable data space, to generate more relevant results.

- Our objective is to increase relevancy in existing web search by gathering useful information from the Web based on user's interest.
- To classify web search data according to the interest and need of web search engine user.
- The aim of introducing semantics and structures to the Web is to enhance the precision of search engines and to enable the use of logical reasoning in Web documents to answer users' queries.

A. Disambiguation

- By giving the query "Jaguar" user may mean either Jaguar car or Jaguar animal.
- By giving the query "Apple" user may mean either Apple company or Apple as a fruit.

B. Proposed Solution

We will develop a search engine which improves search accuracy by taking input from the user about what actually he/she wants to search (ex. "Apple" is either a fruit or a Company). Then our system will filter the contents of web pages, reject irrelevant web pages and displays the web pages which are most relevant and close to query condition to the top of the list.

III. SCOPE

We will develop web search engine that will do filtering and sorting of web pages [3]. We will develop a web search engine that becomes capable of analyzing all the data on the Web – the content, links, and transactions between people and computers.

IV. OBJECTIVE OF PROJECT

- Our objective is to improve retrieval effectiveness in web search by gathering useful information based on user's interest. That is, to identify that portion of the web that is truly relevant to one user's interests.
- To classify web search data according to the interest and need of web search engine user.
- The aim of introducing semantics and structures to the Web is to enhance the precision of search engines.
- The goal is to deliver the precise information queried by a user rather than have a user sort through a list of loosely related keyword results.

V. MY WORK

- I have made enhancement in search engine, in which I have captured search results of Yahoo!, Google and MSN and then filter and refine and then display those search results according to user's interests and inputs.
- A user profile is made and software gives accurate feedback to the information queried by user according to user logged in (web search personalization).

VI. TECHNIQUE USED- RELEVANCE FEEDBACK

- Relevance feedback is one of the techniques for improving retrieval effectiveness [5].
- Steps:

(1) The user first identifies some relevant and some irrelevant documents in the initial list of retrieved documents

(2) The system expands the query q by extracting some additional terms from the sample relevant and irrelevant documents to produce new query q'.

(3) Perform a second round of retrieval.

An enormous improvement can be seen after a single iteration of this technique.

I have used Google's Web search API and Yahoo! BOSS API to display and use search results for my project. It gives the ability to take the search results and re-use them in anyway we want.

VII. SYSTEM ARCHITECTURE

- To implement our system we divide our complete system into two parts; these two different parts work together to form a complete system:

(i) In the first phase we design search engine, knowledge base, and knowledge retrieval and adaptation system.

(ii) In the next phase we work on the results of the system. For this purpose we take input from the user and perform filtering and sorting on the search results [3]

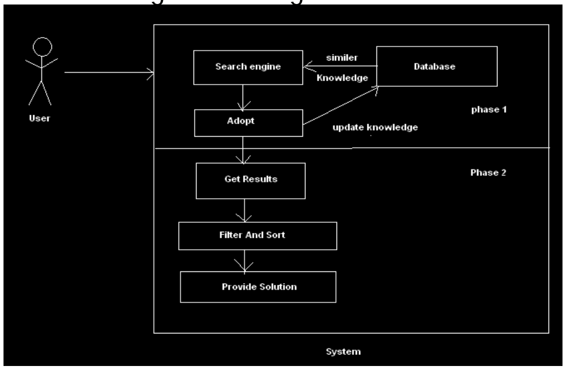


Fig 1 :System Architecture

A. Subsystem Architecture

Above given system is the combination of two phases to achieve the desired goal. These phases are created by the small and complete subsystems. In this section we discuss each of them.

Phase 1: This phase involve making a better search by writing the effective query using semantic knowledge.

a. Search engine: it is an existing web search engine to make a search for the user. User directly interacts with this.

b. Database: it is a knowledge base designed by us to store the user query and its use. This is for knowledge storage and retrieval for the system.

c. Adopt: it is an algorithm to fetch the similar knowledge from the database and provide help to the user to write a effective query for it. It is also used for make changes over the knowledge to design the representation of search results [6, 7].

Phase 2: in this phase we are concentrating on the result produced by the search engine and make it easy to use for the user.

d. Get results: we are collecting all output results on the temporary table to make effective and fast searchable results.

e. Filter and sort: in this step we get input a table that holds results temporarily and categorized and sort according to user need [4].

f. Provide solution: this is final step of our application system in this step we are map our results on the user screen.

B. Algorithm Used

1. Add Knowledge to database

```
(i.)   Get a phrase by user (User Query).
(ii.)  Add to different search engine
(iii.) Check the phrase in knowledge base
         Check page hits on phrase
         Else
        Add to knowledge base and page hits.
```

2. pseudocode

```
(i.)    Get inputphrase(User Query)
(ii.)   Parse into words
(iii.)   Apply on existing search engine
```

```
if (WordFound)
        {
          if (PageFound)
           {
             ListAllPages.Rankwise ();
           }
          else
           {
             add PageToKnowledgeBase();
           }
                update PageRanks ();
          }
        else
          {
            add new word to KnowledgeBase ();
            update pages to KnowledgeBase ();
          }
```

(v.) List predictive results as well as existing search results.

C. Implementation Steps

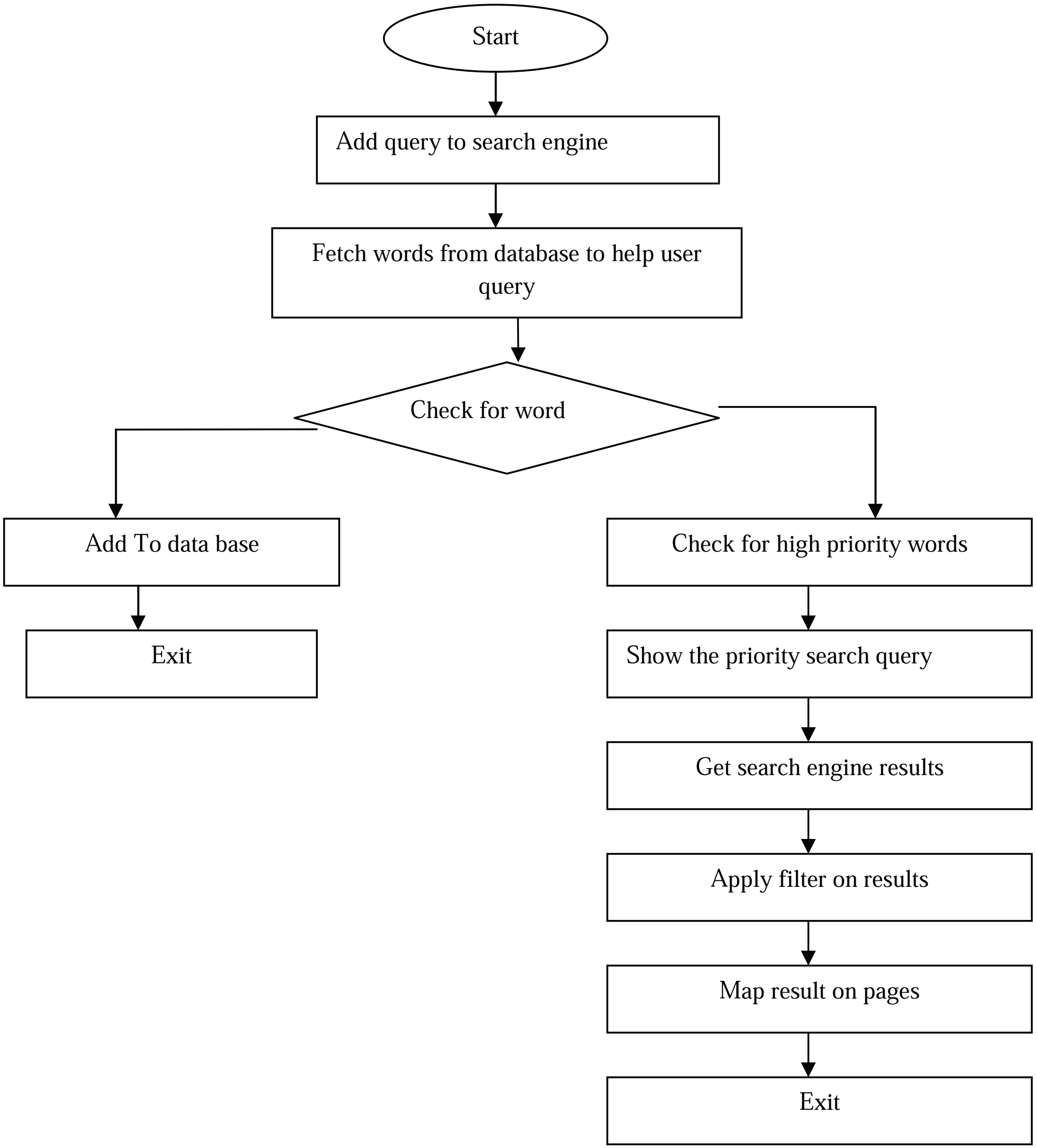
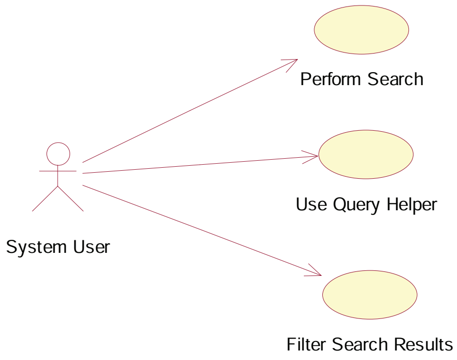

Start
Add query to search engine
Fetch words from database to help user query
Check for word
Add To data base
Exit
Check for high priority words
Show the priority search query
Get search engine results
Apply filter on results
Map result on pages
Exit

Fig 2 :Implementation Steps

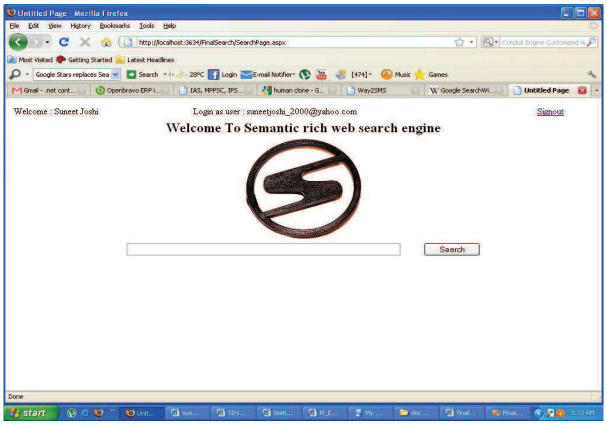
In above diagram we describe the basic steps to implement our application. When user want to make query to the search engine then according to the user query we will show the most common searches related to these word. These words are stored in database and using the probability function we get the direction of search. If word related to the user defined query is not found in database we add it to the database for future use. After that search engine return a result from web we store it to the temporarily and sort and categorized according to the user's last query.

VIII. USE CASE DIAGRAM



IX. SCREEN AND REPORT FORMATS

1. Home Page: there is an option for user to choose any search engine from yahoo, google, MSN and/or or optimize search engine.

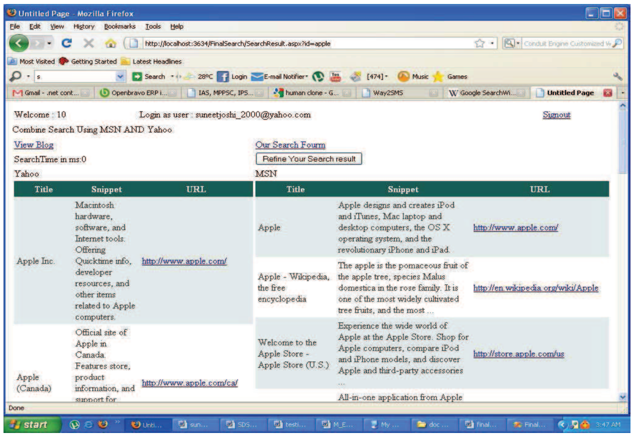

Untitled Page - Mozilla Firefox
File Edit View History Bookmarks Tools Help
http://localhost:3634/FinalSearch/SearchPage.aspx
Most Visited Getting Started Latest Headlines
Google Stars replaces Sea Search 28°C Login E-mail Notifier [474] Music Games
Gmail - .net cont... Openbravo ERP i... IAS, MPPSC, IPS... human clone - G... Way2SMS Google SearchWi... Untitled Page
Done
start

Welcome : Suneet Joshi Login as user : suneetjoshi_2000@yahoo.com Signout

Welcome To Semantic rich web search engine

Search

Snapshot 1: our optimized search engine


Untitled Page - Mozilla Firefox
File Edit View History Bookmarks Tools Help
http://localhost:3634/FinalSearch/SearchResult.aspx?id=apple
Most Visited Getting Started Latest Headlines
Search 28°C Login E-mail Notifier [474] Music Games
Gmail - .net cont... Openbravo ERP i... IAS, MPPSC, IPS... human clone - G... Way2SMS Google SearchWi... Untitled Page
Done
start
3:47 AM

Welcome : 10 Login as user : suneetjoshi_2000@yahoo.com Signout

Combine Search Using MSN AND Yahoo

View Blog Our Search Fourm

SearchTime in ms:0 Refine Your Search result

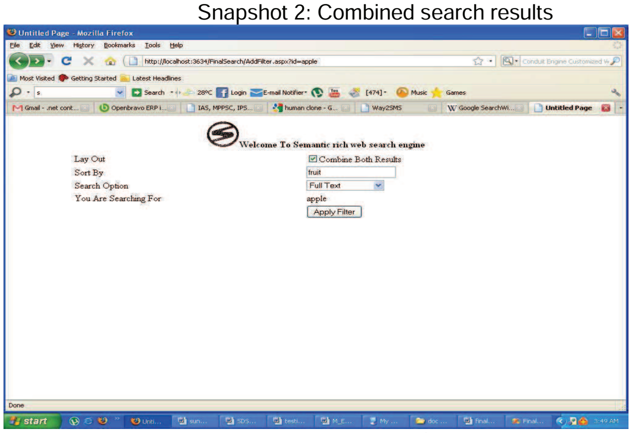
Yahoo

Title	Snippet	URL
Apple Inc.	Macintosh hardware, software, and Internet tools. Offering Quicktime info, developer resources, and other items related to Apple computers.	http://www.apple.com/
Apple (Canada)	Official site of Apple in Canada. Features store, product information, and support for	http://www.apple.com/ca/

MSN

Title	Snippet	URL
Apple	Apple designs and creates iPod and iTunes, Mac laptop and desktop computers, the OS X operating system, and the revolutionary iPhone and iPad.	http://www.apple.com/
Apple - Wikipedia, the free encyclopedia	The apple is the pomaceous fruit of the apple tree, species Malus domestica in the rose family. It is one of the most widely cultivated tree fruits, and the most ...	http://en.wikipedia.org/wiki/Apple
Welcome to the Apple Store - Apple Store (U.S.)	Experience the wide world of Apple at the Apple Store. Shop for Apple computers, compare iPod and iPhone models, and discover Apple and third-party accessories ...	http://store.apple.com/us
	All-in-one application from Apple	

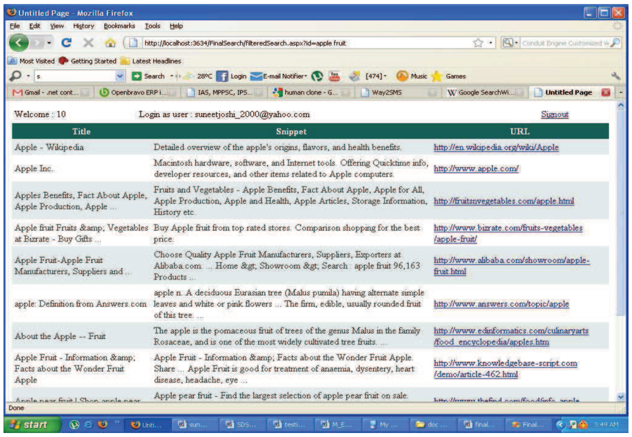
Snapshot 2: Combined search results


Untitled Page - Mozilla Firefox
File Edit View History Bookmarks Tools Help
http://localhost:3634/FinalSearch/AddFilter.aspx?id=apple
Most Visited Getting Started Latest Headlines
Search 28°C Login E-mail Notifier [474] Music Games
Gmail - .net cont... Openbravo ERP i... IAS, MPPSC, IPS... human clone - G... Way2SMS Google SearchWi... Untitled Page
Done
start

Welcome To Semantic rich web search engine

Lay Out ☑ Combine Both Results
Sort By fruit
Search Option Full Text
You Are Searching For apple
Apply Filter

Snapshot 3: Refine results


Untitled Page - Mozilla Firefox
File Edit View History Bookmarks Tools Help
http://localhost:3634/FinalSearch/filteredSearch.aspx?id=apple fruit
Most Visited Getting Started Latest Headlines
Search 28°C Login E-mail Notifier [474] Music Games
Gmail - .net cont... Openbravo ERP i... IAS, MPPSC, IPS... human clone - G... Way2SMS Google SearchWi... Untitled Page
Done
start
3:49 AM

Welcome : 10 Login as user : suneetjoshi_2000@yahoo.com Signout

Title	Snippet	URL
Apple - Wikipedia	Detailed overview of the apple's origins, flavors, and health benefits.	http://en.wikipedia.org/wiki/Apple
Apple Inc.	Macintosh hardware, software, and Internet tools. Offering Quicktime info, developer resources, and other items related to Apple computers.	http://www.apple.com/
Apples Benefits, Fact About Apple, Apple Production, Apple ...	Fruits and Vegetables - Apple Benefits, Fact About Apple, Apple for All, Apple Production, Apple and Health, Apple Articles, Storage Information, History etc.	http://fruitsnvegetables.com/apple.html
Apple fruit Fruits &amp; Vegetables at Bizrate - Buy Gifts ...	Buy Apple fruit from top rated stores. Comparison shopping for the best price.	http://www.bizrate.com/fruits-vegetables/apple-fruit/
Apple Fruit-Apple Fruit Manufacturers, Suppliers and ...	Choose Quality Apple Fruit Manufacturers, Suppliers, Exporters at Alibaba.com. ... Home &gt; Showroom &gt; Search : apple fruit 96,163 Products ...	http://www.alibaba.com/showroom/apple-fruit.html
apple: Definition from Answers.com	apple n. A deciduous Eurasian tree (Malus pumila) having alternate simple leaves and white or pink flowers ... The firm, edible, usually rounded fruit of this tree. ...	http://www.answers.com/topic/apple
About the Apple -- Fruit	The apple is the pomaceous fruit of trees of the genus Malus in the family Rosaceae, and is one of the most widely cultivated tree fruits. ...	http://www.edinformatics.com/culinaryarts/food_encyclopedia/apples.htm
Apple Fruit - Information &amp; Facts about the Wonder Fruit Apple	Apple Fruit - Information &amp; Facts about the Wonder Fruit Apple. Share ... Apple Fruit is good for treatment of anaemia, dysentery, heart disease, headache, eye ...	http://www.knowledgebase-script.com/demo/article-462.html
Apple pear fruit | Shop apple pear	Apple pear fruit - Find the largest selection of apple pear fruit on sale.	http://www.thefind.com/food/info-apple

Snapshot 4: Final Result

X. CODING FOR FINAL SEARCH (FILTER SEARCH RESULTS):

```
Imports net.bing.api
Imports System.Data
Partial Class filteredSearch
    Inherits System.Web.UI.Page
    Const AppId As String =
"A68DE51D46856A4F8B0E4AE5C6708E2DAFF68810"
    Public m_Resuls1 As New DataTable("MSN")
    Public dt As New DataTable("a")
    Public dtl As New DataTable("b")
    Public dtN As New DataTable("c")
    Protected Sub Page_Load(ByVal sender As Object, ByVal e
As System.EventArgs) Handles Me.Load
        Try
            If Session("User") = Nothing Then
                Response.Redirect("Login.aspx")
            End If
            Dim str As String
            str = Request.QueryString("id")
            Dim str1 As String
            str1 = Request.QueryString("type")
            Label1.Text = Session("User").ToString
            Label2.Text = Session("UName").ToString
            Using service As BingService = New BingService
                Try
                    If str1 = "Relavent Search" Then
                        Dim request As SearchRequest =
BuildRequest(str, str1)
                        m_Resuls1.Columns.Add("title")
                        m_Resuls1.Columns.Add("snippet")
                        m_Resuls1.Columns.Add("url")
                        Dim response As SearchResponse =
service.Search(request)
                        Dim builder As New System.Text.StringBuilder
                        Dim result As WebResult
                        For Each result In response.Web.Results
                            Dim row As Data.DataRow
                            row = m_Resuls1.NewRow
                            row("title") = result.Title
                            row("snippet") = result.Description
                            row("url") = result.Url
                            m_Resuls1.Rows.Add(row)

                        Next
                        GridView1.DataSource = m_Resuls1
                        GridView1.DataBind()


    End Sub
```

XI. CONCLUSION

Search results of Google, Yahoo! And MSN search engines are filtered and refined. The precision of search engine is enhanced. The information queried by the user is precisely and unambiguously delivered. We are making enhancement according to user interest. Adding feed back to the system. Using intelligent data base refine search according to user data.

XII. FUTURE ENHANCEMENT

We can develop an improved rank algorithm for search engines based on domain ontology and categorization technology in order to make the algorithm simulate the actual user behaviors in browsing web pages more accurate. More iteration can be provided in order to improve search results (tree form) [until user satisfaction].

REFERENCES

[1] *Rakesh Agrawal, Thomas Imielinski, and Arun Swami. Mining associations between sets of items in massive databases. In Proc. of the ACM SIGMOD Int'l Conference on Management of Data, 1993.*

[2] *Sergey Brin and L. Page. The anatomy of large-scale hypertextual web search engine. In Proceedings of the Seventh International World Wide Web Conference, 1998.*

[3] *Robert Cooley, Pang-Ning Tan, and Jaideep Srivastava. Websift: The web site information filter system. In Proceedings of the Web Usage Analysis and User Profiling Workshop, 1999.*

[4] *J Dean and M. R. Henzinger. Finding related pages in world wide web. In Proceedings of the Eighth International World Wide Conference, 1999.*

[5] *Ronny Kohavi and Carla Brodley. Knowledge discovery and data mining cup part of SIGKDD 2000*

[6] *V.N. Padmanabham and J.C. Mogul. Using predictive prefetching to improve world wide web latency. Computer Communication Review, 1996..*

[7] *M. Deshpande and G. Karypis. Selective markov models for predicting web page accesses. Transactions on Internet Technology, 4(2):163–184, 2004.*

[8] *N. R. Mabroukeh and C. I. Ezeife. Using domain ontology for semantic web usage mining and next page prediction. In Proceedings of the 18th ACM Conference on Information and Knowledge Management (CIKM 2009), November, 2-6 2009.*