Multi-Agent Based Search Engine for Researchers and Scientists

Naveed Dastgir Chheena and Muhammad Naeem Ahmed Khan

Department of Computer Science
Shaheed Zulfikar Ali Bhutto Institute of Science and Technology (SZABIST)
Islamabad, Pakistan
chheena@acroloigx.com , mnak2010@gmail.com

Abstract

The World Wide Web (WWW) comprises enormous websites located around the world, and it is the cheapest source to publish information that instantly becomes accessible across the globe. The information available on the Internet is invariably increasing on the daily basis, which in turn is resulting in serious challenges for the researchers to retrieve the relevant information from huge bulk of data. The precious time of researchers and scientists goes wasted as they have to sift through massive amount of irrelevant and unnecessary material returned by the conventional search engines. Due to this fact, the task of searching focused and relevant material has become a tedious activity in the information retrieval process over the Internet. This very unfocused and nonspecific type of the Internet surfing often results in overlooking the important contents due to monotonous and sloppy structure of the retrieved information, in addition to the tiresomeness and grimness it causes to the end user. In this paper, we present a framework for multi-agent based search engine tool that can help the researchers to find their desired information in an efficient way and with the minimal pointer clicks. Through this methodology, the researchers would be able to obtain a summarized report along with the search details that closely match to the supplied search phrases. The same summary report can also be generated based on the research topics inscribed in the researchers' profile or based on the history of their previous WWW usage.

Keywords: *Wild Wide Web (WWW), Search Engine, Decision Support System, Context Driven Search, Web Crawler*

1. Introduction

Academic and research resources on the web are increasing day by day. The plethora of information on the web though could be more beneficial to the researchers/scientists to find specific and detailed information in their concerned area of interest, but at the same time it also raises new concerns for them to locate the more relevant information. The huge information silos not only add complexity to the information retrieval process but also waste a lot of researchers' precious time in the quest for accumulating the relevant information. Researchers have to sift through loads of web pages that occasionally are not of their concern — a phenomenon pertaining to peculiar situations that augment the research complexity. Web based search engines like Google, Yahoo, AltaVista, Ask, Bing, Swoogle (Ding, et al., 2004), to name a few, though help researchers to find the desired information, but they return a colossal amount of hits. Most of the search results returned by the search engine are irrelevant, causing researchers to ponder upon them for a substantial amount of time to filter the meaningful stuff. Hence, search engines alone do not appear to be a viable approach. To discover an appropriate solution for this problem, we have made an attempt to present

an efficient search engine framework that is capable of minimizing the search time as well as enhance the search precision by sorting and summarizing the retrieved information and making it available in the form of an abridged report.

The idea to develop an agent-based search engine for researchers is based on the concept to produce the summary report against the search phrase entered by a researcher in the browser. The search results shown in the summary report are based on the context of researcher's area of interest along with the geographical location specific to the retrieved information. The search engine is designed in a way that it generates search results in multiple phases, thus decreasing the overall search time. In the first phase, it provides a summary of the search result and in the next step, it presents details of the search results. Afterwards, the search engine sorts the information according to incongruent criteria, including: location, technology, institution, research groups and research centers, etc. This approach helps researchers not only to get hold of the existing knowledge in their desired area, but also assists in finding gaps and potential future work in the specific research area. Agent-based search engine first finds the information available on WWW by crawling the Internet and then the accumulated information is evaluated both in a quantitative and qualitative way through the specialized search functions. Subsequently, the sorting function sifts the information and the report generation function formulates information in an abridged form. Since data mining techniques are based on certain context (Lohani & Jeevan, 2007), therefore, the report is organized based on researcher's context by applying the data mining techniques. Attaching context with the sought information helps the search engine to locate the more appropriate and meaningful information from the bulk data and filter out the garbage data automatically.

Multi-agent based search engine provides unique search results to individual researchers according to their area of interest. The search engine also summarizes the search results in a concise way so that these results can be quickly viewed in one sight by the user. In addition, the search engine finally makes comparisons of search results retrieved from different locations. This helps the researcher to refine the search criteria based on specific or multiple locations. This is a distinctive feature of our proposed search engine as other contemporary search engines do not offer such a functionality. The detailed description and functionalities of our proposed search engine are described in Section III.

2. Related Work

WWW has emerged as the biggest source of information as well as a vibrant platform to run the businesses across the globe. A number of search engines have been designed and developed in the past to facilitate Internet users to locate meaningful information from the websites scattered all around the world. According to Peristeras, et al., (2009), the governments across the world have induced huge investments in information and communication technologies (ICT), but still users face difficulties to successfully run the automated application. Now e-government and e-participation applications help governments focus on their citizens and businesses. There is still a challenge for the governments to determine how the ICT can be beneficial to their citizens and businesses through better use of Intelligent Technologies such as the semantic web, service oriented architecture (SOA), Web 2.0, etc. Rich Site Summary (RSS) is a format for delivering regularly changing web contents, and it has gradually developed novel services, e.g., podcast — an audio-based online service or a broadcasting service on demand. E-learning 2.0 (Li, et al., 2009; Ma, 2007) focuses on interaction between

people and has brought an innovative method for users to learn. Social computing leaders argue that the conventional intelligent technologies should be combined with the latest searching techniques to enhance and improve the efficiency of the information retrieval process. For instances, Natural Language Processing (NLP) with the support of argument representation, visualization and opinion processing can be more beneficial for efficient use of web services (Peristeras, et al., 2009). NLP can provide more meticulous and accurate results, which will save a lot of time of researchers by elaborating the search result in a precise manner.

Delineating the search query dimensions is also useful on the web for understanding the user's intents (Liliana & Ricardo, 2010). The search results of most of the queries rest in one or two dimensions; therefore, their relationships and dependencies too are needed to be explored. Liliana & Ricardo (2010) maintain that the intent of a user behind formulating the search query is a big challenge for the modern information retrieval systems. With particular reference to our proposed search engine, much of the user intents are automatically filtered as it accounts for researchers and scientists as the only users of the system.

Durgin & Sherif (2008) endorse that information skittered on the web requires autonomous evolution by encoding procedures of semantic web that are understandable by the machine; therefore, the author suggests using XML to present the syntax structure for the semantic web comprising subject, verb and object — a practice known as serried of triple. Hence, an ontology based standard is required in the semantic web technology. Currently, such functionalities are being supported through XML. Since URI presents the subject and verb, thus by using URI, the uniqueness of an object is defined (Durgin & Sherif, 2008). Semantic web applications require more recent data to be collected against each query even though every single context of a user would require new information crawling.

Cho & Tomkins (2007) researched on the intelligent agents for searching contents on the social media along with focusing on the contextual challenges faced due to diverse nature of the search criteria formulation. Most of the social-networking platforms carry information that is based on the visitors' interest as well as hot topics related to politics and social issues. Crawling for information on the social media remains a challenging issue for the search engines to draw any context from the available contents on these social-networking websites. However, social media websites and web pages usually do not pose many concerns to the researchers or scientists, therefore, can be simply ignored in the information retrieval process.

Website aging is an important factor in search engine optimization. Websites that are already online for a long time can progress upward swiftly than a new website because Google and other search engines carry out indexation of older websites more frequently, which ultimately helps in easy and fast crawling. Social media plays an important role in influencing the search engine results by giving preference based on the authority of the author and the number of times a piece of content is shared on social-networking sites (Patel, 2012).

Sivashanmugatn, et al., (2005) describe a scenario to build standards on the semantic web technology. Semantic web pages change the contents based on the semantic process templates to accommodate the participation of activities, controls, calculations and environment settings. Search engines are required to be updated based on the changes made on the web services which are often periodic. Tho, et al., (2006) investigated the searching mechanisms exclusively for research publications. Birukov, et al., (2005) recommend an agent-based search engine using data mining techniques

that are capable of ascertaining users browsing habits. Dikaiakos, et al., (2009) consider cloud computing as the future technology and predict that desktop PCs would become large data centers due to lesser processing and cost along with a huge amount of data storage. Heymann, et al., (2007) discuss spam associated with the social media searches and propose to delete the spam trough bookmarking. Hepp (2006) makes a critical review of publishers for not considering semantic web or semantic web service as the part of publication. Lond, et al., (2003) propose using soft computing for building intelligent machine. Farooq, et al., (2007) compare social bookmarking and tagging, which are used to retrieve the users intended information from the bulk of data. In general, the majority of researchers considers the web information as a data or metadata for the search engine along with some kind of information shortening algorithms to furnish the search results.

3. Proposed Framework and Methodology

In view of the aforementioned challenges associated with the existing search engines, we propose a multi-agent based search engine in this paper that consists of multiple components. The architecture of the proposed search engine is illustrated in Figure 1 that shows how different components of the model interact to provide the desired search results to the users. A described earlier, the model is initially designed and developed for use by the researchers and the scientists; however, it can be easily tailored to suit needs of distinctive interest groups or groups of user communities. A brief description of the modules that constitute our model is described in the subsequent paragraphs.

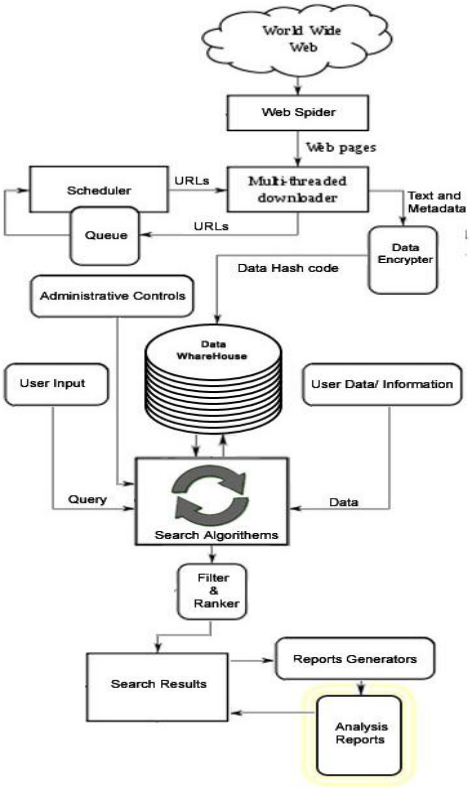


Figure 1. Architecture of the Multi-agent based Search Engine

3.1 Search Query

The search query module takes user input and passes it to the search engine to find the desired results. The search engine fetches the results from the web as per the query and returns them back to the user as an output. In the search query formulation, the user specifies his or her area of interest. Further, the query string formulates the returned results according to user's interest.

3.2 User Information/Data

The proposed model also maintains a detailed record of the user in the User Information/Data module which primarily consists of the user's preferences and areas of interest. The information held in this repository is passed to the search engine to filter the fetched results according to user's area of interest. User Information/Data is passed to the search engine in the following ways.

a. ***Direct Input*:** The user provides input through the interface called *user registration form* that contains his or her name, email, qualification and main areas of interest.

b. ***Indirect Input*:** The indirect input is obtained by the system itself from the user's previous browsing habits. For this purpose, multi-agents that run in the background on the user's machine are used to capture and maintain a history of the user's earlier searches all the time.

Researchers browsing history is captured and maintained based on different activates performed by him/her while using the search engine. The browsing history ordinarily consists of previous search queries, URLs of the surfed websites, etc. The prior browsing history of a researcher is always made available to him/her at the run time. The data pertaining to browsing history of researchers is archived into a database which is also used for further refining the search results.

3.3 Administrative Controls

These controls are supposed to be monitored by an administrative body within an organization where the proposed search engine is deployed. The administrative body controls the search operations and can confine the search areas as per the organization's policy. The administrative control ensures that the search engine operates in the customized manner, and undesired websites are barred. The websites restricted by the administrator remain inaccessible to the search engine. In addition, the administrator may add a new search location by adding the Top Level Domain (TLD) and can also change the format of the analysis report. Moreover, an administrator can manage functionality of the search engine by controlling the depth level up to what an individual website can be searched.

3.4 Search Algorithms/Functions

Search algorithms/functions — also called *search bots* — are commonly used on the Internet by the search engines like Google, Yahoo, Bing, etc. Though these search engines provide efficient search facility, however, the algorithms used by these search bots are not yet made public. These *search bots* apply the search algorithms on the search results in the background so that functionality of any algorithm could not be determined by anyone.

In our proposed model, the search algorithm that caters for the specific search requirements of researchers is divided into the following parts:

a. ***Web Spider*:** Web spider collects URLs of the available websites from the Internet and stores these URLs into the database. Web spider updates the database periodically on the basis URLs availability and non-availability. There are about 206 Top Level Domains (TLDs) [17] on the Internet such as *.com, .net, .ac, .gov, .org,* etc. and the total number of websites as of the end of year 2011 were more than 366 million [16]. The functionality of the web spider is not at all affected by the user's search query because it merely collects websites information from the World Wide Web for further processing by the web crawler.

b. ***Web Crawler*:** Web crawler is central to the search engine, and the quality of a search engine is measured in terms of the quality of its crawler. Web crawler, as the name suggests, crawls through the URLs supplied by the web spider up to the lowest level of depth as set by the administrator of the search engine. Web crawler may either work in conjunction with the web spider or independently based on the search query provided by the user. This particular linkage of crawler could be defined in terms of the available capacity of the data-storage devices — if a large amount of storage space is available, then it would be better to link the web crawler with the web spider to increase the efficiency by eliminating the time factor. A sample of the web crawler's code is shown in Figure 2.

```
$contactdetails = array();
$ch = curl_init() or die(curl_error());
$params=array( 'name' => 'chheena.com');
curl_setopt($ch, CURLOPT_POST,1);
curl_setopt($ch, CURLOPT_POSTFIELDS, $params);
curl_setopt($ch, CURLOPT_URL,"chheena.com");
curl_setopt($ch, CURLOPT_RETURNTRANSFER, 1);
curl_setopt ($ch, CURLOPT_HEADER, 0);
$data1=curl_exec($ch) or die(curl_error());
$locationstring = addslashes($data1);
echo curl_error($ch);
curl_close($ch);
```

Figure 2. Sample Code of Web Crawler

c. ***Data Warehouse*:** Data warehouse is a back-end repository which is used to store the user profiles, availability or non-availability of different websites, and the results fetched by the crawler in response to the user's search queries. Moreover, the information stored in the data warehouse also pertains to history of users' activities and meticulous websites' niceties along with the relevant metadata information.

Due to growing size and complexity of data sets, the data mining techniques like clustering, association rules, classification, pattern matching and data visualization, etc. are ordinarily used for knowledge discovery [18]. In our proposed methodology, the web spider scans the WWW and accumulates all the necessary URLs belonging to the specific TLDs (Top Level Domains). Afterwards, the gathered URLs are classified according to the categories of TLDs. Likewise, when the web crawler creeps through the requisite metadata and the corresponding web pages' information becomes available, then another crawler function is called to collect additional pages associated to the main URLs. A unique ID is assigned to each word found on the web pages and is stored into a lookup table. Specific ID information of the entire web page, that is stored into the lookup table, is coded to make the

future retrieval process efficient. Words stored in the lookup table are used for correcting spelling mistakes and are also used to provide *tool tips* at the time of query formulation.

3.5 Search Results (Level-I)

Search results (Level-I) component of the model acts as a first-level filter to further refine the search results obtained by the web crawler. In fact, it acts as a placeholder for the search results that do not satisfy the desired criteria as explicated in the user's preference and choices. These search results need further sifting to bring the search results up to the nearest level of user's expectation.

3.6 Filters and Rankers

In this module, the search results obtained in the previous level are filtered and ranked as per the user's preferences and become second or final level results. The data filters primarily act based on the user's information provided through the registration forms and history of user's activities. The filters are also applied to the search results in accordance with the administrative controls predefined for the search engine.

Page rankers define the sequence of search result and are used to sift the search results up to the top levels. There are many rankers available through different web services that rank web pages up to the 10^{th} level (http://www.prchecker.info/check_page_rank.php); and this process is also called the *Page Rank*. The page rank of a web page highlights the popularity of web page and, likewise, the ranks of multiple pages of a website collectively determine rank/popularity of that website. The websites are placed in higher and lower levels among the search queues as per their rank. Alexa.com is another useful website that provides global rankings and associated information of thousands of websites. The page rank of a website depends on the following factors:

a. Availability Level of Search Phrase
b. Frequency of Search Phrase
c. Placement of Search Phrase
d. Response Time of the Website
e. Visitors History of the Website
f. Scripting Language of the Web Page

The abovementioned factors are used Search Engine Optimization (SEO). The SEO experts usually used them to rank-up their websites in the search engines.

3.7 Search Results (Level-II)

After cleansing by the filters and page rankers, the search results obtained at the level-I become the second or final level results for the user query.

3.8 Interaction of Multi-Agents

Tasking, data capturing, intelligent and interface agents are the primary agents in our proposed model to find search results against a user's query. The role of the tasking agent consists of the following two functionalities:

a. ***Searching the query***: We have developed code in PHP for searching the user's query. Besides searching, our code sorts the acquired search results to display them in an organized way. In addition, the search results are also stored into the databases for future references. This approach helps in boosting the prospective searches. Due to the aforesaid mechanism, whenever a user initiates a search query, the required information is made available to him/her in the form of a summarized report within a short span of time.

b. ***Searching the WWW:*** As discussed earlier, web spider and crawler search websites on the Internet. The activities of web spider and crawler are dealt separately in our proposed model. The job performed by web spider and crawler are handled by the schedule tasks. Usually, a maximum of 500 web pages is crawled in one go; however, the number of pages to be crawled are decided based on load on the search engine as well as the number of available websites on www.

The code for data capturing agent is illustrated in Figure 2. Data captured by the web spider and crawler are stored into a database after coding it with the help of word lookup table. The next agent used in our model is the intelligent agent. On the basis of the data that becomes available, this agent identifies user behavior in terms of user's area of interest, locations, educational level, organization, etc.

3.9 Reports Generator - Analysis Reports

The final level search results are summarized in the form of reports by the reports generator module. The reports generator module produces reports in the following formats which help researchers to find the desired web pages of their interest.

a. ***Report by Categories*:** This type of report defines the major categories of websites. This report is calculated on the basis of formulation of URL registered for a website, for example, the *.edu* domain is specific to educational institutes, *.org* is pertains to organizations, *.gov* corresponds to the government institutes and *.com* is a general-purpose domains dedicated to the commercial or multinational entities.

b. ***Report by Locations*:** This type of report is categorized on the basis of locations, either by hosted server or by the owner's address. The location of server is calculated based on the server's IP address and owner's location is collected through the registrar of URL, e.g., *verisigninc.com*, *who.is* and *name.com* — among these, Verisign provides online APIs to crawl the information of website and its owner.

Snapshots of the sample output produced for "*Report by Category*" and "*Report by Location*" are provided in Figure 3 and Figure 4 respectively.

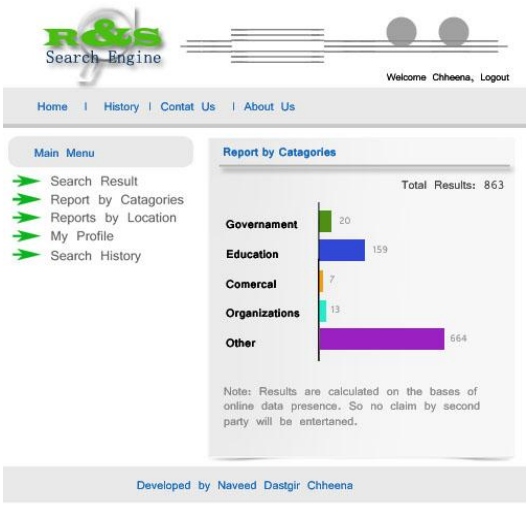


Figure 3. Screenshot of *Report by Category*

Figure 4. Screenshot of Report by Location

4. Discussion

Most of the agents used in our proposed search engine perform their functions independently; however, some agents work in collaboration. The web spider agent is responsible to collect the valid URLs which are transferred to the web crawler agent. These URLs are also stored into a database by the storage agent. The web crawler crawls through the web pages to find contents of interests and additional URLs. The URLs found by the crawler are again verified by the web spider. Afterwards, the crawled data contents are passed to the encryption agent that passes data to storage agent and control of the search engine algorithm is shifted to searching agents. The search query formulated by a researcher is transferred to the searching agents for

necessary processing. The search results obtained by the search agents are passed to reports generator agent for producing the analysis reports.

A comparison of the proposed search engine with the existing search engines can be drawn by understanding the capabilities and behavior of the existing search engines. Some search engines only return search results from the scientific resources like academia, research journals and conferences' websites and ignore the other sources of information on the web which may serve a meaningful purpose to the researchers. On the other hand, most of the search engines do not account for authenticity of the information resources. Our proposed search engine offers a range of different search patterns and provides summarized reports based on distinctive sorting, sifting and filtering techniques.

Our proposed search engine offers correct, precise, targeted and easy to use feature set. And these features act as the distinguishing parameters for evaluation of the efficiency, robustness and correctness of a search engine. Liliana & Ricardo (2010) report different dimensions of the user intent behind the search query, but their technique just covers the correct dimensions and only iterates once on the user supplied query. Whereas our proposed methodology not only covers all the four parameters of the search query but also iterates recursively to produce more precise and reliable results to the users. Cho & Tomkins (2007) explored intelligent agents to search contents on the social media and made an attempt to address contextual challenges faced due to the search criteria. However, their work suffers from limitations about the authenticity of resources, while our framework addresses the issue of authenticity of resources along with determining context of the user's query. Tho et al. (2006) focused on requisites of the search engine to search contents from research publications only. On the contrary, our search engine accounts for almost all types of formal contents for generating meaningful summarized reports.

5. Conclusion and Future Work

The proposed search engine is primarily developed for consumption by the researchers and scientists; however, it could be of much use for the ordinary users; particularly, students, analysts and the specialists working in the fields of science, engineering and technology can also be the prospective users of it. Moreover, the proposed search engine could potentially be a useful tool for the web researchers to monitor and analyze the dropped out topics and web pages on the WWW. Since search engines are part of the e-marketing (e.g., pay per click marketing for products and services), the proposed search engine could offer a platform to the vendors and web-developers to understand the faults and omissions due to which their product-related web pages are overlooked by the specialized search engines.

As a future dimension to our research, we intend to make our search engine part of an intelligent system that would be able to find deficiencies associated to the returned search results and would also suggest remedial measures to overcome these deficiencies.

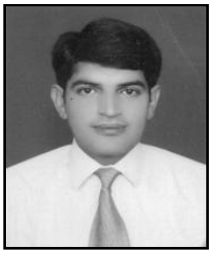
References

[1] L. Ding, T. Finin, A. Joshi, R. Pan, R. S. Cost, Y. Peng, P. Reddivari, V. C. Doshi and J. Sachs, "Swoogle: a search and metadata engine for the semantic web", In: 13th ACM Conference on Information and Knowledge Management, **(2004)**.

[2] M. Lohani and V. K. J. Jeevan, "Intelligent software agents for library applications Library Management", Emerald Group Publishing Limited, vol. 28, no. 3, **(2007)**, pp. 139-151.

[3] J. Cho and A. Tomkins, "Social Media and Search", IEEE Computer Society. IEEE Internet Computing, vol. 11, no. 6, **(2007)**, pp. 13-15.
[4] Q. T. Tho, A. C. M. Fong and S. C. Hui, "A scholarly semantic web system for advanced search functions", Emerald Group Publishing Limited, vol. 31, no. 3, **(2007)**, pp. 353-364.
[5] A. Birukov, E. Blanzieri and P. Giorgini, "Implicit: An Agent Based Recommendation System for Web Search", Association for Computing Machinery (ACM), Utrecht, Netherlands, **(2005)**.
[6] V. Peristeras, G. Mentzas, A. Tarabanis and A. Abecker, "Transforming E-government and E-participation through IT", IEEE INTELLIGENT SYSTEMS, **(2009)**, pp. 1541-1672.
[7] C. Liliana and L. Ricardo, "Towards a Deeper Understanding of the User's Query Intent", Workshop on Query Representation and Understanding Geneva, Switzerland, SIGIR'2010, **(2010)**.
[8] K. Durgin and S. Sherif, "The semantic web: a catalyst for future e-business", Emerald Group Publishing Limited, vol. 37, no. 1, **(2008)**, pp. 49-65.
[9] K. Sivashanmugatn, J. A. Miller, A. P. Sheth and K. Verma, "Framework for Semantic Web Process Composition", International Journal of Electronic Commerce, vol. 9, no. 2, **(2005)**, pp. 71-106.
[10] Q. Li, W. H. Lau, W. C. Leung, F. Li, V. Lee, W. Wah and H. Ashman, "Emerging Internet Technologies for E-Learning", IEEE Computer Society, **(2009)**, pp. 780-1089.
[11] D. Dikaiakos, G. Pallis, D. Katsaros, P. Mehra and A. Vakali, "Cloud Computing Distributed Internet Computing for IT and Scientific Research", IEEE Internet Computing, **(2009)**.
[12] P. Heymann, G. Koutrika and H. Garcia-Molina, "Fighting Spam on Social Web Sites: A Survey of Approaches and Future Challenges", IEEE Internet Computing, vol. 11, no. 6, **(2007)**, pp. 36-45.
[13] M. Hepp, "Semantic Web and Semantic Web Services Father and Son or Indivisible Twins?", IEEE Internet Computing, IEEE 2006, vol. 10, no. 2, **(2006)**, pp. 85-88.
[14] A. Lond, "The role of soft computing in intelligent machines", Physical and Engineering Sciences, vol. 361, **(1809)**.
[15] U. Farooq, Y. Song, J. M. Carroll and C. L. Giles, "Social Bookmarking for Scholarly Digital Libraries", IEEE Computer Society, **(2007)**, 1089-7801/072007.
[16] Webhosting.info, www.webhosting.info/domains/global_stats/total_domains/.
[17] IANA, http://www.iana.org/domains/root/db/.
[18] B. S. Qin and S. J. Yi, (Department of Computer Science and Technology, Xi'an Jiaotong University, Xi'an 710049), Journal of Computer Research and Development, 2001-03, DOI: CNKI:SUN:JFYZ.0.2001-03-015, **(2001)**.
[19] Y. H. Ma, "Research on Individualized Online Learning System Based on Web Mining", Journal of Henan Agricultural Sciences, vol. 3, **(2007)**, pp. 40–41.
[20] N. Patel, "How Social Media Affects SEO", http://www.quicksprout.com/2011/08/24/how-social-media-affects-seo/, **(2012)** August 6.

Authors



Naveed Dastgir Chheena

Naveed Dastgir Chheena obtained MS degree in Computing from Shaheed Zulfikar Ali Bhutto Institute of Science and Technology (SZABIST), Islamabad. Presently, he is working as Senior Software Engineer in Acroloigx. His research interests include in multi-agent systems, search engine optimization and software engineering.

Muhammad Naeem Ahmed Khan

Muhammad Naeem Ahmed Khan obtained D.Phil. degree in Computer System Engineering from the University of Suusex, Brighton, England, UK. Presently, he is associated with Shaheed Zulfikar Ali Bhutto Institute of Science and Technology (SZABIST), Islamabad, Pakistan. His research interests are in the fields of software engineering, digital forensic analysis and machine learning techniques.