A Survey on Various Developments In The Field Of Computer Science

Anik Shah, Animesh Shah
12bce084
Computer Science, Nirma University.
12bce084@Nirmauni.Ac.In
12bce085
Computer Science, Nirma University.
12bce085@Nirmauni.Ac.In

Abstract-- The World Wide Web (WWW) allows the people to share the information (data) from the large database repositories globally. In order to manage this data efficiently, we need to use specialized technologies to improve the user experience. This survey paper superficially explains the various web technologies such as HTML5 - Multimedia, Intelligent Semantic Web Search Engine, Web Data Mining, Distributed/Grid Computing and Image Processing.

Index Terms- HTML5, Multimedia, Intelligent Semantic, Web Search Engine, Web Mining, Distributed Computing, Image Processing.

I. INTRODUCTION

The problem of managing digital multimedia documents including audio and video in a Web environment is still a matter to look upon. Often this problem is solved by requiring the installation of extensions and third-party plugins into Web browsers.[2] HTML5, along with all its Multimedia attributes, plans to remove all such proprietaries and create one, simple web language.[1]

The Semantic Web is an extension of the current Web [3] that allows the metadata (meaning/information about data) to be precisely described in terms of well-defined vocabularies that are understood by users and computers. [4]

The web is huge, diverse and dynamic and thus, raises the scalability, multimedia data and temporal issues respectively. Due to those situations, we are currently drowning in information and facing information overload. Information users could encounter, among others, the following problems when interacting with the web:

- Finding relevant information,
- Personalization of the information, etc. [5] [6]

The Grid is a secure and coordinated resource sharing (direct access to computers, software, data, and high performance equipment) among groups of individuals, institutions and resources. These resources have rules which define:

- what is shared
- who is allowed to share resources and
- How resources can be shared.

Because Grid computing uses Internet to enable resource sharing of geographically diverse computational resources, we can use Web Services (WS) to ensure the communication between Grid applications. [7]

Printing technology identification and associated problems in document forensics have been projected as challenges in image processing application. [8] Various image processing approaches based on textures, spatial variation, HSV color space, spatial correlation, and feature based on histogram and some of the pattern recognition methods, like gray level co-occurrence matrix, roughness of the text, perimeter of edge are highlighted.

II. HTML5 – MULTIMEDIA

In this section, some related works based both on HTML 5 or other platforms are described. It also outlines the various features provided in HTML5 language for developing multimedia-related webpages and applications.

In the absence of a standard way to include video in an HTML page, video on the Web has been mostly the prerogative of browser-specific implementations or third-party plug-ins such as the VLC plug-in, Apple QuickTime or Adobe Flash Player. These plug-ins are activated through the use of the <object> tag in HTML. [9]

Relying on third-party plug-ins to render the video works fine in a variety of use cases, but they do have their drawbacks.

Thus, CSS (Cascading Style Sheet) does not have the feature for styling the video, or to apply transformations. SVG cannot be used to apply masks and filters on the video. [9]

The <video> tag introduced by HTML5 alleviates these problems. Since video becomes a regular tag such as <p> or <div>, it is directly integrated within the rest of the Web page. CSS may thus be used to style and transform the video box.

The HTML5 specification does not restrict the list of video formats and codecs that a browser may support, leaving the door open for future formats.

The <video> tag also provides us with the facility of adding subtitles to our video in the web page.

STREAMING VIDEO ON THE WEB - The easiest way to publish a video on the Web is to use HTTP progressive delivery. Progressive delivery is supported by all Web browsers. [11]

HTML5 is easier to integrate across multiple platforms: HTML5 is beneficial as web developers only need to write once and it can run on other systems while Flash and other plug-ins always require different versions for each platform.

In HTML5, developers will be allowed to design more friendly applications to users without being dependent on third-party software. For instance, embedding video, sound, high-quality graphics, charts and other motion pictures with abundant contents instead of adding any plugins such as window media player, flash player. [10]

But on the other side, HTML5 has its disadvantages as it is still under development. It cannot stream live video, or interact online with a webcam.

It also has the bitter side in perspective of its popularity. [11] As compared to Flash, HTML5 is used in a very lesser proportion. About 97% of the devices across multiple platforms use Flash.

Hence, by summarizing the topic, we can say that overall HTML5 will make the life of developers a lot easier by providing simple attributes to design web applications in to context to the multimedia, but it will take some years before it completely eradicates its competitors.

III. WEB SEARCH ENGINE

Information retrieval by searching information on the web is not a new idea but has different challenges when it is compared to general information retrieval.

Different search engines return different search results due to the variation in indexing and search process. Google, Yahoo, and Bing have been out there which handles the queries after processing the keywords.

These search engines only search information given on the web pages. Some recent research groups have already took a step in delivering results from their semantics based search engines, but most of them are in their initial stages. None of the search engines have come close to indexing the entire web content, much less the entire Internet yet.

Current web is the biggest global database that lacks the existence of a semantic structure and hence it makes it difficult for the machine to understand the information provided by the user.

When the information was distributed in web, two kinds of research problems interrupted the progress in [14] search engine i.e.

- How can a search engine map a query to documents where information is available but does not retrieve in intelligent and meaning full information?

- The query results produced by search engines are distributed across different documents that may be connected with hyperlink. How search engine can recognize efficiently such a distributed results?

Semantic web [13] can solve the first problem in web with semantic annotations to produce intelligent and meaningful information by using query interface mechanism and ontologies.

The other problem can be solved by the graph-based query models [15]. The Semantic web would be required to solve extraordinarily difficult problems in the areas of knowledge representation, and natural language understanding.

Hakia [16] is a general purpose semantic search engine that does the same searching work of information as Wikipedia. Hakia calls itself a "meaning-based (semantic) search engine". They're trying to provide search results based on meaning match, rather than by the popularity of search terms to avoid the previous problems.

Semantic web is being developed to overcome the following main limitations of the current Web [17]:

- The web content lacks a proper structure regarding the representation of information.
- Confusion of information resulting from poor interconnection of information.
- Lack of automatic information transfer.
- Unable to deal with infinite many number of users and content ensuring security at all levels.
- Incapability of machines to understand the provided information due to lack of a universal format.

IV. WEB MINING

With huge amount of information available on the internet, the World Wide Web is a fertile area for Web Mining research. The web mining research is the superset that involves and calls for research from several research communities, such as database, information retrieval, and within AI, especially the sub-areas of machine learning and natural language processing.

For the survey, we focus on the representation issues, on the process, on the learning algorithm and on the application of the recent works on the criteria in web mining.

Web mining is the use of data mining techniques to automatically discover and extract information from web documents and services. [22] Web mining can be decomposed into the following subtasks:

- Resource finding: the task of retrieving intended web documents
- Information selection and pre-processing: automatically selecting and pre-processing specific information from retrieved Web resources
- Generalization: automatically discovers universal patterns at individual Web sites and across multiple sites
- Analysis: validation and/or interpretation of the mined patterns [5]

We categorize Web mining into three areas of interest based on which part of the web to mine:

- Web content mining: describes the discovery of useful information from the Web contents
- Web structure mining: tries to discover the model underlying the link structure of the web. It is based on the topology of the hyperlinks with or without description of the links. Used to categorize Web pages. [23]
- Web usage mining: tries to make sense of the data generated by the Web surfer's sessions or behaviors. [24]

In the database view of the Web content mining, the structure within web documents (intra-document structure) will be main focus.

In Web structure mining, the structure of the hyperlinks within the Web itself (inter-document structure) will be main point to look upon.

Web usage mining focuses on techniques that could predict user behavior while the user interacts with the Web. The mined data within this category are the secondary data on the Web as the result of interactions.

To sum this category up, we first point out the area in which mining is carried out. Then we coin the term Data mining from which we bring the term Web data mining in the light. We later superficially list out the subtasks into which it can be divided. We also suggest the three categories into which Web mining can be divided and then situate some explanation with respect to these categories and their respective needs.

V. GRID COMPUTING

Grid Computing is a platform for coordinated resource sharing and problem solving on a global scale among virtual organizations.

The benefit of increasing the computational power by combining geographically diverse systems and high performance computational equipment raises the issues of heterogeneity and scalability [18]. The use of a middleware, such as Globus Toolkit, addresses that of heterogeneity, thus providing uniformity through standard interfaces to shared resources.

Similar to WS [19], GS needs to apply resource discovery to be functional. However, sometimes the conventional search of services does not provide good results, in these cases, Semantic Web tries to solve the problem adding meaning to GS and thus, the search is improved and better results are obtained [20].

Web Services describe a distributed computing paradigm that is focused on Internet-based standards to facilitate heterogeneous distributed computing [18].

Grid Services are software components which provide access to a set of Grid resources such as data sources, high performance equipment and computational resources. GS are the base of the Open Grid Services Architecture (OGSA) [21], which aims to standardize all services on a Grid application. OGSA defines the GS concept based on technologies of Grid computing and WS. It also defines standard mechanisms for naming, creating and discovering GS.

Even though, the goal was the convergence of the standards for WS (WSDL) and GS (OGSI), this convergence has not achieved. New standard (WSRF) was created for this convergence of WS and GS, but it did not bring a significant change. In contrast, it was only an evolution of the OGSI.

Semantic Web is defined as an extension of Web technology in which the information has its own meaning. It makes possible for the web to understand and satisfy the requests of people and machines using web content.

VI. IMAGE PROCESSING

Recent research publications demonstrate various approaches suggested for discriminating printing techniques. Research activities on characterization of electro photographic printers in [22] gray level co-occurrence feature in and most frequently occurring letter "e" and Gaussian mixture model (GMM) in [23] are the techniques used for printer identification.

Identification of printed document and photocopies is based on the techniques of Color image processing. Color image processing techniques used for identification of printing process, employed by the HSV color space. In this work, hue histogram is used to identify between the printed or photocopied document.

While the hue histograms, generally, are bi-modal and wider for photocopied documents, this is uni-modal and narrower for printed document.

Another method for identification of the electro-photographic printers is application of principal component analysis and Gaussian mixture models (GMM) [22]. In this identification method each printer is represented as Gaussian distribution.

The techniques and methods discussed in above sections are based on the high resolution images. Most of the tools developed are for color images and few categories of printers. However, one should also focus and develop tools for identification of offset printers, gray scale images in place of

color images, enhancement in GVM approach in particular to low resolution environmental setup.

VII. SUMMARY

The above survey paper provides with five miscellaneous developments in the field of computer science, especially Web technologies. These developments are quite important with respect to creation of new web-based applications and their management. Hence, to conclude, we can say that the paper defines the topics and explains the general idea behind each category.

References:

[1] HTML5, W3C Working Draft,
http://www.w3.org/TR/html5/

[2] MANAGING MULTIPLE MEDIA STREAMS IN HTML5
The IEEE 1599-2008 Case Study

[3] Berners-Lee, T., Hendler, J. and Lassila, O. "The Semantic Web", Scientific
American, May 2001.

[4] Intelligent Semantic Web Search Engines: A Brief Survey
International journal of Web & Semantic Technology (IJWesT) Vol.2, No.1, January 2011

[5] Web Mining Research: A Survey
SIGKDD Explorations. Copyright 2002

[6] P. Maes. Agents that reduce work and information overload. Communications of the ACM.

[7] I. Foster, C. Kesselman, and S. Tuecke, "The Anatomy of the Grid: Enabling Scalable Virtual Organizations", Proceedings of the International Journal of Supercomputers Applications, Vol. 15, 2001.

[8] A Survey of Image Processing Techniques for
Identification of Printing Technology in Document
Forensic Perspective.
IJCA Special Issue on "Recent Trends in Image Processing and Pattern Recognition"
RTIPPR, 2010.

[9] Towards Video on the Web with HTML5
author: François Daoust, W3C/ERCIM, 2004

[10] Application of HTML5 Multimedia
2012 International Conference on Computer Science and Information Processing (CSIP)

[11] HTML5 Multimedia
DEVELOP AND DESIGN
Ian Devlin

[12] Combining HTML 5 with MVC Framework to Simplify Real-Time Collaboration for Web Development
2009 International Conference on Multimedia Information Networking and Security.

[13] T.Berner-Lee and M. Fishetti, Weaving the web "chapter Machines and the web."Chapter Machines and the web,

[14] D.Fensal, W. Wahlster, H. Lieberman, "Spanning the semantic web: Bringing the worldwide web to its full potential, "MIT Press 2003.

[15]G. Bholotia et al.: "Keyword searching and browsing in database using BANKS," 18th Intl. conf. on Data Engineering (ICDE 2002), San Jose, USA, 2002.

[16] D. Tümer, M. A. Shah, and Y. Bitirim, An Empirical Evaluation on Semantic Search Performance of Keyword-Based and Semantic Search Engines: Google, Yahoo, Msn and Hakia, 2009 4th International Conference on Internet Monitoring and Protection (ICIMP '09) 2009.

[17] Sanjib Kumar, Sanjay Kumar Malik "TOWARDS SEMANTIC WEB BASED SEARCH ENGINES" National Conference on "Advances in Computer Networks & Information Technology (NCACNIT-09) .

[18] I. Foster, J. Nick, C. Kesselman, and S. Tuecke, "The Physiology of the Grid: An Open Grid Services Architecture for Distributed Systems Integration", 2002.

[19] E. Al-Masri, and Q. Mahmoud, "Discovering the Best Web Service: A Neural Network-based Solution", Proceedings of the 2009 IEEE International Conference on Systems, Man, and Cybernetics, 2009, pp. 4250- 4255.

[20] M. Muhammad, L. Yuan, and Z. Jianqiu, "Web 3.0: A real Personal Web", 2009 Third International Conference on Next Generation Mobile Applications, Services and Technologies, 2009, pp. 125 – 128.

[21] I. Foster, K. Czajkowski, D. Ferguson, J. Frey, S. Graham, T. Maguire, D. Snelling, and S. Tuecke "Modeling and Managing State in Distributed Systems: The Role of OGSI and WSRF", Proceedings of the IEEE, Vol. 93, pp. 604-612, 2005.

[22] Nitin Khanna, Aravind K, Mikkilineni, Anthony F.Martone, Gazi N. Ali, George T.C. Chiu, Jan Allebach,and Edward J. Delp. A survey of forensic characterization methods for physical devices., 2006

[23] Ali G. N., Chiang P. J., Mikkilineni A. K., Chiu G.T.- C, Delp E.J., and Allebach J. P. "Application of Principal Components Analysis and Gaussian Mixture Models to Printer Identification". In Proceedings of the IS & T"s NIP20: International Conference on Digital Printing Technologies, Nov 2004.

[22] O. Etzioni. The World wide web: Quagmire or gold mine. Communications of the ACM, 39(11):65-68

[23] S. Chakrabarti, B. Dom, D. Gibson, J. Klienberg, S. Kumar, P. Raghavan, A. Tomkins.
Mining the link structure of the world wide web. IEEE Computer, 32(8):60-67, 2004

[24] Proceedings of the 9th IEEE International Conference on Tools with AI
R. Cooley, B. Mobasher and J. Srivastava: Web Mining: Information and Pattern discovery on the world wide web