

Mining Approach for Web Engineering on Development of Web based Business Intelligence Applications

G. Sirisha^{#1}, M. Sridevi^{*2}, Dr. A.V. Krishna Prasad^{#3}

^{#1}*Department of Computer Science, CVSR, JNTU, Hyderabad, India*

^{*2}*Associate Professor, Department of Computer Science, CVSR, JNTU, Hyderabad, India*

^{#3}*Associate Professor, Department of Computer Science, MVSR, OU, Hyderabad, India*

¹sirisha.godavari@gmail.com

²siri_235@yahoo.com

³kpvambati@gmail.com

Abstract— Over the last few years, there has been a remarkable increase in use of World Wide Web (WWW) for a wide and variety of applications. The web plays a central role in e-commerce applications. The Business progresses as Product View, Managerial View, Business Intelligence view and Next Generation Business Intelligence View, in this consequence, there is a need to improve Intelligence of Web Engineering Applications in the context of Business and IT industry. To achieve this objective web engineering tasks must be able to identify some useful insights for business intelligence. The present research work attempts to initiate Business Intelligence from Web Based Applications..

Keywords— Web Mining, Web Engineering, Business Intelligence, Web site errors, Web objects, Web logs..

I. INTRODUCTION

Web Engineering is the evolution of software engineering which focuses on the methodologies, techniques and tools that are the foundation of Web application development and which support their design, development, evolution, and evaluation. Web engineering aims in the development of web based systems in a manner that result in high quality, robust, secure, scalable and maintainable applications. Web engineers face the same traditional concerns as software engineers: the risks of failure to meet business needs, project schedule delays, budget overruns and poor quality of deliverables. Web Engineering addresses the problems associated with shorter lead times which require rapid prototyping and agile methods, the interactivity and visual nature of the medium which makes Human Computer Interaction aspects highly significant, and multimedia features of Web applications. Web Engineering has an interdisciplinary approach covering Web development concepts, methods, tools and techniques, useful for Web Software developers, Web designers and project managers. These mining techniques eventually enhance insights of web engineering applications pertaining to the areas of web site design, consumer behaviour and security architectures.

Business intelligence applications are of vital importance as they help organizations manage, develop and communicate intangible assets such as information and knowledge. Organizations that have undertaken business intelligence initiatives have benefitted from increases in revenue as well as significant cost savings. BI technologies provide historical, current and predictive views of business operations. The Business progresses as Product View, Managerial View, Business Intelligence view and Next Generation Business Intelligence View. In 19th century we have product views as business information as information which assists the means of buying and selling through the application of economics and management techniques of the production of goods and services. In the last decades of the 20th century we have managerial view which includes large business corporations, government agencies and non-profit organizations. This includes not only its vendors, employees, and customers, but even members of a community where its offices or factory may affect the local economy or environment. Next view is business intelligence view which comes over the period of 1960's to 1990's. Business intelligence view indicates that business information embraces all information flows within the enterprise and between the enterprise and the external parties: customers, suppliers, the general public government etc. This views business information as ideas, messages or intelligence a business person requires to improve the day to day operation of a business. Such information assists in business activities such as selling and buying goods, establishing opportunities, nature of competition, simple management techniques, and the knowledge of available services from government and the Non Governmental Organization (NGO's) sector. The next Generation intelligence view includes Suppliers, Customers, Owners, Employees, Competitors, Government, Environment, Online communities, news, media, International Partners, &Multinational employees. In 21st century, a large portion of the global data infrastructure is built upon World Wide Web technology. We are informed, educated, and entertained by

web applications and services. The web provides us with an international mall where we can shop for every imaginable item.

II. MINING FOR WEB ENGINEERING

Web mining is the usage of data mining techniques to extract interesting information from web data. Patterns extracted from applying web mining techniques on web data can be used to maintain websites by improving their usability. A mining approach for web engineering application is necessary to study the role of various components of web applications. The Web Mining techniques that are applied on Web data produce various tasks related to Web Engineering frame work. These tasks associated with web services, web architectures, web configuration management data, web application classification, web testing, web communities, website navigation, web security etc. The patterns can also used to study user behaviour and interests, facilitate support and services introduced to the website navigator and improve the structure of the website.

III. MINING APPROACH FOR WEB ENGINEERING

Web Engineering must be a systematic discipline and quantifiable approach for development, operation and maintenance of web based business applications. A successful Web Engineering Applications is one which can improves insights of web based business intelligence by sharing results among various modules in the process of effective decision making [6]. To achieve the goal of business intelligence in web based applications, one has to concentrate on various aspects of web content. A mining approach for web engineering application is necessary to study the role of various components of web application in business intelligence.

In this part of research work, an algorithm was developed as a web mining approach which can investigate the components of web that are used in business intelligence.

An algorithm consists of three phases. They are

Phase 1: Extracting Web Data from web application

Phase 2: Mining Techniques to extract features of web application

Phase 3: Web Engineering tasks associated with Web Mining

Phase 1: Extracting Web Data from web application: An algorithm was developed using the web program in extracting the components of web such as Web site Structure, Web site Error Reports, Web site objects used in web design, and contents of Web log. The Web Program retrieves all necessary web data elements using a standard set of web tools. These tools include Web site Extractor, Web page Analyzer, W3C HTML Validator, Power Mapper, and Web log Expert. The overall Architecture of the algorithm is shown in figure 1. The description of the algorithm is shown in table 1.

Algorithm Description: The URL address of Web site is thoroughly scanned using web site extractor to get all web pages of the web site and each web page is stored in a separate

database. A snap shot of website extractor is shown in figure 2. The web objects of each web page of web site are gathered using web page analyzer. All these objects are stored in a separate data base. A snap shot of web page analyzer is shown in figure 3. The web site structure shows organization of web pages in the web site starting from home page. A Web tool Power Mapper constructs a site map for the web site in a hierarchical tree structure. The structure of a website is stored in a database for further process. The World Wide Web consortium (W3C) developed a standard web tool W3C HTML Validator to generate error reports of web site. All error reports are then stored in a separate database. The snapshot of website errors report is shown in figure 4. The standard web tool web log expert is used to extract information of web log. A web server log file records user's transactions in the web. Usually, the web log file contains information about the user IP address, the requested page, time of request, the volume of the requested page, its referrer, and other useful information. Web log file is the main source of data analysis in web mining. Web log file contains data about requested URL, time and date of request, method used, etc. The snap shot of web log file is shown in figure 5.

Phase 2: Mining techniques to extract features of web application: Web mining is the usage of data mining techniques to extract interesting information from web data. Patterns extracted from applying web mining techniques on web data can be used to maintain websites by improving their usability. The patterns can also used to study user behaviour and interests, facilitate support and services introduced to the website navigator, improve the structure of the website, and facilitate personalization and adaptive websites. The web engineering data is extracted from various components using the procedure described in phase 1. The overview of various components that are used to extract web engineering data is shown in the figure 6.

After extracting web engineering data, various mining techniques are applied on data base which consists of web information.

Phase 3: Web Engineering Tasks associated with mining: The Mining techniques that are applied on Web data produce various tasks related to Web Engineering frame work. These tasks associated with web services, web architectures, web configuration management data, web application classification, web testing[29], web communities, website navigation[71], web security[72] etc. These services are very much essential in any type of web application.

Insights of Web Engineered Business Intelligence Applications: General Business objectives and their functionality for Rich Security Model that users can administer are: Provide more effective mechanisms to move work between business entities, such as self-service for customers or partners or enabling outsourcing by providing business partners a collaborative environment or business data

on an extranet. The valuable insights from this approach include ease of use, scalability, disconnected from processes for these developed applications, improved customer satisfaction, increased business agility, reduced time to market, increased revenue and operational efficiency and improvements.

CONCLUSION

The research work can still find the scope to improve web applications not only as informative but also as a tool for web engineered on-line applications. In this direction the work further provides scope for future work in the areas of Semantic Web and Knowledge Management. Hence in this research work, valuable insights are identified and web engineers can use the findings found in this research work to develop well justified web based business applications as full-fledged e-commerce applications.

REFERENCES

- [1]. Dr. A.V.Krishna Prasad, Dr. S.Ramakrishna "Mining for Web Engineering", International Journal of Computer Trends and Technology (IJCTT), pp. 151-156, December 2011.
- [2]. Ajith Abraham, "Business Intelligence from Web Usage Mining", Department of Computer Science, Oklahoma State University, USA.
- [3]. Anupam Joshi and Pranam Kolari, "Web Engineering Column: Web Mining: Research and Practice", IEEE Computing and Science and Engineering, July/August 2004 pp. 49-53.
- [4]. Anupam Joshi, Tim Finin, akshay Java and Pranam Kolari, "Web 2.0 Mining: Analyzing Social Media", Taylor & Francis Group, 2008.
- [5]. Athula Ginge and San Murugesan, "Web Engineering: A Methodology for Developing Scalable, Maintainable Web Applications", Cutter IT Journal, Vol.14, No.7, July 2001, pp.:24-35.
- [6]. Bart Czernicki, "Silver light 4 Business Intelligence Software", Apress, USA 2010.
- [7]. Beibei Li and Jiabin Le "Application of Web Service in Web mining" CIS 2004, LNCS 3314, 2004, pp 989-994.
- [8]. CenZic Inc., "Web Application Security Trend Reports", 2009.
- [9]. Chan Gaik Yee, G.S.V. Radha Krishna Rao, "An Introductory study on Business Intelligence Security", Idea Group Inc(2007): pp. 204 - 217.
- [10]. Collin White, "Next Generation of Business Intelligence: Operational Business Intelligence", BI Research, July 2006.
- [11]. Gondi Sreedhar, Dr. A.A.Chari, Dr. V.V.Venkata Ramana, "A Qualitative and Quantitative Frame work for effective Web Site Design", IJCA, Vol 2, May 2010, PP 48-56.
- [12]. Halvard Skogsrud," Modeling Trust Negotiation for Web Services", IEEE February 2009.
- [13]. Jaideep Srivastava, Prasanna Desikan, Vipin Kumar, "Web Mining - Accomplishments & Future Directions" IEEE conference, Argentina
- [14]. James Governor, Dion Hinchcliffe, Duane Nickull, "Web 2.0 Architecture", Oreilly Publishers, USA 2009.
- [15]. Jiawei Han, Kevin Chen-Chuan Chang, "Data Mining for Web Intelligence", IEEE Computer November, 2002 pp. 64-70.
- [16]. Jim Highsmith, Alistair Cockburn "Agile Software Development: The Business of Innovation", IEEE Computer September 2001 pp: 120-122.
- [17]. Jing Dong, Yajing Zhao, Tu Peng "A Review of Design Pattern Mining Techniques".
- [18]. Ladislav Burita-Vojtech Ondryhal, "Extending Uml For Modelling Of Data Mining Cases"
- [19]. Liu, B. 2007, "Web Data Mining", Springer - Verlag, New York, ISBN - 13: 9783540378815.
- [20]. Martha Koutri and Sophia Daskalaki. "Improving Web Site Usability Through a Clustering Approach", In Proceedings of the 10th International Conference on Human-Computer Interaction HCI, Crete, Greece, 2003, PP. 11-19.
- [21]. Massimiliano Albanese, Antonio Picariello, Carlo Sansone, and Lucio Sansone. "A Web Personalization System Based on Web Usage Mining Techniques", In Proceedings of the 13th international World Wide Web conference on Alternate track papers and posters, New York, USA, 2004. ACM Press, PP. 288-289.
- [22]. Mohammad A. A. Alhawamdeh, "Web Mining: Strategic Web Site Design for small business" Proceedings of the world congress on Engineering 2007 Vol I.
- [23]. Nakayama.T, H. Kato, and Y. Yamane, "Discovering the gap between website designers' expectations and users' behavior" In Proceeding of the Ninth Int'l World Wide Web Conference, Amsterdam, May 2000.
- [24]. Paola Britos, Oscar Dieste, and Ramon Garcia "Requirements Elicitation in Data Mining for Business Intelligence Projects", IFIP 2008, pp.139-150.
- [25]. Pranam Kolari and Anupam Joshi, "Web Engineering Column: Web Mining: Research and Practice", IEEE Computing and Science and Engineering, July/August 2004 pp. 49-53.
- [26]. Rosa Meo, Maristella Matera "Designing and Mining Web Applications: A Conceptual Modeling approach" Idea Group Publishing 2006.
- [27]. Tao Xie, Suresh Thummalapenta, David Io, Chao Liu,"Data Mining for Software Engineering", IEEE Computer, August 2009, pp. 55-62.
- [28]. Wingyan Chung "Designing Web-based Business Intelligence Systems: A Framework and Case Studies: In DESRIST", February 24-25, California CA USA 2006, pp. 147 - 171.

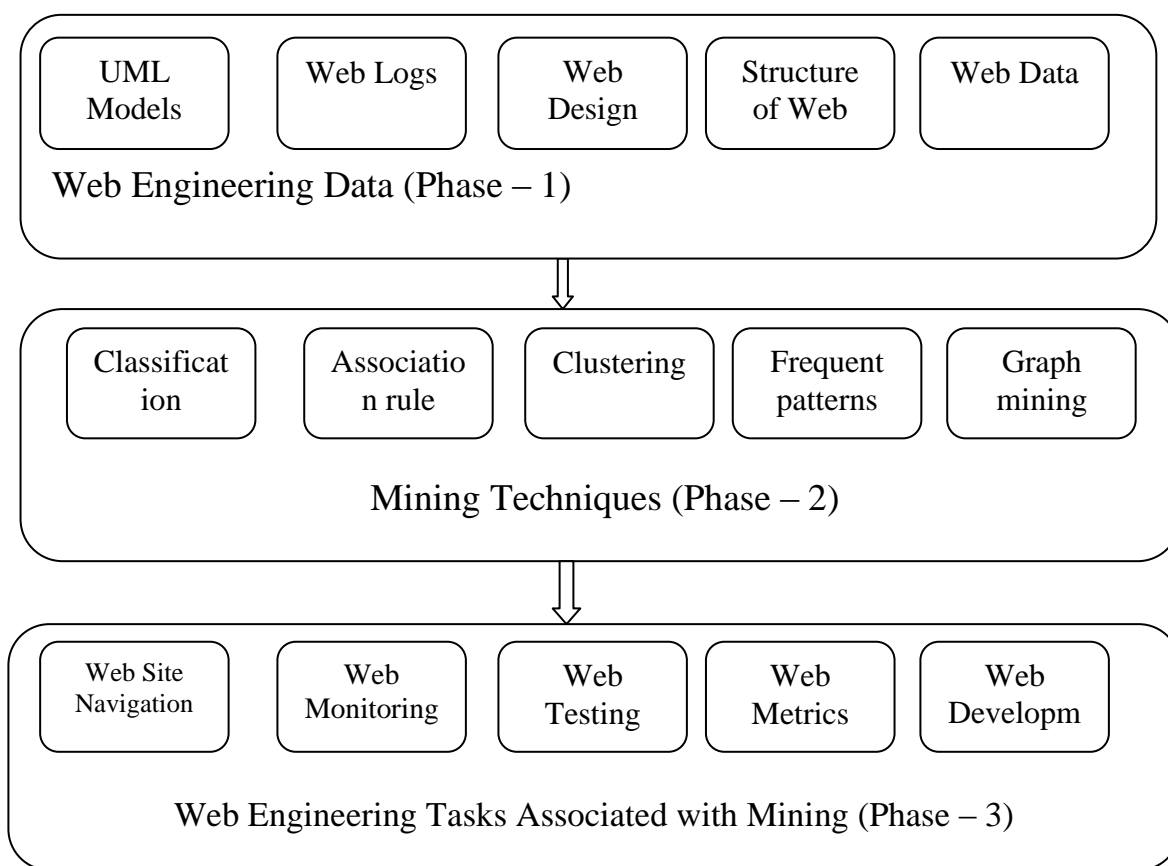


Figure 1 Phases of Mining for Web Engineering

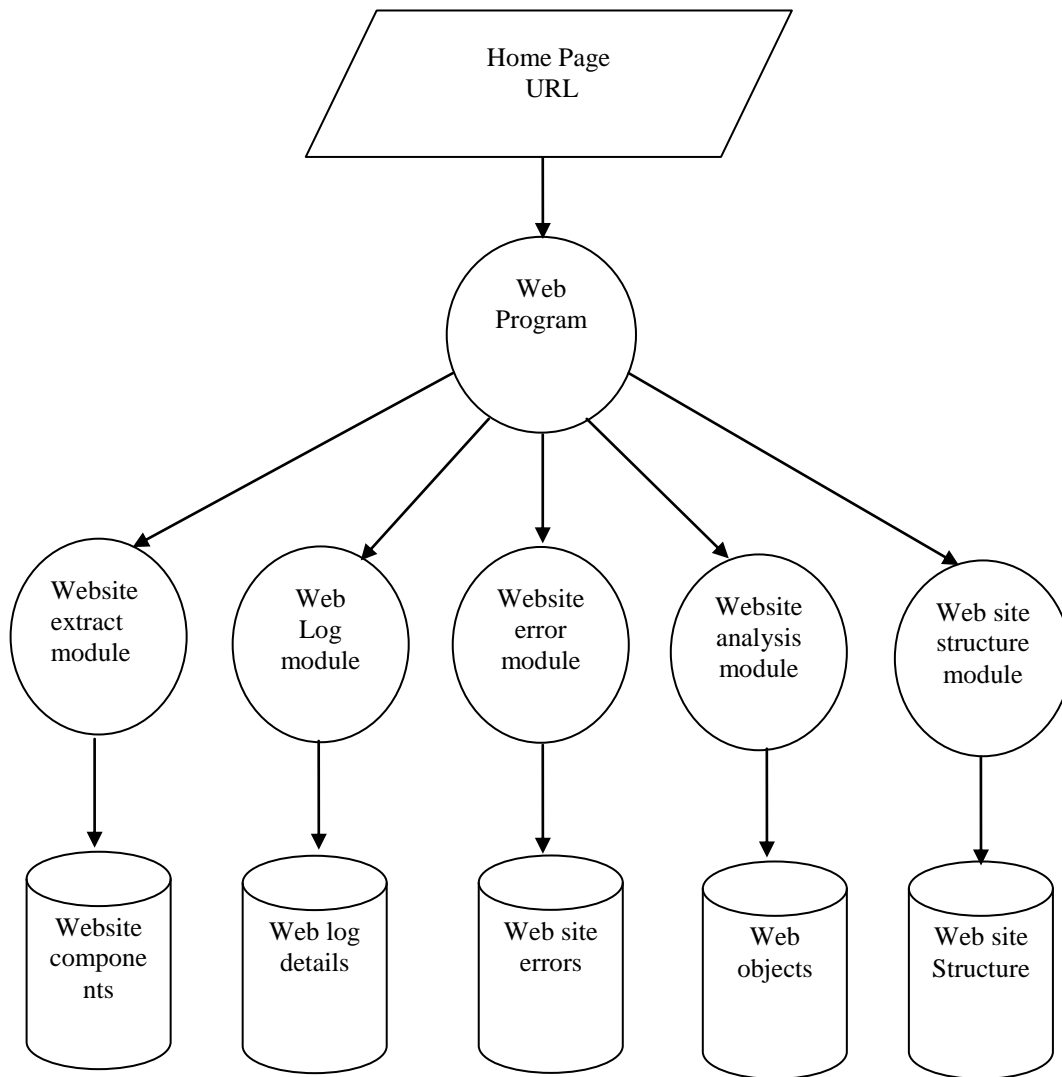


Figure 1: Architecture of Algorithm

```

Input: Home Page (URLH)
Begin
  Accept URLH
  Retrieve list_of_pages using WebsiteExtractor(URLH)
  PageCount = count(list_of_pages)
  SiteMap = PowerMapper(URLH)
  Web log = WeblogExpert(URLH)
  Store SiteMap into WEBSITE_STRUCTURE
  For i = 1 to PageCount do
    Begin
      Identify errors(i) using WebPageValidator(i)
      Store errors(i) into WEBSITE-ERRORS
      Page_Dt(i) = webPageAnalyzer(i)
      Broken_Link(i) = LinkChecker(i)
      Retrieve web objects(i) using WebsiteExtractor
      Store web objects(i) into WEB_COMPONENTS
      Store Page_Dt(i) into WEBOBJECTS_DT
    End
  End
Output: Website errors, web site structure, web objects, web log data
    
```

Table 1: Pseudo code for Web Algorithm

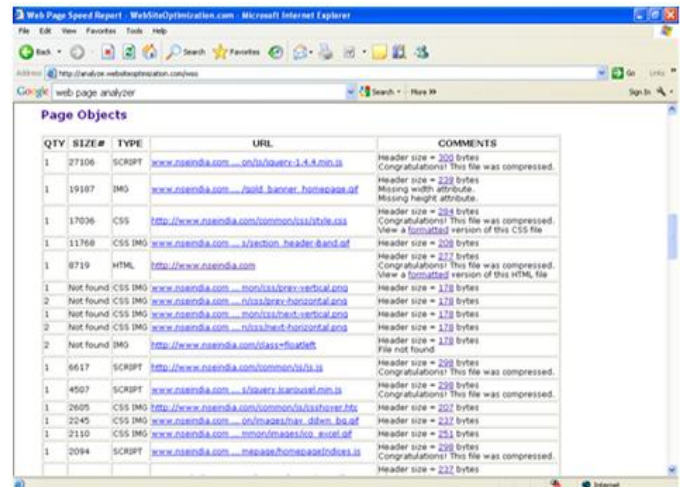


Figure 3: Web Page Analyzer



Figure 2: Website Extractor

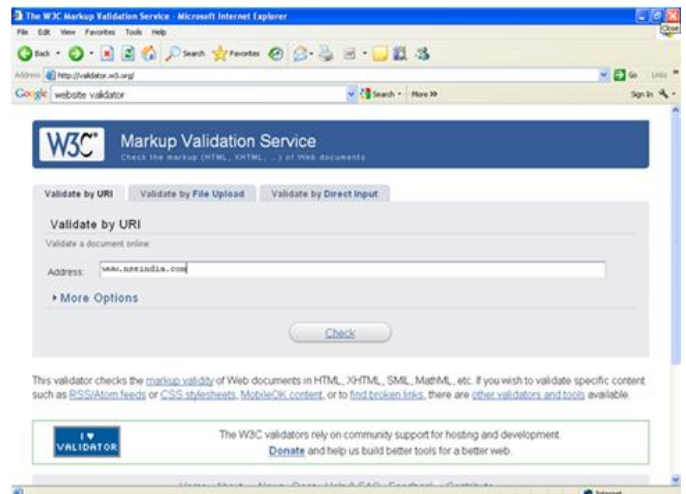


Figure 4: Website Errors

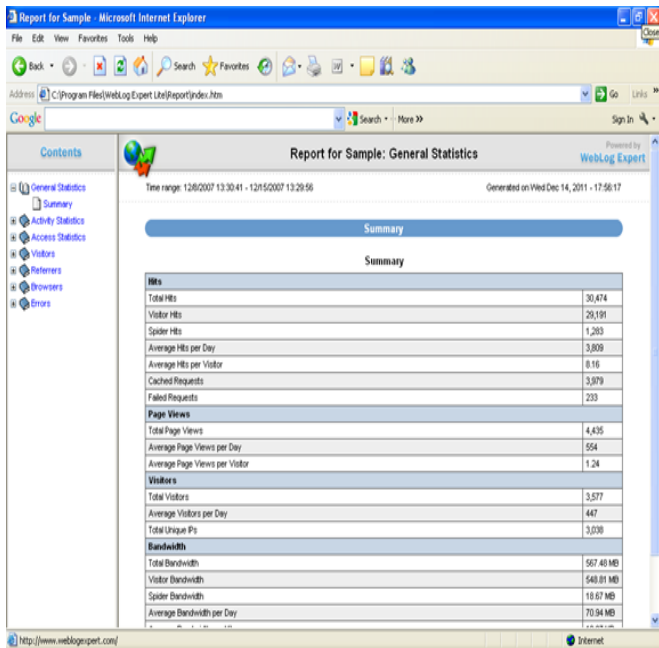


Figure 5: Web log details

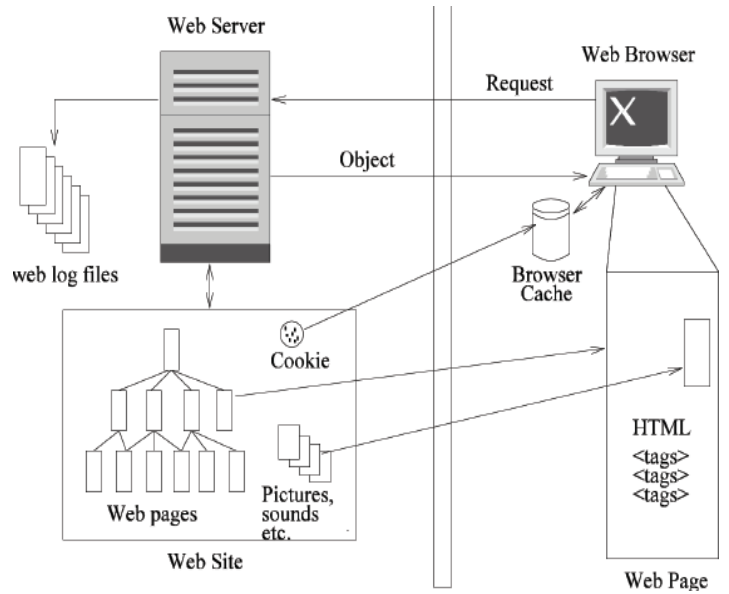


Figure 6: Web Components involve to extract web engineered data