An Unstructured Web Mining For The Recognition About Women Policies of Jaipur Urban Areas In Rajasthan

Mrs. Pooja Jain
Department of Computer Science & Information Technology
JayotiVidyapeeth Women's University, Jaipur (Rajasthan)
Pooja1981jain@yahoo.com

Prof. S. Lal
Department of Science & Technology
JayotiVidyapeeth Women's University, Jaipur (Rajasthan)
dean.fet@jvwu.ac.in

***Abstract*— Web mining is one of the distillate techniques across the World. It is the cornerstone of internet searching. Technology is going to advance observation around web and it pioneers the massive learning. When one thinks about the policies of women, one has to explore for plentiful websites. Using the World Wide Web (W.W.W.), policies can be collected by Content Mining to extract the full data for implement over the integration of data. It will explore to be aware about the perception of securing total satisfaction for a woman. So for obtaining it is grasping the way how one can identify the navigation behavior using the unstructured mining. This will make the women to the necessity of recognition as a reality. This paper focuses on Unstructured Interviews of urban areas for getting data collection. It is analyzed about the policies for Text Summarization using Python. After analysis in the selected urban area of Jaipur, it is revealed that women have very little knowledge about their policies and need to be more and more aware by using web mining. At last implementation phase describes an easy way to access the women's policies by creating codes and generated web page.**

Keywords*— *Web Technique, Execution, Generated Codes, Generated Webpage

I. INTRODUCTION

Web Mining is the technique to extricate knowledgeable data from Web sites. It penetrates the overall wealth of intelligent data from the World Wide Web. Web mining opens the path for searching dynamic, interrelated Web Pages. It uncovers all the information saved behind the W.W.W. It evaluates valuable patterns. Web mining encompasses three ways as Web Content Mining, Web Structure Mining and Web Usage Mining. Web Content Mining follows different methods for excavation such as Unstructured Mining, Structured Mining, Semi-Structured Mining and Multimedia Mining. The hidden Web sources are usually acquired with Unstructured Web Mining. Unstructured or Text Mining can be done in many forms like Information Extraction, Topic Tracking, Summarization, Categorization, Clustering and Information Visualization. Jaipur is the 10th most popular district of India and the capital of Rajasthan. In this paper, it has mentioned 6 out of 13 urban areas of Jaipur as Jaipur, Chomu, Kotputli, Chaksu, Shahpura, and Phulera. Data about acknowledgement of policies from 6 urban areas are being collected by unstructured interview. After applying Text Mining, it is concluded that Rajasthan government has included many points of women's policies which can be accessed and understood using Generated codes and Web pages like to promote the standards of health and survival of women, allowing opportunities for a safe and effective environment,make employment, skilled development and entrepreneurship successful, give equal status in political and social areas for decision and representation, provide legislation to end discrimination and violence, elevate awareness about solar energy and bio-gas, ensuring shelter and property control through legislative and administrative measures, granting opportunities to women in schemes, programs, or institutes and confirming the right to respect and equality to women.

II. OBJECTIVE

The web is a way to store massive data. It is necessary to make the women of the urban area to be aware of policies through web mining. It is essentialfor finding an objective, having a systematic knowledge and searching idea to collect the data. There must be kept about the requirement to enhance women's participation and the need in Web Mining. This research paper contains objective of achievingawareness among women and get satisfaction to access the government schemes. For framing this to the depth and effective direction it will be established to apply text mining and then formulate into generated web page.

III. LITERATURE SURVEY

A. *Khalil, Mohammed Farooque (*2016) [9]

The author describes data warehousing which is beneficial in decision-making. Data mining analyzes the problem and leads us to the appropriate result. The Data warehouse develops and contains the Client Information Management (CIM) and System Development Life Cycle (SDLC) of the project. Data mining applies to stored information of the data warehouse. The data warehouse accesses stakeholders for the future by using data mining.

B. *Gopalakrishnan T (2017)[11]*

In this paper, it elaborates on the classification of web pages for improving the clarity of Web search. It uses the Sentiment classification for the extraction of knowledge. Here the research stresses three phases of Web page classification techniques. It shows the highest degree of similarity and dissimilarity according to clusters in the first phase with reduced computation time and training time. Fuzzy clustering suggests to all data having minimum distances from all centers in the second phase. Escalation of the Fuzzy C-Means algorithm performs in the third phase.

C. *Leila Shahmoradi (2014)[12]*

The author has examined that information should retrieve according to the similarities of documents. We seem that web pages have clustered by using algorithms and Structures as Cross-page links and In-page links have implemented. The experimented consequences are seen effectual to the structures of Web pages in institution's websites. Author centers of attention on virtual structures of Websites.

D. *Pooja Sharma, RupaliBhartiya (2012)[13]*

This paper explains a new algorithm for locating data clustering. It is one of the web mining techniques. The author aims to become betterment of accessing Web log files. It proposes an algorithm for improving time complexity concerning maintain log data. It seems to be typical without the use of log data of a particular user. It has implemented by an algorithm according to the access pattern of web pages for user identification. Web Usage Mining has obtained through the proposed algorithm.

E. *S.Vijayarani, E. Suganya, M. Prakathambal (2018)[14]*

This paper has impacted on varieties of log files for web usage mining research. Through a web browser, when anyone searches for web pages, it creates a log file with the user's time, URL, etc. Using Data Processing, log files pre-processes for getting applicable and inapplicable data. In this, a survey reports about Web server data, Web server logs, etc.

F. *Peter Svec, LubomirBenko, MiroslavKadlecik, JanKratochvil, Michal Munk (2020)[22]*

The author has concluded about the importance of pre-processing data. It has also described about the mistakes in the Web Usage Mining. As pre-processing phase has affect on the quality. If one makes underestimation about pre-processing then it will be directly affect on gained knowledge.

G. *RituBeniwal, Vandana Tanwar (2014)[16]*

The author mentions the Web Personalization to solve the needs of a particular user. Personalization adopts the requirement, navigational behavior, correlation with Web techniques, etc of user's choice by analyzing. Web Personalization has implemented using four techniques as Decision Technique, Hyperlink Based Technique, Content-Based Filtering, and Collaborative Filtering.

H. *AndemariamMebrahtu, BaluSrinivasulu (2017)[20]*

Researchers have worked on Web Content Mining Techniques and Tools. These are in the form of unstructured and heterogeneous data on www. Users randomly mine to the remarkable information from the internet. Further, data organizes into a structured and Semi-structured form through web content mining. Exploratory Mining Tools and Techniques have used for huge information collection.

I. *R.Malarvizhi, K.Saraswathi (2013)[19]*

Researchers use the techniques and algorithms of Web Content Mining. These are related to information retrieval, Database Management Systems, and Artificial Intelligence. It depicts the importance of classification algorithms such as Decision Tree, K-Nearest Neighbor, Naive Bayes, Support Vector Machine, and Neural Network. It presents how to increase information by using more and more data.

IV. EXECUTION OF WOMEN POLICIES

In the execution of women's policy, it is initialized the variables as Unstructured Interviews, Government Women Policies-Websites and Awareness of Women. It is applied Unstructured Text Summarization Techniques using Python. Then codes and pages for women are created using html. If one gets a satisfactory result, it will be stopped. Otherwise, there will be searched for other Websites and continued the same procedure for an acceptable result.

V. COMPARISION STUDY OF JAIPUR URBAN AREAS AND TEXT SUMMARIZATION OF POLICIES

This comparison study is concerned with the knowledge about Women Policies. The Fig. 1 shows the comparison code of recognition about policies. It is presented using 5 urban areas of Jaipur and percentage of awareness about policies (see Fig. 2). This is examined that after collection of secondary data and analysis among urban areas. There seems that it is a crucial matter to grow awareness and it is essential to build software for getting benefits from Policies. The survey has been reached to the decision that percentage of awareness among women is very low by using Fig. 3 and Fig. 4.

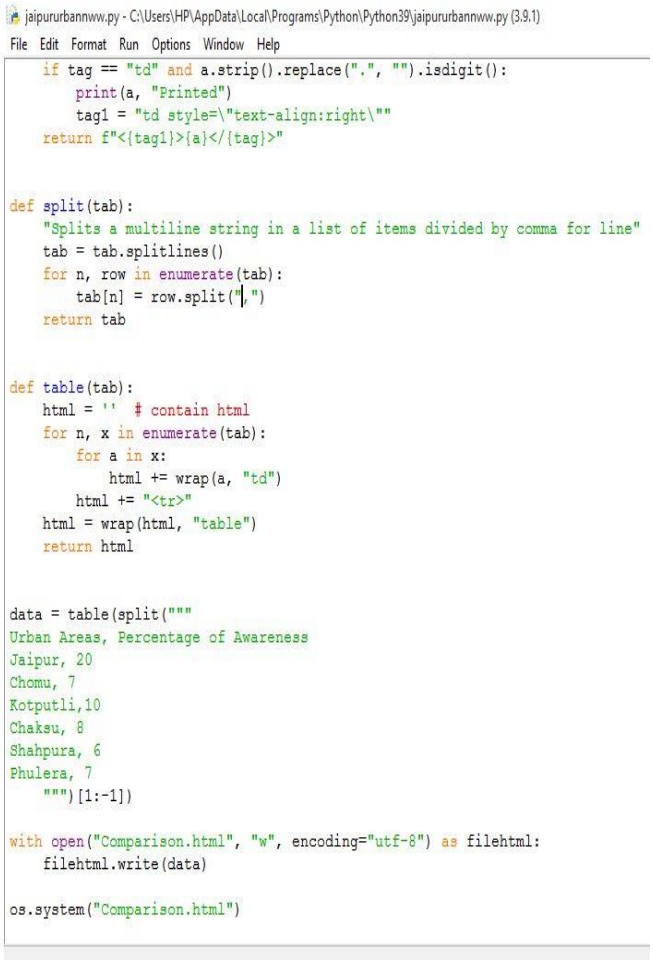
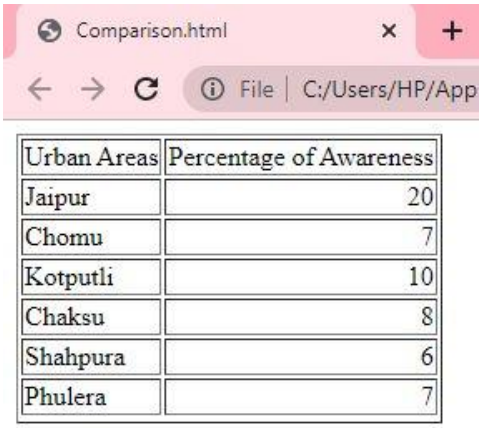


Fig. 1. Comparison Code of Recognition Table for Women Policies



Urban Areas	Percentage of Awareness
Jaipur	20
Chomu	7
Kotputli	10
Chaksu	8
Shahpura	6
Phulera	7

Fig. 2. Table for Comparison of Recognition for Women Policies

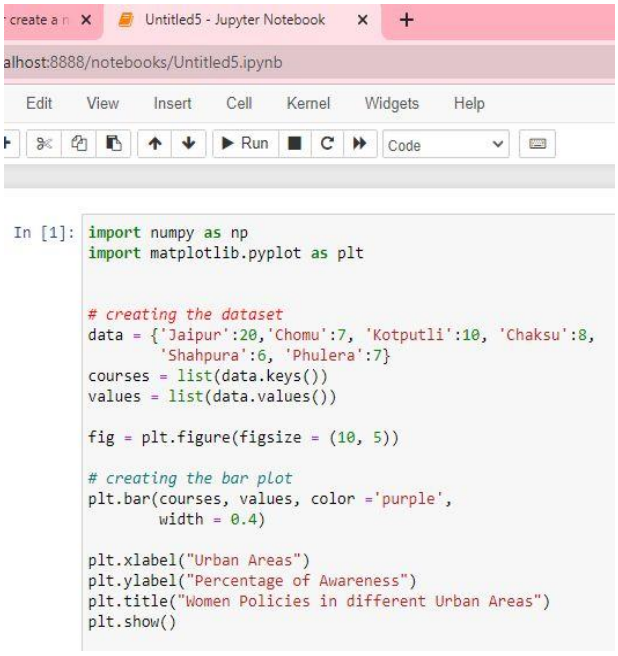


Fig. 3. Comparison Code of Recognition Graph for Women Policies

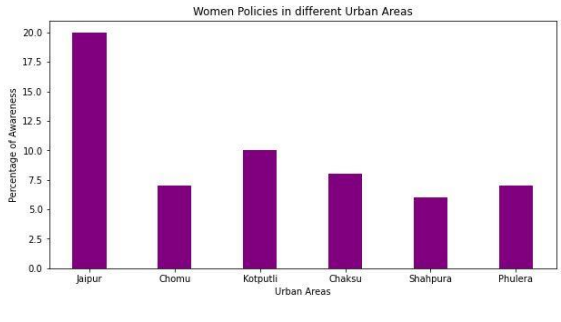


Fig. 4. Comparison Graph of Recognition for Women Policies

After collecting data from the urban area, the technique of increasing acknowledgement about the policies of women using text mining will be resolved. Text Summarization converts unstructured text of Web data about Policies into documents and databases for observation. It is shown below an extraction using spaCy, a Natural Language Processing tool in Python (see Fig. 5).

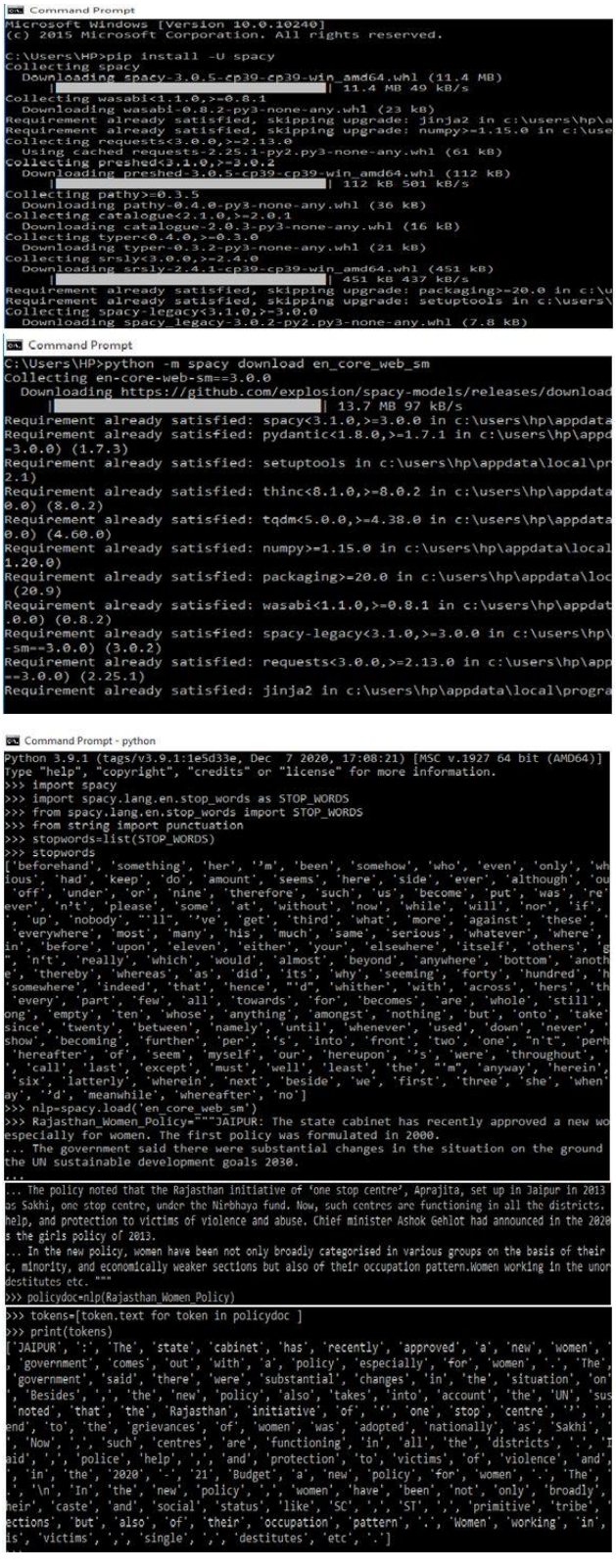


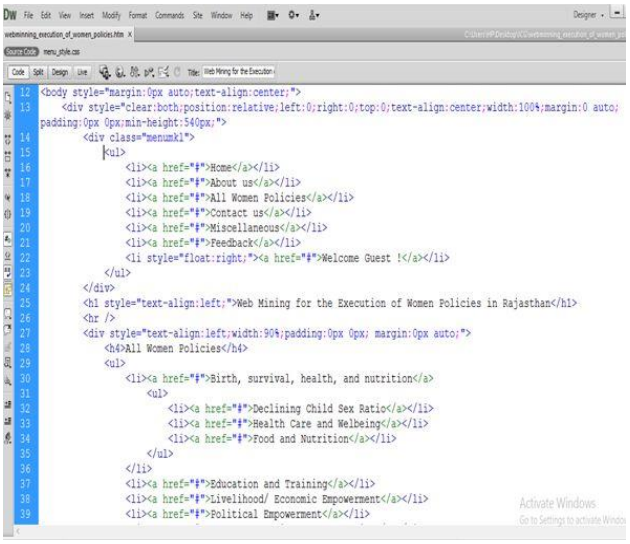
Fig. 5. Text Summarization of Women Policies using Python

VI. ACCESS OF WOMEN POLICIES BY GENERATED CODES

Today there have women policies in Rajasthan but due to the lack of appropriate access to it, women are unable to make use of these Policies because of the inappropriate education. For creating easily access it can be developed by the Website for the implementation. Here the generated

codes and webpage for retrieving particular Policies are shown below (see Fig. 6 and Fig. 7).

Accordingly, it has the coding of the Webpage for accessing the usable data. There have the hyperlinked Policies here. When one clicks on the links another page of schemes will be opened. It shows the easy access of WebPages. It gives the proper execution for Mining.



```
<body style="margin:0px auto;text-align:center;">
    <div style="clear:both;position:relative;left:0;right:0;top:0;text-align:center;width:100%;margin:0 auto;
padding:0px 0px;min-height:540px;">
        <div class="menumk1">
            <ul>
                <li><a href="#">Home</a></li>
                <li><a href="#">About us</a></li>
                <li><a href="#">All Women Policies</a></li>
                <li><a href="#">Contact us</a></li>
                <li><a href="#">Miscellaneous</a></li>
                <li><a href="#">Feedback</a></li>
                <li style="float:right;"><a href="#">Welcome Guest !</a></li>
            </ul>
        </div>
        <h1 style="text-align:left;">Web Mining for the Execution of Women Policies in Rajasthan</h1>
        <hr />
        <div style="text-align:left;width:90%;padding:0px 0px; margin:0px auto;">
            <h4>All Women Policies</h4>
            <ul>
                <li><a href="#">Birth, survival, health, and nutrition</a>
                    <ul>
                        <li><a href="#">Declining Child Sex Ratio</a></li>
                        <li><a href="#">Health Care and Welbeing</a></li>
                        <li><a href="#">Food and Nutrition</a></li>
                    </ul>
                </li>
                <li><a href="#">Education and Training</a></li>
                <li><a href="#">Livelihood/ Economic Empowerment</a></li>
                <li><a href="#">Political Empowerment</a></li>
```

Fig. 6. Generated Codes for Women Policies

Fig. 7. Generated Webpage of Women Policies

VII. RESULTS

The result is the format that leads the research problem to a solution after inputting, analysis and implementation. Through this research, it was seen that in which way the use of government policies should be made easy. Attitude and awareness of women revealed after data collection from unstructured interviews. To make the women aware of the policies in the urban area, the work became easier by adopting Text Summarization Technique. Web pages provided an easy way to increase women's satisfaction and awareness.

VIII. SCOPE OF FUTURE RESEARCH

This research work will be helpful in taking web mining forward. By this, web mining tools and websites will be developed in different languages using ASP.NET, C#, PHP, Python, and SQL Server etc. so it will fulfill the requirement of women appropriately. It will certainly provide automated accessibility of a secured, filtered, more analyzed and manipulated processing.

CONCLUSION

This paper reaches the closure of web mining that it can be mined for getting fruitful and practical knowledge to extract the use of Women Government Policies. The results come to the decision that after interacting among urban areas for gaining the current scenario of women it is needed to implement a clear and easy approach by creating a tool and website for theWomen policies. After summing up it defines that if one clicks on the corresponding sub-link- Declining Child Sex Ratio on the webpage, it is obtained the one scheme by the hyperlink. When one uses this type of Generated Webpage, it will be found the accurate results. In this way, it's beneficial for women to attain the Women Policies. Make use of Web Content Mining and designed into the shape of Web Structure is executed more efficiently. Further, it will explore to the techniques for automating the policies using navigation.

REFERENCES

[1] Inamdar S.A., G.N. Shinde.2011.Web data mining using an intelligent information system design, Int. journal of computer tech, and applications, PP.280-283, ISSN 2229-6093.

[2] Shipra Saini, Hari Mohan Pandey. 2015.Review on web content mining techniques, International journal of computer applications (0975 – 8887) Volume 118 – No. 18.

[3] Ramajayam G., Soundharya V., Likitha M.S.2018. A survey on web mining and web usage mining, Vol-6, E-ISSN: 2347-2693.

[4] Bhanu Bhardwaj. 2012.Extracting data through web mining, International journal of engineering research & technology (IJERT), Vol. 1, ISSN: 2278-0181.

[5] A. A. Barfourosh et al. 2002.Information retrieval on the world wide web and active logic: A survey and problem definition.

[6] Ali Khalili.2008.A semantic web service-oriented model for project management, IEEE 8th International conference on computer and information technology workshops, pp 667 – 672.

[7] Gilles Fedak.2009.BitDew: A data management and distribution service with multi-protocol file transfer and metadata abstraction, Journal of network and computer applications,,Volume 32, pp 961–975.

[8] Harish Kumar, Anuradha, A.K. Solanki, Krishna Kant Singh.2019.Progressive machine learning approach with webastro for web usage mining, ICCIDS, Procedia Computer Science 167 (2020) 1400–1410 .

[9] Mohammed Farooque Khalil.2016. A research study on data mining and warehousing for comprehensive web-based project management software, shodhganga.inflibnet.ac.in

[10] Rosario Girardi, Leandro Balby Marinho, Ismenia Ribeiro de Oliveira.2005.A system of agent-based software patterns for user modelling based on usage mining, Volume 17, pp 567–591.

[11] Gopalakrishnan T. 2017. Certain investigations on web document classification using swarm intelligence, shodhganga.inflibnet.ac.in

[12] Leila Shahmoradi. 2014. Structure-based web pages clustering, international journal of scientific & engineering research, Volume 5, ISSN 2229-5518.

[13] Pooja Sharma, Rupali Bhartiya. 2012. An efficient algorithm for improved web usage mining, Vol 3 (2), 766-769, ISSN: 2229-6093

[14] S.Vijayarani, E. Suganya, M. Prakathambal. 2018. Web log files in web usage mining research – a review, Vol 5, ISSN 2394-2320.

[15] A. Jebaraj Ratnakumar. 2005 – 2010. An implementation of web personalization using web mining techniques, journal of theoretical and applied information technology, JATIT

[16] Ritu Beniwal, Vandana Tanwar. 2014. Evaluation of web personalization, IJIRST, Volume 1, ISSN: 2349-6010

[17] Ayesha Ameen, Khaleel Ur Rahman Khan, B.Padmaja Rani. 2012. Semantic web personalization: a survey, information and knowledge management, Vol 2, ISSN 2224-5758

[18] Tsuyoshi, M and Saito, K. 2006. Extracting user's interest for web log data, IEEE, pp. 343-346, ISBN: 0-7695-2747-7

[19] R.Malarvizhi, K.Saraswathi. 2013. Web content mining techniques tools & algorithms – a comprehensive study, IJCTT, volume 4, ISSN: 2231-2803

[20] Andemariam Mebrahtu, Balu Srinivasulu. 2017. Web content mining techniques and tools, ijcsmc, Vol. 6, pg.49 – 55, ISSN 2320–088X

[21] Qingyu zhang, Richard s. Segall. 2008. Web mining, a survey of current research, techniques, and software,IJITDM, Vol. 7, 683–720

[22] Peter Svec, Lubomir Benko, Miroslav Kadlecik, Jan Kratochvil, Michal Munk. 2020. Web usage mining: data pre-processing impact on found knowledge in predictive modelling, CoCoNet'19, Procedia Computer Science 171, 168–178