

## **WEB PAGE RECOMMENDATION SYSTEM BY INTEGRATING ONTOLOGY AND STEMMING ALGORITHM**

Mohamed Uvaze Ahmed Ayoobkhan  
Department of Computing,  
Westminster International University in Tashkent,  
Uzbekistan.  
*mohamed.sha33@gmail.com*

Liayakath Ali Khan Subair Ali  
Department of Computing,  
Westminster International University in Tashkent,  
Uzbekistan.  
*sliayakath@wiut.uz*

**Submitted:** Dec, 15, 2021 **Revised:** Feb, 18, 2022 **Accepted:** Feb, 24, 2022

**Abstract:** In this research, we offer a customized-recommendation system that uses item representations and user profiles based on the ontologies that provide personalized services to semantic applications. To develop and implement the personalized-recommendation system, a system that uses the representations of the items and the user profiles based on the ontologies to provide the semantic applications with personalized services. Recommendation systems can use semantic reasoning capabilities to overcome present system limits and increase the quality of recommendations. The recommender makes use of domain ontologies to improve personalization: on the one hand, a domain-based inference method is used to model user interests more effectively and accurately; on the other hand, a semantic similarity method is used to improve the stemmer algorithm, which is used by our content-based filtering approach, which provides a measure of the affinity between an item and the user. In recommender systems and web personalization, Web Usage Mining is crucial. This study presents an effective recommender system based on ontology and web usage mining. The approach's first step is to extract features from online documents and build on related ideas. Then, they create an ontology for the website using the concepts and relevant terms retrieved from the records. The semantic similarity of web documents is used to group them into multiple semantic themes, each with its own set of preferences. The suggested solution incorporates ontology and semantic knowledge into Web Usage Mining and personalization procedures, as well as a stemming algorithm, and gets an overall accuracy of 90%.

**Keywords:** Personalized recommendation, user profiles, ontology, stemming algorithm, feature extraction and semantic knowledge.

### **I.INTRODUCTION**

The ability to extract proper knowledge from the Web usage statistics and to portray that knowledge in a way that allows for good Web-page suggestions is critical and challenging. Existing systems provide a method for offering enhanced Web-page suggestions by way of semantic improvement. This is accomplished by combining the domain knowledge of a website with information about how the website is used on the Web [1-2]. In this article, we provide two fresh models for describing domain knowledge. The ontology is used to represent subject matter expertise in the first model. The second model uses a semantic network that has

been artificially formed to describe domain names, Web sites, and the links between them. The conceptual prediction model is an additional innovative way that was created for automatically building a semantic network of the semantic Web use knowledge [3-4]. This method combines the information about the domain with the information about how people use the Web. There have been a lot of different queries developed to look into these information stores. A set of recommendation algorithms for producing Web-page candidates has been provided as a result of these enquiries. The outcomes of the proposal were evaluated and contrasted with those of a more sophisticated earlier Web Usage Mining (WUM) technique [5-6].

Web user pages are suggested to users depending on the queries they enter by analyzing the data that users save on their own websites. An efficient web page recommendation system is produced by skillfully combining a number of different strategies. Latent Semantic Analysis is used to the User Method-Term Matrix and the Term Approach-Frequency Matrix, both of which are constructed with data collected from Web Usage, in order to produce a Term Prioritization Vector [6-7]. Techniques such as Latent Dirichlet scheduling for Topic-based Segregation of the URLs and the Normalized Pointwise Mutual data methods are applied in order to create web pages that are dependent on the searches that visitors do. The Prioritization Vector is used to achieve personalization via the use of ranking websites. In addition to that, the system makes use of an internal algorithm to locate websites that may be trusted [8].

The phrase "web page suggestion" refers to a strategy that is used to propose online user sites to other users based on their prior navigational behaviour, which is determined by their utilization of web data. This approach is used on the internet. Any approach for mining web use may be compared to sequential rule mining, which is applied to extract information about the behaviour of users based on how the data they generate is utilized [9-10]. To make improvements to the online user page suggestions, it is required to add semantic information of the web pages into the process of making recommendations. A knowledge representation that is capable of successfully conveying the real meaning of a web page is called ontology. This study proposes strategies for the partial-automatic generation of ontologies within the context of a customized system framework for web pages that is based on web mining technologies [11].

A filtering approach known as a recommendation system calculates the user's priority or rating of an item based on the user's input. The individualized system is used for a variety of purposes, including but not limited to the generation of audio or video playlists for users of media streaming applications, the proposition of social media posts, and the recommendation of items on e-commerce platforms. The process of web mining involves removing current-affairs-related terms from the user's archived online content. WordNet is put to use in order to provide a consistent knowledge base with appropriate synonyms for the data that was gathered. On the knowledge base, classification methods such as Random Forest are used to classify the standardized dataset, and Ant Colony Optimization is used to discover the shortest route across the graphs. Both of these methods were developed using artificial intelligence. These data are then fed into machine learning algorithms, which are used to categories the standard dataset [12-13].

In order to give patients with incredibly relevant web sites, web page health recommender systems are introduced with the employment of specific agents. The primary characteristic of Particle Agent Swarm Optimization (PASO) is that the algorithm is created by a group of Particle agents who work together to achieve the task's objective, User behaviour analysis for enhancing website performance interests many researchers since the number of websites grows daily. In this paper, the web log feature of web mining is used to introduce the web page

recommendation model. For the purpose of determining the association rule using a single data iteration technique, the Feed Forward Counter Model (FFC) has been introduced in this study [14-15].

To recommend these tasks, a hybrid technique of collaborative filtering and content-based filtering is used. The goal is to compile and recommend suitable employment to job seekers, particularly in the engineering field [16]. Creating a model to produce the best list of recommendations that maximizes the Click-Through Rate (CTR) of the entire homepage is a difficult job. It has been suggested to use an entity enrichment approach to recommend web pages. The method includes entities enhanced by Resource Description Framework generation, semantic frame matching, and background information from the Linked Open Data cloud. Social awareness is added by adding entities from the Twitter API [17-18].

## **II. PROPOSED SYSTEM**

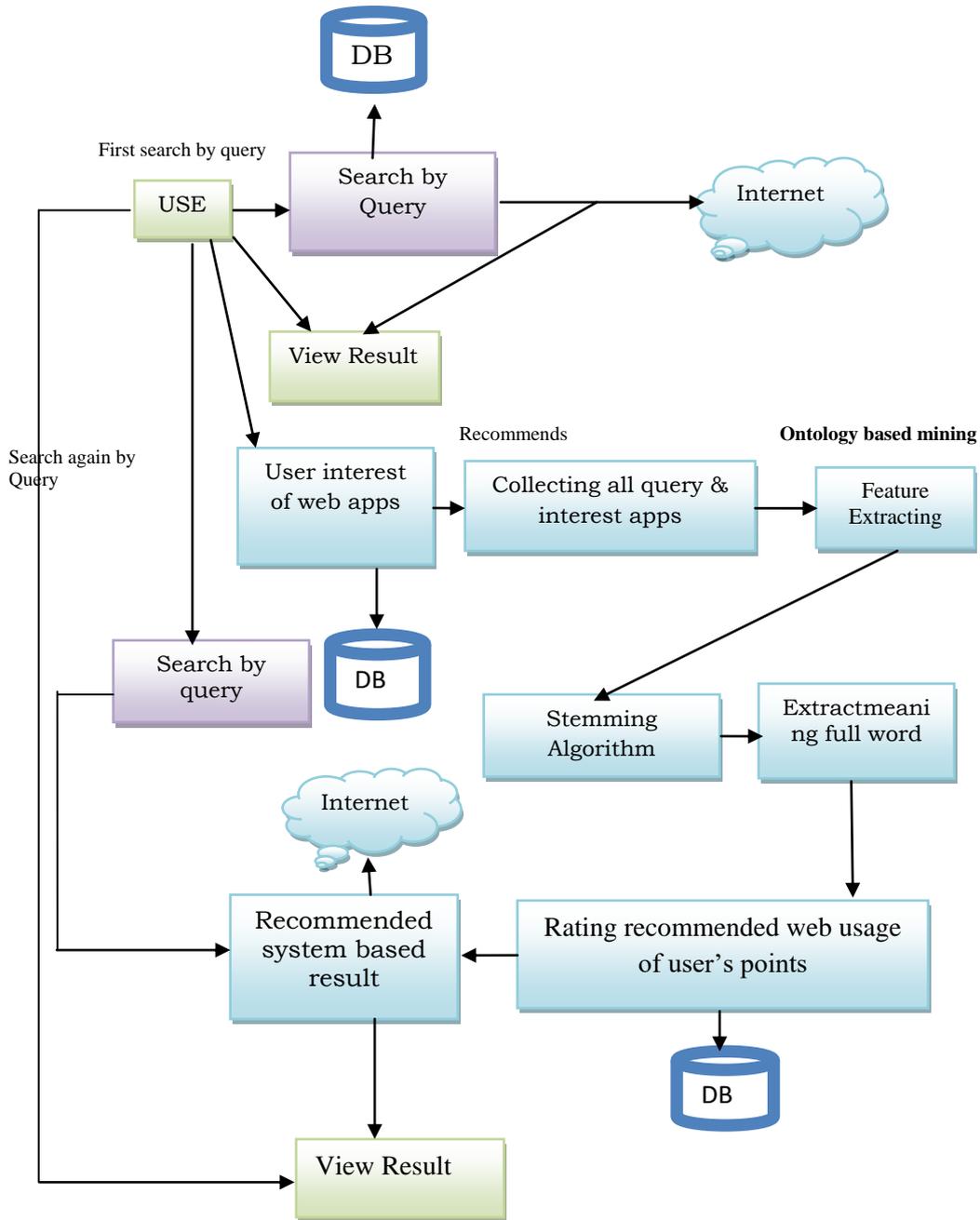
The suggested method for recommending web pages to users use ontology-based data mining in conjunction with the stemming algorithm to construct user profiles. All of the interesting queries and applications are taken from the database and compiled by a recommendation engine that is based on ontology. The process of mining ontologies produces knowledge representations that are reusable and shareable and also establishes a connection between the various representations. It manages the millions of records that are searched by the query and are saved on the server that is accessed over the internet. The next step is the extraction of features, which involves minimizing the quantity of data that was previously collected and obtaining valuable information from the server. In this research, we coupled ontology with feature extraction and semantic analysis with a stemming algorithm to see the best result at the top of the web page depending on the user's query. This was accomplished by combining these three aspects.

The stemming method offers a robust model and normalizes data by eliminating repeated terms and reducing them to their root words that are kept in the database. This helps the algorithm produce a more accurate representation of the data. Because of this, we are able to boost the efficiency of general online searches based on individualized recommendations. The extraction of the base form from a large quantity of data is the goal of the stemming algorithm, which aims to reduce the amount of time and memory space required for the process. The goal of semantic analysis, which is a subfield of natural language processing (NLP), is to comprehend the connotations associated with natural language. The ability to immediately comprehend natural language is unique to humans; nonetheless, it is a challenging endeavor for computers. The use of semantic analysis is beneficial to machines. The proposed system is shown in Figure 1. It illustrates that the users initially search by query to view the result, which is stored in the database through the internet.

### **A. Lovins**

For its age, the stemmer is regarded as sophisticated, with an extensive list of ends that can be eliminated and a list of the exceptions and the recording. The stemmer makes advantage of this enormous finishing list in an effort to steer clear of an iterative stemming operation while still removing more than a single order class of endings from the output. Because the list has all of the common word suffices, such as *ing* and *ly*, as well as valid matches of these ends, like *ly* and *ing*, there is no need to iterate because the list contains both. The efficiency of the

process is improved by sorting the ends first according to length and then by alphabetical order. Since the stems are ordered according to their length, the longest matches will be found first when the stemmer is used. This eliminates the need to do any more research since the stemmer is designed to find the longest possible match.



**Fig. 1 Web page recommendation using Ontology based mining with Stemming Algorithm**

## **B. Porter**

Martin Porter, a graduate of the University of Cambridge, is credited with inventing the Porter stemmer, which made its debut in 1980. The stemmer is a method for removing suffixes that takes into account the surrounding context. It is the stemmer that is implemented in the most languages, making it the most widely used of all the stemmers. The stemmer is split into five or six linear stages, depending on how a step is defined, that are utilized to produce the final stem. Depending on how a step is defined, the stemmer is divided into five or six linear steps.

## **C. Paice/Husk**

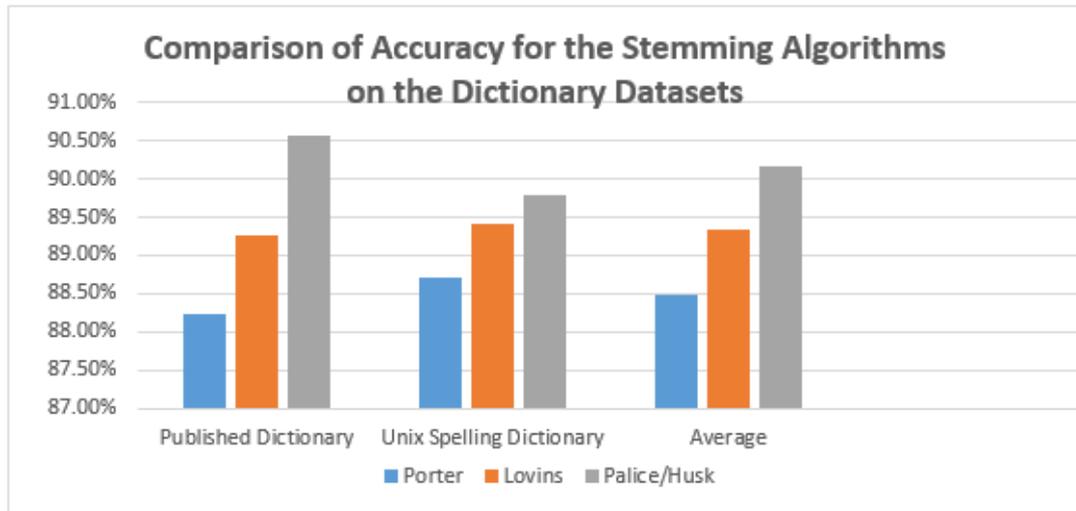
The Paice/Husk stemmer was originally released in 1990 and was the result of a collaborative effort between Chris Paice and Gareth Husk. This is an iterative conflation stemmer, which is being used. The stemmer has a well-deserved reputation for being incredibly powerful and aggressive while yet managing to maintain its effectiveness and simplicity. This replacement procedure is essential in order to get around the issue of spelling text exceptions that was covered in the preceding paragraph. The English stemmer eliminates the requirement for a separate step in the stemming process by altering the ends rather than just removing them. This means that no recoding of words or partial matching is required once the stemming process has been completed. Because of this, the algorithm is able to keep its efficiency while still performing its intended function.

A recommendation system that makes use of ontology-based mining compiles all of the fascinating applications and questions that are asked. In order to accomplish browsing, suggestion, and profile generation, it takes use of the hierarchical structures that are present in both the users and the things. The meaningful word will be eliminated by the system once the feature has been extracted with the help of a stemming algorithm. In order to remove unnecessary information such as the precise form of the word, a stemming algorithm is used. This generates word bases, which are then used as features for the classification of user points and recommended web usage, and allows users to view the results that are stored in the database.

One strategy for stemming is to maintain a database that contains all of the index words and their respective stems. After that, the words from the queries and the indexes might be stemmed with the use of hash tables in a lookup table. The study of structural linguistics, which makes use of the distribution of words to determine word and morpheme boundaries, provides the foundation for a variety of successor stemmers. Affix removal stemmers are used to remove any prefixes or suffixes from the words, leaving just the stem.

## **III. RESULTS AND DISCUSSIONS**

The Moby Common Dictionary Dataset, which includes 74,500 words from 25 texts, is used to assess performance. The frequent or meaningless words, acronyms, hyphenated terms, and other items are deleted using simple preprocessing processes like tokenization and stop word removal. Finally, 49,650 general English words with an average length of 8.05 characters are in the final word list.



**Fig. 2 Comparison of Accuracy for Porter, Lovins&Palice/Husk**

To evaluate the accuracy of three stemming algorithms, we use the following:

$$Accuracy = \frac{\text{Number of correctly stemmed words}}{\text{Number of unique words}} \quad (1)$$

This is based on the number of unique words in the given datasets and the number of words correctly stemmed by the stemming algorithm.

**TABLE. 1 Comparison of Accuracy for Stemming Algorithms**

Datasets	Accuracy		
	Porter	Lovins	Palice/Husk
Published Dictionary	88.23%	89.25%	90.56%
Unix Spelling Dictionary	88.72%	89.42%	89.78%
Average	88.48%	89.34%	90.17%

We get the average accuracy of these three stemming algorithms using these three datasets. Porter performs with an average accuracy of 88%, Lovins performs 89%, and Husk performs 90%. Figure 2 and Table 1 show that Paice/Husk performed better than the other two types, i.e., 90%.

#### IV. CONCLUSIONS

In this paper, an efficient web page recommendation system by integrating ontology and stemming algorithms is proposed. First, the user will submit the requests via the search engine in the form of queries. The server provides a response with the pertinent data, which is then saved in the database. After that, ontology-based mining using a stemming algorithm is done to decrease the unnecessary information and cluster all of the important data that is recorded in the database in order to provide the precise result at the top of the web page that was advised by the user. Our suggested technique got better results than theirs

did after being tested in a number of tests on large-scale datasets, which we carried out. In this paper, we have presented a new framework for a web page recommendation system from the user's viewpoint of internet user preference learning. This framework incorporates the combination of ontology with feature extraction and semantic analysis with a stemming algorithm to achieve a better performance of recommendation system when compared to the existing system, which achieves a performance of 90%. In addition, this framework is based on the user's perspective of internet user preference learning. Using this procedure, the user profile rating matrix's missing values may be filled in with the quality of the internet rating that has been developed. In the future, we are going to investigate both functional and non-functional assessment as potential user models for the issue of web page suggestion.

### **REFERENCES**

- [1]. D. Surya, G. Deepak and A. Santhanavijayan, "KSTAR: A knowledgebased approach for socially relevant term aggregation for web page recommendation," International Conference on Digital Technologies and Applications, 2021, pp. 555-564.
- [2]. G. Matošević, J. Dobša and D. Mladenčić, "Using machine learning for web page classification in search engine optimization," Future Internet, vol. 13, no. 1, 2021, pp. 1-9.
- [3]. K. Kikuchi, M. Otani, K. Yamaguchi and E. Simo-Serra, "Modeling Visual Containment for Web Page Layout Optimization," Computer Graphics Forum, vol. 40, no. 7, 2021, pp. 33-44.
- [4]. S. Singh and M. S. Aswal, "Constructing an effective ontology for web page recommendation," International Journal of Web Engineering and Technology, vol. 16, no. 2, 2021, pp. 86-112.
- [5]. G. Rani, V. S. Dhaka, U. Pandey and P. K. Tiwari, "Intelligent and Adaptive Web Page Recommender System," International Journal of Web Services Research, vol. 18, no. 4, 2021, pp. 27-50.
- [6]. J. Kiesel, L. Meyer, F. Kneist, B. Stein and M. Potthast, "An Empirical Comparison of Web Page Segmentation Algorithms," European Conference on Information Retrieval, 2021, pp. 62-74.
- [7]. C. Dang, H. Randrianarivo, R. Fournier-S'Niehotta and N. Audebert, "Web Image Context Extraction with Graph Neural Networks and Sentence Embeddings on the DOM tree," arXiv preprint arXiv: 2108.11629.
- [8]. S.E. Oliver, "The advisory committee on immunization practices' Interim recommendation for use of moderna COVID-19 vaccine-United States, December 2020," MMWR Morbidity and mortality weekly report, vol. 69, 2021, pp. 1-4.
- [9]. X. Chamberland-Thibeault and S. Hallé, "An Empirical Study of Web Page Structural Properties," Journal of Web Engineering, vol. 20, no. 4, 2021, pp. 971-1002.
- [10]. C. Lo, H. Yu, X. Yin, K. Shetty, C. He, K. Hu, J. M. Platz, A. Ilardi and S. Madhvanath, "Page-level Optimization of e-Commerce Item Recommendations," Fifteenth ACM Conference on Recommender Systems, 2021, pp. 495-504.
- [11]. G. Deepak, B. N. Shwetha, C. N. Pushpa, J. Thriveni and K. R. Venugopal, "A hybridized semantic trust-based framework for personalized web page recommendation," International Journal of Computers and Applications, vol. 42, no. 8, 2020, pp. 729-739.

- [12]. X. Zhang and H. Wu, "PageRank Algorithm and HITS Algorithm in WebPageRanking," International Conference on Application of Intelligent Systems in Multi-modal Information Analytics, 2021, pp. 389-395.
- [13]. T. Khede and A. Sharma, "Bio-Geography Based Page Prediction Using Web Mining Feature," International Journal of Science, Engineering and Technology, vol. 7, no. 6, 2019, pp. 1-5.
- [14]. R. Manikandan and V. Saravanan, "A novel approach on Particle Agent Swarm Optimization (PASO) in semantic mining for web page recommender system of multimedia data: a health care perspective," Multimedia Tools and Applications, vol. 79, no. 5, 2020, pp. 3807-3829.
- [15]. B. Pavithra and M. Niranjanamurthy, "Web page recommendation using genetic and feed forward association rule on web-log features," Journal of Computational and Theoretical Nanoscience, vol. 17, no. 9-10, 2020, pp. 4462-4467.
- [16]. Q. Hao, T. Luo and G. Huzhang, "Re-ranking with constraints on diversified exposures for homepage recommender system," arXiv preprint arXiv:2112.07621. 2021, pp. 1-10.
- [17]. N. Kumar, M. Gupta, D. Sharma and I. Ofori, "Technical Job Recommendation System Using APIs and Web Crawling," Computational Intelligence and Neuroscience, vol. 2022, 2022, pp. 1-13.
- [18]. G.S. Chhatwal and G. Deepak, "IEESWPR: An Integrative Entity Enrichment Scheme for Socially Aware Web Page Recommendation," Data Science and Security, 2022, pp. 239-249.