ORIGINAL RESEARCH ARTICLE UHD JOURNAL OF SCIENCE AND TECHNOLOGY

Semantic Web Recommender System over Different Operating Platforms

Halo Khalil Sharif, Kamaran Hama Ali. A. Faraj

Department of Computer Science, College of Science, Sulaimani University, Sulaimani, KRG, Iraq

ABSTRACT

Semantic-Web Recommender System (SWRS) evaluation over different operating systems (OSs) used to facilitate and improve human electronic recommendation management (HERM). The HERM is address the needs of user and dataset of movie in our proposed system through internetworking means which increase the speed of automated recommendation and enhance the goodness of SWRS and services also electronically to select right movies-title to user demand. Furthermore, it will be a benefit for selection a right favor by user for right selection from (i.e., 3000 records in dataset of movie-Lens) in the backend. There are a direct relation between time-consume of selection movie-title, also the time-consume, and accuracy. The two-mentioned parameters, namely, time-consume and accuracy over two different operation system (OSs) which designed by web technology Python. In our research, SWR system is proposed; it is provide with some recommendation methods. The system designed and improved using content-based algorithm (CBA). Investigational results indicate that the developed algorithm technique confident a reasonable performance such as accuracy and time consuming compared to other existing works with a testing average accuracy of 85.63 for windows and 88.35 for Linux operating system. In conclusion, SWRS investigated on two different operating platforms and could be seen that the Linux is faster than windows in accuracy and time consuming.

Index Terms: Semantic Web, E-Recommender System, Content-based, RDF, SPARQL, Python

1. INTRODUCTION

Semantic-Web (as a recommender system) together with all the necessary tools and methods required for creation, maintenance, and application. In actual history, the Semantic-Web is usually future as an heightening of the present World Wide Web (WWW) or (3W) with machine-justifiable data (rather than a large portion of the ongoing Web, which is generally focused on at human utilization), together with services – intelligent agents [1]. Nevertheless, in our proposed system used to facilitate and improve human electronic recommendation management (HERM) is mean that the current web became semantic web (SW) with recommender system (RS). The human consumption artificial intelligent (AI) modify to Semantic-Web Recommender System (SWRS). Our contribution is SW instead of human consumption and RS instead AI and combine to SWRS. However, our proposed system is SW with the cosine similarity is a method and part of content-based algorithm (CBA) for filtering all title-movie in dataset of movie-Lens [2]. The resource description framework (RDF) suggest to graph-based data model, which became part of the Semantic-Web vision [3], the RDF in our proposed system is very necessity with a view to represent data that recommended the title-movie and store into the RDF file. The RDF is much more accurate than the ontology file due to: (1) Easy to use, (2) easy to understand also, nd (3) accurate. Apart from one parameter that used two parameter to enhance accuracy and execution consume-time. The ontology modified from only one parameter to two parameters in propose of

Access this article online

DOI:10.21928/uhdjst.v6n2y2022.pp21-19-24

E-ISSN: 2521-4217
P-ISSN: 2521-4209

Corresponding author's e-mail: halo.sharif@univsul.edu.iq

Received: 14-07-2022 Accepted: 28-07-2022 Published: 10-08-2022

the system. Content-based RS make suggestions that consider the users the ratings that users give to items according to their preferences and the content of the items (e.g., extracted keywords, title, pixels, and disk space). The content based algorithms with using the filtering technique is a main idea of our proposed system. The training algorithm is start first for training all dataset to predict the movie-title that situated between limitation and after that, the TEST algorithm is start to filtering of training output. The activates are depend on training algorithm and TEST algorithms between the user's demands and movie's title (plus demands) to build the SWRS decisions. Semantic-Web utilizes the Resource Description Framework (RDF) and the Simple-Protocol and Query/Update Languages (SPARQL) as uniform logical data illustration and handling models, permitting machines to straight interpret data from the Web. As Semantic-Web, applications is growing progressively popular, new-fangled and stimulating threats of security arise [4], it is impossible to achieve our proposed system or any evaluation without RDF because in RDF is store and transfer data to web application through SPARQL. The two parameters that mentioned namely tag line and original title. Nevertheless, the only used parameter is overview parameter used in Cami *et al.* [5]. The contribution in our proposed system is two parameters. While the deployment of (www) and the internet was swiftly increasing, the recommendation outfits become electronic to support e-commerce (EC) business. Usually, the concept of E-recommender is relevant with all kinds of digitalizes businesses and it uses three-tier architecture [6]. Regardless of the fantastic measure of data that is accessible in the reality or on the Web, it is difficult for the searcher to track down items or services that he may be interested in. Decision-making is an essential part that the traditional and electronic recommendation should do. The vast amounts of digitally available candidate information denote a sizeable opportunity for improving matching quality and it leads to better web semantic recommendation performance [7]. This paper proposes a new procedure for recommending movie-titles using a content-based filtering algorithm and generally used dataset (MovieLens). The whole of the paper is arranged as follows. Section 2 places forward a literature review. Section 3 shows a complete SWRS for the recommending of movie-titles, containing units like an outline of system architecture, MovieLens dataset description, data preprocessing, feature extraction, and performance metrics. Section 4 discusses the experimental results achieved after applying different feature extractors and comparing them from different platforms with the existing methods. Finally, Section 5 deals with the conclusion of the work.

2. LITERATURE SURVEY

In Semantic-Web Recommender System (SWRS), techniques have conveyed exceptional outcomes; these techniques are regularly acted in the recommendation on movie-titles dataset. Recently, various works were executed with the assistance of various content-based methodology to distinguish and predict of movie-titles. A short audit of a few significant contributions from the current literature is given.

Soumya Prakash Rana (2020) [8] proposed arrangement, health recommender systems (HRS) have arisen for patient-situated decision-making to suggest better medical care guidance in light of profile health records (PHR) and patient data sets. The HRS can upgrade medical services frameworks and at the same time oversee patients experiencing a scope of various sicknesses utilizing prescient investigation and suggesting fitting therapies. A content-based recommender system (CBRS) is a tweaked HRS approach that focuses on the assessment of a patient's set of experiences and "learns," through AI (ML), to produce forecasts. Moreover, CBRS plans to offer personalized and believed data to the patient's with respect to their health status.

Donghui Wang (2018) [9] they fostered a content-based diary and meeting recommender framework for software engineering and innovation. To the extent that, there is no comparative recommender system or distributed strategy like what they have presented here. Besides, there was no dataset to utilize. Hence, the web crawler has been intended to gathering information and creates preparing and testing informational indexes. Then, unique component determination techniques and played out few trials used to choose a decent system and recreate include space. Despite the fact that accomplishing 61.37% exactness for paper proposal.

Ibukun Tolulope Afolabi (2019) [10], in this examination, showed a semantic-web content digging approach for recommender frameworks in web based shopping. The strategy depends on two significant stages. The primary stage is the semantic preoperational of text-based information utilizing the blend of a created cosmology and a current metaphysics. The subsequent stage utilizes the Naïve Bayes calculation to make the proposals. The result of the framework is assessed utilizing accuracy, review and f-measure.

Carlos Luis Sanchez Bocanegra (2017) [11] this shows the practicality of utilizing a semantic content-based recommender framework to enhance YouTube health

recordings. Assessment with end-clients, notwithstanding medical services experts, will be expected to distinguish the acknowledgment of these suggestions in a no simulated data looking for setting. Most of sites suggested by this framework for health recordings were pertinent, in view of evaluations by health experts.

Albatayneh (2018) [12], this examined to present an original proposal engineering that can prescribe intriguing post messages to the students in an e-learning on the web conversation gathering in view of a semantic content-based separating and students' negative appraisals. We assessed the planned e-learning recommender framework against leaving e-learning recommender frameworks that utilization comparable sifting methods concerning suggestion exactness and students' exhibition. The got exploratory outcomes display that the suggested e-learning recommender framework beats other comparative e-learning recommender frameworks that utilization non-semantic content-based separating strategy (CB), non-semantic content-based sifting method with students' negative appraisals (CB-NR), semantic content-based sifting procedure (SCB), concerning framework precision of around 57%, 28%, and 25%, separately.

3. PROPOSED METHODOLOGY

3.1. System Architecture

TF-IDF is used for the vectorization of the information and cosine similarity is utilized to compute the similarity measure between the vectors. TF-IDF is normally used as a portion of content-based algorithm recommendations systems in proposed system. It contains of two positions: Term-Frequency (TF) and Inverse-Document-Frequency (IDF). TF deals-with the occurrence of interests and preferences in user profile. Whereas, IDF deals with inverse of the word frequency among the entire data provided by user profile. These two theories are joint together to present the recommendation for a user based on the data's presented by user profile. Cosine similarity be able to catch the similarity among two attribute or more from the dataset found by determining cosine value between two vectors or more. Use of cosine similarity can be executed on any two texts such as documents, sentences, attributes or paragraph. Occasionally through the similarity measurement between the vectors which produce unstable results. Finally, the SWRS are build using famous algorithm content-based (CB) and RDF. The important steps in proposed structure design are shown in Fig. 1. In the below figure shoe all steps as an instruction of our system. ROW one show all main steps, but the underneath RAW is subset of first RAW. RAW one and two are complete each other's for the sake of processes of the system.

3.2. Dataset Explanation

The proposed system was trained as well as tested on the MovieLens dataset. The dataset consists of movies released on or before July 2019. Information focuses incorporate cast, group, plot, watchwords, spending plan, income, banners, delivery dates, dialects, creation organizations, nations, TMDB vote counts, and vote midpoints. The Complete MovieLens Datasets comprising 26 million evaluations and 750,000 label applications from 270,000 clients on every one of the 45,000 motion pictures. This dataset is a troupe of information gathered from TMDB and GroupLens. The Movie Detail, Credit and Keyword have been gathered from the (TMDB) open an API. This item utilizes the TMDb API however is not embraced or affirmed by TMDb. Their API likewise gives admittance to information on numerous extra motion pictures, entertainers and entertainers, group individuals, and TV shows. The Movie Links and Ratings have been gotten from the Official GroupLens site. A portion of the things you can do with this dataset: Predicting film income or potentially film achievement in view of a specific measurement. What motion pictures will generally get higher vote counts and vote midpoints on TMDB? Building Content-Based and Collaborative Filtering Based Recommendation Engines [13].

3.3. Preprocessing

To be capable handling information concurring appropriately, really, and productively, that it requires the capacity as far as

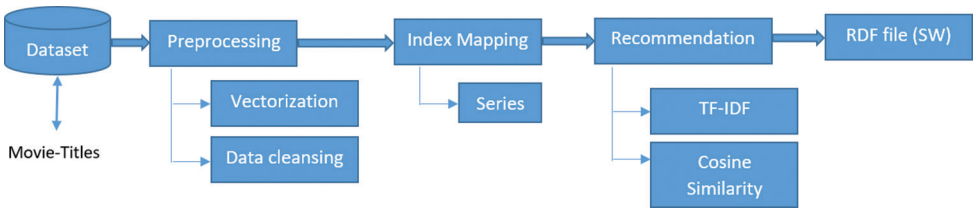

Fig. 1. Architecture of proposed system.

a specific programming language that is explicitly devoted to handling information or data in many place of origin in the association or the web to turn into a valuable information researcher for associations or organizations [14], because of in the proposed technique (fillna) method is used to cleansing data from the dataset to achieve the best result. Scikit-learn is a permitted software (utility) machine-learning library for the Python programming language. It assists python numerical and scientific libraries, in which Tfidf-Vectorizer is one of them. It alters a group of raw documents to a matrix of TF-IDF structures. As tf–idf is extremely frequently used for text sorts, the class Tfidf-Vectorizer merges all the options of Count-Vectorizer and Tfidf-Transformer in a particular model. Tfidf-Vectorizer uses an in-memory vocabulary (a python dict) to map the most recurrent words to features indices and henceforward calculate a word occurrence frequency (sparse) matrix, the class of TfidfVectorizer used to vectorizing the two attribute from the dataset (MovieLens) in our proposed system.

3.4. Recommendation Engine

Term-Frequency Inverse-Document-Frequency (TF-IDF) is utilized to yield recommendations to the user's favorites. Each data attribute from the datasets is converted into a vector by applying the TF-IDF vectorization algorithm described before. For each vector, a similarity measure is calculated using the cosine similarity method. When a user requires number of recommendations for a certain movie, the correspondence quantities are produced for the movies with concern to that movie. Individually similar movie detected will have a confident score of how similar it is to the represented movie, which is sorted into descending order, because of list the movies with high to low similarity. Conferring to the amount of recommendations demanded by the user, the indices of those movies are gathered and showed to the user as a list of movies. The recommendations created by the engine are displayed over a user interface to the user; the engine is trained to yield similarity measures using the training data. The backend is scripted using Python language, whereas the calculations performed from Equations 1 to 4 to find Cosine-Similarity and TF-IDF [15].

Cosine similarity, Based on vector similarity, similarity among vectors can be denoted as Eq. 1:

$$cos(\theta) = \frac{\sum_{i=1}^{n} AiBi}{\sqrt{\sum_{i=1}^{n} Ai^2}\sqrt{\sum_{i=1}^{n} Bi^2}} \quad (1)$$

Where, Ai and Bi are components of vector A and B respectively:

TF, i.e. word frequency, indicates the frequency of terms in the text showed in Eq. 2.

$$tf_{i,j} = \frac{n_{i,j}}{\sum_{k} n_{i,j}} \quad (2)$$

Where, n is the frequency of terms in the movie-title.

IDF, i.e. inverse document frequency, represents the reciprocal of the quantity of movie-title containing words in the mass displayed in Eq. 3.

$$idf_{j=log\left[\frac{n}{df_j}\right]} \quad (3)$$

Where, n is the frequency of movie-title containing words

Thus, the TF-IDF weight for catchphrase in record can be composed in Eq. 4

TF-IDF = (Frequency of words/Total words of sentences) × (Total documents/Documents containing the word) (4)

3.5. Evaluation

Evaluation is used to assessment the consideration space and results from various models or algorithms. For the recommendation of movie-titles, so when it comes to a classification problem, can be counted on an AUC - ROC Curve. Because of needed to scan or imagine the performance of the proposed system, It is denoted by the AUC (Area Under The Curve) ROC (Receiver Operating Characteristics) curve. It is one of the greatest significant estimation metrics for testing any arrangement model's performance. It is as well written as AUROC (Area Under the Receiver Operating Characteristics).The range AUC is between 0 and 1, An brilliant model has AUC proximate to the 1 and that implies it has a moral degree of distinguishability. The unwell model has an AUC near 0 which denoted it has the poorest measure of separability. Three broadly utilized performance metrics were applied to assess the proposed system's performance: TPR (True Positive Rate)/Recall/Sensitivity, Specificity and FPR(False Positive Rate)/Precision. To calculate the metrics specified by Equations 5–7, three distinct performance factors were selected: True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN).

TPR (True Positive Rate)/Recall/Sensitivity:

$$TPR / Recall / Sensitivity = \frac{TP}{TP + FN} \quad (5)$$

Specificity:

$$Specificity = \frac{TN}{TN + FP} \quad (6)$$

FPR:

$$FPR = \frac{FP}{TN + FP} \quad (7)$$

4. RESULTS AND DISCUSSION

Although the Semantic-Web Recommendation System (SWRS) built by other developers have used any technique filtering techniques, they had encountered weaknesses, which were slight disturbing. In our paper, we had implemented (SWRS) by content-based algorithm in two attributes from the MovieLens dataset utilizing Cosine-Similarity and Term-Frequency Inverse Document-Frequency (TF-IDF), after that the algorithm is tested on the windows 10 64-bit and Linux 18.2 64-bit operating system with the different number of records (movie-title), then these results are shown in Table 1 that display the results achieved on windows 10 64-bit operating system in different number of records in our dataset to produce process time, execution time and accuracy form read dataset to create RDF file, furthermore Table 2 that display information as Table 1 but on the real (not virtual) Linux 18.2 64-bit operating system. These marks pointed to that the building of (SWRS) on Linux operating system is better than on windows operating system, moreover the Fig. 2 on windows 10 and Fig. 3 on Linux operating system demonstrates to verify the results that found by Area Under Curve (AUC) to accuracy of creating SWRS. As a result of all the evaluation found out the two parameters are better in Quality of service (QoS), Quality of information (QoI), in spite of the results (accuracy and speed) can be affected by the features of the computer such as (CPU, RAM, Data Bus, Graphic Card) for this situations, so these issues should be handled before any processing to provide the predicted results.

TABLE 1: Results of the SWRS on windows-V10 operating system

Windows 10 64-bit Operating System			
No. Records in Dataset	Process Time (Second)	Execution Time (Second)	Accuracy (AUC)
1000	1.00315	1.01364	88.75%
2000	1.07825	1.15712	88.25%
3000	1.40268	1.52974	87.15%

TABLE 2: Results for the SWRS on linux-V22 operating system

Linux V 18.2 64-bit Operating System			
No. Records	Process Time (Second)	Execution Time (Second)	Accuracy (AUC)
1000	0.7104	0.6034	92.10%
2000	0.7445	0.6499	91.75%
3000	0.8268	0.6726	90.35%

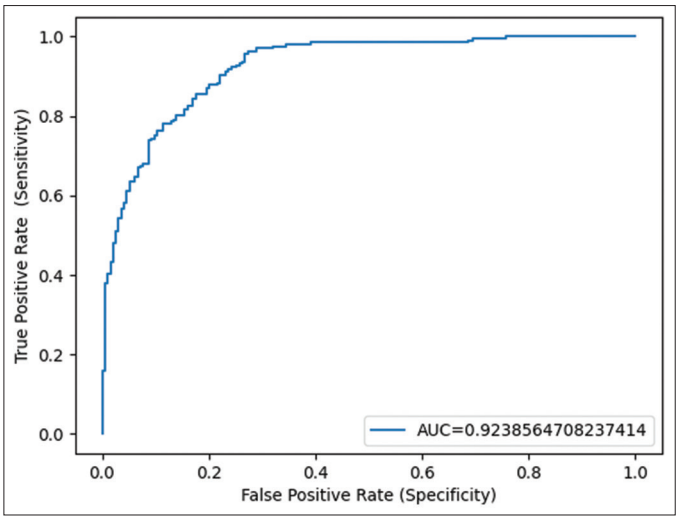

Fig. 2. Display AUC on windows for SWRS.

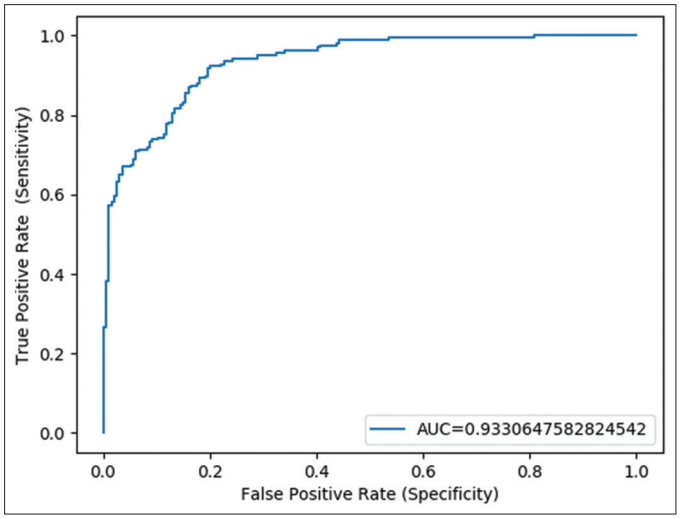

Fig. 3. Display AUC on Linux for SWRS.

5. CONCLUSIONS

Quick recommendations of movie-title, provides the greatest fortuitous to finding the correct titles (movies) to the users, semantic web-based content-based algorithm recommender system able to use in automatically and successfully analyze required data to identify the movie-titles. The main objective of our research is to use TF-IDF and cosine Similarity model to perform recommendations then creating RFD file as semantic web using input data from the amount of output of the data that recommended by the proposed method, after that Simple Protocol and RDF Query Language (SPARQL) used and it is the query language for the Semantic-Web that performs demand information from the databases or any data source that can be plotted to RDF. The proposed system offered a higher average recommendation accuracy approximately (88.5%) for windows operating system and (91.25%) for Linux operating system investigational results exposed that the proposed method is more effective than the previous works.

REFERENCES

[1] P. Hitzler, 2021. "A review of the semantic web field". *Communications of the ACM*, vol. 64, pp.76-83, 2021.

[2] I. Portugal, P. Alencar and D. Cowan. "The use of machine learning algorithms in recommender systems: A systematic review". *Expert Systems with Applications*, vol. 97, pp. 205-227, 2018.

[3] J. E. Gayo, E. Prud'hommeaux, I. Boneva and D. Kontokostas. "Validating RDF data". *Synthesis Lectures on Semantic Web*: *Theory and Technology*, vol. 7, pp. 1-328, 2017.

[4] H. Asghar, Z. Anwar and K. Latif. "A deliberately insecure RDF-based semantic web application framework for teaching SPARQL/ SPARUL injection attacks and defense mechanisms. *Computers and Security*, vol. 58, pp. 63-82, 2015.

[5] B. R. Cami, H. Hassanpour and H. A. Mashayekhi. "A content-based Movie Recommender System Based on Temporal User Preferences". *In: 3rd Iranian Conference on Intelligent Systems and Signal Processing* (*ICSPIS*). pp. 121-125, 2017.

[6] G. M. Zebari, K. Faraj and S. Zeebaree. "Hand writing code-php or wire shark ready application over tier architecture with windows servers operating systems or linux server operating systems". *International Journal of Computer Sciences and Engineering*, vol. 4, pp. 142-149, 2016.

[7] K. Faraj. "*Design of an E-commerce System Based on Intelligent Techniques*". Sulaimani University, Sulaimani, KRG, Iraq, 2010.

[8] S. P. Rana, M. Dey, J. Prieto and S. Dudley. "Content-based Health Recommender Systems". In: *Recommender System with Machine Learning and Artificial Intelligence: Practical Tools and Applications in Medical, Agricultural and Other Industries*. John Wiley and Sons, Hoboken, pp. 215-236, 2020.

[9] D. Wang, Y. Liang, D. Xu, X. Feng and R. Guan. "A content-based recommender System for computer science publications". *Knowledge-Based Systems*, vol. 157, pp. 1-9, 2018.

[10] I. T. Afolabi, O. S. Makinde and O. O. Oladipupo."Semantic web mining for content-based online shopping recommender systems". *International Journal of Intelligent Information Technologies*, vol. 15, pp. 41-56, 2019.

[11] C. L. Bocanegra, J. L. Ramos, A. Civitet and L. F. Luqure. "HealthRecSys: A semantic content-based recommender system to complement health videos". *BMC Medical Informatics and Decision Making*, vol. 17, pp. 1-10, 2017.

[12] N. A. Albatayneh, K. I. Ghauth and F. F. Chua. "Utilizing learners' negative ratings in semantic content-based recommender system for e-learning forum". *Journal of Educational Technology and Society*, vol. 21, pp. 112-125, 2018.

[13] Available from: https://www.kaggle.com/datasets/rounakbanik/the-movies-dataset [Last accessed on 2022 Feb 05].

[14] S. Sardjono, R. Y. Alamsyah, M. Marwondo and E. Setiana. "Data cleansing strategies on data sets become data science". *International Journal of Quantitative Research and Modeling*, vol. 1, pp. 145-156, 2020.

[15] R. H. Singh, S. Maurya, T. Tripathi, T. Narula and G. Srivastav. "Movie recommendation system using cosine similarity and KNN". *International Journal of Engineering and Advanced Technology*, vol. 9, pp. 556-559, 2020.