International Conference on Recent Trends in Data Science and its Applications
DOI: rp-9788770040723.005

International Conference on Recent Trends in Data Science and its Applications
DOI: rp-9788770040723.005

Using Harmonic Progression Convergence Model and Hyperlink Induced Topic Search Prompted For Page Ranking

John Vaseekaran
Research Scholar,
Sathyabama Institute of Science and Technology,
Chennai, India

N Srinivasan
Professor, Department of Computer Science and Engineering,
Rajalakshmi Engineering College,
Chennai, India

***Abstract*—Using the Web's inherent link structure, the Hyperlink-Induced Topic Search (HITS) algorithm identified and ranked sites based on their relevance to a search query. However, it solely considered the hyperlink structure and overlooked the contents of web sites and the reality that various hyperlinks on the Web may have varying degrees of value. In this research, we offer a unique page ranking method that integrates the hyperlink using the triadic closure theory, considering both the Vector Space Model (VSM) as well as the TrustRank algorithm, in an effort to counteract the aforementioned subject drifts. The approach initially determined the degree of relevance between two arbitrarily selected websites by comparing their subject matter and the number of references they shared. Then, a new eigenvector was built on top of that model to iteratively determine each page's authority and hub value. Following this, we used the trust-score technique to determine how trustworthy each page in the first group actually was. With our proposed Web page ranking algorithm against three traditional HITS-based methods, including the Web Page Topic Resemblance, Conventional Benchmark Degree, and Trust-degree algorithm. Based on the experimental findings, it is clear that our suggested method is superior than the other four traditional enhanced techniques and the HITS methodology.**

Keywords* — *HITS, trust degree, vector space model, hub value, web mining.

I. Introduction

The usage of a web search is ubiquitous in the pursuit of knowledge on the Internet. Search engines take a user's keyword input, then utilise ranking algorithms to order results and send them to the user.Three common approaches are used to categorise the relevant information for search engine page rankings.Two well-known algorithms are used to analyse the structure of hyperlinks [1].

They leverage the interconnected network of links between pages to its maximum potential, abstracting the relationships into a knowledge graph and iteratively determining a rating for each page.In order to determine how these search engine results pages are ranked, page rank as well as HITS are extensively used metrics.Using a combination of co-link analysis, social network analysis, and a semantic clustering algorithm, Yang presented a Web site ranking model for identifying subject communities among academic websites (SNA) [2].

Page Rank as well as HITS algorithms have flaws that are becoming more obvious as they are used more widely in search engines.To address these issues, researchers have developed several enhancements for HITS.Using a vector space algorithm to determine the resemblance between the web links and the user requests, Jones's probability ranking and an ontological selection ranking model determined the cardinality for internetpages of search engine results [3].An all-encompassing probabilistic framework for WPSS (web page scoring systems) has been created.

To rank websites for Web crawlers, we used the cell-like transmembrane computing optimization technique (CMCFC), which considered not only the main text of each page but also the content of any anchor links, page titles, and paragraphs immediately around them.Combining the hyperlinks and the word distribution across online pages, using formal concept evaluation to build the concept context graph, and then rating the web sites based on this, is what the semantic ranking does.

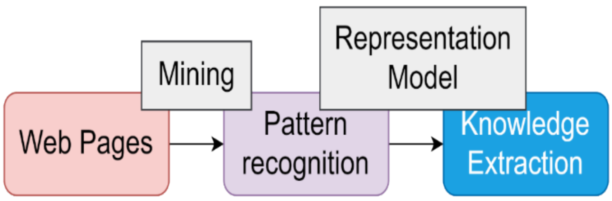

Fig. 1. Overview of web mining process

By examining the semantic relationship between Web sites, a relevance context graph was built from link (hyperlink) context graphs, allowing URLs with the same subject to point to each other.The rankings values for the specified webpages may be predicted using the relevance context graph.In order to direct Web crawlers, the idea context graph coupled the interlinking of web sites with the addition to online pages to create a concept ranking of web pages.HITS, a traditional scoring method, is widely used and successful websites between them. Overview of web mining process is shown in Figure 1.

In this case, the content material on the hub sites is relevant to the authority website that is directed to by those hub pages.The authoritative websites are linked to using hypertext from the hub page.In the HITS method, the

DOI: rp-9788770040723.005

values of the hub and authority pages are positively correlated; that is, a high-quality hub page will link to many high-quality authority sites, and a high-quality authority page will be linked to by many high-quality hub pages [4].

The HITS algorithm, on the other hand, simply uses the internal and external link counts of the correspondingly ranked web pages to determine rankings.

As a result, the symbiotic nature of host- and topic-drift connections will become problematic.Some studies and methods have been developed to address the issue.By giving website content from the same server significantly less weight, they suggested the BHITS(Bharat's enhanced HITS) algorithm.The issue of mutual reinforcement was nearly fixed.

To improve their capacity to anticipate the relevance of web sites, Zong et al. devised the P-HITS (Probability-based HITS) method, which uses probability to choose URLs from a list of URLs and incorporates the meta data of hyperlinks.They proposed the SALSA algorithm, which combines the query-related features of the HITS algorithm with the random walks employed by PageRank.The XHITS model was suggested on the basis of the HITS-hub and authority degrees of web sites, both of which allow for the concealment of semantic knowledge within hyperlinks [5].

The model introduced a unique machine learning technique for calculating web page rank scores, extending previous work on hubs and authority to include additional feature ideas.These techniques, however, rely only on the structure of the links themselves.Some integrated methods based on hyperlink and content analysis were developed to address topic drifts.Results show that including semantic text into the HITS algorithm improves its performance.In contrast, there are additional algorithms that make use of data gleaned from things like user comments or Web server logs.

In this paper, we present a unique ranking algorithm on the basis of HITS by combining the harmonic progression convergence theory with the TrustRank algorithm, which considers both the page content and the links between pages (R-HITS).The testing findings confirmed the efficacy and superiority of our suggested algorithm, demonstrating a decrease in topic drifts and an improvement in the quality of retrieval [6].

II. Literature Survey

Web page rankings are determined by HITS and HITS-based algorithms. Three issues that these algorithms have yet to address are the filtering of irrelevant hyperlinks, the prevention of subject drift, and the existence of mutually reinforcing ties across Web pages [7]. Here, we introduce the original HITS method and various updated versions of the algorithm for ranking Web sites, and then we critically examine their performance. HITS is an algorithm that just analyses links; it does not consider the text of web pages and cannot determine whether connections are more important than others.

For this reason, once Kleinberg published the HITS algorithm, several academics investigated it and offered many enhanced algorithms for addressing the various application criteria. HITS calculate a page's hub value as the total of the authority values of all pages that connect to it. A page's authority score is calculated by adding the hub scores of all the pages that link to it. For this reason, a hub page that links to numerous low-authority pages would have a larger hub weight than one that links to fewer high-authority pages [8].

Borodin et al. presented the Hub-Averaging method to address this shortcoming. At first, Lempel and Moran attempt to merge the best features of Page Rank with the HITS algorithm, creating a system that takes use of HITS's strengths where they pertain to the user query's topic and also incorporates Page Rank's random walk model. In reality, several experimental data demonstratethat SALSA provides superior search results to Page Rank and HITS [9].

Today, it ranks high among the most effective algorithms for link analysis. The Hub-Averaging algorithm is a combination of the HITS and SALSA procedures. This is an operation that can be performed using the algorithm. The purpose of ARC is to collect quantitative a set of credible web sites on any subject [10]. The core of the system is an algorithm that analyses text and links locally before reaching a "global agreement" on the finest resources available.

The HITS algorithm has been implemented by a wide variety of systems for the purpose of categorising websites. As a result, Teoma was able to apply a similar method to rank websites. It has been claimed that these methods can produce more relevant results in a search than either text-index search engines like AltaVista or directory-index search engines like Yahoo. Despite the fact that several trials have demonstrated that HITS yields useful search results over a broad spectrum of query topics, this approach may run into the following issues due to its utter disregard for textual contexts throughout its implementation [11]. It is often accepted as true that a website's pages and links are all the work of a single contributor or group.

A single "user's" opinion on quality is all that should be reflected in them. But HITS algorithm doesn't take that into account when determining a page's importance and authority. Unfortunately, it is not uncommon for many pages on the first host to all refer to the same page on the second site, p1.In each of the examples above, one "person's" perspective lends disproportionate weight to the respective website.Tools that create web pages sometimes include extraneous hyper connections like navigation links, commercial links, and other regularly created useless links.However, HITS do not include in how closely the page's content relates to the subject matter of the user's inquiry.Furthermore, it gives equal weight to all hyperlinks, despite the fact that the relevance of numerous hyperlinks may vary [12].

As a result, the HITS algorithm unfairly prioritises the less-relevant or irrelevant hyper link (links).The most popular authority (hub) websites have nothing to do with the subject matter of the user's search [13].Distraction from the

DOI: rp-9788770040723.005

original topic might occur for two main reasons.There are a couple of possible explanations for this: first, it's possible that the basic set B was obtained by include irrelevant or less relevant web page(s) in the root set R expansion process, and second, hub web sites contain several themes [14].

HITS algorithm might give more weight to the TCK page if it detects a cluster of unrelated web sites.In this research, we use HITS and the traditional HITS-based algorithms to propose that the degree of subject similarity between web sites, the degree to which they share references, and the degree to which they can be trusted are all crucial in determining where on the web each page should be ranked [15].Web page concept resemblance is a semantic feature used to prevent subject drift and filter irrelevant pages; a higher degree of shared reference can increase the value added by a common web page(hyperlink) while decreasing the impact of a mutual hyper link on a page's rank.A higher trust level may eliminate certain unhelpful connections [16].

I. PROPOSED SYSTEM

A targeted crawler's primary objective is the rapid indexing of as many relevant web pages as feasible. We need to give a more efficient means of locating pertinent websites. The adjacency matrix's building blocks are believed to be these connections that have some level of significance. In this research, we examine how three factors - web page subject similarity (brief as PTS), common reference degree (brief as CRD), and Trust Rank of Web Page - affect the relevance between any two given web pages.

Crawled websites need to be analysed so that keywords may be extracted from the complete text, title, anchor text, and surrounding paragraph language. There is now widespread usage of the Vector Space Model (VSM) in the conventional information retrieval sector, and manysearch engines utilise similarity calculations relying on this approach to rank websites. The following is a breakdown of how VSM is used to determine the degree of similarity between a web page's content and the subject matter of a user's search query as:

$$\vec{e_y} = (x1j, x2j, x3j \dots j), \vec{r_k} = (r1, r2k, r3k \dots rnk) \quad (1)$$

where $\vec{e_y}$ represents the web page and $\vec{r_k}$ defines the query initiated by the particular user. xnj and rnk such that $1 \leq l \leq n$ where these are the relative importance of the terms ul on web page ej and subject rk in the user's query, and n refers to total elements on script ej and topic rk. The Maximum - likelihood Document Frequency formula is used to assign relative importance to individual words in a corpus such that:

$$x_{lj} = ueg_{lj} \cdot je\ lj \ \frac{gl}{gmax} \cdot log \ \frac{O}{ol} \quad (2)$$

Where O is the cumulative quantity in documents over C regarding the manipulatorrequest discussion, ol refers to quantity of scripts such as ul, ugl. and $jeglj$. are the term intensity and reversetext recurrence of the descriptor ul in ej or a request response issue rk, respectively, and gl is the happenings of the term ul in web page ej or the query response context rk. User query subject phrase weights are also calculated.

Cosine likeness is used to calculate the degree of similarity between the topics covered on the web page ej and also the user's query subject rk such that:

$$S(\vec{e_y}, \vec{r_k}) = \cos(\vec{e_y}, \vec{r_k}) = \frac{\sum_{i=1}^{n} xlj.xlk}{\sqrt{\left(\sum_{i=1}^{n} x_{lj}^2\right)\left(\sum_{i=1}^{n} x_{lj}^2\right)}} \quad (3)$$

If page P has a link to page Q that reads $qP \rightarrow qQ$, then page Q is being promoted by page P. Furthermore, it indicates that the writer of qP endorses the ideas expressed in qQ. As a result, we may conclude that the contents of qP and qQ are highly comparable if there is a hyper connection $qP \rightarrow qQ$ between them. Let's suppose that the likeness among qP and q is tP and that the resemblanceamong qQ and manipulatorrequestissue q is qQ.

We may therefore interpret the sum of qP and qQ as the degree to which propositions P and Q are same in terms of q. In this work, we refer to this similarity as the subject similarity between websites. It's seen as a potential component that impacts the significance of a pair of seemingly unrelated pages. The significance of the hyperlink $qP \rightarrow qQ$ increases as the degree of topic similarity between the two web pages increases. Without such context, the connection is irrelevant.

So, thelikenessamong web pages qP and qQ built overrequest r is derived as the result of multiplying tj and tk, where tj and tk are the similarities of web pages tj and tk linked to the query q. The similarity between web pages is represented as follows:

$$S(qP, qQ) = \begin{cases} tj \cdot tk, & if qP \rightarrow qQ (qP \neq qQ) \\ 1, & qP = qQ \\ 0, & otherwise. \end{cases} \quad (4)$$

Here, we'll incorporate the concept of triadic closure into our study on social networks. Please begin by explaining and critiquing the concept of triadic closure. There is a higher chance of friendship between two people in a social network if they share a friend or two in common. This is the first and foremost rule of triadic closure.

When two nodes Y and Z share a friend X, the resulting edge between them forms a triangular shape in the social network, with edges coming from and connecting all three nodes X, Y, and Z. This triangle "closes" at its third edge, which connects points Y and Z. If we take two pictures of a social network, one after theother, we will often uncover many new edges connectingpeople who had a shared neighbour in the first image dueto triangle-closing.

DOI: rp-9788770040723.005

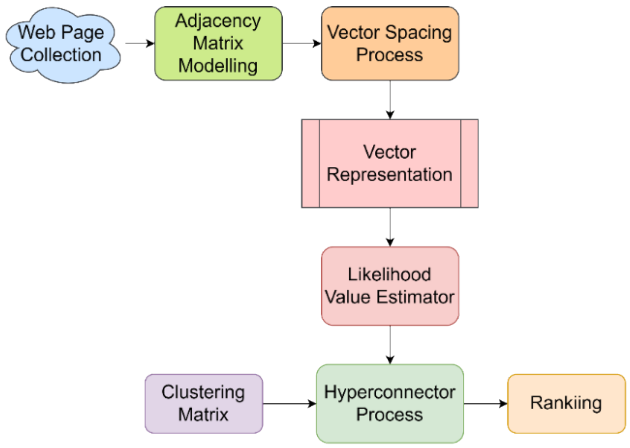

Fig. 2. Process flow of proposed model

The theory of harmonic progression is extended by two theorems. The initialleewaydeduction may be stated as follows: if a set of people has a large number of mutual friends, then it follows that the chances of those people becoming friends themselves increase. There are only two people who both person X and person Y know in common. All three of Yand Z's mutual pals are named A1,B1, and C1. The following inferences can be made with little effort.

The likelihood of A1 and B1 becoming friends in the imminent is developed that of A and B. As an illustration of the second extension theorem, we may say that the likelihood that two people will become friends increases if their common friends have deeper ties with one another. The web itself may be seen as a directed graph, as is common knowledge. Hyperlinks between websites are represented by edges in a directed graph, while nodes represent individual pages.

We already established that if page P has a hyper link to page Q, then P endorses Q. By including a link to another work, the author of work P implicitly endorses the work in question. If we think of the Internet as a social network and each webpage as a person, then the friendship between two websites is analogous to the friendship between two people. In particular, if two nodes P and Q are connected by an edge may both signify the friendship between the two people in the social network, and the hyper link between the two web pages in the web network. The process flow of proposed model is shown in Figure 2.

Because of this, the theorem proving the applicability of triadic closure theory in social networks may be applied to the World Wide Web. Here's a more in-depth explanation: Web sites P and Q seem to be more likely to expound on the same topic if the ordinary in-degree between them is bigger and the unusual in-degree between them is lower. The evidence suggests that websites P and Q may be more pertinent. Consequently, the degree of common reference is another aspect that impacts the significance of any two pages chosen at random.

In order to identify spam websites, the TrustRank algorithm described by Gyongyi et al. was implemented. A website's credibility on the Internet is determined by the algorithm. The basic premise of this technique is that more trustworthy pages will have their trust scores slightly lowered but remain extremely high if they are linked to from a page with a higher TrustRank. The TrustRank algorithm was at first created to help identify spammy websites. However, TrustRank was more commonly utilised in the search engine ranking algorithm. Websites' overall rankings are frequently impacted by this.

Values of TrustRank are often calculated at the level of the domain in the search engine ranking algorithm. The more the host's trustworthiness, the greater the aggregate ranking's strength. Like the PageRank algorithm, the TrustRank algorithm relies on a massive amount of information available on the World Wide Web. For this reason, Asano et al. suggested a trust-score method based on the TrustRank algorithm. A trust-score is assigned by the trust-score algorithm to every page in the HITS basic set Q.

For the vast majority of user query themes, Asano et al. conducted exploratory studies, discovering that more than half of the root set of web pages are credible and relevant to the particular algorithm also relies on this premise: This leads to numeroussides in the underlyingcause set v is a trustworthy; and if page v is connected from many dependable, then page v is a dependable authoritative web page. Denote these sets by their initials: $R$for the collection of roots, Was vertices, and $F$for the edges.

We present a ranking system that considers a web page's trustworthiness in addition to its topical similarity and degree of shared referencing. The suggested approach offers two improvements over the HITS algorithm. By merging the page topics that are comparable and the degree of common reference, a new clustering matrix is formed. The updated adjacency matrix is then used to calculate the core and authority values. The second is calculating the credibility of each website in the B. To determine a page's rating, we consider both its core or credibility value and its trust-degree. First, we create the new transformation matrix B by linearly combining the matrixes article theme resemblance and common evaluation level. Also, we can denote as:

$$B = \mathrm{P} + \gamma \text{where } 0 \le \mathrm{P} \le 1, 0 \le \mathrm{P} \le \gamma \text{ and } \mathrm{P} + \gamma = 1. \quad (5)$$

In this approach, the parameters and define how adaptable the resulting adjacency matrix $B$is, whereBmn = $\mathrm{P}mn$ + $\gamma mn$is a matrix element. Next, the new transitive matrix $B$is used to determine hub and authority values. Also, the entire technique of calculating is identical to that used by HITS. This is how the new iterative formulae are written:

$$b = B^T i \quad (6)$$

$$i = Bb \quad (7)$$

$$bj = \sum k \in (j)\ Bnmik = \sum k \in JK(j)(\mathrm{P}mn + \gamma mn\ )ik \quad (8)$$

$$ik = \sum k \in (j)\ Bnmbk = \sum k \in PM(j)(\mathrm{P}mn + \gamma mn\ )bk \quad (9)$$

In this approach, the parameters and define how adaptable the resulting adjacency matrix $B$is, whereBnm =

DOI: rp-9788770040723.005

$Pmn + \gamma mn$is a matrix element. Next, the new transitive matrix $B$is used to determine hub and authority values. Also, the entire technique of calculating is identical to that used by HITS. This is how the new iterative formulae are written.

II. Results and Discussion

Here, we examine the proposed system to three traditional advancing algorithms based on HITS and choose a large number of subjects with MAP, F-Measure, and DCG to evaluate the algorithm's performance. The experimental data shows that compared to three competing algorithms, our suggested approach for ranking web pages yields much better results.In our trials, we compare our proposed method to the HITS algorithm and three other well-known enhancement algorithms (BHITS, SALSA, the Hub-Averaging, and ARC).You may shorten their names to HITS, SAHITS, and ARCHITS.

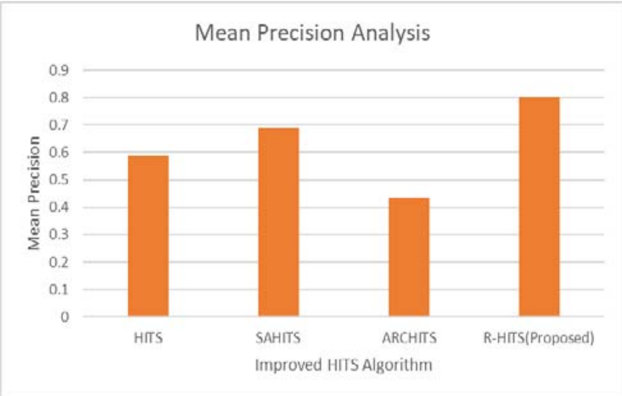

Fig. 3. Mean Precision Analysis

Virtualization, wall climbing, drinking, the Olympics, the railroad, the kayak, the yard, cheddar, Hepatitis, and the Kabuki theatre are only ten of the numerous categories from which we selected user query topics to showcase the experimental outcomes.Our investigations make use of root sets Bs that are composed of the top 200 results from a subject.For example, the ARC method's group B includes webpages that are link-distance three separate or less from at minimum, but in the HITS algorithm this was done just once.

We discovered that 20 iterations is a good number to utilise in tests since iteration quickly converges.Using indices such as MAP (Mean Average Precision), F-Measure, as well as DCG, the effectiveness of the aforementioned six algorithms may be assessed (Discounted Cumulated Gain).For each algorithm, in particular, the average value across all user query subjects represents its precision.

Accuracy is how many highly relevant results an algorithm returns in comparison to how many results are found in the top 10 for a given algorithm.The recall measures how many of the most relevant results returned by an algorithm are included in its top ten list.If the VSM-calculated similarity among the page's subject and the query response topic is more than the threshold value of 0.5, then the page is considered relevant.DCG's key notion is that a page's score should be deducted if it has a better grade than one that is higher in the ranking.

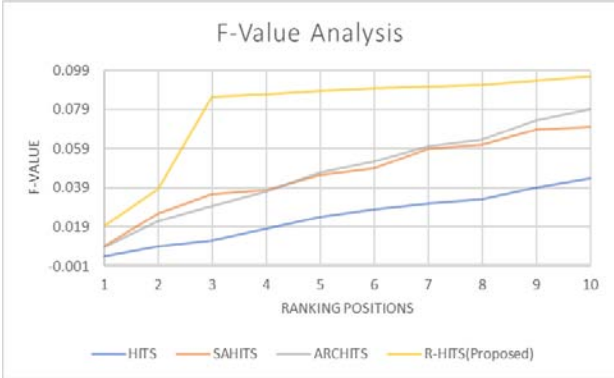

Fig. 4. F-Value analysis

First, the proper relevance degree is assigned to the website that falls inside the Top10 websites discovered by each algorithm.In particular, the page receives a score of 0 if the likeness between the homepage and the user querytopic falls within the range [0.0, 0.25]; if the range [0.25, 0.5] is crossed, the page receives a score of 1; if the range [0.5, 0.75] is crossed, the page receives a score of 2; and if the range [0.75, 1.0] is crossed, the page receives a score of 3.In this case, n is set to 10, which indicates that we are analysing ten different web sites for this result rank.

We calculate themes for each method in the tests.It's easy to spot similarities between the three figures:Compared to the other five methods, the values of the proposed method are clearly greater.In particular, the HITS method has a MAP of 0.6716, the BHITS algorithm has a MAP of 0.5107, SALSA has a MAP of 0.8258, Hub-Averaging has a MAP of 0.7043, and the suggested approach has a MAP of 0.9609.

Based on these results, it can be concluded that the suggested method has a higher MAP value than the HITS algorithm (by 43.0%), the BHITS algorithm (by 88.2%), the SALSA algorithm (by 30.9%), the Hub-Averaging algorithm (by 16.3%), and the ARC algorithm (by 36.4%).Overall F as well as DCG scores of the proposed method show clear superiority over the RC algorithms over a range of search engine rankings.When x = 0.5, they display how the proposed algorithm's values vary with varying.Each of the three characters shares some traits with the others.

Nonetheless, the MAP (F and DCG) values develop at varying rates, with the highest rates occurring when = 0.On the other hand, when = 0, the values are practically identical.Furthermore, the smallestvalues occur at = 0, while the largest occur at = 1.Examples like these shows how subject similarity across web pages may significantly improve a page's position in search results.When = 0.5, they illustrate how the suggested method performs in terms of MAP, F, and DCG.

Fig. 5. Performance analysis for different ϑ values

The values of MAP (F and DCG) in all three figures tend to decrease as x increases, which indicates a common denominator.When going from x = 1 to x = 1, however, the rates at which the values decrease are quite large.Values are also maximal at x = 0 and minimal at x = 1.These examples show that t(vi), the trust degree, had a minor impact on the ranking of websites.Indicating that the suggested algorithm is superior in its capacity to place high-quality websites at the top of search engine results.In addition, they validate our contention that the subjectsimilarity of websites is a significant factor in the suggested website rank.

III. Conclusion

Using harmonic progression closure concept, vector space models, and the TrustRank algorithm, we provide a new page ranking system in this study. Subject similarity, degree of sharedorientation, and are all defined. To iteratively determine authority and hub values, a novelmodel is built according to the subject similarity of web pages and the sharedorientationgrade. In addition, by integrating the trust-score method, we establish the degree of confidence we have in each website in the baseline collection. Last but not least, the score has been derived by combining thehub value along with degree of trust. Experimental findings reveal that our suggested page ranking method locates higher-quality Top10 websites than the HITS, BHITS, and ARC algorithms. In the past, there may have been some theme drifts, but HITS can help fix that. In conclusion, we have created a HITS-based algorithm that successfully returns the top 10 web sites that best answer user query subjects.

References

[1] T.Sekiya, Y. Matsuda, and K. Yamaguchi,“Improvements of a Hybrid Syllabus Search Tool by Syllabus- related Heuristics,” In 2022 IEEE Frontiers in Education Conference (FIE), IEEE, pp. 1-8, October 2022.

[2] M.R. Baker, and M.A.Akcayol,“A novel web ranking algorithm based on pages multi-attribute,” International Journal of Information Technology, vol. 14, no. 2, pp.739-749, 2022.

[3] S.Ang, H. Wu, M. Chen, and F. Yang,“Social network analysis for cross-evaluation in data envelopment analysis,” Expert Systems, vol. 39, no. 9, p.e13063, 2022.

[4] L. Yan, Y. Wei, Z.Gui, and Y. Chen,“Research on PageRank and hyperlink-induced topic search in web structure mining,” In 2011 International Conference on Internet Technology and Applications, IEEE, pp. 1-4, August, 2011.

[5] R.Prajapati,“A survey paper on hyperlink-induced topic search (HITS) algorithms for web mining,” International Journal of Engineering, vol. 1, no. 2, 2012.

[6] W.Y. Zhang, S.S. Guo, and S. Zhang,“Combining hyperlink-induced topic search and Bayesian approach for personalised manufacturing service recommendation,” International Journal of Computer Integrated Manufacturing, vol. 30, no. 11, pp.1152- 1163, 2017.

[7] Gomathy, V., Janarthanan, K., Al-Turjman, F., Sitharthan, R., Rajesh, M., Vengatesan, K., &Reshma, T. P. (2021). Investigating the spread of coronavirus disease via edge-AI and air pollution correlation. ACM Transactions on Internet Technology, 21(4), 1-10.

[8] B. Zhao, S. Hu, Z. Zhang, C. Xu, and X. Tang,“Hyperlink induced topic search-based method to predict essential proteins,” International Journal of Data Mining and Bioinformatics, vol. 22, no. 3, pp.250-264, 2019.

[9] P.R. Kumar, and A.K. Singh,“Web structure mining: Exploring hyperlinks and algorithms for information retrieval,” American Journal of Applied Sciences, vol. 7, no. 6, p.840, 2010.

[10] T.T. Hoa, and N.N. Ha,“Edge-weighting hyperlink- induced topic search (E-HITS) algorithm,” In Proceedings of the 2017 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, pp. 925-930, July, 2017.

[11] A. Jain, R. Sharma, G. Dixit, and V.Tomar,“Page ranking algorithms in web mining, limitations of existing methods and a new method for indexing web pages,” In 2013 International Conference on Communication Systems and Network Technologies, IEEE, pp. 640-645, April, 2013.

[12] Rajesh, M., &Sitharthan, R. (2022). Introduction to the special section on cyber-physical system for autonomous process control in industry 5.0. Computers and Electrical Engineering, 104, 108481.

[13] S. Chawla,“Intelligent web search system for personalised web search based on recommendation of web page communities,” International Journal of Intelligent Systems Design and Computing, vol. 3,no. 1, pp.12-30, 2019.

[14] B. Liu, and Y. Zhu,“ProtDec-LTR3. 0: protein remote homology detection by incorporating profile-based features into learning to rank,”IEEE Access, vol. 7, pp.102499-102507, 2019.

[15] K.Kanathey, R.S. Thakur, and S.Jaloree,“Ranking of web pages using aggregation of page rank and hits algorithm,” International Journal of Advanced Studies in Computers, Science and Engineering, vol. 7, no. 2, pp.17-22, 2018.

[16] L. Yan, W. Du, Y. Wei, and H. Chen,“A Novel Heuristic Search Algorithm Based on Hyperlink and Relevance Strategy for Web Search,” In Advances in Electronic Commerce, Web Application and Communication, Springer, Berlin, Heidelberg, pp. 97-102, 2012.