Personalized Search Engine Using Binary Tree Traversal (BTT) - A Survey

Jose Triny K, Dharani RK, Pavithra GR, Priyadharshini R, Revathi G

Abstract: Web pages have an increasing number of been used because the patron interface of many software programstructures. The simplicity of interplay with internet pages is an idealbenefit of the usage of them. However, the character interface also can get extracomlicatedwhilegreatercomplexnet pages are used to construct it. Understanding the complexity of net pages as perceived subjectively with the resource of clients is thereforecrucial to betterlayout this sort of consumer interface. Searching is one of them not unusual placeassignmentachievedon the Internet. Search engines are the essentialtool of the net, from wherenyou willcollectassociatedstatistics and searched in keeping with the favoredkey-word given by the character. The recordson theinternet is developing dramatically. The consumer has to spend extra time with inside theinternet case youneed to find outthe correctfactsthey may befascinated in. Existing net engines like Google do now no longerundergo thoughtsuniqueneeds of character and serve eachpatron similarly. For this ambiguous query, some offiles on wonderfulsubjects are decreaselower backby engines like Google. Hence it will become difficult for the consumer to get the requiredcontent materialfabric. Moreover it additionally takes extra time in searching a pertinent content materialfabric. In this paper, we are able to survey the numerous algorithms for decreasing complexity in internetweb page navigations.

Keywords: Web pages, Search Engine, Web page complexity, Page navigation, Ambiguous question.

1. Introduction

A seek engine is a software programstructurethis is programmed to behaviornet searches (Internet searches) and, as a end result, to go looking the World Wide Web in a systematicway for precisefacts set out in a textual netseekquestion. Theare looking foroutcomes are regularlyproven in a line of outcomes, that'scalled a are looking for engine outcomes tab (SERPs) The recordssought to consist ofa combination of hyperlinks to netweb sites, photos, videos, data graphics, posts, studies papers, and differentstyles of documents. Some serpsadditionally scour libraries and open directories for facts. Unlike net directories, that areaspeciallycontrolledby human editors, engines like Google regularlyhold actual-time recordsthroughmanner of on foot an set of rules on a web crawler. The deep net is a time period used to explaiin netcontent material that cannot constantly be diagnosed the usage of a seek engine.

With the giant enlargement of the Internet, maximum contemporary serps, inclusive of Google, Yahoo, and MSN, offercustomers with an unbroken, prepared linear listing of web sites, every with partial content material ranked through relevance to the questquestion. The query-listing paradigm is utilized by the giant majority of serps. Customers sat thenet are compelled to sift viav protractedlisting and study the titles with the intention to locate the facts they want. It is believed that serps will now no longer back the maximum not unusual placedocuments that correspond to a question. It is likewise anticipated to offer correct records for the whole united states of america. Clustering the search outcomes into distinctive businesses has been defined as a precious approach to the hassles stated above.

Customers clearly want to pick the right cluster and examine for the favored report if the results were delivered on this way. Considering the constraint of time enforced on the systems used for seek and personalization being a way regarding extra time, the patron profiles get better only with extra time and utilization. Personalization structures which give new rating to the filesobtained from retrieval commonly employ consumer profile on the customer facet. Also, in location of acquiring all outcomes from the source, they re-rank most effective fantastical pinnacle ranked documents. Due to this overtime required, the manner turns into significantly sluggish but a immoderatediploma of personalization can be acquired. In question alternate approach, only query instance may be altered with inside the profile of the patron. As a consequence, it is much less possibly to effect end result lists. Web crawling from web website online to web website online is how serps like Google get their outcomes. The "spider" appears for filename robots which might be normal. It obtained a textual content material addressed to it. The machines, this is. The directives with inside the txt recordsonform seek spiders which pages to move slowly. After checking for robots.Txt and both finding it or no longer, the spider sends surefacts lower back to be indexed relying on many elements, inclusive of the titles, JavaScript, headings, web
Cascading Style Sheets (CSS) or its metadata in HTML meta tags. After a fantastic extensive kind of pages crawled, amount of facts indexed, or time spent on the internet site, the spider stops crawling and moves on. O(n) net crawler can also additionally more over honestly circulate slowly the complete reachable internet. Due to endless net sites, spider traps, junk mail, and different exigencies of the actual net, crawlers as a substitute exercise a move slowly coverage to determine while the crawling of a domain need to be deemed enough. Some web sites are crawled exhaustively, whilst others are crawled most effective partially.

2. Literature Review

[1] A. Paranjape, et.al,….. navigate throughlinks, however preserving a first-rate connection shape is hard. Human editors can locate it hard to understand pairs of pages that should be associated, specially if the internet site is huge and modifications often. Furthermore, given a fixed of beneficial link candidates, the project of integrating them into the internet sitemap may be costly, because it commonly calls for human beings to make modifications to web sites. Expanding facts-pushed strategies for automating hyperlink placement is a plausible preference in mild of those challenges. We gift a technique for locating beneficial hyperlink in a domain automatically. We use those indicators to expect the ability software of connections that do not exist but. We outline the hassle of connection placement beneath economic constraints and advise a green set of rules for fixing it primarily based totally on our model. We display the efficacy of our manner through checking out it on Wikipedia, a giant database for which we've admission to every server logs (used for coming across beneficial new links) and the whole revision history (which gives a floor fact of all modifications).

[2] H. Kao, J. Ho, et.al,…. Investigate the hassle of mining intrapage informative shape in information Web pages with the intention to pick out and put off redundant facts. It's really well worth noting that the intrapage edifying form stays subsection unique Trap folio too is made from sequence good-grained then edifying slabs. Maximum effective anchors linking to ne are contained within the intrapage informative systems of pages in a information website online. WISDOM is an intrapage edifying shape pulling out approach that applies Information Theory to DOM tree knowledge so you can create the form. WISDOM divides a DOM tree into numerous smaller sub timber and makes use of a group of pinnacle-down descriptive block searching regulations to pick a fixed of candidate informatics.

[3] H. Kao, S. Lin, J. Ho, et.al,…. studied the problem of appealing out the revealing construction of an facts website online entails masses hyperlinked files. Outline edifying shape information website online as per difficult and rapid catalogue folios (or else called TOC, that is, slab innards, folios) then conventional artefact folios related through manner of those TOC folios. Grounded taking place HITS set of regulations, We endorse entropy-grounded evaluation (LAMIS) apparatus used for studying entropy of broadcaster manuscripts too hyperlinks in the direction of eradicate severance hyperlink shape in order complicated form website online canister stay refined. Nevertheless, on the way to upsurge charge then user-friendlyness folios, utmost gratified material netweb put commonly generally have a habit on the way to place up folios thru meddling laid off statistics, along with steering panes, commercials, reproduction proclamations, etc.

[4] P. Loyola, G. Martinez, et.al,…. targeted on the usage of Web usage logs. Only in recent times has using statistics from clients’ natural responses emerged as an opportunity to beautify the assessment. In thoseart work, a model is proposed to understand Website Key Objects that now no longer most effective takes under consideration seen gaze hobby, collectively with fixation time, but additionally the impact of scholar dilation. Our
foremost hypothesis is that there can be a strongcourting in phrases of the scholar dynamics and the Web patronopportunities on a web page.

[5] M. Butkiewicz, H. Madhyastha, et.al,…. diagnosed fixed of difficult and rapid metrics to mirror the problem of web sites at each the content material and carrier levels (e.g., axetensivekind of servers/origins). We located that the distributions of these metrics are absolutelyimpartial of a website online/popularityrating. Some groups, inclusive of News, are extracomplex than others. While the developing intricacy Trap folios then hers bearing taking place normaloverall performance has been properlysaid anecdotally, no systematic studies has been carried outat the subject. We proposed a number onetemptton this paper to symbolizeweb page complexity and degree its effects. We graded the complexity of Web pages primarily based totally on the quantity of content material they include and the offerings they offer. The recognition of a internet siteat thenet is a poor indicator of its complexity, while its magnificence is significant. News web sites, for example, load some distanceexramerchandise from many extra servers and reassets than different groups.

[6] P. Yin and Y. Guo, et.Al,…. studied of character perceptions approximately netweb site discsloses that the maximumcrucial layout skills for distinctivenetweb website onlinedomains consist of navigations, timeliness, clarity, visualization, accuracy, and protection. The clean-toc-navigate characteristic is ranked a number of the pinnacle3 for all domains. Web customers appearance beforehand to extracomfybrowsingtales which require the WWW surroundings to be everypowerful and green. Effective browsingmethod that the clients can with outtroubles are looking for the maximumexcitingnetweb website onlinethroughmanner of specifying relevant keywords, whilst greenbrowsing indicates the customers can obtain the purpose internet site in a netweb website online with clearly few clicks. Both necessities can be facilitated via the usage of the net mining techniques with inside the formatphase. In this have a take a observe we recommend a contemporary approach for the netweb website onlineshape optimization (WSO) problem primarily based totally on a whole survey of gift works and exercise concerns.

[7] M. Chen and Y. Ryu, et.al,…. superior a mathematical programming (MP) model of a internet site that aids consumer navigation with minimummodifications to its contemporary format Our version is designed for informational web sites with static content material that has remained fairlysolid over time. Universities, visitor destinations, hospitals, federal agencies, and sports activities departments are all examples of agencies which have informational web sites. However, our model might stay apt meant for trap putstmost effective routine go-aheadfolios or includerisky content material. Our version, on the opposite hand, might not stay apt meant for trap spotsmost effective use dynamic pages or have risky content material. Although numerous techniques for relinking webpages to beautify navigability via the usage of consumer navigation facts were proposed, the wholly modernized newfangled form may stay rather erratic, then value customers being disoriented because of the modifications has but to be determined. This broadside lectures the manner near beautify an internet spot with out introducing giantmodifications. Unambiguously, recommend accurate software design archetypalnear enhance character steering proceeding online even as curtailing modifications near the aforementioned contemporary format. Failouts as of significant assessments finished happening overtly to be had tangible facts customary imply archetypal notmost effective substantially rallies consumer triangulation thru just a scarce adjustments, however additionally cannister stayefficiently unraveled. We've additionally placed archetypal via its paces taking place massivevunrealstatisticsdevicesso the way to peer how properly it scales.

Furthermore, we pick outsizesstandards and custom on the way to degreepermanence of the superiormetweb online even as usage of the actualfacts collection. The character navigation on the superior forms is likewise appreciably better, in line with the assessment outcomes.

[8] C. Kim and K. Shim, et.al,… finished stencil exposure then abstraction performances partake acquired massesshobby presently near enhance overall recital internet programs, along with statistics integration, serps, class of internet documents, and so on. Thus, template detection strategies have obtained edan entire lot of hobby in recentABOBE to enhance the overall performance of serps like Google and yahoo, clustering, and class of netfiles. Inside this document, we present original algorithms intended for extract template as of a massive type of internet papers to be generate as of varied template. We band netfile constructed scheduled parallel causal stencil systems with inside pamphlets in order stencil meant for every hand stays haul out in chorus. We mature a unique golly diploma thru the aforementioned debauched guessimate meant for huddling then afford complete evaluation set of rules. Our trial effects thru actual-natural liferecordssunitssanction use then heftiness set of rules in comparison to the United States of America of the artwork for template detection algorithms.

[9] Y. Yang, Y. Cao, et.Al,… introduces a hybrid version HCRF then prolonged Semi-Markov (Semi-CRF) on the way to take benefit of web folio shapes outcomes cutting-edge abletextual content breakdown then marking. The choice of the HCRF model can direct the choice of the Semi-CRF version on this stop-down integration version. The disadvantage of the pinnacle-down integration strategy, but, is that the Semi-CRF version's selection couldn't be utilized by the HCRF model to direct its selection-making. This paper proposed WebNLP, a singularmachine that
permits for iterative bidirectional integration of netweb pageformknowledge and textual contentknowledge. We have finished the proposed framework to close byemployer entity extraction and Chinese character and employercall extraction. Experiments display that the WebNLP framework executedappreciablybetteroverall performance than contemporarytechniques.

[10] J. Hou and Y. Zhang, et.al,… proposed algorithms for findingassociated pages primarily based totally on netweb page similarity. The essentialhomes are constructed into the brand new netweb pagedeliver on which the algorithms are constructed. The estimation and outline of netweb page similarity is absolutelydepending on the linkrecords of a number of the Web pages. The first set of regulations, Extended Cognition set of regulations, is a cogitation set of rules outspreads conventional co-quotation principles. The aforementioned stays innate then succinct. The subsequent solitary, baptized LLI set of regulations, revealsrelevante pages extraefficaciously and exactlythroughtroughmanner of the usage of rectilinear algebra philosophies, in particular curious fee putrefaction of milieu, toward show unfathomable dealings some of folios. This paper giveshyperlinkevaluation-grounded set of rules near bagainingermane folios intended for prearranged trap folio (URL). The foremost set of regulations arises as of stretched deliberation evaluation Trap folios. The aforementioned stays innate then cleanon the way to place into impact. The subsequent solitary revenues gain of in lines algebra philosophies to show profounder associations most of Trap folios then near end upaware aboutapplicable pages extrained then effectually. The investigational effectsdisplay likelihood then efficacy set of rules.

These set of rules is probably cast-off used for innumerable Trap packages, inclusive ofpleasing to the eye Trap seek. The mind besides strategies in thoseart exertionmay staybeneficial to different Trap-interrelated inquiries.

3. Comparative Analysis

| S.NO | TITLE                                                                 | TECHNIQUES                      | PROS                                      | CONS                                      |
|------|-----------------------------------------------------------------------|---------------------------------|-------------------------------------------|-------------------------------------------|
| 1    | Improving Website Hyperlink Structure Using Server Logs               | Greedy marginalbenefithyperlink placement set of rules | Refining the connectivity of the Web seek  | Limited in database                       |
| 2    | WISDOM                                                               | Trap Intraplate Enlightening Edifice Pulling out primarily built totallyat the DOM | Useful for indexing, extracting           | Outliers can be occurred                  |
| 3    | Mining Web Informative Structures and Contents Based on Entropy Analysis | Entropy-primarily based totallyevaluation | Mine beneficialsystems and contents from Web webweb sites | Time eatingmanner                         |
| 4    | Characterizing Web Page Complexity and Its Impact                     | Website's recognition rank set of rules | Transfer and renderal Web web page        | Need large set of consumer profiles       |
| 5    | Coalescing sense monitoring toolpupillary distention evaluation near pick out Website Vital Stuffs | Web item mapping approach        | Tough definition of the organization of Web Objects | Page load time is high                    |
| 6    | Optimization of multi-standardsinternet siteshapeprimarily based totally on more suitablyabuse and netutilization mining | Enhanced tabuseek (ETS) set of rules | Progressive seek features                 | Computationally inefficient               |
| 7    | Easing Operative Handler Steering via Website Edifice Upgrading       | Mathematical programming version | Significant enhancements to consumer navigation | Difficult to pickoutcustomers’ targets    |
| 8    | MANUSCRIPT: Reflex Stencil Mining beginning Mixt Trap Folios           | Template detection strategies   | Speed up the retrievalmanner              | Need to educate the big database          |
| 9    | Closing the Loop in Webpage Understanding                               | Markov Conditional Random Fields | Extract more than coincident features     | Manual methodmay be needed                |
| 10   | Well Verdict Germane Trap Folios as of Relation Statistics            | Successfully Verdict Trap Folios and of Link Info | Finds out applicable pages                | Static server may be needed               |
4. Proposed System
The current framework consists of K-Means clustering set of rules and Page rank set of rules to extract the net pages primarily based totally on click onviafacts.

4.1. K-Means set of rules:
The K method set of rules is easy to enforce, requiring aseasrecordsshape to holdsomefacts in eachera to be used in thenext new release. The idea makes k-mannerextragreen, particularly for dataset containing large extensive kind of clusters. Since, in each new release, the k-method set of rules computes the distances among facts component and all facilities which might be computationally very expensive particularly for large datasets. Therefore, we do can use from previous new release of okay-approach set of regulations. K-Means is one of the top ten clustering algorithms which may be broadly utilized inrealglobal programs. It is a totally clean unsupervised analyzing set of rules that discovers actionable knowledge through the usage of grouping comparable devices into various clusters. However, it needs the wide variety of clusters to be mentioned priori. We can calculate the distance for every facts to nearest cluster. At the subsequent new release, we compute the gap to the preceding nearest cluster. The factor remains in its cluster, if the brand new distance is much less than or identical to the preceding distance, and it is not required to compute its distances to the opposite cluster centers. The K-method set of regulations is the most customarily used partitioned clustering set of regulations due to the fact it could be with out troubles applied and is the most inexperienced one in terms of the execution time.

The primary set of rules pseudo code as follows:
Input: X = be the set of facts factors, Y = be the set of facts factors and V = be the set of facilities.
Step 1: Select ‘c’ cluster facilities arbitrarily.
Step 2: Compute the gap among every facts and cluster cores using the Euclidean Distance metric as follows
\[ Dist(X, Y) = \sqrt{\sum_{j=1}^{n} (X_{ij} - Y_{ij})^2} \]  
Eqn(1)
\(X, Y\) are the set of facts factors
Step 3: Pixel is assigned to the cluster middle whose distance from the cluster middle is minimal of all cluster facilities.
Step 4: New cluster middle is calculated the usage of
\[ V_i = \frac{1}{C_i} \sum_{j=1}^{C_i} x_j \]  
Eqn(2)
Where Vi denotes the cluster middle, Ci denotes the wide variety of pixels within the cluster
Step 5: The distance among each pixel and new acquired cluster centers is recalculated
Step 6: If no pixels have been reassigned then stop. Otherwise repeat steps from three to 5
The flowchart of the set of rules is proven in fig 3.1

**Fig2. Flow chart of K-Means clustering**
4.2. Page Rank Algorithm

PageRank (PR) is a fixed of regulations used by Google Search to rank websites in their rare looking for engine effects. One of the founder of Google, Larry Page modified the PageRank. It isn't always the most effective set of rules utilized by Google to reserve seek engine effects, but it is the primary set of regulations that modified into utilized by the organization, and it's miles the best-mentioned. The above centrality diplomats not implemented for the multigraphs. The PageRank set of regulations outputs a chance distribution used to symbolize the chance that someone randomly clicking on links will arrive at any uniquenetwork page. It is believed in several studies papers that the distribution is flippantly divided amongst all files in the collection on the begin of the computational way. The PageRank computations require numerous passes, acknowledged as “iterations”, via the gathering to adjust approximate PageRank values to extraintently reflect the theoretical right charge. The length of each query is proportional to the general length of the alternative faces which might be pointing to it. The pseudo code for the set of rules is:

- Assign every node an preliminary web page rank
- Repeat till convergence calculate the web page rank of every node (the usage of the equation within the preceding slide)

\[ \text{PR}(A) = (1-d) + d \sum \left( \frac{\text{PR}(T_i)}{C(T_i)} + \ldots + \frac{\text{PR}(T_n)}{C(T_n)} \right) \]

Subsequently wholly, summation slanted web page ranks wholly folios Ti stays increased thru curbing component d may be located customary among zero in addition 1. So, expand web page rank gain for a web page through any other web page linking to it's miles reduced.

![Graph Model](image)

Fig 3: Page Rank set of rules

4.3. Greedy Algorithm

Grounded taking place solidarity badly-behaved, we use a grasping set of rules. Implicit facts consists of past sports activities as recorded in Web server logs through cookies otherwise consultation stalking segments. Overt records commonly hail from as of recordkeeping formulae too evaluation opinion poll. Additional record which includes demographic and alertness records (as an instance, e-trade transactions) additionally may stay castoff. Trendy a few gears, Trap gratified material fabric, shape, also alertness statistics can beadd as extrabelongings of facts, to shed extravagand on the following levels. Facts be located often pre-deal with to place the aforementioned properright hooked on aplanlikeminded thru evaluation approach for usein the subsequent step. Preprocessing can also additionally moreover embody cleaning records of inconsistencies, filtering out outside the factor facts in keeping with the goal of assessment (instance: mechanically engendered desires on the way to entrenched pixmay be located chronicled hip internet waitperson kindling, notwithstanding the reality that they add little facts approximately patron interests), and finishing the mislaid families (owed on the way to hoarding) cutting-edge half-finished clunk on nonexistent routes. Most importantly, precise classes prerequisite on the way to be situated recognized as of the exceptional requests, primarily constructed totally taking place a empirical, which include appeals instigating beginning an indistinguishable IP deal withinside a prearranged stretch old-fashioned. Scrutiny of Trap facts - As well called Trap Convention Pulling out, this footstep rub on contraption getting to know otherwise Facts Pulling out performances on the way to find out thought provoking utilization forms too algebraic parallels amonggent folios too consumer businesses. This pace outcomes trendy automated character describing, too stays commonly pragmatic on-line, just thus the aforementioned see to now no longer add a burden on the net server. The last phase in personalization uses...
the effects of the preceding evaluation step to supply tips to the consumer. The advice machine commonly involves producing go-ahead Trap pleased material fabric taking place the sail, inclusive of which include hyperlinks in the direction of the former net trap folio asked via the character. Hip the begin, a consumer silhouette be situated erratically determined on because the pit contemporary gathering. The bordering consumer silhouette be located constantly decided on too mixed per pit till band mollifies p-congeniality or else dimensions gathering |Gi| mollifies limit |Gi| ≥ |U|avg . Next to subsequent footstep, consumer contour per elongated aloofness on the way to preceding pit stays chosen because pit brand newfangled band.

declaration

C ← Ø
seed ← a randomly picked consumer profile from S
while |S| > zero do
seed ← the furthest consumer profile (with the min similarity value) to seed
while C does NOT fulfill p-likeability AND |S| > zero do upload the nearest consumer profile (with the max similarity value) to C
endwhile
if C does fulfill p-likeability then
result ← result∪ C;
C ← Ø
end if
endwhile
for every consumer profile in C do assign it to the nearest cluster cease for

The issue to defend privateness is producing an internet profile this is placed into impact on a seek proxy walking on a consumer gadget itself. This proxy can have the hierarchical consumer profile and custom designed privateness necessities. Phases on this Architecture is composed on line and offline segment. Hierarchical era of consumer profile on consumer aspect and custom designed privateness necessities exact through the consumer are handled. The above stated operating and question managing is determined in on linesegment as:

1. User troubles a question Q1 at the consumer, seek proxy will generate a consumer profile in runtime ensuing the generalized consumer profile G1 pleasurable the privateness necessities.
2. Both the question and generalized consumer profile are despatched to the server for the customised seek to retrieve the applicable outcomes.
3. The end result is personified with the profile and is despatched to the question proxy in which the proxy will gift the outcomes or re-ranks them in line with consumer profile.

Fig 4: Greedy Search Algorithm

5. Conclusion

Personalized netseek modifies the quest outcomes to development the quest first rate for net customers. However, consumer’s non-public facts is probably uncovered with inside the consumer profile that's the inspiration in customized netseek. In this survey, mentioned numerous set of rules and associated paintings for decreasing net page
complexity in netseek engine. Based in this survey, K-Means clustering desire guide intervention to extract the facts from database. And additionally Page rank set of rules desires a big wide variety of click on via datasets. Finally grasping set of rules is used to put in force private ness primarily based totally customized seek in greenway.

References

A. Paranjape, R. West, L. Zia and J. Leskovec, "Improving Website Hyperlink Structure Using Server Logs," in Proceedings of the Ninth ACM International Conference on Web Search and Data Mining, 2016.

1. H. Kao, J. Ho and M. Chen, "WISDOM: Web Intrapage Informative Structure Mining Based on Document Object Model," IEEE Transactions on Knowledge and Data Engineering, vol. 17, no. 5, pp. 614-627, 2005.

2. H. Kao, S. Lin, J. Ho and M. Chen, "Mining Web Informative Structures and Contents Based on Entropy Analysis," IEEE Transactions on Knowledge and Data Engineering, vol. 16, no. 1, pp. 41-55, 2004

3. P. Loyola, G. Martinez, Muñoz, V. J. D. K., Maldonado and C. A. P., "Combining eye tracking and pupillary dilation analysis to identify website key objects," Neurocomputing, vol. 168, pp. 179-189, 2015.

4. Murugesan, M., Thilagamani, S., "Efficient anomaly detection in surveillance videos based on multi layer perception recurrent neural network," Journal of Microprocessors and Microsystems, Volume 79, Issue November 2020, https://doi.org/10.1016/j.micpro.2020.103303

5. P. Yin and Y. Guo, "Optimization of multi-criteria website structure based on enhanced tabu search and web usage mining," Applied Mathematics and Computation, vol. 219, no. 24, pp. 11082-11095, 2013.

6. M. Chen and Y. Ryu, "Facilitating Effective User Navigation through Website Structure Improvement," IEEE Transactions on Knowledge and Data Engineering, vol. 25, no. 3, pp. 571-588, 2013

7. Thilagamani, S., Pandhakumar, C., " Implementing green revolution for organic plant forming using KNN classification technique", International Journal of Advanced Science and Technology, Volume 29, Issue 7S, pp. 1707–1712

8. J. Hou and Y. Zhang, "Effectively Finding Relevant Web Pages from Linkage Information," IEEE Transactions on Knowledge and Data Engineering, vol. 15, no. 4, pp. 940-951, 2003.

9. P. Gupta, A. Goel, J. Lin, A. Sharma, D. Wang, and R. Zadeh, "Wtf: The who to follow service at twitter," in WWW, 2013, pp. 505–514.

10. H.-N. Kim and A. El Saddik, “Personalized pagerank vectors for tag recommendations: inside fokrank,” in RecSys, 2011, pp. 45–52.

11. Thilagamani, S., Shanti, N., “Gaussian and gabor filter approach for object segmentation”, Journal of Computing and Information Science in Engineering, 2014, 14(2), 021006, https://doi.org/10.1115/1.4026458

12. P. Lofgren, S. Banerjee, and A. Goel, “Personalized pagerank estimation and search: A bidirectional approach,” in WSDM, 2016, pp. 163–172.

13. P. A. Lofgren, S. Banerjee, A. Goel, and C. Seshadhri, “Fast-ppr: Scaling personalized pagerank estimation for large graphs,” in KDD, 2014, pp. 1436–1445.

14. Rhaqini, A., Thilagamani, S., "Women defence system for detecting interpersonal crimes", International Journal of Advanced Science and Technology, 2020, Volume 29, Issue 7S, pp. 1669–1675

15. S. Luo, X. Xiao, W. Lin, and B. Kao, “Efficient batch one-hop personalized pageranks,” ICDE, 2019.

16. T. Maehara, T. Akiba, Y. Iwata, and K.-i. Kawarabayashi, “Computing personalized pagerank quickly by exploiting graph structures,” VLDB, vol. 7, no. 12, pp. 1023–1034, 2014.

17. N. Ohsaka, T. Maehara, and K.-i. Kawarabayashi, “Efficient pagerank tracking in evolving networks,” in KDD, 2015, pp. 875–884.

18. L. Page, S. Brin, R. Motwani, and T. Winograd, “The pagerank citation ranking: Bringing order to the web.” Stanford InfoLab, Tech. Rep., 1999.

19. K. Deepa, S. Thilagamani, “Segmentation Techniques for Overlapped Latent Fingerprint Matching”, International Journal of Innovative Technology and Exploring Engineering (IJITEE), ISSN: 2278-3075, Volume-8 Issue-12, October 2019. DOI:10.35940/ijitee.L2863.1081219.

20. Z. Qin, Y. Yang, T. Yu, I. Khalil, X. Xiao, and K. Ren, “Heavy hitter estimation over set-valued data with local differential privacy,” in CCS, 2016, pp. 192–203.
21. Deepa. K, LekhaSree. R, Renuga Devi. B, Sadhana. V, Virgin Jenifer. S, “Cervical Cancer Classification”, International Journal of Emerging Trends in Engineering Research, 2020, 8(3), pp. 804–807. https://doi.org/10.30534/ijeter/2020/32832020.
   A. D. Sarma, A. R. Molla, G. Pandurangan, and E. Upfal, “Fast distributed pagerank computation,” Theoretical Computer Science, vol. 561, pp. 113–121, 2015.
22. Santhi, P., Priyanka, T., Smart India agricultural information retrieval system, International Journal of Advanced Science and Technology, 2020, 29(7 Special Issue), pp. 1169–1175.
23. J. Tang, J. Sun, C. Wang, and Z. Yang, “Social influence analysis in large-scale networks,” in KDD, 2009, pp. 807–816. [28] H. Tong, C. Faloutsos, and Y. Koren, “Fast direction-aware proximity for graph mining,” in KDD, 2007, pp. 747–756.
24. Santhi, P., Lavanya, S., Prediction of diabetes using neural networks, International Journal of Advanced Science and Technology, 2020, 29(7 Special Issue), pp. 1160–1168
25. S. Wang, Y. Tang, X. Xiao, Y. Yang, and Z. Li, “HubPPR: effective indexing for approximate personalized pagerank,” VLDB, vol. 10, no. 3, pp. 205–216, 2016.
26. S. Wang, R. Yang, X. Xiao, Z. Wei, and Y. Yang, “FORA: Simple and effective approximate single-source personalized pagerank,” in KDD, 2017, pp. 505–514.
27. Santhi, P., Mahalakshmi, G., Classification of magnetic resonance images using eight directions gray level co-occurrence matrix (8dglcm) based feature extraction, International Journal of Engineering and Advanced Technology, 2019, 8(4), pp. 829–846.
28. F. Zhu, Y. Fang, K. C.-C. Chang, and J. Ying, “Incremental and accuracy-aware personalized pagerank through scheduled approximation,” VLDB, vol. 6, no. 6, pp. 481–492, 2013.
29. R. Cooley, B. Mobasher and J. Srivastava, “Data Preparation for Mining World Wide Web Browsing Patterns,” Knowledge and Information Systems, vol. 1, p. 1–27, 1999.
30. R. Srikant and Y. Yang, “Mining Web Logs to Improve Web Site Organization,” in Proceedings of the 10th International Conference on World Wide Web, Hong Kong, 2001.
31. Vijayakumar, P, Pandiaraja, P, Balamurugan, B & Karuppiah, M 2019, ‘A Novel Performance enhancing Task Scheduling Algorithm for Cloud based E-Health Environment’, International Journal of E-Health and Medical Communications, Vol 10, Issue 2, pp 102-117
32. M. Morita and Y. Shinoda, "Information filtering based on user behavior analysis and best match text retrieval," in Proceedings of the 17th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, Ireland, 1994.
33. P. Pandiaraja, N Deepa 2019, “ A Novel Data Privacy-Preserving Protocol for Multi-data Users by using genetic algorithm”, Journal of Soft Computing, Springer, Volume 23, Issue 18, Pages 8539-8553.
34. M. Chen, "Improving website structure through reducing information overload," Decision Support Systems, vol. 110, pp. 84-94, 2018.
35. N Deepa, P. Pandiaraja, 2020, “ Hybrid Context Aware Recommendation System for E-Health Care by merkle hash tree from cloud using evolutionary algorithm”, Journal of Soft Computing, Springer, Volume 24, Issue 10, Pages 7149-7161.
36. H. Liu and V. Keselj, “Combined mining of web server logs and web contents for classifying user navigation patterns and predicting users’ future requests,” Data and Knowledge Engineering, vol. 61, no. 2, p. 304–330, 2007.
37. B. Berendt, B. Mobasher, M. Spiliopoulou and J. Wiltshire, “Measuring the accuracy of sessionizers for web usage analysis,” in Proceedings of the Web Mining Workshop at the 1st SIAM International Conference on Data Mining, Chicago, 2001.
38. N Deepa , P. Pandiaraja, 2020, “ E health care data privacy preserving efficient file retrieval from the cloud service provider using attribute based file encryption “, Journal of Ambient Intelligence and Humanized Computing, Springer, https://doi.org/10.1007/s12652-020-01911-5.
39. M. Claypool, P. Le, M. Waseda and D. Brown, "Implicit interest indicators," in Proceedings of the 6th International Conference on Intelligent User Interfaces, 2001.
40. K Sumathi, P Pandiaraja 2019, “ Dynamic alternate buffer switching and congestion control in wireless multimedia sensor networks”, Journal of Peer-to-Peer Networking and Applications, Springer, Volume 13, Issue 6, Pages 2001-2010.
41. P. Loyola, G. Martínez, Muñoz, V. J. D. K., Maldonado and C. A. P., "Combining eye tracking and pupillary dilation analysis to identify website key objects," Neurocomputing, vol. 168, pp. 179-189, 2015.
42. M. Butkiewicz, H. Madhyastha and V. Sekar, "Characterizing Web Page Complexity and Its Impact," IEEE/ACM TRANSACTIONS ON NETWORKING, vol. 22, no. 3, pp. 943-956, 2014.

43. Y. Zhang, H. Zhu and S. Greenwood, "Website Complexity Metrics for Measuring Navigability," in In Proceedings of the Fourth International Conference on Quality Software, 2004.

D. Chi, P. Pirolli and J. Pitkow, "The Scent of a Site: A System for Analyzing and Predicting Information Scent, Usage, and Usability of a Web Site," in Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Hague, Netherlands, 2000.

E. Chi, P. Pirolli, K. Chen and J. Pitkow, "Using information scent to model user information needs and actions on the Web," in Proceeding of the ACM Conference on Human Factors in Computing Systems, Seattle, WA, 2001.