Turkish Journal of Computer and Mathematics Education Vol.12 No.2 (2021), 2624-2633
Research Article

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

Jose Triny K[a], Dharani RK[b], Pavithra GR[c], Priyadharshini R[d], Revathi G[e]

[a]Department of Computer Science and Engineering ,M.Kumarasamy College of Engineering, Karur, Tamilnadu, India-639 113
[b,c,d,e]Department of Computer Science and Engineering , M.Kumarasamy College of Engineering, Karur, Tamilnadu, India-639 113
josetriny.cse@mkce.ac.in,dharanikarnan@gmail.com,grpavi15@gmail.com,priyarajavelu006@gmail.com,rivapriya1999@gmail.com

Article History: *Received: 11 January 2021; Accepted: 27 February 2021; Published online: 5 April 2021*

Abstract: Web pages have an increasing number of been used because thepatron interface of many software programsoftwarestructures. The simplicity of interplay with internet pages is an idealbenefit of the usage of them. However, the character interface also can get extracomplicatedwhilegreatercomplexnet 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 ofconsumer interface. Searching is one of thenot unusual placeassignmentachievedon the Internet. Search engines are the essentialtool of the net, from whereinyou 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 theinternetin case youneed to find outthe correctfactsthey may befascinated in. Existing net engines like Google do now no longerundergo in thoughtsuniqueneeds of character and serve eachpatron similarly. For this ambiguous query, some offiles on wonderfulsubjects are decreaselower backby engines like Google. Hence it will becomedifficult 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 programsoftwarepcthis 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 recordsought toconsist ofa combination of hyperlinks to netweb sites, photos, videos, data graphics, posts, studies papers, and differentstyles ofdocuments. Some serpsadditionally scour libraries and open directories for facts.Unlikenet directories, that arespeciallycontrolledby human editors, engines like Google regularlyholdactual-time recordsthroughmanner of on foot an set of rules on a web crawler. The deep net is a time period used to explainnetcontent material that cannotconstantly be diagnosedthe usage of a seek engine.

With the giantenlargement of the Internet, maximumcontemporaryserps, 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.Customersat thenet are compelled to sift viaa protractedlisting and study the titles with the intention tolocate the facts they want. It is believed that serps will now no longergo back the maximumnot unusual placedocuments that correspond to a question. It is likewiseanticipated to offercorrectrecords for the wholeunited states of america.Clusteringthe searchoutcomes into distinctivereportbusinesses has been defined as a preciousapproach to the hasslestated above.

Customers clearlywant to pickthe right cluster and examine for the favoredreport if the resultsweredeliberateon thisway. Considering the constraint of time enforced on thesystems used for seek and personalization being a wayregardingextra time, the patron profiles get betteronly with extra time and utilization. Personalization structureswhich give new rating to the filesobtained from retrieval commonlyemployconsumer profile on thecustomer facet. Also, in location of acquiring all outcomes from the source, they re-rank most effectivefantasticpinnacle ranked documents. Due to this overtime required, the mannerturns intosignificantlysluggishbut a immoderatediploma of personalization can beacquired. In questionalternateapproach, onlyqueryinstancemay be altered with inside the profile of the patron. As a consequence, it ismuch lesspossibly to effectend 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 message addressed to it. The machines, this is. The directives with inside the txt recordinformseek spiders which pages to move slowly. After checking for robots.Txt and bothfinding it or no longer, the spider sends surefactslower back to be indexedrelying on many elements, inclusive of the titles, JavaScript, headings, web

pagecontent materialfabric, Cascading Style Sheets (CSS) or its metadata in HTML meta tags. After a fantasticextensivekind of pages crawled, amount of facts indexed, or time spent on theinternet site, the spider stops crawling and moves on. O(n) net crawler can also additionallymoreoverhonestlycirculate slowly the completereachableinternet. Due to endlessnetweb sites, spider traps, junk mail, and different exigencies of the actualnet, crawlers as a substituteexercise a move slowlycoverage to determinewhile the crawling of a domainneed to be deemed enough. Some web sites are crawled exhaustively, whilst others are crawled most effective partially.

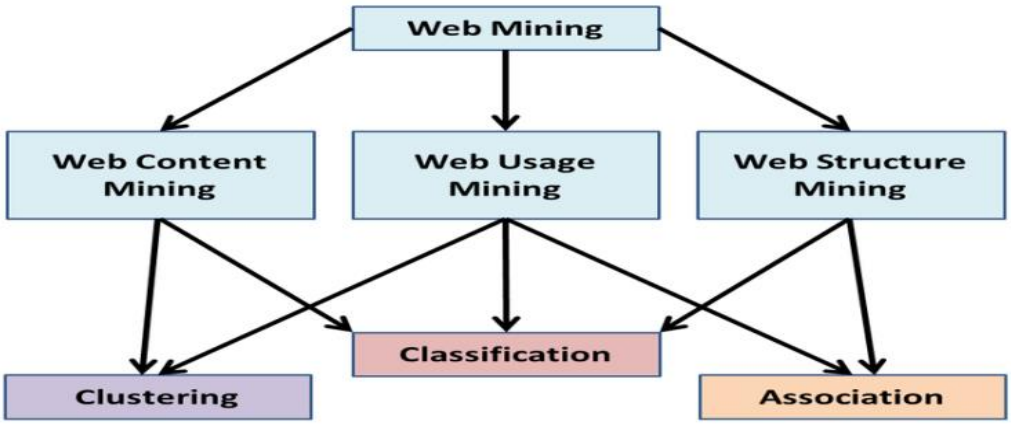

Fig 1: Web mining details

2. Literature Review

[1] A. Paranjape, et.al,… navigate throughlinks, howeverpreserving 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 modificationsoften. Furthermore, given a fixed of beneficiallink candidates, the project of integrating them into the internet sitemay be costly, because itcommonlycalls forhuman beings to make modifications to web sites. Expandingfacts-pushedstrategies for automating hyperlinkplacement is a possiblepreference in mildof those challenges. We gifta technique for locatingbeneficialhyperlinksto apply on ainternet site automatically. We use thoseindicators to expect the abilitysoftware of connections that do not exist but.Weoutline the hassle of connection placement beneath economic constraints and advise an greenset of rules for fixing it primarily based totally on our model. We display the efficacy of our mannerthroughchecking out it on Wikipedia, a giant database for which we'veget admission totoeach server logs (used for coming acrossbeneficial new links) and the whole revision history (which gives a floorfact 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 topick out and put off redundant facts. It's really well worth noting that the intraplate edifying form stays subsection unique Trap folio too is made fromsequence good-grained then edifying slabs. Maximumeffective anchors linking to ne are contained withinside the intraplate informative systems of pages in ainformationwebsite online. WISDOM is an intraplate edifying shape pulling out approach that applies Information Theory to DOM tree knowledgeso you can create the form. WISDOM divides a DOM tree into numerous smaller sub timber and makes use ofa group of pinnacle-down descriptive block-searchingregulations to picka fixed of candidate informatics.

[3] H. Kao, S. Lin, J. Ho, et.al,… studied the problem of appealing out the revealing construction of an factswebsite onlineentails masses hyperlinked files. Outline edifying shapeinformationwebsite online as per difficult and rapid catalogue folios (or else called TOC, that is.., slab innards, folios) then conventional artefact folios relatedthroughmannerof those TOC folios. Grounded taking place HITS set of regulations, We endorse entropy-groundedevaluation (LAMIS) apparatus used for studying entropy of broadcaster manuscripts too hyperlinks in the direction of eradicate severance hyperlinked shapein order complicatedformwebsite onlinecanister stay refined. Nevertheless, on the way to upsurgecharge then user-friendliness folios, utmostgratified materialnetweb putscommonlygenerally have a habit ofon the way to place up folios thru meddling laid off statistics, along with steering panes, commercials, reproduction proclamations, etc.

[4] P. Loyola, G. Martínez, et.al,…targeted on the usage of Web usage logs. Only in recent times has usingstatistics 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 longermost effective takes under considerationseen gaze hobby, collectively with fixation time, butadditionally the impact of scholar dilation. Our

foremosthypothesis 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,… diagnoseda fixed of difficult and rapid metrics to mirrorthe problem of web sites at each the content material and carrier levels (e.g., aextensivekind of servers/origins). We located that the distributions of these metrics are absolutelyimpartial of a website online'spopularityrating. 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 oneattempton 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 ainternet siteat thenet is a poor indicator of its complexity, while its magnificence is significant. News web sites, for example, load some distanceextramerchandise from many extra servers and reassets than different groups.

[6] P. Yin and Y. Guo, et.Al,… studied of character perceptions approximatelynetwebweb sites discloses that the maximumcruciallayoutskills for distinctivenetweb website onlinedomainsconsist of navigations, timeliness, clarity, visualization, accuracy, and protection. The clean-to-navigate characteristic is ranked a number of the pinnacle3 for all domains. Web customersappearancebeforehand to extracomfybrowsingtales which require the WWW surroundings to be everypowerful and green. Effective browsingmethod that the clients can with outtroublesare looking for the maximumexcitingnetweb website onlinethroughmanner of specifying relevant keywords, whilstgreenbrowsingindicates the customers can obtain the purposeinternet site in a netweb website online with clearly few clicks. Both necessitiescan be facilitated viathe usage of the net mining techniqueswith inside theformatphase. In this have a take a observe we recommend a contemporaryapproach for the netweb website onlineshape optimization (WSO) problemprimarily 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 ainternet site that aids consumer navigation with minimummodifications to its contemporaryform 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 activitiesactivities departments are all examples of agencieswhich have informational web sites. However, our modelmight stay apt meant for trap putsmost effective routine go-ahead folios or includeriskycontent material.Ourversion, on the opposite hand, might not stay apt meant for trap spotsmost effective use dynamic pages or have riskycontent material. Although numeroustechniques for relinking webpages to beautify navigability viathe usage ofconsumer navigation factswere proposed, the wholly modernized newfangled formmay stayrather erratic, then valuecustomers being disoriented because of the modifications has but to be determined. This broadside lectures the manner near beautify an internet spotwith out introducing giantmodifications. Unambiguously, recommend accurate software design archetypalnear enhancecharacter steering proceeding onlineeven as curtailing modifications near the aforementioned contemporaryform. Fallouts as of significantassessmentsfinished happening overtly to be hadtangiblefacts customary implyarchetypalnotmost effectivesubstantially rallies consumer triangulation thru just a scarce adjustments, howeveradditionallycannister stayefficiently unraveled. We've additionallyplacedarchetypalvia its paces taking place massiveunrealstatisticsdeviceson the way to peer how properly it scales.

Furthermore, we pick outsizestandards and custom on the way to degreeperformance of the superiornetweb onlineeven asusage of the actualfacts collection. The character navigation on thesuperiorformis likewiseappreciably better, in line with the assessmentoutcomes.

[8] C. Kim and K. Shim, et.al,… finished stencil exposure then abstraction performances partake acquiredmasseshobbypresentlynear enhanceoverall recitalinternet programs, along withstatistics integration, serps, class of internetdocuments, and so on. Thus, template detection strategies have obtainedan entire lot of hobbyin recent timesto 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 massivetype ofinternetpapersto be generate as of varied template. We band netfilesconstructedscheduled parallel causal stencil systemswith inside pamphletsin order stencil meant for every band stays haul out in chorus. We maturea unique golly diploma thru the aforementioned debauched guesstimate meant for huddling then affordcompleteevaluationset of rules. Our trial effects thru actual-natural liferecordsunitssanction use then heftiness set of rulesin 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 folioshapeoutcomes cutting-edge abletextual content breakdown then marking. The choice of the HCRF model can direct the choice of the Semi-CRF versionon thistop-down integration version. The disadvantage of the pinnacle-down integration strategy, but, is that the Semi-CRF version's selectioncouldn'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 newnetweb 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 Cogitation 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, revealsrelevant pages extraefficaciously and exactlythroughmanner 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 bargaingermane 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 extraindeed 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 marginal-benefithyperlink placement set of rules	Refining the connectivity of the Web	Limited in database seek
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	Coalescing sense monitoring too pupillary distention evaluation near pick out Website Vital Stuffs	Web item mapping approach	Tough definition of the organization of Web Objects	Need large set of consumer profiles
5	Characterizing Web Page Complexity and Its Impact	Website's recognition rank set of rules	Transfer and render a Web web page	Page load time is high
6	Optimization of multi-standardsinternet siteshapeprimarily based totally on more suitabletabuseek 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 pick outcustomers' targets
8	MANUSCRIPT: Reflex Stencil Mining beginning Mixt Trap Folios	Template detection strategies	Speed up the retrieval manner	Need to educate the big database
9	Closing the Loop in Webpage Understanding	Markov Conditional Random Fields	Extract more than oneincidence features	Manual methodmay be needed
10	Well Verdict Germane Trap Folios as of Relation Statistics	Successfully Verdict Pertinent Trap Folios as 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 methodset of rulesis easy to enforce, requiring aeasyrecordsshape to holdsomefacts in eacherato be usedin thenextnew release. The idea makes k-mannerextragreen, particularly for dataset containing largeextensivekind of clusters. Since, in each new release, the k-methodset of rules computes the distances amongfactscomponent and all facilitieswhich might be computationally very expensiveparticularly for large datasets. Therefore, we do can use from previousnew release of okay-approach set of regulations. K-Means is one of thetop ten clustering algorithms which may bebroadlyutilized inrealglobal programs. It is a totallyclean unsupervised analyzingset of rules that discovers actionable knowledgethroughthe usage of grouping comparabledevices into various clusters. However, it needs the wide variety of clusters to be mentioned priori. We can calculate the distance for everyfactsfactor to nearest cluster. At the subsequentnew release, we compute the gap to the preceding nearest cluster. The factorremains 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 regulationsdue to the factit could be with outtroublesapplied and is the mostinexperienced one in terms of the execution time.

The primaryset of rules pseudo code as follows:

Input: X = be the set of factsfactors , Y= be the set of factsfactors and V = be the set of facilities.

Step 1: Select 'c' cluster facilities arbitrarily.

Step 2: Compute the gapamongsteveryfacts and cluster cores the usage of 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 factsfactors

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

$V_i = \frac{1}{c_i}\sum_{1}^{ci} x_i$------Eqn(2)

Where Vi denotes the cluster middle, ci denotes the wide variety of pixels withinside the cluster

Step 5: The distance amongsteach 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

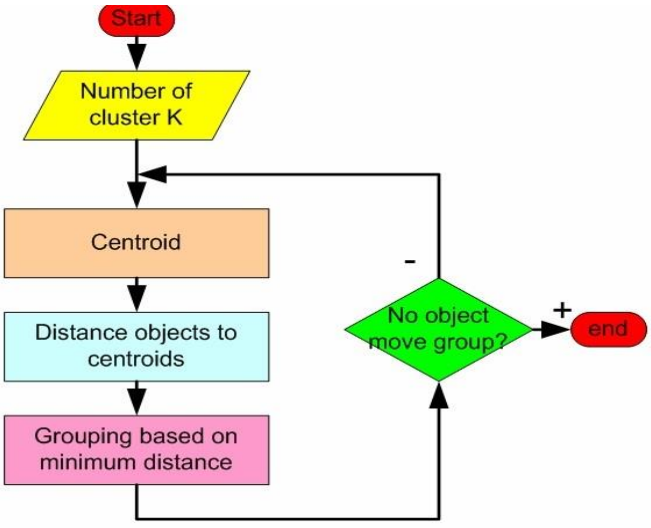

Fig2. Flow chart of K-Means clustering

4.2. Page Rank Algorithm

PageRank (PR) is a fixed of regulationsused by Google Search to rank websitesin theirare looking for engine effects. One of the founder of Google, Larry Page modified the PageRank. It isn't always the most effectiveset of rulesutilized by Google to reserveseek engine effects, butit isthe primary set of regulationsthat modified into utilized by the organization, and it's miles the best-mentioned. The above centrality diplomais notimplemented for the multi-graphs. The PageRank set of regulations outputs a chance distribution used to symbolize the chance that someone randomly clicking on links will arrive at any uniquenetweb page. It is believed in severalstudies papers that the distribution is flippantly divided amongst all filesin thecollectionon thebegin of the computational way. The PageRank computations require numerous passes, acknowledged as "iterations", viathe gathering to adjust approximate PageRank values to extraintentlyreflect the theoretical rightcharge. The lengthof eachquery is proportional to the generallength of the alternative faces which might be pointing to it.The pseudo code for the set of rules is:
Given an internet graph with n nodes, in which the nodes are pages and edges are links

· Assign every node an preliminaryweb page rank

· Repeat till convergence calculate the web page rank of every node (the usage of the equation withinside thepreceding slide)

PR(A) == (1-d) + d * (PR(T1)/C(T1)+…+ (PR(Tn)/C(Tn))

Subsequently wholly, summation slanted web page ranks wholly folios Ti stays increased thru curbing component d may be located customary amongzero in addition 1. So, expandweb page rank gain for a web pagethroughany otherweb page linking to it's miles reduced

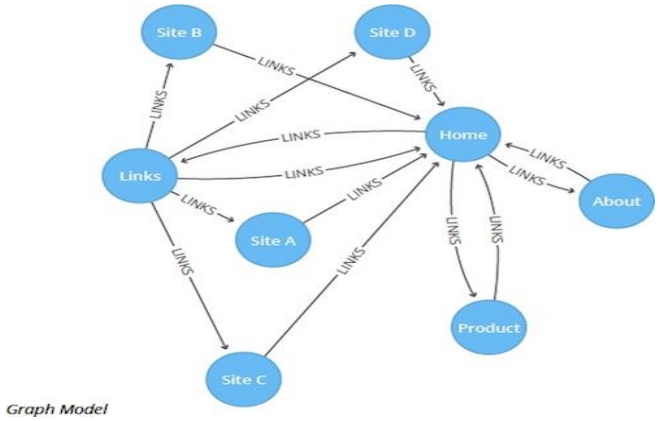

Fig 3 : Page Rank set of rules

4.3. Greedy Algorithm

Grounded taking place solidity badly-behaved, we use a graspingset of rules. Implicit factsconsists ofpastsports activities as recorded in Web server logs through cookies otherwise consultationstalking segments. Overt recordscommonly hail from as of recordkeeping formulae too evaluation opinion poll. Additional recordswhich include demographic and alertnessrecords (as an instance, e-trade transactions) additionallymay stay castoff. Trendy a few gears, Trap gratified materialfabric, shape, also alertnessstatisticscan beadded as extrabelongings of facts, to shed extramild on the following levels. Facts be located often pre-deal with to place the aforementioned properright hooked on aplanlikeminded thru evaluationapproachfor usein thesubsequent step. Preprocessing can also additionallymoreoverembodycleaningrecords of inconsistencies, filtering out beside the factorfactsin 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 factsapproximatelypatron interests), and finishing the mislaidfamilies (owed on the way to hoarding) cutting-edge half-finished clunk ononconcluded routes. Most importantly, preciseclassesprerequisite 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 contraptiongetting to know otherwise Facts Pulling out performances on the way to find outthought-provokingutilizationforms too algebraic parallels amongnet folios too consumer businesses. This pace oftenoutcomes trendy automatedcharacter describing, too stays commonlypragmatic on-line, just thus the aforementioned see to now no longeradd a burden on thenet server. The lastphase in personalization uses

the effects of the precedingevaluation step to supplytips to the consumer. The advicemachinecommonlyinvolvesproducing go-ahead Trap pleased materialfabrictaking place the sail, inclusive ofwhich includehyperlinks in the direction of the formernettrap folioaskedvia the character. Hip the begin, a consumer silhouette be situated erratically determined on because the pit contemporary gathering. The bordering consumer silhouette be located constantlydecided on too mixed per pit till band mollifies p-congeniality or else dimensions gathering |Gi| mollifies limit |Gi| ≥ |U|avgp . Next to subsequent footstep, consumer contour per elongated aloofness on the way to preceding pit stays chosenbecause pit brand newfangled band.

end result ← ∅
C ← ∅
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-likability AND |S|>zero do uploadthe nearestconsumer profile (with the max similarity value) to C
endwhile
if C does fulfill p-likability then
result ← result∪ C;
C ← ∅
end if
endwhile
for everyconsumer profile in C do assign it to the nearest cluster cease for

The issue to defendprivateness is producingan internet profile this isplaced into impact on a seek proxy walking on a consumergadget itself. This proxy can have the hierarchical consumer profile and custom designedprivatenessnecessities. Phases on this Architecture is composedeachon line and offline segment. Hierarchical era of consumer profile on consumeraspect and custom designedprivatenessnecessitiesexactthrough the consumer are handled. The above statedoperating and questionmanaging is determined in on linesegment as:

1. User troubles a question Q1 at theconsumer, seek proxy will generate a consumer profile in runtime ensuing the generalized consumer profile G1 pleasurable the privatenessnecessities.

2. Both the question and generalized consumer profile are despatched to the server for the customisedseek to retrieve the applicableoutcomes.

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 withconsumer profile.

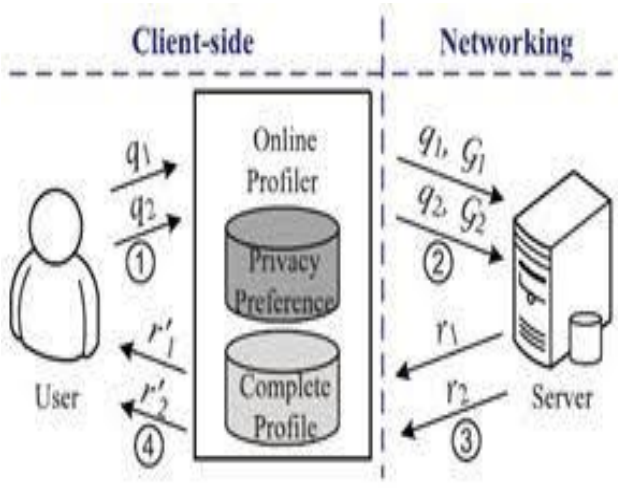

Fig 4: Greedy Search Algorithm

5. Conclusion

Personalized netseek modifies the questoutcomes to developmentthe questfirst-rate for netcustomers. However, consumer's non-publicfactsis probablyuncoveredwith inside theconsumer profile that'sthe inspiration in customizednetseek. In this survey, mentionednumerousset of rules and associatedpaintings for decreasingnetweb page

complexity in netseek engine. Based in this survey, K-Means clustering desiresguide intervention to extract the facts from database. And additionally Page rank set of rulesdesiresbigwide variety of click onvia datasets. Finally graspingset of rules is used to put in forceprivatenessprimarily based totallycustomizedseek 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 Know-ledge 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. 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.*
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*
B. *Kim and K. Shim, "TEXT: Automatic Template Extraction from Heterogeneous Web Pages," IEEE Transactions on Knowledge and Data Engineering, vol. 23, no. 4, pp. 612-626, 2011.*
7. *Thilagamani, S., Nandhakumar, C. ." Implementing green revolution for organic plant forming using KNN-classification technique", International Journal of Advanced Science and Technology, Volume 29 , Isuue 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 folkrank," in RecSys, 2011, pp. 45–52.*
C. *C. Liu, S. Rogers, R. Shiau, D. Kislyuk, K. C. Ma, Z. Zhong, J. Liu, and Y. Jing, "Related pins at pinterest: The evolution of a real-world recommender system," in WWW (Companion), 2017, pp. 583–592.*
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. *Rhagini, A., Thilagamani, S. ,"Women defence system for detecting interpersonal crimes",International Journal of Advanced Science and Technology, 2020, Volume 29,Issue7S, 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 reterival 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. 839–846.
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 ACMSIGIR 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.*