Global Journal of Pure and Applied Mathematics.
ISSN 0973-1768 Volume 15, Number 4 (2019), pp. 383-397
© Research India Publications
http://www.ripublication.com/gjpam.htm

Stochastic Query Covering Using Greedy Muticover Algorithm

Padma Priya G[1], Dr. Hemalatha M[2]

[1]PhD Research Scholar, Bharathiar university, Coimbatore, Tamil Nadu, India.

*[2]Professor, Sri Ramakrishna College of Arts and Science, Coimbatore,
Tamil Nadu, India.*

Abstract

The general thought is to populate the cache with archives that add to the outcome pages of countless, rather than reserving the top records for each query. Things being what they are, the issue is hard and comprehending it requires learning of the structure of the questions and the outcomes space, just as information of the information query dispersion. We define the issue under the system of stochastic advancement; hypothetically it tends to be viewed as a stochastic general form of set multicover. While the issue is NP-difficult to be understood precisely, we demonstrate that for any dissemination it tends to be approximated utilizing a straightforward greedy methodology. Our hypothetical discoveries are supplemented by exploratory movement on genuine datasets, demonstrating the possibility and potential enthusiasm of query-covering approaches by and by. In this paper proposed to stochastic query covering utilizing in greedy multicover calculation.

Keywords: Query, Greedy, Multi Cover, Cache Memory, Recall.

1. INTRODUCTION

With the enormous development of information accumulations, the region of surmised query noting has gotten expanding consideration in the database network over the past 10– 20 years. The thought is that at whatever point a client presents a query, the framework may restore an answer that isn't the best one, as per the chose positioning measure, yet is near it. This is a financially savvy technique, since the rough answer is quick to process, getting reserve funds reaction time. It is valuable in settings where: (I) an accurate answer isn't really required, and a rough reaction gets the job done to give the essential data; (ii) a starter reaction is returned while the precise answer is determined; or (iii) quick criticism is given to the client about the nature of the query. A typical strategy to help inexact query noting is the making of a reasonable rundown (or summary or sketch) of the whole dataset that is suitable for specific applications. These are made either by figuring a few insights over the dataset or by testing a portion of its substance. We present the stochastic query covering as an appropriate model of the situation sketched out above. It takes into account clever examination though a most pessimistic scenario approach does not give any instinct. The drawback is that the examination turns out to be in fact increasingly included. In particular, we expect a structure in which clients submit inquiries to a report recovery framework after some time. The framework utilizes a cache of constrained size to hold a subset of the archive accumulation. Preferably, records in this subset will in general be applicable for inquiries that clients are well on the way to submit. At whatever point a client presents a query, the record recovery framework must restore a lot of reports pertinent to the query. The framework either builds the outcomes utilizing records put away in cache, or it bombs; in the last case, the framework acquires a cache miss, that is, a punishment mirroring the way that developing an outcome page will require a tedious task (e.g., getting to optional capacity). This methodology gives an adaptable method to plan choice systems that are touchy to various improvement criteria by reasonably characterizing loads. To feature the adaptability of our methodology, we present two weight definitions that represent query– record significance or broadening of cached reports. With regards to the last mentioned, more or less, the customary outcome broadening implies boosting the likelihood that, after presenting a query, a client acquires probably some significant archives for the goal behind the query she submitted.

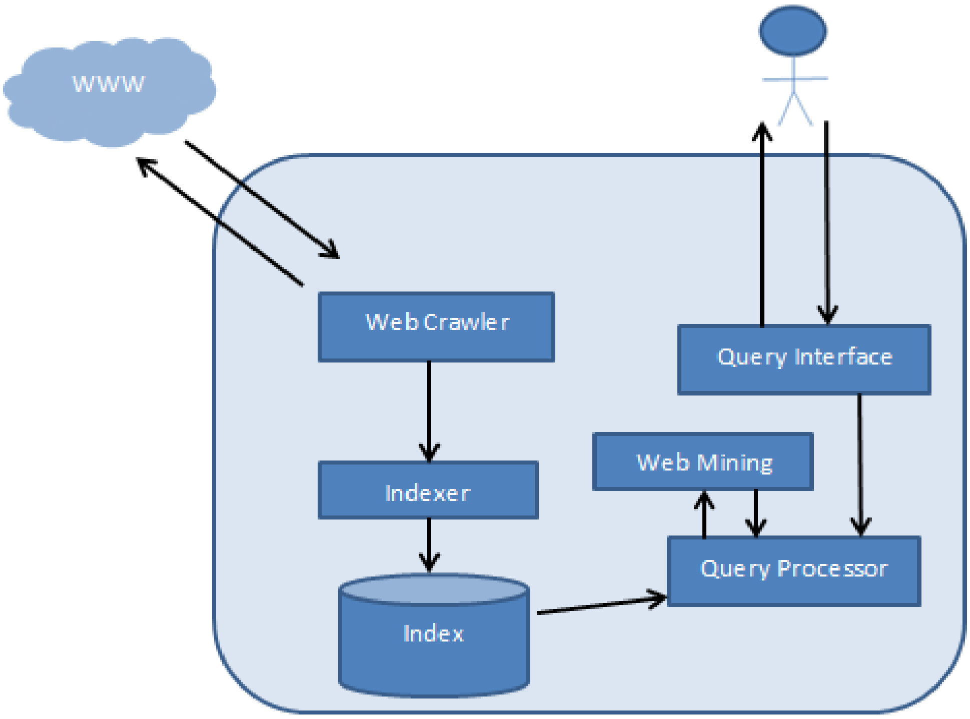

Figure 1: Query Process

Applications Our methodology is general enough and is appropriate at whatever point we have an expansive archive accumulation that we have to query. We accept that a query result is a lot of various reports pertinent to the query, and that we know (or we have enough measurements to learn) the query conveyance. It is roused by situations where we have little stockpiling capacities, for example, on account of little compact gadgets, huge correspondence costs, or when the query-preparing time increments altogether with the extent of the gathering. In the rest of present some solid potential applications in which we trust that the thoughts behind query covering could be useful. Current web indexes need to process a great many questions for every second over accumulations of billions of records, and clients expect low reaction times. To this end, web indexes utilize an assortment of reserving methods as a way to give results convenient and insignificant decrease in quality. We expect to have a static cache putting away an accumulation of records ordered suitably, and to react to a query the web index can get to the cache and thusly, whenever required, the primary memory. This technique introduces the favorable position that a similar record can be utilized to serve multiple inquiries, with the goal that extra room is utilized all the more proficiently. A cache of this sort can be utilized with different kinds of cache, for instance, one can utilize a first-level cache putting away the total page of query results for the most incessant questions, while a second-level cache stores single reports, which are utilized to remake the outcome page on the off chance that that it was not found in the principal level cache. Our methodology can be utilized as a methods for a static reserving in settings where it is admissible not to restore unquestionably the top-k reports (this may be the situation in a huge corporate internet searcher). Despite the fact that it is a sort of file pruning (we can make pruned files from the subset of reports that we select) our inspiration is not quite the same as the standard utilization of file pruning. More often than not, the procedures for file pruning proposed in the writing have the prerequisite that all the top-k archives of the gathering are returned for each query so as to protect the nature of indexed lists. Rather, in our methodology, we permit (really we frequently lean toward), for certain inquiries, to return profoundly pertinent records yet not really all the top-k. This brings significantly more funds; besides, for questions where a few reports are essential (e.g., navigational inquiries), we can set the loads so that we uphold these archives to be incorporated into the outcomes at whatever point a client plays out these questions. Condensing, despite the fact that query covering can't be utilized fundamentally for standard web look, it can give valuable thoughts and instincts to improving storing procedures of web crawlers.

2. LITERATURE SURVEY

1. Debahuti Mishra and Niharika Pujari (2011) proposed giving completely computerized support to multi space questions. This model (a) coordinates diverse sort of administrations into a worldwide metaphysics (b) covers inquiry formulation perspectives over worldwide cosmology and question changing in terms of nearby administrations (c) Several web related administrations that vanquishes the issue and the bundle that we have offered here is XAMPP Control board. the client question is reformulated to a few revamped benefits by utilizing information coordination approaches GAV, LAV and GLAV. Coherent access plan/design is started to decide the order of execution of the literals in the body of the inquiry with the goal that all the entrance examples of summoned administrations are regarded. Inquiry improvement happens by extricating question related information from the web scrapper database, which has incorporated all the distinctive areas of the client question. Web Scrapper is a PC programming strategy of removing information from sites. It helps in separating a lot of information from different sites when manual reorder activities make the errand dreary. It enables us to export the separated substance into EXCEL, My-SQL, MS-SQL, CSV, XML or TEXT format. It is finished by making system or content written in any programming language that forms the unstructured or semi-organized web information of an objective site to separate substance or information for changing over unstructured substance into organized format. With assistance of web extraction you can interface with a site's pages and solicitation information or a pages, precisely as your program would do. The web server will send back the html website page which you would then be able to separate explicit information from that page. Web information mining is otherwise called web content mining, web text mining, on the grounds that the substance or text is the most broadly inquired about zone in world of web. Our Web server utilized is XAMPP Control Panel. XAMPP Control board is a free and open source cross platform web server bundle, comprising predominantly of the Apache HTTP server, My SQL database and mediators for contents written in PHP or Perl language.2. **Chichang Jou, Yucheng Cheng** (2011) proposed Heuristics-based profound web question interface Schema Extraction framework (HSE) that distinguishes labels, elements, mappings among labels and elements, and relationships among elements. There are four parts in the blueprint model of HSE: (1) elements (2) labels (3) relationships among elements (4) mappings among labels and elements. The initial two are bases for the third and fourth parts. For relationships among elements, we distinguish the accompanying four relationships in our pattern demonstrate: 1) Parent-youngster relationship: There exists one to numerous various leveled relationships among one and a few different elements. Instances of this relationship are: (1) a vehicle's image and model, (2) the state, city, and road address, and (3) the class and sub-classifications of items. 2) Group relationship: for radio and checkbox elements, ordinarily the content in the correct hand side of the radio catches and checkboxes could be treated as their comparing esteems. In this way, for these elements, we bunch them together, and treat them as a unit. 3) Range relationship: This is for sets of elements with relational word labels, as "previously" and "after", "from" and "to", and so on. The info esteems for the primary component would force requirements on those for the second component. The inquiry states of these sets of elements will be "AND" together to form a range condition. 4) Part-of relationship: This is for related sets of elements not with range relationship. For mapping among labels and elements, the accompanying two sorts of mappings are characterized in our diagram show: 1) Basic mapping 2) Advanced mapping.**3. Kuldeep R. Kurte, Surya S. Durbha, Roger L. King, Nicolas H. Younan, and Rangaraju Vatsavai** (2016) A spatial IIM (SIIM) system is proposed, which coordinates a rationale based thinking component to separate the covered up spatial relationships (both topological and directional) and empowers picture recovery dependent on spatial relationships. In SIIM, the learning driven semantic information demonstrating approach is embraced, which comprises of 1) improvement of space and spatial connection philosophy and 2) connected information portrayal of the EO information. A spatial design speaks to spatial items in a RS picture and the topological and directional relations among them. Papadias and Sellis have instituted the expression "spatial information," which portrays the spatial setup among particular spatial elements. The real focal point of this work is to formally characterize this spatial information in the RS symbolism to such an extent that it will encourage the spatial connection based questions onto the EO information. When all is said in done, spatial relationships are delegated pursues 1) Topological relations depict neighborhood and frequency (e.g., EC, DC, and nontangential appropriate part (NTPP). 2) Directional relations portrays relative headings of articles (e.g., North and South). 3) Ordinal relations portrays the separation relationships. In this work, we have thought about the demonstrating of topological and directional relations among picture areas and furthermore among geo-MBRs of districts.4. **Rabah A. Al-Zaidy, C. Lee Giles** (2018) proposed a framework to extricate semantic relations among substances in insightful articles by utilizing outer syntactic examples and an iterative student. Since our methodology does not depend on a current information base, we bootstrap our calculation by separating sets utilizing a lot of hand-created designs. The utilization of lexicosyntactic designs for bootstrapping hyponymy extraction is a typical practice since they have generally high accuracy while trading off review, which is middle of the road for a bootstrapping step. We utilize the Hearst designs from that characterize syntactic examples that are regularly used to mean a hyponymy relationship. Table I demonstrates the examples utilized in our framework. NP represents Noun Phrase, that are first recognized and afterward coordinated against the example. The Stanford lexical parser parses the sentence to extricate thing phrases. These are thusly passed to a lexical tree matcher that distinguishes the examples as per the thing expression recognized at the most elevated amount in the hub. will remove NP1 as the whole expression of high and low dimension programming dialects rather than low dimension programming dialects. The Stanford token matcher is utilized to apply the examples to the tokenized sentence to get this outcome. The removed competitor sentences are then parsed by distinguishing the arrangements of X and Y words containing every one of the hypernym and hyponym phrases, separately. Stemming and evacuating introductory stop words that are regular descriptive words is performed to purge the expression preceding the hypernym recognition step. B. Semantic Unit Extraction The bootstrapping approach depends on syntactic examples that misuse grammatical form labels to recognize thing phrases. Be that as it may, the subsequent expressions can be extremely boisterous on the off chance that they contain identifiers and particular kinds of modifiers. So as to improve the nature of separated idea phrases, we utilize the thought of semantic units characterized as "the maximal-length thing phrase speaking to a solitary element particularly and completely, paying little respect to the printed contextâăi. We receive an empirical methodology dependent on huge scale corpus experimentation to determine heuristics. We characterize two arrangements of standards to help us linguistically categorize a semantic unit, outskirts leads and parsing rules appeared table II. The wilderness rules are utilized to strip the front of the expression by recursively coordinating the principal token in the expression against the standard until none of the principles can be connected. The parsing rules exploit the pos labels to additionally diminish thing expressions to as essential a unit as could be expected under the circumstances. 5. **Ali Jalali Santanu Kolay Peter Foldes Ali Dasdan** (2013) proposed a circulated blame tolerant framework that can produce such gauges quick with great precision. The primary thought is to keep a little representative example in memory over numerous machines and formulate the estimating issue as questions against the example. The stratified testing gives the advantage of improved precision over the uniform inspecting given a similar example measure requirement. Anyway the exactness is as yet reliant on the example estimate. So as to deal with a bigger example estimate, we circulate the example over a lot of servers. As the questions are trivially parallelizable, the general framework dormancy is essentially decreased. We are additionally ready to help numerous synchronous inquiries over a similar example. We portray the necessities of the circulated framework. Guaging is key for promoters to pinpoint their precise focusing on requirements and spending plan spend in order to achieve the best sections of the group of onlookers with the optimum valuation. In that capacity, anticipating must be exact. Besides, anticipating should be quick to enable promoters to perform intelligent investigation while finishing their campaign parameters amid campaign setup. Anticipating is likewise valuable to assess the DSP accomplice of the sponsor to think about a when picture of what was determined and how the campaign returned. A similar evaluation is additionally utilized by DSPs amid debugging campaign conveyance and execution issues.

3. PROPOSED WORK

3.1 Greedy Multicover Algorithm

To begin with, we remark on the assumption that queries are drawn from a circulation. Note that to acquire a solid hypothetical outcome the stochastic assumption is important to accomplish a little estimation proportion; notwithstanding for the straightforward widespread set spread issue (where $k = 1$) there is a lower bound of $\Omega(\sqrt{n})$ if no stochastic assumptions for the info conveyance are made. Then again, expecting that queries are inspected from a dispersion is a standard and sensible assumption (depicting the outfit of queries from all clients) and can furnish us with a lot more keen limits. An issue with the stochastic assumptions is that outcomes are all the more demanding in fact; which is as of now in fact included, can't be connected here and we need further specialized work. Practically speaking, be that as it may, we don't have a clue about the appropriation, rather we watch tests from it. In our analyses we will utilize the learning from a given period (it very well may be multi day, or a given timeframe of multi day) to get an estimate of the circulation and apply the calculation for the following time frame (a similar period following day, or a similar period one week from now, and so forth.). Another issue is that the dispersion changes after some time, anyway for a large portion of the queries, between two back to back timeframes we expect that a large portion of the queries will have comparative likelihood to show up. This is one reason that to become familiar with the circulation we utilize just the past timeframe, albeit another sensible choice is amass queries over a more drawn out period—we expect that subjectively our discoveries will be the equivalent. In our trial results we think about the cover between queries in back to back timespans; of course, there is cover between queries with high recurrence, while queries in the tail of the circulation basically don't cover. A cautious examination of the verification can give us likewise some instinct of the fundamental picture. Specifically, Lemma 3 gives us data about the issue structure that can enable us to apply the insatiable calculation even for the situation that the store estimate is restricted. At an abnormal state, what the lemma states is that there exists a lot of reports that covers a vast portion of the queries k times.

```
Function multiCoverGreedy( k )
  /* Initially all queries are uncovered. */
  foreach (q ϵ  Q)
    results (q) ← Ø
  while (exists q s.t. |results(q)| < k)
    /* If there exists an uncovered query, find the most cost-effective doc */
    doc ← most Cost EffectiveDoc()
    foreach (q ϵ  queries(doc))
      results(q) ← results(q) ∪ {doc}
  end while

Function mostCostEffectiveDoc()

  /* In the unweighted case, the most cost-effective doc covers most queries */
  foreach (u ϵ  U)
    costEffectiveness(u) ←
      1/ | {q ϵ  Q : results(q) < k and u ϵ  q}|
    if (costEffectiveness(u) < costEffectiveness(doc))
      doc ← u
  return doc
```

Figure 2: Multicover Greedy Algorithm

Moreover, the insatiable calculation can discover it. In this way, regardless of whether the reserve measure isconstrained, we can apply the insatiable calculation and populate the store with those records, utilizing pre-getting. Note however that the extent of queries that are not secured by Lemma 3 relies upon the estimation of the ideal arrangement (for t demands), Copt. On the off chance that the proportion Copt/t is little this suggests there is an extensive number of archives that can cover a few queries. This is both due to cover between the important archives among different queries, just as a result of the skewness of the question circulation. By and by, both of these circumstances are valid: the outcome sets of web seek queries, for example, "lodging rome,""cheap inn rome," "rome convenience, etc, are relied upon to be profoundly covering, and comparable is the circumstance in different settings where we can apply inquiry covering thoughts (or possibly where question covering is reasonable). Likewise question frequencies pursue control law circulations (see for instance [14, 18]). In this manner we anticipate the proportion Copt/t to be little enough and in this way the eager calculation to have the capacity to cover numerous queries that will show up later on by embeddings records in the reserve.

4. EXPERIMENTAL RESULTS

Recall

Table 1: Comparison of Recall

Existing 1	Existing 2	Proposed
20	17.5	26.6
28.67	23.9	30.56
33.9	27.1	36.66
35.2	30.12	38.26
37.8	32.4	42.14

The table 1 explains the comparison of recall of two existing method and a proposed method. The recall options in existing 1 has minimum 20 query test and maximum 25th training set to 37.8 query test in 125th training set, existing 2 has minimum 17.5 query test in 25th training set and maximum 32.4 in 125th training test, but in proposed method minimum 26.6 query test in 25th training set and maximum 42.14 in 125th training set and it is assumed that the proposed method is better to recall the query set in large training set.

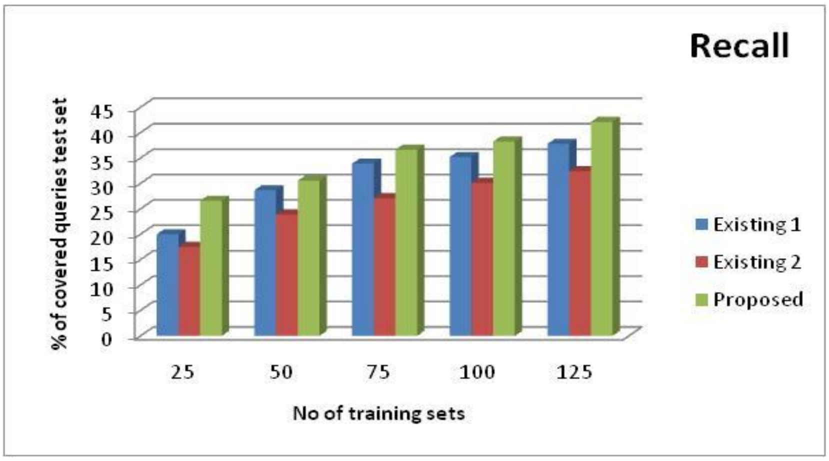

Figure 3: Recall

The graph 3 explains the recall sets of existing and proposed method. The percentage of covered queries test set and number of training sets are used to compare the capability to recall of both existing and proposed method. The recall options in existing 1 has minimum 20 query test and maximum 25th training set to 37.8 query test in 125th training set, existing 2 has minimum 17.5 query test in 25th training set and maximum 32.4 in 125th training test, but in proposed method minimum 26.6 query test in 25th training set and maximum 42.14 in 125th training set and it is assumed that the proposed method is better to recall the query set in large training set.

Dim Cache

Table 2: Comparison of Dim Cache

Existing 1	Existing 2	Proposed
6.6	10	3.5
8.56	13.67	6.9
10.21	16.9	8.1
12	17.2	10.12
15.11	19.8	13.4

The table 2 explains the comparison of dim cache usage in existing method and proposed method to specifies the intended type of a variable. The cache ratio of existing 1 is minimum 6.6 in 30th training set and maximum 15.11 in 150th training set and in existing 2 the cache ratio is minimum 10 in 30th training set to maximum 19.8 in 150th training set. The cache ratio of proposed method is minimum 3.5 30th training set to maximum 13.4 in 150th training set. This shows that the proposed method have more intended type of variable than compared to existing methods.

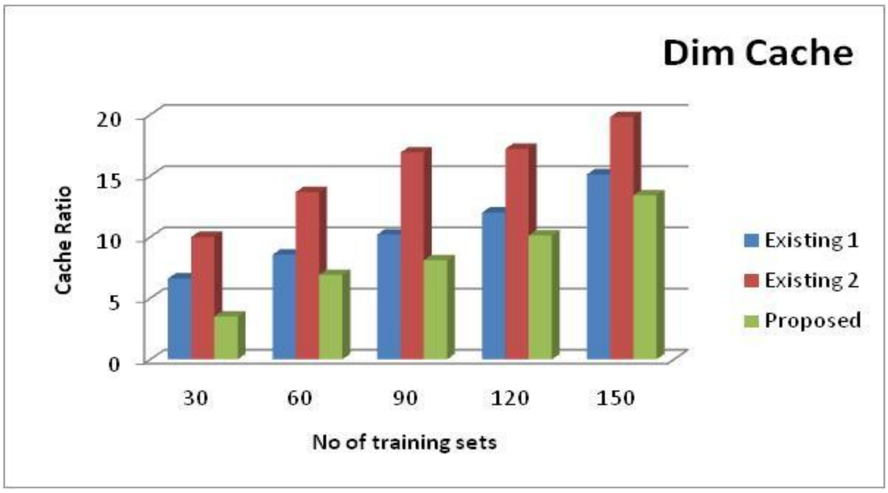

Figure 4: Dim Cache

The graph 4 explains the dim cache ratio of two existing method and a proposed method. . The cache ratio of existing 1 is minimum 6.6 in 30th training set and maximum 15.11 in 150th training set and in existing 2 the cache ratio is minimum 10 in 30th training set to maximum 19.8 in 150th training set. The cache ratio of proposed method is minimum 3.5 30th training set to maximum 13.4 in 150th training set. This shows that the proposed method have more intended type of variable than compared to existing methods.

Effective Usage of Cache

Table 3: Comparison of effective usage of cache

Existing 1	Existing 2	Proposed
3.2	2	5.2
5.7	4.9	8.1
6.8	5.7	10.3
8.2	7.1	12.4
10	9.5	13.9

The table 3 explains the comparison of effective usage of cache of existing and proposed method. In existing method 1 the percentage of cache usage is minimum 3.2 in 20th training set and maximum 10 in 100th training set and in existing method 2 the percentage of cache usage is minimum 2 in 20th training set and maximum 9.5 in 100th training set. But in proposed method the percentage of cache usage is minimum 5.2 in 20th training set and maximum 13.9 in 100th training set. It is assumed that the proposed method is effective one and stores maximum datum for future use than the other two existing methods.

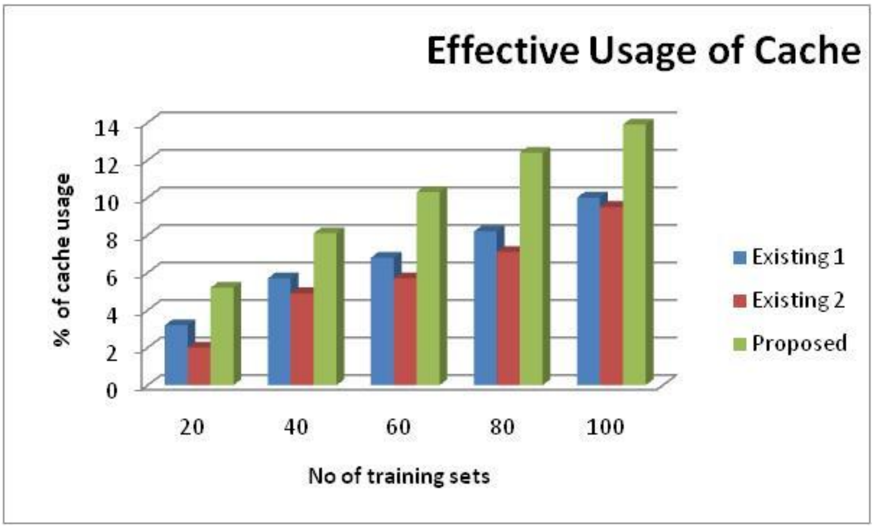

Figure 5: Effective usage of cache

The graph 5 explains the effective usage of cache in percentage and by calculating number of training sets. It is shown that in existing method 1 the percentage of cache usage is minimum 3.2 in 20th training set and maximum 10 in 100th training set and in existing method 2 the percentage of cache usage is minimum 2 in 20th training set and maximum 9.5 in 100th training set. But in proposed method the percentage of cache usage is minimum 5.2 in 20th training set and maximum 13.9 in 100th training set. It is assumed that the proposed method is effective one and stores maximum datum for future use than the other two existing methods.

Precision

Table 4: Comparison table of Precision

Existing 1	Existing 2	Proposed
5.2	3.5	6.6
8.1	6.9	8.56
10.3	8.1	10.21
11.4	10.12	12
13.9	13.4	15.11

The table 4 explains the comparison table of precision of two existing method and one proposed method. In existing 1 the accuracy of data is minimum 5.2 in 25th dataset and maximum 13.9 in 125th data set and in existing 2 the presicion ratio of data is minimum 3.5 in 25th data set and maximum 13.4 in 125th data set. The precision ratio of data set is minimum 6.6 in 25th data set and maximum 15.11 in 125th data set. It is assumed that the data accuracy ratio is better in proposed method than existing method.

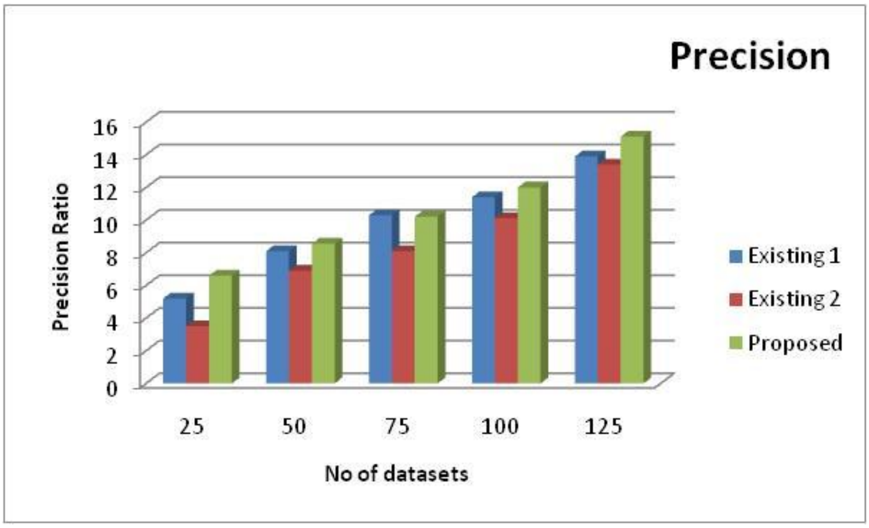

Figure 6: Comparison graph of Precision

The graph 6 explains the comparison of better precision of two existing methods and one proposed method using the precision ratio and by number of datasets. In existing 1 the accuracy of data is minimum 5.2 in 25th dataset and maximum 13.9 in 125th data set and in existing 2 the presicion ratio of data is minimum 3.5 in 25th data set and maximum 13.4 in 125th data set. The precision ratio of data set is minimum 6.6 in 25th data set and maximum 15.11 in 125th data set. It is assumed that the data accuracy ratio is better in proposed method than existing method.

Accuracy

Table 5: Comparison table of accuracy

Existing 1	Existing 2	Proposed
3.5	5.2	10
6.9	8.1	13.67
8.1	10.3	16.9
10.12	12.4	17.2
13.4	13.9	19.8

The table 5 explains the comparison table of accuracy of two existing method and one proposed method. In existing 1 the accuracy ratio is minimum 3.5 in 20th datatset and maximum 13.4 in 100th data set. In existing 2 the accuracy ratio is minimum 5.2 in 20th dataset and maximum 13.9 in 100th data set. In proposed method the accuracy ratio is minimum 10 in 20th data set and maximum 19.8 in 100th data set. It is assumed that the accuracy ratio of proposed method is better when compared to existing methods.

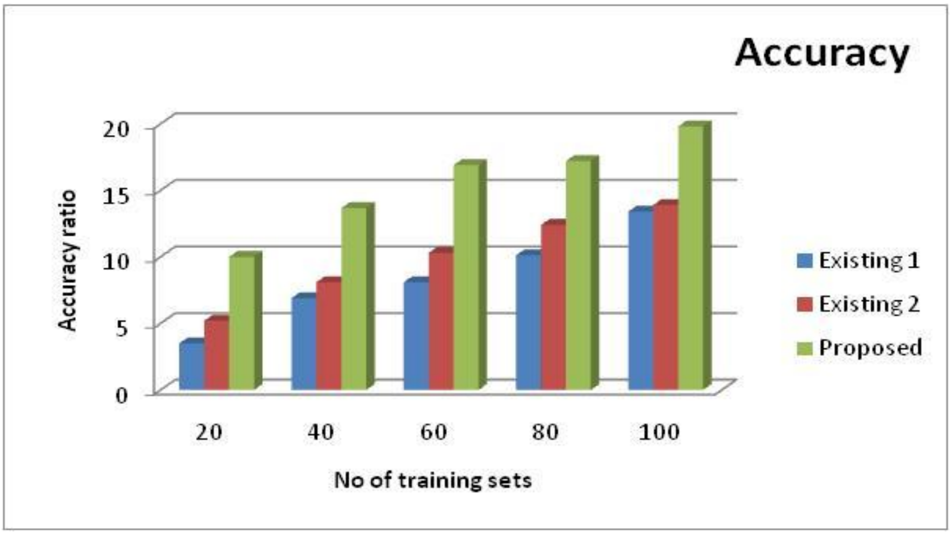

Figure 7: Comparison graph of accuracy

The graph 5.5 explains the comparison table of accuracy of existing methods and proposed methods. In existing 1 the accuracy ratio is minimum 3.5 in 20th datatset and maximum 13.4 in 100th data set. In existing 2 the accuracy ratio is minimum 5.2 in 20th dataset and maximum 13.9 in 100th data set. In proposed method the accuracy ratio is minimum 10 in 20th data set and maximum 19.8 in 100th data set. It is assumed that the accuracy ratio of proposed method is better when compared to existing methods.

Robustness

Table 6: Comparison of robustness

Existing 1	Existing 2	Proposed
17.5	10	26.6
23.9	13.67	30.56
27.1	16.9	36.66
30.12	17.2	38.26
32.4	19.8	42.14

The table 6 explains the comparison of robustness in existing method and proposed method. In existing 1 the ratio of robustness is 32.4 maximum and in existing method 2 the ratio of robustness is 19.8 maximum. In proposed method the ratio of robustness is 42.14 maximum. It is assumed that the proposed method has better robustness to control the data than the two existing methods.

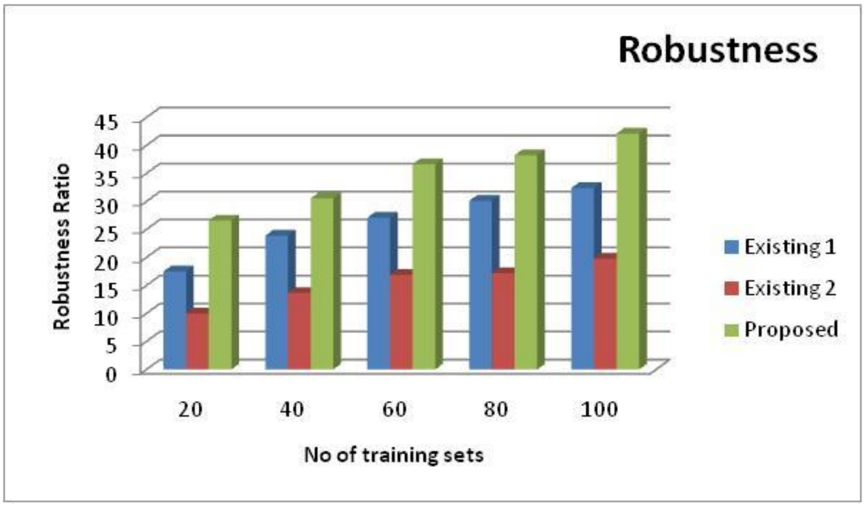

Figure 8: Comparison chart of robustness

The graph 8 explains the comparison chart of robustness of two existing methods and a proposed method. In existing 1 the ratio of robustness is 32.4 maximum and in existing method 2 the ratio of robustness is 19.8 maximum. In proposed method the ratio of robustness is 42.14 maximum. It is assumed that the proposed method has better robustness to control the data than the two existing methods.

CONCLUSION

We have presented the issue of stochastic question covering, which is of potential enthusiasm to the motivation behind effectively and efficiently storing archives showing up in the outcome rundown of numerous inquiries. We have introduced covetous multi spread calculations, which we have broke down in a stochastic framework, and for which we have given guess ensures. We have demonstrated that our methodology permits consolidating distinctive advancement criteria by reasonably characterizing loads. Specifically, we have portrayed two distinctive weighting schemes to catch the archive importance and to consider expansion of stored reports. In the two cases the superiority of the covetous calculations has been affirmed by the trial results.

REFERENCES

[1] Debahuti Mishra and Niharika Pujari, “Cross-Domain Query Answering: Using Web Scrapper and Data Integration”, 978-1 -4577-1386-611 $26.00©2011 IEEE.

[2] Chichang Jou, Yucheng Cheng, “Heuristics-Based Schema Extraction for Deep Web Query Interfaces”, 978-0-7695-6243-8/17 $31.00 © 2017 IEEE.

[3] Kuldeep R. Kurte, Surya S. Durbha, Roger L. King, Nicolas H. Younan, and Rangaraju Vatsavai, “Semantics-Enabled Framework for Spatial Image Information Mining of Linked Earth Observation Data”, 1939-1404 © 2016 IEEE.

[4] Rabah A. Al-Zaidy, C. Lee Giles, “Extracting Semantic Relations for Scholarly Knowledge Base Construction”, 0-7695-6360-0/18/$31.00 ©2018 IEEE.

[5] Ali Jalali Santanu Kolay Peter Foldes Ali Dasdan, “Scalable Audience Reach Estimation in Real-time Online Advertising”, 978-0-7695-5109-8/13 $31.00 © 2013 IEEE.

[6] Niharika Pujari, Debahuti Mishra, Kaberi Das: Query Reformulation, Data Integration to Multi Domain Query Answering System. Int. J. of Computer and Communication Technology, 2(1), 2010.

[7] Furche, T., Gottlob, G., Grasso, G. et al. (2013). The ontological key: automatically understanding and integrating forms to access the deep Web. The VLDB Journal, 22(5), pp. 615-640.

[8] Saissi, Y., Zellou, A., and Idri, A. (2016). Towards XML schema extraction from deep web. Proceedings of 4th IEEE International Colloquium on Information Science and Technology, pp. 94-99.

[9] Su, W., Wu, H., Li, Y., Zhao, J., Lochovsky, F.H., Cai, H., and Huang, T.

(2013). Understanding Query Interfaces by Statistical Parsing. ACM Transactions on the Web, 7(2), Article No. 8.

[10] Yu, H., and Ye, F. (2015). Research on Extract the Schema of Query Interfaces. Proceedings of the 10th International Conference on Intelligent Systems and Knowledge Engineering, pp 442-447.

[11] M. Koubarakis et al., "Building virtual earth observatories using ontologies and linked geospatial data," in Web Reasoning and Rule Systems. New York, NY, USA: Springer, 2012, pp. 229–233.

[12] F. Wan and F. Deng, "Remote sensing image segmentation using mean shiftmethod," in Advanced Research on Computer Education, Simulation and Modeling, New York, NY, USA: Springer, 2011, pp. 86–90.

[13] J. Lee, W.-S. Han, R. Kasperovics, and J.-H. Lee, "An in-depth comparison of subgraph isomorphism algorithms in graph databases," in Proc. VLDBEndowment,RivadelGarda,Trento,Italy,Dec.2012,vol.6,no.2, pp. 133–144.