

Review on Advanced Prediction of Difficult Keyword Queries Over Databases

Sruthi S.¹ Swathi Hariprasad² Dhanya P.³ Sruthi Muralidharn⁴
^{1,2,3,4}B.Tech Student

^{1,2,3,4}Department of Computer Science & Engineering
^{1,2,3,4}Nehru College of Engineering and Research Centre Thrissur, India

Abstract— The discovery of patterns from a large data set is known as data mining. In this, information is being tapped from a big collection of data and is being transferred to other structure that will be familiar to the users. These days, many organizations are seen to use the keyword search more often than before. Keyword queries over databases provide effortless reach to data and information. Keyword Query Interface is being used to furnish resilience and effortless access to complete the search of data. Goal of this paper is to prove that the system of predicting the keyword queries on databases is adequate than the existing systems for which a survey was done. Based on the selected parameters viz, speed, flexibility, robustness, noise, accuracy few papers were surveyed and outputs were recorded. The papers taken for surveying are studied thoroughly and a comparison was done on the basis of parameters chosen.

Key words: Advanced Prediction, Difficult Keyword Queries over Databases

I. INTRODUCTION

Anticipating the difficulty of keyword query plays a crucial role in retrieval of information. Users issue queries in response to which the interface gives a list of answers that are ranked. If the answer doesn't satisfy then user is requested to rephrase the query in order to enhance the search. Data extracting can return the use of computerization on accessible software and hardware platforms which can be implemented on systems. When business data were first safely kept on computers this evolution began consigning with improvements in data's right of entry, more newly generated technologies that makes users able to find the way by which the data is valid in time. Knowledge extraction or data mining is basically meant for applications in the community of business. The most commonly and widely used form of querying is by using keywords. Keyword queries can be used to easily access the data from the database. In Information Retrieval (IR), a useful variety is the query performance prediction. Keyword Query Interfaces (KQIs) for the databases have afflicted on the consolidation over the past decade bounding to their flexibility of having used to analyze the data.

These must at first mark the information requirements by the keyword queries and then rank the answer set by which the user is returned with preferred answers with the most relevant one on the top of the list.

Here in this paper, a few papers are studied in detail with respect to their methodologies used and are compared based on the common parameters. The conclusions are recorded and tabulated. Parameters include speed, flexibility, robustness, noise and accuracy. Each paper is reviewed in detail to prove that the system of anticipating the keyword queries over database is more efficient than the existing systems. The parameters, 'speed'

says how fast each of the systems taken under review performs compared to the proposed system, 'flexibility', is how well can the system perform with varied kind of queries, and 'robustness' tells about strength of each system. The parameter 'noise' tells about well the system can perform with presence of some noise in the data, finally 'accuracy', says how precise the answer set after applying different methods in the papers under review.

The rest of the paper is organized as follows. In Section II, the literature survey is introduced. The various paper taken for the review are briefly presented followed by a comparison table. Finally conclusion is given in the Section III.

II. LITERATURE REVIEW

Literature review is a process of gathering information from other sources and documenting it. It is going into the depth of the literature surveyed. It is a process of re-examining, evaluating or assessing the short-listed literature. Review of literature gives a clarity & better understanding of the research or project. A literature review can be a precursor in the introduction of a research paper, or it can be an entire paper in itself, often the first stage of large research projects. It is a summary and synopsis of a particular area of research, allowing anybody reading the paper to establish why you are pursuing this particular research program. A good literature review expands upon the reasons behind selecting a particular research question.

A. Anticipating Query Proficiency

A mechanism for anticipating query performance is by computing the entropy between a query and corresponding language model [1]. The coherence of the usage in documents are measured by the resulting clarity score. This measures the ambiguity with respect to documents and shows the average precision. Thus to identify ineffective queries, on averages, without relevance information clarity score may be used. An algorithm for automatically setting the clarity score threshold has been developed. An important issue in information retrieval is how well the poorly performing queries are dealt with. If a query of highly ranked document is just about a single topic it will have a model characterized by unusually large probabilities for a relatively smaller number of topical terms. On the contrary, when a query returns a mix of articles on different topics will have a model that will be smoother and more likely the collection as a whole. Therefore a query that has a high coherence will get a high score whereas the one with lower coherence will have a low score. Thereby we can say that there is a possibility to predict the performance of a query to an extent without relevance information.

B. A Probabilistic Extraction of Semi Structured Data

Either a structured query such as XPath or a keyword query that doesn't structure into account is required to typically extract the structured data [2]. This mapping probability is being used as a means to combine the language models. Although, there is huge amount of unstructured text data on web, the structured databases support most of the web services. XML being the popular one as the format for semi structured data. Hence for many applications the effective retrieval of both structured and semi structured data poses significant search issues. Retrieval of semi structured data in the form of XML often has considerable overlap in the content and use of element. Users entering queries specifications into multiple fields lead to search interfaces. In such a case, providing structural information can be more effective. From the user point of view it would be a burden as they are willing to express the queries in the simplest form. The observation shows that the "advanced" search functions are not used by most users. Here it is focused to move the burden of specifying the query for semi structured from users to the system. This is done by making the system "guess" the structured information which is implicit in the keywords.

C. Review on Enhanced Anticipation of Difficult Keyword Queries over Database

Keyword query interface on database give effortless entry to facts, but experience moderate sorting standard that is, less accuracy [3]. It is useful to identify questions that are acceptable to have moderate sorting standard to exceed the user fulfilment. The main objective of this material is to forecast the features of callous queries and introduce a novel structure to estimate the amplitude complications for the keyword queries over the databases and it is granted for twain design content of the database and the by-products of query. In addition, the material presents, a current uptake,

which is nothing but continuing the synopsis in which, before determining the structural robustness score, the K-means congregation to partition the input dataset into a number of congregates those having authorized statistics are applied.

D. Composite Keyword Queries Exploration and Prediction over Databases

Users query in search agents using keyword and it extensively utilizes the anatomy of querying in the answer to a user requirement (or) query the test of a data recouping structure is to recover useful data from a broad storehouse of data[4]. Explore agents generally do keyword approximation only; whereas the query term is present in the record then it is collected and presented to the user as a by-product of a given query. If a user query is common and debatable, the query agents discover the query as a confused query as it cannot clip the conscionable result which the user believes. Explore agents do not recognize the needs behind the query therefore; ultimate of the pertinent report or record are not recovered. The complication of this task will be influenced by diverse elements, associated to the structure or algorithm used, to the premises of the information to be recouped or to the characteristics complication of the user's data need. The effect of these components upon retrieval performance is often mentioned to as query complexity or drawback and the query is said to be composite or complex query examine and scan about the powerful query recouping and forecast over the database is done. The query engine should recognize and inspect the craved attributes analogous with one by one word given in the query. Thus, it is vital to distinguish such queries that are responsible to have moderate location standard with a certain and objective to magnify the user satisfaction level. The search engine or explore agents can also help to the user to filter their query by appending more accurate keyword to it.

Survey papers	Speed	Flexibility	Robustness	Noise	Accuracy
Anticipating Query Proficiency	High	Moderate	Relatively Stable	Lesser	High
A Probabilistic Extraction of Semi Structured Data	High	Moderate	High	Moderate	High
Review on Enhanced Anticipation of Difficult Keyword Queries Over Database	High	Low	Moderate	Moderate	Moderate
Composite Keyword Queries Exploration and Prediction over Databases	High	High	Moderate	Less	High

Table 1: Comparison Table

III. CONCLUSION

This paper focuses on studying in detail the various algorithms proposed to prove that the system of anticipating the keyword queries over database is more efficient than the existing systems. For the same, a few papers were studied in detail with respect to their methodologies used and compared against the common parameters. The conclusions recorded and tabulated are shown. From all the study conducted, it is evident that the system of predicting the hard keyword queries over the database is efficient than any other previously proposed systems.

REFERENCES

- [1] J. Han, M. Kamber, and J. Pei, "Data Mining: Concepts and Techniques". San Francisco, CA: Morgan Kaufmann, 2011.
- [2] Mr Sameer P Mamadapure, Prof. P.D.Lambhate, "Survey On Optimized Preditions Of Hard Keyword Queries Over Databases", International Journal of Computer Engineering and Applications, Volume IX, Issue VI, June 2015
- [3] Nikita D Pattanshetti, Prof. Savitha Bakare, "Complex Keyword Queries Search And Prediction Over Databases", International Journal of Advance Research In Science And Engineering, IJARSE, Vol. No.4, Special Issue(01), May 2015

- [4] Nikita D Pattanshetti, Prof. Savitha Bakare, "Complex Keyword Queries Search And Prediction Over Databases", International Journal of Advance Research In Science And Engineering, IJARSE, Vol. No.4, Special Issue(01),May 2015.

