International Research Journal of Education and Technology
Peer Reviewed Journal
ISSN 2581-7795

Fetching Secured Information using Fuzzy Set Searching Algorithm

SHARMA RAHUL, Lokendra singh Songare

PG Scholar, CSED, Dr. APJ Abdul Kalam University Indore, M.P., India

Assistant Professor, CSED, Dr. APJ Abdul Kalam University Indore, M.P., India

ABSTRACT

Instant searching technique finds answers to a query instantly when user types in keywords character-by-character. Fuzzy searching is advancement in instant searching which finds perfect match keywords of query keywords. User expects fasts results within few milliseconds and perfect match. This is the main computational challenge results in the high-speed requirement, i.e., each query needs to be answered within milliseconds to achieve an instant response and a high query throughput.

In this research a fuzzy search is carried by doing ranking to obtain results in efficient time and more accurate. Number of studies have been done like computing all answers which is slow and requires more space. Early termination technique may solve the above problem up to some extent. An approach that focuses on common phrases in the data and queries, so in proposed searching is carried on synonymous of keywords to find relevant answers by using edit distance function.

Keywords: Instant search, Fuzzy search, Auto completion, Synonymous words, Edit distance, tree index

INTRODUCTION

Computers and computer networks are being used in each and every field. Information storing on large capacity is possible. Network provides resource sharing feature. Information is stored on various storage devices at different sites. As we store large amount of data it is necessary to have techniques to find out the required information from these storage devices/databases. Information searching must be fast and exact one. The user expects results as soon as he requests a query. Instant search is a system which finds answers to a query immediately while a user types in

keywords character-by-character. Fuzzy improves user search experiences by finding answers with keywords similar to query keywords. A main thing in this that each query needs to be answered fast within less time to achieve an instant response.

Instant Search:

Many recent studies have been focused on instant search, also known as type- ahead search. The studies proposed indexing and query techniques to sup-port instant search. It uses trie based techniques to do instant search.

Fuzzy Search:

The studies on fuzzy search can be classified into two categories, gram-based approaches and trie-based approaches. In the former approach, substrings of the data are used for fuzzy string matching . The second class of approaches indexes the keywords as a trie, and relies on a traversal on the trie to find similar keywords. This approach is especially suitable for instant and fuzzy search. Auto Completion:

In auto-completion, the system suggests several possible queries the user may type in next. There have been many studies on predicting queries. Many systems do prediction by treating a query with multiple keywords as a single pre x string.

Early Termination:

Early-termination techniques have been studied extensively to support top-k queries efficiently. Adopted existing top-k algorithms to do instant-fuzzy search [4] Most of these studies reorganize an inverted index to evaluate more relevant documents FIRST. Persin et al proposed using inverted lists sorted by decreasing document frequency. Hwang et al [5] the effect of term independent features in index reorganization.

Edit Distance:

The similarity between two keywords can be measured using various metrics such as edit distance, Jaccard similarity, and cosine similarity. In this work the focus is on the commonly used function, edit distance. The edit distance between two strings is the

minimum number of single-character operations (insertion, deletion, and substitution) to transform one string to the other. For example, the edit distance between the keywords “Rin” and “Tin" is 1, because the former can be transformed to the latter by substituting the character “R" with “T”. Let ed(wi,p) be the edit distance between a query keyword wi and a prex p from a record, and be a threshold. We say p is similar to wi if ed (wi,p) . Our techniques can be extended to other variants of the edit distance function, such as a function that al-lows a swap operation between two characters, a function that uses different costs for different edit operations, and a function that considers a normalized threshold based on the string lengths.

Ranking:

Ranking is used to rank documents based on distance between query key-words. The entire set of documents which have all query keywords or words similar to query keywords are considered for ranking. If all words are in same bin or adjacent bin then the document is ranked higher else the document is ranked lower.

OBJECTIVE OF THE WORK

It can be achieved by implementing instant search, fuzzy search, ranking and fat finger problem by using edit distance method. Many recent studies have been focused on instant search, also known as type-ahead search.

The similarity between two keywords can be measured using various metrics such as edit distance, Jaccard similarity, and cosine similarity. The edit distance between two strings is the minimum number of single-character operations (insertion, deletion, and substitution) to transform one string to the other. Ranking is used to rank documents based on distance between query keywords. The entire set of documents which have all query keywords or words similar to query keywords are considered for ranking. If all words are in same bin or adjacent bin then the document is ranked higher else the document is ranked lower.

International Research Journal of Education and Technology
Peer Reviewed Journal
ISSN 2581-7795

Result:-

1. Precision Graph

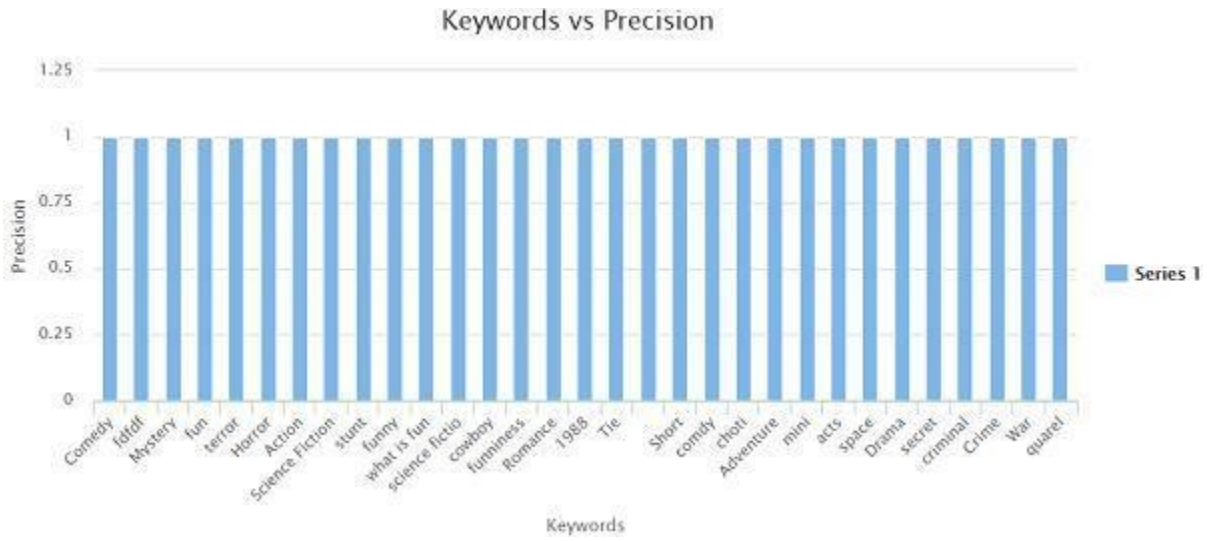

Figure 1: Precision Graph

2. Recall Graph

International Research Journal of Education and Technology
Peer Reviewed Journal
ISSN 2581-7795

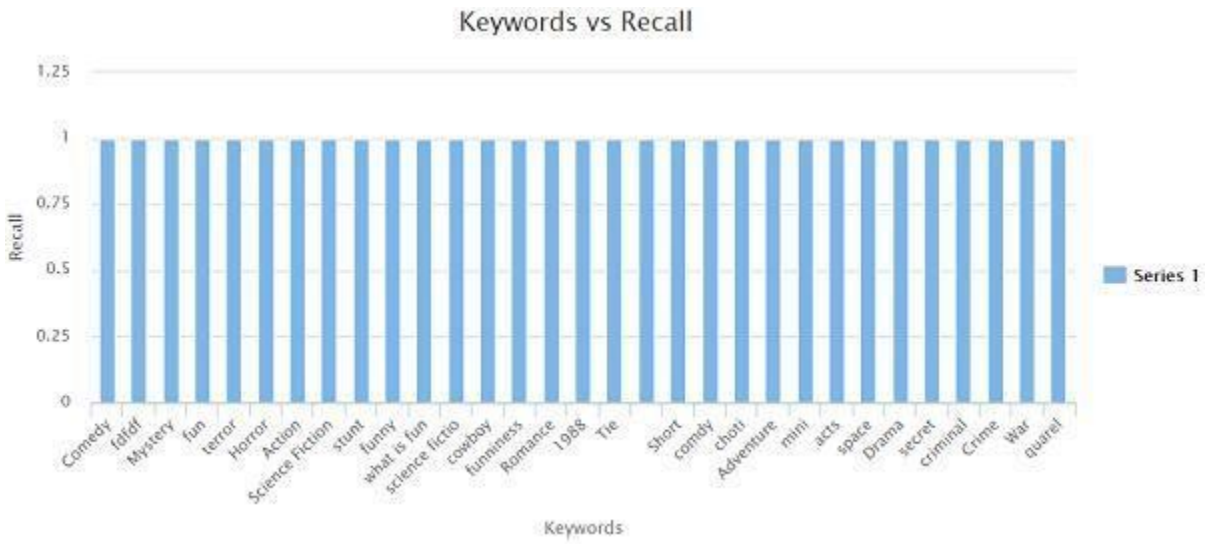

Figure 2: Recall Graph

3. Time Graph

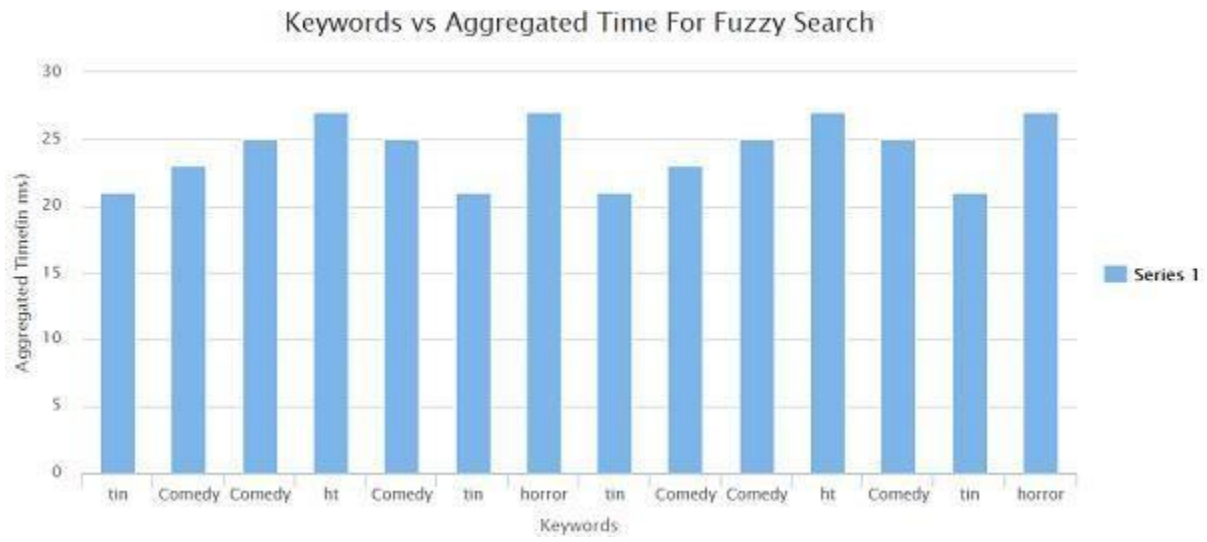

Figure 3 Time Graph

Conclusion-

A very useful application has been studied in accordance of all aspects and results have been studied. This approach retrieves relevant results even if there are few typographical errors in the query keywords. Edit distance is used in fuzzy search. Ranking makes use of modified inverted list by storing word positions in the form of bin Id. There are other advanced search features like page ranking, segmentation etc. which will be considered in future for implementation.

REFERENCES

[1] I. Cetifindil, J. Esmaelnezhad, C. Li, and Chen Li. “E cent in- stant fuzzy search with proximaty ranking", in Data Engineering (ICDE),2014 IEEE 30th International Conference on IEEE, Chicago, IL, March 31-April 4 2014, pp. 328-339.

[2] I. Cetifindil, J. Esmaelnezhad, C. Li, and D. Newman, “Analysis of instant search query logs," in 15th International Workshop on the Web and Databases (WebDB 2012), May 20, 2012 - Scottsdale, AZ, USA, pp. 7{12.

[3] C. Silvefirstein, M. R. Henzifinger, H. Marais, and M. Moricz, “Analysis of a very large web search engine query log", NewsletterACM SIGIR Forum, Vol- ume 33, Issue 1, Fall 1999, ACM New York, NY, USA SIGIR Forum, pp. 6{12.

[4] G. Li, J. Wang, C. Li, and J. Feng, “Supporting efficient top-k queries in type- ahead search", in SIGIR ’12 Proceedings of the 35th international ACM SIGIR conference on Research and development in information retrieval, Portland,Oregon, USA.,August 12-16, 2012, pp. 355{364.

[5] R. Schenkel, A. Broschart, S. won Hwang, M. Theobald, and G.Weikum, “Efficient text proximity search" , in 14th International Symposium, SPIRE 2007Santiago, Chile, October 29-31, 2007 , pp.287{ 299.

[6] H. Bast and I. Weber, “Type less, find more: fast autocompletion search with a succinct index", in SIGIR ’06 Proceedings of the 29th annual international ACM SIGIR conference on Research and development in information retrieval, Seattle, Washington, USA.,August 6-11, 2006, pp. 364{371.

[7] H. Bast and I. Weber, “ESTER: Efcient Search on Text, Entities, and Relations", in SIGIR ’07 Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval ,Amsterdam, The Netherlands, July 23-27,2007, pp.671-678.

[8] S. Ji, G. Li, C. Li, and J. Feng, “Efficient interactive fuzzy keyword search", in WWW ’09 Proceedings of the 18th international conference on World wide web,Madrid, Spain, April 20-24, 2009, pp. 371{380.

[9] i G. Li, S. Ji, C. Li, and J. Feng, “Efficient type-ahead search on relational data: a tastier approach", in SIGMOD ’09 Proceedings of the 2009 ACM SIGMOD International Conference on Management of data, Providence, Rhode Island, USA, June 29-July 02,2009, pp. 695{706.

[10] M. Hadjieleftheriou and C. Li, “Efficient approximate search on string collections", in VLDB Efindowment VLDB Efindowment , Volume 2 Issue 2, August 24-29 ,2009, pp. 1660{1661.