Ranked Keyword Search in Cloud Computing: An Innovative Approach

[1,] Vimmi Makkar [2,] Sandeep Dalal
[1,] *(M.Tech)* [2],*(Assistant professor)*
[1,2,] *Department of Computer Science and Applications,*
M.D. University, Rohtak, Haryana, India

ABSTRACT
Cloud computing has become an integral part of IT industry. Amount of information available on World Wide Web is increasing at an exponential pace. In such a vast collection it becomes difficult for the user to query something out of the whole collection. Great efforts have been made for facilitating users via keyword search. However, there are a few researches about entertaining the exact user query and presenting a ranked URL list according to it. In this paper, We give an overview of our framework for keyword searching with summaries, besides we describe a ranking algorithm for ranked keyword search and their results. Keyword searches are typically done so that users can actively utilize clouds to query a collection. This paper discusses our study design on user presenting a query in a cloud.

KEYWORDS: *Cloud computing, Ranked keyword search*

I. INTRODUCTION

Cloud computing is the use of computing resources (hardware and software) that are delivered as a service over a network .It has the potential to change the IT industry. It enables cloud customers to remotely store their data into the cloud so as to enjoy the on-demand high quality application and services from a shared pool of configurable computing resources [7]. Cloud Computing is the result of evolution and adoption of all the existing technologies and paradigms. The goal of cloud computing is to allow users to take benefit from all of these technologies, without the need for deep knowledge about or expertise with each one of them. Clouds enable customers to remotely store and access their data by lowering cost of hardware ownership while providing robust and fast services [6]. As Cloud Computing becomes prevalent, sensitive information are being increasingly centralized into the cloud. In this, data owners may share their outsourced data with a large number of users, who might want to only retrieve certain specific data files they are interested in during a given session. One of the most popular ways to do so is through keyword-based search. Such keyword search technique allows users to selectively retrieve files of interest and has been widely applied in plaintext search scenarios [7]. In a cloud the service providers offer their resources as services to the general public. Public clouds offer several key benefits to service providers, including no initial capital investment on infrastructure and shifting of risks to infrastructure providers. However, public clouds lack fine-grained control over data, network and security settings, which hampers their effectiveness in many business scenarios

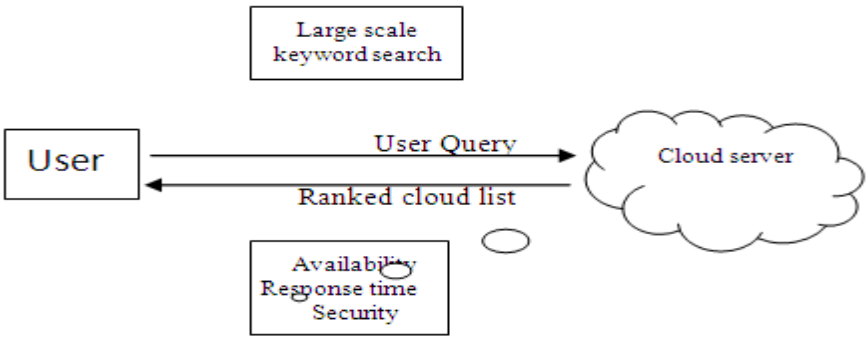


Fig.1 Architecture of keyword search

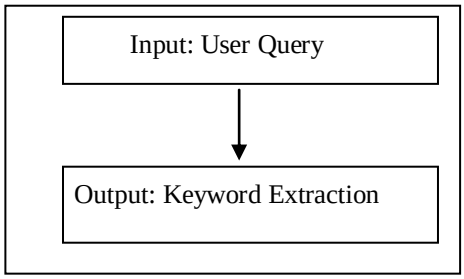
II. RELATED WORK

In this section, main research areas related to keyword search are presented. When we talk about cloud service the work is more specific and the parametric. Many researchers performed a lot of work in the same direction. In year 2007, Byron Y.L. Kuof presented a tag based summarization approach for the web search. The presented work is suggested on the public cloud. In which the integration of the web architecture and the database extraction is integrated. The work includes the refinement of the user query based on the cloud tags. The words extracted from the query are been summarized and this summarized query is passed to the public cloud. The cloud interface enabled the extraction of new and required information [1].In year 2009, Hang Guo presented personalization architecture for the cloud services. The work includes the individual access to the cloud to perform the user query. The author work is presented in two main parts one is client side and other is cloud side. The client side basically fetches the periodic information from the system where as the cloud data search engine presents the data for the modeling [2]. In year 2011, Ju-Chiang Wang presented a content oriented tag based search for the database search. In his work the music database is selected for the query analysis. The query performed by the user is analyzed and divided to different colors or the levels to perform the effective content based retrieval. The probabilistic fusion model was defined based on Gaussian mixture model and the multinomial mixture model. The author evaluated the proposed system for the effectiveness of the user query and the related results [3].In Year 2012; Cengiz Orencik presented a rank based keyword search on the data cloud. In this work the document retrieval is performed on the cloud server based on the keyword analysis and the information search is performed relative to the defined information. The presented work is performed on the encrypted data that has improved the security and the reliability of the retrieval. On this basis a secure protocol is suggested called Private Information Retrieval. The system will performed the query and present the final results on the basis of parametric ranking. The presented work is the efficient computation and communication of the requirement analysis [4]. Another cloud search is suggested by Daniel E. Rose in 2012 which is based on the information retrieval. The author presented his work on Amazon cloud service. The work is tested under different criteria such as scalability, configuration etc. The presented search reduce the barrier to allow a person or the organization to perform the content oriented search and the search is tested under the enterprises environment as well as on web search[5].

III. PROPOSED WORK

The cloud architecture is widely spread in the form of public cloud and is available to all the users. The presented work is the agent based work performed by the middle layer to perform service level agreement. In this, at first the user query will be analyzed respective to the user requirement specification. Now a parametric match will be performed based on the availability and the requirements. For this kind of match a cloud search will be performed to identify the keyword occurrence and will assign an initial rank based on the keyword match. The second generic criteria defined here is the security, the clouds that provides the communication in secure way will be assigned with some higher rank. The security will be identified based on the SSL availability and the https transmission for cloud access. The first level is completely the based on server side features. The second level of ranking will be done based on the availability and the response time factors. This layer will work for the middle layer architecture that will exist between the user and the cloud server. The middle layer will maintain a database to identify the minimum, maximum and average response time for all cloud services and based on which the rank will be affected. The parameters taken will be the user interest and the user visit over a cloud. More the user interest, higher the cloud rank will be. Finally these three levels will be collected together and will generate an effective ranking formula. As the approach is rule based, the results here are more reliable. The work will perform a segmented search that will also increase the efficiency of the search mechanism.

In this, user will interact with the web for his topic based query to retrieve the web pages. As the page is query performed it will send request to the web and generate a basic list of all the URLs related to the topic. Now it will retrieve the data from the web. For the URL collection it will use the concepts of indexing and ranking. Indexing will provide a fast access to the web page and ranking will arrange the list according to the priority. The analysis includes the keyword extraction by removal of stop words. Once the keyword is extracted the next work is to perform the keyword summarization based on frequency of keywords. Once we get the summarized keywords it will be used as the content based analysis. In this, at first the query is made by the user and on this a query analysis is performed.



Now this extracted keyword will work as input to the cloud search architecture and based on the algorithmic approach it will return the effective URL list along with ranking.

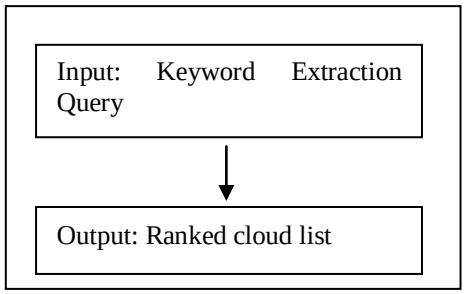


Fig 2: Basic concept of the Ranked Keyword Search

Ranking Algorithm:

[1] Define the list of available clouds on any public cloud server called Cloud(1),Cloud(2)…..Cloud(n)

[2] For i=1 to n
{Identify parameters for Cloud (i) called Availability (i), Response Time(i), Security(i) }

[3] Accept the User Query called Req under the specification ReqKeyword, ReqSecurity, ReqDeadLine,

[4] Activate the Middle layer to provide the best service selection

[5] Accept the user query and filter it to retrieve the keywords under the following step

- Remove the stop list words from the query list
- Rank the different keywords respective to category
- Find the frequency of keywords
- Keep the most occurring keywords and present as relevancy measure

[6] As the keywords retrieve perform query on each public cloud and perform the content and tag based match.

[7] Find the list of M clouds that satisfy the relevancy criteria as well as identify the other cloud parameters like response time, security measure

[8] For i=1 to M
[Perform the Content based similarity measure as]

[9] RelevencyVector = 0
For j=1 to Length (User Keywords)
{RelevencyVector=RelevencyVector + Keyword Occurrence(Clound(i) ,Keyword(j)) /Total Keywords(Clound(i),Keyword(j)); }

[10] Security Vector=0; If (UserSecurityReq=Security (Cloud(i))
Security Vector=1;

[11] ResponseTimeVector=0 If (User Deadline>Response Time(Cloud(i))

{ResponseTimeVector=UserDeadline-ResponseTime(Cloud(i) }

[12] Rank (Clound(i))= RelevencyVector*w1 +Security Vector*w2 +ResponseTimeVector*w3;

[13] As user get Ranked list of clouds, selection can be performed for best cloud service provider respective to user interest.

IV. RESULTS

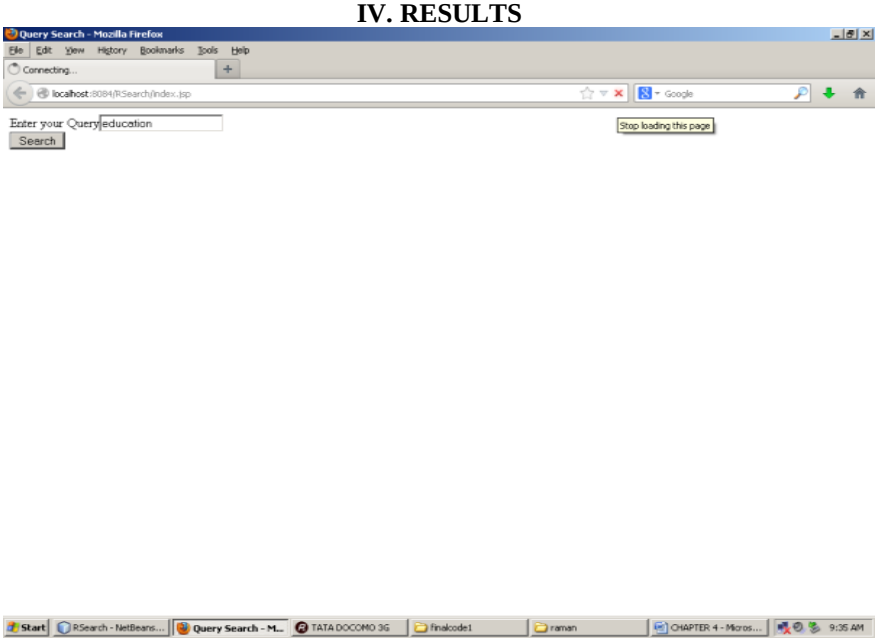


Fig. 3 Basic GUI of Proposed Work

This figure shows the basic GUI of proposed work. The application will accept the keyword from user and perform the proposed approach on it. On server side the page is processed via database to get the optimized results.

Fig 4. Output of Keyword Search (Education)

Figure 4 is showing the results driven based on the search performed for keyword Education. The output is shown in the form of crawled Cloud Service. The Cloud Services are presented with few options to show the visited count, page rank and the like/dislike count for each Cloud Service. The ranking will be changed respective to the user interest and the number of visits to different pages. If visitor count is say 4 then, it will be considered as 0.4. If page-rank is 60 then, it will be considered as 6.

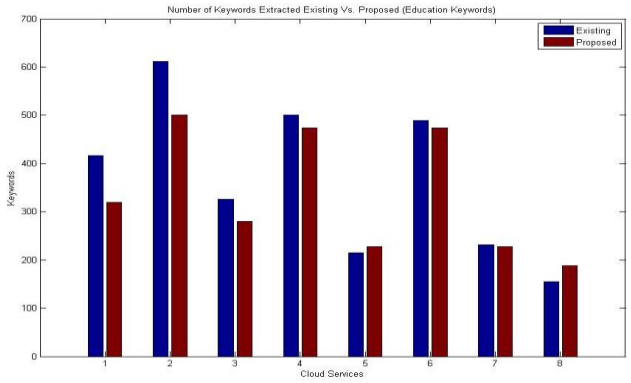


Fig. 5 Comparison of Keyword Search Education (Proposed Vs. Existing)

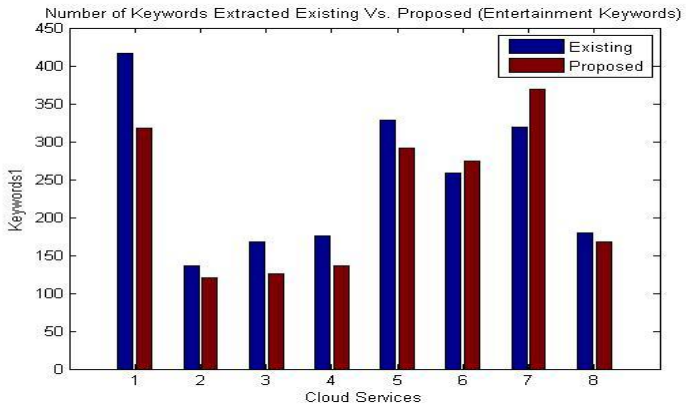


Fig.6 Comparison of Keyword Search Entertainment (Proposed Vs. Existing)

Here, the comparison is shown of the keyword analysis performed in existing work and the proposed work. The existing work represents the query performed on the cloud search without keyword extraction where as the proposed approach shows the keyword extraction after the keyword analysis. Here the outcome of the keyword analysis of education is shown. As we can see, the presented approach gives more filtered relevancy so that the comparison can be performed easily.

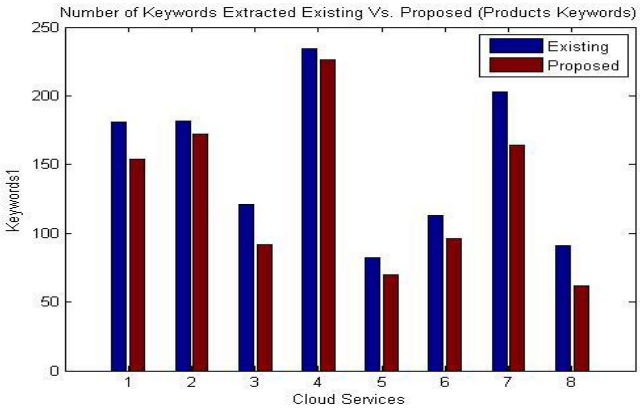


Fig.7 Comparison of Keyword Search Products (Proposed Vs. Existing)

Here, the comparison is shown for the keyword products.

V. CONCLUSION AND FUTURE SCOPE

In this paper, as an initial attempt, we motivate and solve the problem of supporting efficient ranked keyword search in a cloud. We proposed a framework that allows thinking of keyword searches more naturally. In this present work, the GUI interface will be created to pass the user query like a search engine. The first output will be drawn in terms of query filtration and extraction of keyword from query analysis. Once the keyword analysis is performed, keyword reduction will be done and finally the keywords will be drawn as output. Now, this extracted keyword will work as input to the cloud search architecture and based on algorithmic approach, it will return the effective URL list along with ranking. The proposed ranking method proves to be efficient in analyzing user query and returning highly relevant document corresponding to submitted search terms. We prove that are proposed method satisfies all the requirements of keyword analysis. Following the current research, we propose several possible directions for ranked keyword search. There is large agenda of interesting issues to be tackled. We are interested in exploring different criteria for selecting words and ranking and indexing them. The efficiency in this work is the major issue, as the repository is larger and not having the service database, the search contents takes time. In future, the work can be improved in this direction.

REFERENCES

[1] Byron Y-L. Kuo (2007),” Tag Clouds for Summarizing Web Search Results”, WWW 2007, May 8–12, 2007, Banff, Alberta, Canada.pp. 1203-1204

[2] Hang Guo (2009),” Personalization as a Service: the Architecture and a Case Study”, CloudDB’09, November 2, 2009, Hong Kong, China. Pp. 1-8

[3] Ju-Chiang Wang (2011),” Colorizing Tags in Tag Cloud: A Novel Query-by-Tag Music Search System”, MM’11, November 28–December 1, 2011, Scottsdale, Arizona, USA.ACM p 293-302

[4] Cengiz Orencik (2012) ,” Efficient and Secure Ranked Multi-Keyword Search on Encrypted Cloud Data”, PAIS 2012, March 30, 2012, Berlin, Germany. ACM, p 186-195

[5] Daniel E. Rose (2012),” Cloud Search and the Democratization of Information Retrieval”, SIGIR’12, August 12–16, 2012, Portland, Oregon, USA.PP. 1022-1023

[6] http://enWikipedia.org/wiki/ cloud _computing

[7] C.Wang , N.Cao,J .Li, K.Ren, and W.Lou. Secure ranked keyword search over encrypted cloud data. In ICDCS’10,Pages 253-262,2010.