International Journal of Emerging Technology in Computer Science & Electronics (IJETCSE)
ISSN: 0976-1353 Volume 21 Issue 1 – APRIL 2016.

EXTRACTION OF WORD FRAGMENT USING INCREMENTAL KEY INDEX FAST SEARCH (IKIFS)

V.Haripriya[#1] And Mrs Ramya Devi[*2]

[#1]*PG Scholar, Deptof CSE, Velammal Engineering College, Chennai, India.*
[*2]*Assistant Professor, Dept of CSE,Velammal Engineering College, Chennai,India.*

***Abstract*--Data Mining is mainly used for storing and retrieve needed information. Reading and summarizing the contents of large entries of text into a small set of topics is difficult and time consuming for a human, so that it becomes nearly impossible to accomplish with limited manpower as the size of the information grows.Thus evaluation of keyword search helps to retrieve the document with less time. To overcome the hurdle we proposes, Just In Time retrieval algorithm which displays the relevant document in the exact time depends on the keyword present in the document or how many times it is present in an document. The appropriate document will be retrieved from the database and document storage (such as pdf, word, images, videos, music, etc). Using the Incremental Key Index Fast Search, the documents will be retrieved faster with the key feature of the query. Filtering option is used to extract the file in a particular format, then the extracted document will be stored in Mongo database. It is a document database that provides high performance, high availability and easy scalability so that search will be fast and time consumption will be less. User Frequency Suggestions (UFS) is to retrieve a document which is searched by the user again and again that is the UFS will have a list of frequently searched document. So if the user start to type the keyword, UFS will match the keyword and it helps to extract the document soon. Thus it will not display all the documents, but based on the IKIFS and UFS algorithm, it displays the User requirement documents.**

Index terms---Extraction of word fragment, Retrieving document, Less time consumption.

I. INTRODUCTION

Data Mining is a process to turn raw data into useful information. By using software to look for patterns in large batches of data, businesses can learn more about their customers and develop more effective marketing strategies as well as increase sales and decrease costs. Data mining depends on effective data collection and warehousing as well as computer processing. Data Mining play a vital role in my project since collection of Keywords are stored in the database.

HUMANS are surrounded by lots valuable information, available as documents, databases, or multimedia resources. Access to this information is conditioned by the availability of suitable search engines, but even when these are available, users often do not initiate a search, because their current activity does not allow them to do so, or because they are not aware that relevant information is available. Inorder to provide a fast search we are using Keyword extraction technique. Keywords play a crucial role in extracting the correct information as per user requirements. Everyday thousands of books, papers are published which makes it very difficult to go through all the text material, instead there is a need of good information extraction or summarization methods which provide the actual contents of a given document. This is done by fragmenting the queries of the users to extract the recommended document. These fragments are stored in the Database to avoid the Keyword frequency is to avoid the duplication of data. Keyword extraction is not that much simple because it has to relate the keyword with the user information to retrieve the recommended document. Lots of algorithm is used to extract the document by counting its frequency, word occurrence, etc. Keywords are index terms that contain most important information. Automatic keyword extraction is the task to identify a small set of words, key phrases or keywords from a document that can describe the meaning of document. Keyword extraction is considered as core technology of all automatic processing for text materials.

Now a days Keywords are extracted based on the expressions posed by human activities (like facial expressions, gaze, and their participations among groups). Thus Technology has improved a lot in extracting a keyword for searching purpose. To display the corresponding files to the user helps them to analysis the evaluation of time they spend in searching. Thus by using Keyword search even web searches becomes familiar to get the appropriate links and results for their searches. Instead of textual way of searching Keywords, advance way have implemented to search the Keyword by using voice.

II. LITERATURE SURVEY

There are various methods for locating and defining keywords have been used both individually and in concert. Despite their

differences, most methods have the same purpose and attempt to do the same thing: using some heuristic (such as distance between words, frequency of word use, or predetermined word relationships), locate and define a set of words that accurately convey themes or describe information contained in the text. Certain methods are mentioned as follows

- **Word Frequency Analysis**
- **Word Co-Occurrence Relationships**
- **Using a Document Corpus**
- **Frequency-Based Single Document Keyword Extraction**
- **Content-Sensitive Single Document Keyword Extraction**
- **Keyword Extraction Using Lexical Chains**
- **Key phrase Extraction Using Bayes Classifier**

Here we are going to discuss about the survey made to extract Keyword which relate to the corresponding document.

A. Just In Time Retrieval

In this paper reviews about the Just In Time Retrieval system(JIT)[1] which helps to extract the relevant document according to user information.

B. Implicit Queries

In this reviews about the implicit queries based on the context of the user information. The Implicit Query (IQ) prototype [2] automatically generates queries based on user activity, and presents results in the context of ongoing work.

C. Collaborative Query Reformation

Query Reformation [3] is discussed basedon users reformulate or modify the queries when they engage in searching information particularly when the search task is complex and exploratory.

D. Small Groups

This illustrates about the emergent leaders among the group. The emergent leaders [4] are the member of the team, based on their behavior and equal participants in the groups their leadership quality rise.

E. Indefinite Ranking

This reviews about the rank list that is encountered during the searches and discuss about priority rank [5] given by the user while searching. A measure of the similarity between incomplete rankings should handle non-conjointness, weight high ranks more heavily than low, and be monotonic with increasing depth of evaluation.

F. Learning by Reflect

The table designed in the way to reflect [6[, the issue of unbalanced participation during group discussions. Its displaying on its surface, a shared visualization of member participation, Reflect, is meant to encourage participants to avoid the extremes such as over and under particiation.Reflect leads to more balanced collaboration, but only under certain conditions. Reflectis designed tosupport collaboration between small groups.

G.Temporal Templates

A novel representation and recognition technique for identifying movements. The approach is based upon temporal templates [7] and their dynamic matching in time. It develops a recognition method matching temporal templates against stored instances of views of known actions.The method automatically performs temporal segmentation, is invariant to linear changes in speed, and runs in real-time on standard platforms.

H. Conversational Patterns

This address the problem of nonverbal cues extracted from conversational patterns. The proposed bag of group NVPs [8] allows fusion of individual cues and facilitates the eventual comparison of groups of varying sizes. Analyzation of group based on conversational patterns are derived only from the audio modality, the bag approach can be extended to include multimodal features e.g combining prosodic cues andvisual attention-based cues, among others.

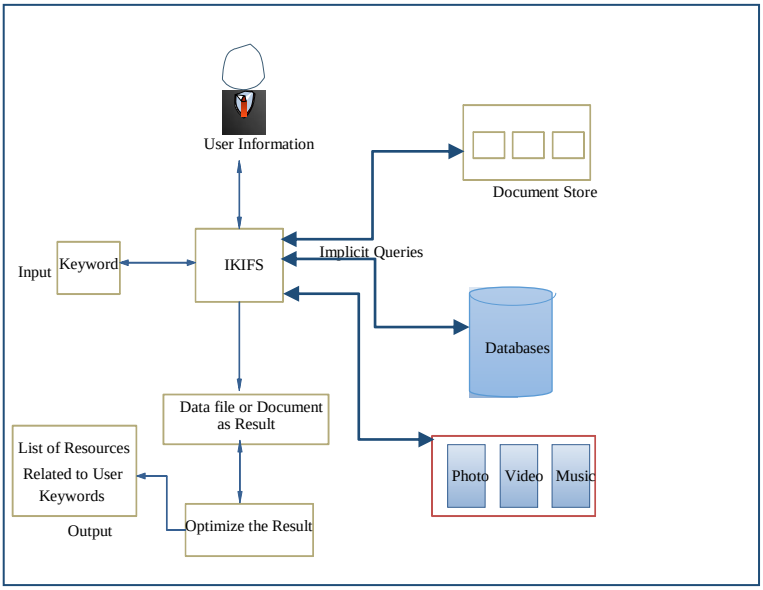
I. Meeting Segmentation

Multiparty meetings are a ubiquitous feature of organizations, and there are considerable economic benefits that would arise from their automatic analysis and structuring. In this paper [9], we are concerned with the segmentation and structuring of meetings (recorded using multiple cameras and microphones) into sequences of group meeting actions such as monologue, discussion and presentation.

III. PROPOSED SYSTEM

Keyword Extraction means extracting the keyword from the implicit queries of the user. Extracting Keywords from the document helps the user to browse in fast manner, where time consumption is reduced. Keyword Extraction widely used many places like organization, college database, mail server, group discussion and internet etc. Thus extracting of keywords in effective way using voice search is the proposed system being discussed. Various algorithms like JIT (Just in Time), IKIFS (Incremental key index fast search), UFS (User frequency search) are used to describe the extraction of keywords for the web based searches.

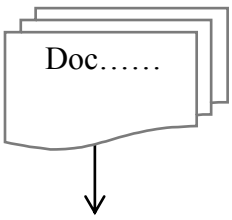
IV. ARCHITECTURE



The working principle of the system is described with the help of the architecture diagram. To initialize the extraction process the system needs to study about the user's context and their information. From where keywords are extracted and searches for recommended document. Keyword extraction is considered as core technology of all automatic processing for text materials. Humans are surrounded by lots valuable information, available as documents, databases, or multimedia resources. Access to this information is conditioned by the availability of suitable search engines, but even when these are available, users often do not initiate a search, because their current activity does not allow them to do so, or because they are not aware that relevant information is available.

In order to provide a fast search we are using Keyword extraction technique.This explains about the extraction of keywords and relating it to the corresponding document need for the user to download and helps to retrieve the relevant document in fast manner and in less time consumption. The working of the system using IKIFS(Incremental Key Index Fast Search) and UFS (User Frequency Suggestion) to derive a documents based upon the user requirements.

A. Flow Diagram



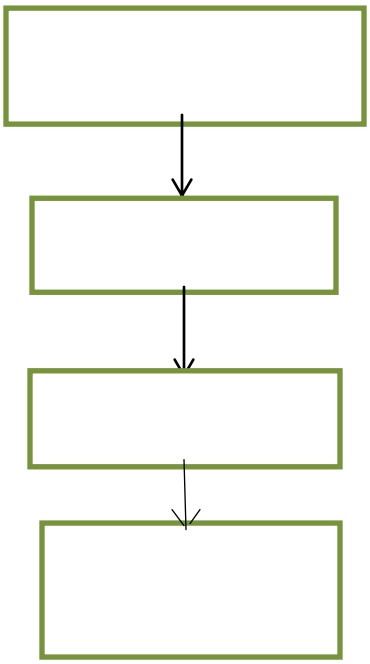


Fig1- Flow Diagram for keyword extraction

This diagram [14] depicts the work flow of keyword extraction. To initialize the extraction process the system needs to study about the user's context and their information. From where keywords are extracted and searches for recommended document. Keywords are index terms that contain most important information. Automatic keyword extraction is the task to identify a small set of words, key phrases or keywords from a document that can describe the meaning of document. Keyword extraction is considered as core technology of all automatic processing for text materials. A Survey of Keyword Extraction techniques have been presented that can be applied to extract effective keywords that uniquely identify a document. After theextraction process Clustering and Ranking of keywords is used to discover the required documents. By placing the key phrase the user can extract the document easily. Thus by using keyword extraction the retrieval of required documents is done within less time consumption.

V. ALGORITHM

Various algorithm are used to extract the keywords in effective manner, but each methods faces certain failure in extracting keywords. Now we are using new algorithms like JIT, IKIFS, UFS.

JIT is an algorithm to extract keywords from Retrieval system (or a manual transcript for testing),which helps to extract the relevant document according to user information. Based on the context of the user Keywords are extracted to retrieve the recommended document. Example of JITIR system is

handling a speech to a group, sharing ideas by lecture on specific topic.Theory and design lessons learned from these implementations are presented, drawing from behavioral psychology, information retrieval, and interface design. They are followed by evaluations and experimental results.Which displays whether the keyword is present in a document and how many times it is present in a document, and it shows the position of the keyword. The key lesson is that the users of JITIR agents are not merely more efficient at retrieving information, but actually retrieve and use more information than they would with traditional search engines.

Then a method IKIFS is to extract the documents faster with the key feature of the query based upon the filtering option, not all the documents will be derived only the user required document will be extracted. Filtering option is used to extract the file in a particular format, then the extracted documents will be stored in database. So that search will be fast and time consumption will be less.UFS is to retrieve a document which is searched by the user again and again that is the UFS will have a list of frequently searched document so if the user start to type the keyword, UFS will match the keyword and it helps to extract the document soon. In order to avoid the duplication of files storage, a key feature is set, to store in the database.

So that when the keywords are found to occur it will relate to the key and displays the need document. By optimizing the result the need document is retrieved to the user.

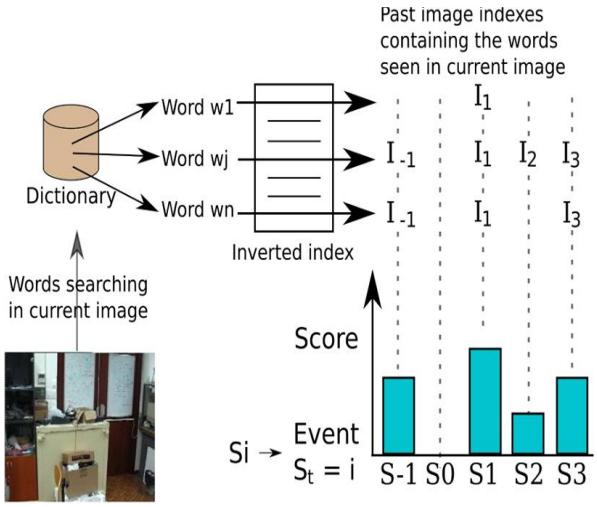


Fig2- Example of IKIFS

From this diagram, it's reviewed that by reading the information from the image and collecting the words in the dictionary. Where the words are placed in the inverted index (in secure way by placing keys to each words for future reference).The score event is measured for relating the history of the image to the current feature and the words are used. Likewise using IKIFS extracting the document to relate user needs.

WORKING OF IKIFS

- IKIFS(Incremental Key Index Fast Search) derive a documents based upon the user requirement
- Selection of keyword and relating to the document should be fast
- After keyword extraction, IKIFS assigns key feature for each query and to their related documents
- Key feature includes any characters, alphabets, etc.

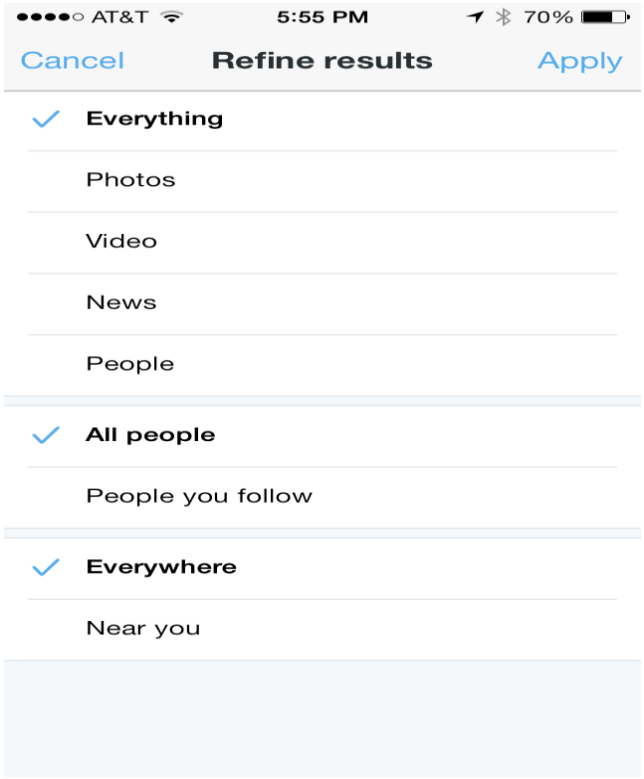


Fig3- Example of UFS

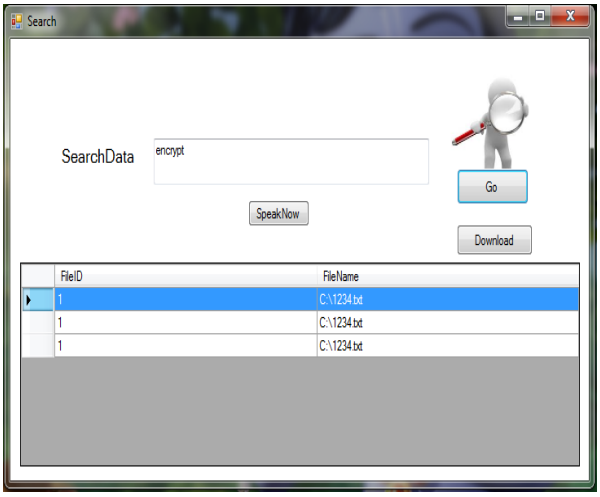
From this diagram, it'sreviewed that based on the user suggestions the documents that are extracted are relatedto the information based on the user. This shows the users information is more needed to extract the document.

WORKING OF UFS

- UFS(User Frequency Suggestions) is to retrieve a document which is searched by the user again

- UFS will have a list of frequently searched document
- UFS will match the keyword and it helps to extract the document soon
- Priority rank searching

VII. EXPERIMENTAL RESULTS



The system focus on the extracting a keywords, cluster them into topic-specific queries ranked by importance the diversity of keywords increases the chances that at least one of the recommended documents answers a need for information. It retrieve the document based on the query by the help of just-in-time-retrieval systems.

Thus from the analysis of experimental results shows the keyword extract model encapsulates the knowledge of which extraction of keyword using voice search. The Extraction of Keyword is based on the context of user behavior. The user will search for any document. Where based on the context behavior of the user the keywords and non-keywords will be separated. Thus based on this keyword is extracted to relate the need documents for the user. The extraction of keyword is based on the algorithm JITIR for user information where the discussion is done based on the context of the user behavior, environment and their needs.

Recommended document are listed by using the IKIFS and UFS algorithms. IKIFS is mainly used for fast search and UFS is used to select the user required documents. These algorithms work effectively to extract the keywords and displays the recommended document. The method IKIFS is to extract the documents based upon the filtering option, not all the documents will be derived only the user required document will be extracted in the fast manner by locating the key feature assigned to the corresponding document. UFS is to retrieve a document which is searched by the user again and again that is the UFS will have a list of frequently searched document so if the user start to type the keyword, UFS will match the keyword and it helps to extract the document soon. Thus UFS uses the priority ranking to the retrieved document.

Thus, the system Analysis is made using the different algorithms to effectively retrieve the documents based the implicit queries posed by the users using keyword search.

VIII. CONCLUSION

After lots of survey made on Extraction of Recommended document, it is concluded based on the efficiency of the algorithm, there will be perfect extraction of recommended documents. Thus using the newly generated algorithm IKIFS, UFS to extract the appropriate Document in accurate and Fast manner. These algorithms are used to overcome the failures posed from the survey made. These algorithms never uses the frequency of words or the word count or no matrix evaluation etc. Thus the working of these algorithm are explained very briefly.And their current goal is to process the explicit queries, and to rank document results with the objective of maximizing the coverage of all the information needs. By integrating these techniquesin a working prototype should help the users to find valuable documents immediately and effortlessly.

IX. FUTURE WORK

Using Automatic Content Linking Device (ACLD) a just-in-time document recommendation system for human users of the system within real-life meetings is related to the semantics search that takes into consideration the conjunction of keywords, sequence of keywords, and even the complex natural language semantics to produce highly relevant search results and search ranking that sorts the searching results according to the relevance criteria.

X. REFERENCES

[1]B. J. Rhodes and P. Maes, “Just-in-time information retrieval agents,”IBM Syst. J., vol. 39, no. 3.4, pp. 685–704, 2000.

[2] S. Dumais, E. Cutrell, R. Sarin, and E. Horvitz, “Implicit queries (IQ)for contextualized search,” in Proc. 27th Annu. Int. ACM SIGIR Conf.Res. Develop. Inf. Retrieval, 2004, pp. 594–594.

[3] A. S. M. Arif, J. T. Du, and I. Lee, “Examining collaborative queryreformulation: A case of travel information searching,” in Proc.37th Int. ACM SIGIR Conf. Res. Develop. Inf. Retrieval, 2014, pp.875–878.

[4] D. Sanchez-Cortes, O. Aran, M. Schmid Mast, and D. Gatica-Perez,“A nonverbal behavior approach to identify emergent leaders in smallgroups,” IEEE Trans. Multimedia, vol. 14, no. 3, pp. 816–832, Jun.2012

[5] W. Webber, A. Moffat, and J. Zobel, "A similarity measure for indefinite rankings," *ACM Trans. Inf. Syst. (TOIS)*, vol. 28, no. 4, pp 20:1–20:38, 2010.

[6] K. Bachour, F. Kaplan, and P. Dillenbourg, "An interactive tablefor supporting participation balance in face-to-face collaborativelearning," IEEE Trans. Learn. Technol., vol. 3, no. 3, pp. 203–213,Jul.–Sep. 2010.

[7] A. F. Bobick and J. W. Davis, "The recognition of human movementusing temporal templates," IEEE Trans. Pattern Anal. Mach. Intell.,vol. 23, no. 3, pp. 257–267, Mar. 2001.

[8] D. Jayagopi and D. Gatica-Perez, "Mining group nonverbal conversationalpatterns using probabilistic topic models," IEEE Trans. Multimedia,vol. 12, no. 8, pp. 790–802, Dec. 2010.

[9] A. Dielmann and S. Renals, "Automatic meeting segmentation usingdynamic Bayesian networks," IEEE Trans. Multimedia, vol. 9, no. 1,pp. 25–25, Jan. 2007