Impact Factor (SJIF): 5.301

International Journal of Advance Research in Engineering, Science & Technology

e-ISSN: 2393-9877, p-ISSN: 2394-2444

Volume 6, Issue 12, December-2019

# DOCUMENT SUGGESTION DURING CONVERSATION USING KEYWORD EXTRACTION AND CLUSTERING

Shruti Bhavsar[1], Sanjana Khairnar [2], Pauravi Nagarkar [3], Sonali Raina [4]
Prof. Amol Dumbare [5]

[1]shrutsbhavsar@gmail.com, [2]sanjana4690@gmail.com , [3]pauravinagarkar5@gmail.com, [4]sraina1998@gmail.com, [5]amol.dumbare@pccoer.in,

*[1]Computer Engineering, Pimpri Chinchwad College of Engineering and Research, Pune University*
*[2]Computer Engineering, Pimpri Chinchwad College of Engineering and Research, Pune University*
*[3]Computer Engineering, Pimpri Chinchwad College of Engineering and Research, Pune University*
*[4]Computer Engineering, Pimpri Chinchwad College of Engineering and Research, Pune University*
*[5]Computer Engineering, Pimpri Chinchwad College of Engineering and Research, Pune University*

***Abstract:***

This paper states the problem statement of keyword extraction from conversations, based on which document gets suggested. With the help of idea described in here user can access the relevant document based on keyword from each short conversation fragment, which can be recommended to users. Sometimes, even a short fragment includes different types of words, which are potentially related to different topics. However, errors can be introducing into system or conversation by using automatic speech recognition system. To extract keyword from the output of ASR technique we use one algorithm. It is based on the sub modular functions which gives range of different words and reduces the noise. Then use a method to obtain a multiple topic from this keyword set only for to access the one relevant recommended document.

***Keywords: Document recommendation, information retrieval keyword extraction, meeting analysis, Local Database, Extraction, Keyword, Clustering.***

## I. INTRODUCTION

The job of suggesting documents to users in small business meetings varies from the task of recommending products to consumers. As in daily life how we suggest small things to our friends like sharing knowledge or giving suggestions as applied to books, videos and the like, attempt to communicate patterns of shared taste or interest among the buying habits of individual shoppers to augment conventional search results But some problems are included like vary of interest and opinions however the idea of recommending can help the users in smaller way too in this type of problems.

Even small variations in search context can weaken the effectiveness of filtering. For example, a Doctor might research in internet database on one side of a medical case today. Like the advantages of performing operation on one case may seem to be dangerous for other similar case. So, providing proper information is important.

Topic-based recommendation systems examine point descriptions to identify items that are of exact interest to the user. It also concludes with a discussion of variants of the approaches, the strengths and weaknesses of content-based recommendation systems, and directions for future research and development. Although the keyword extraction applications normally work on standalone documents, keyword extraction is also used for more complex task (i.e. keyword extraction for the whole collection the entire set of data set, web site or for automatic web summarization.

The function of images content and metadata: In common, related images often acquire similar privacy preferences, especially when people appear in the images. Capturing and analyzing the visual content may not be fruitful to capture users' privacy preferences.

## II. RELATED WORK

Daniel Billsus and Michael J Pazzani of Rutgers University proposed a system that suggests a product or information to a user based upon an explanation of the item and user's interests. These type of recommendation systems helps in recommending web pages, hotels, places, institutes, news articles, restaurants, television programs, and online shopping websites. A way for creating a profile of the user that describes the types of items the user likes, and a means of comparing most usual items to the user profile to determine what to recommend.

This often-selected item's profile is created and updated automatically in response to feedback on the desirability of items that have available to the usefully unsupervised method to extract keywords from meeting speech in real-time. Their approach represents text as a word co-occurrence network and leverages the k-core graph decomposition algorithm and properties of sub modular functions. In related work they have perform multiple baselines in a real-time scenario emulated from the AMI and ICSI meeting.

Prem Melville & Vikas Sindhvani.They discussed the approaches for recommender systems like collaborative filtering, Content Based recommendation and hybrid approaches. Collaborative approaches only use user feedback ratings to recommend items by utilizing machine learning techniques lay k-nearest neighbor. Collaborative filtering includes two methods Neighborhood Based collaborative filtering & Model-based collaborative filtering. Content Based filtering recommends based on topic similarity for example: if the search history of user contains movies of Rajamouli, then it suggests other movies of Rajamouli. Hybrid approach uses both Collaborative and content-based recommendation.

It also discusses the advantages and disadvantages of recommender systems like push attacks and Nuke attacks. Content Based are unaffected by Profile injection attacks.

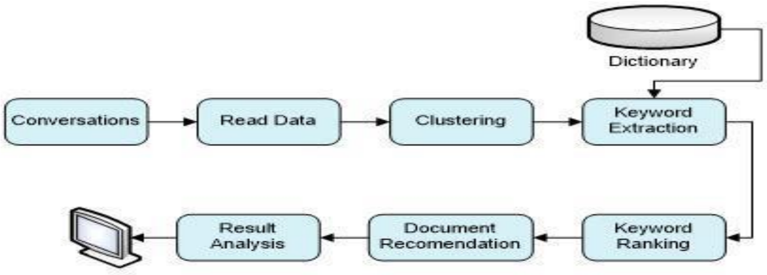
To provide security for the information, automated annotation of images is introduced which aims to create the meta data information about the images by using the novel approach called Semantic annotated Markovian Semantic Indexing (SMSI) for retrieving the images. The proposed system automatically annotates the images using hidden Markov model and features are extracted by using color histogram and Scale-invariant feature transform (or SIFT) descriptor method. After annotating these images, semantic retrieval of images can be done by using Natural Language processing tool namely Word Net for measuring semantic similarity of annotated images in the database. Experimental result provides better retrieval performance when compare with the existing system.

## III. PROPOSED SYSTEM

Here we have proposed a well-organized way for document recommendation system for user using the conversational data. Text file of informal data is given as input. These informal data is partitioned into m clusters. Clusters contain variants of keywords including n number of words. Using Word dictionary only important and useful topic related keywords are extracted.

Keywords are separated based on their number of occurrences weights or occurrences. By selecting the maximum ranked keyword document recommendation approach will be achieved. Propose a calculation to separate decisive words from the yield of an ASR framework (or a manual transcript for testing), which makes utilization of theme demonstrating methods and of a sub modular prize capacity which supports differing qualities in the catchphrase set, to coordinate the potential assorted qualities of points and lessen ASR commotion. At that point, we propose a technique to infer different topically isolated questions from this magic word set, with a specific end goal to expand the shots of making no less than one important suggestion when utilizing these inquiries to pursuit over the English Wikipedia.

## IV. SYSTEM ARCHITECTURE



**MATHEMATICAL MODEL**

S is our proposed system.

S= {U, ASR, D, KC,DKE, QF, O}.

U = User

U = {u1, u2…un}

D = Dataset.

D = {d1, d2...dn}

ASR= Automatic Speech Recognition

DKE = Diverse keyword extraction

KC = Keyword Clustering

QF = Query Formulation

O = Output.

**Procedure:**

**Keyword Extraction:**

Automatic speech recognition converts the speech and provides output to algorithm that extracts and provides keywords from the output of an ASR system.

**Selection of Configurations:**

Using the rank biased overlap as a similarity metric,

based on the fraction of result intersection at different ranks.

$$RBO(S,T) = \frac{1}{\sum_{d=1}^{D} \left(\frac{1}{2}\right)^{d-1}} \sum_{d=1}^{D} \left(\frac{1}{2}\right)^{d-1} \frac{|S_{1:d} \cap T_{1:d}|}{|S_{1:d} \cup T_{1:d}|}$$

Where,

RBO = Rank biased cover Sand T be 2 ranked lists, and Si means keyword at rank i in S.

The Keyword package till rank d in S is {Si : I : <= d}.

RBO is calculating das above Equation.

## V. APPLICATIONS

- This clustering decreases the chances of error into the queries.
- Quick response to the user.
- Content of the will be displayed.

## VI. HARDWARE REQUIREMENT

| | |
|---|---|
| Hard Disk | : 40 GB. |
| System | : Intel I3. |
| Monitor | : 15 VGA Colour. |
| Ram | : 4 GB. |
| Mouse | : Logitech. |

## VII. SOFTWARE REQUIREMENT

|  |  |
|---|---|
| Operating system | : Windows XP Professional/7LINUX. |
| Coding language | : JAVA/J2EE. |
| IDE | : Eclipse Kepler. |
| Database | : MYSQL,XAMPP |

## VIII. CONCLUSION

Our present goals are to practice explicit queries, and to grade document results with the aim of increasing the exposure of all the information requirements, while decreasing redundancy in a shortlist of documents. In our proposed system, we have considered retrieval systems projected for informal environments, in which they suggest to user's documents that are appropriate to their information wants. Enforcing both significance and variety brings unsuccessful progress to Keyword extraction & document retrieval.

## IX. REFERENCES

[1] Khalid Al-Kofahi, Peter Jackson, Mike Dahn*, Charles Elberti, William Keenan, John Duprey.A "Document Recommendation System Blending Retrieval and Categorization Technologies".

[2] Michael J. Pazzani and "Daniel Billsus, Content-based Recommendation Systems "..

[3] Qiang Lu and Jack G. Conrad, "Bringing Order to Legal Documents, An Issue-based Recommendation System via Cluster Association". Thomson Reuters Corporate Research & Development when this work was conducted.

[4] Sangeetha. J 1, Kavitha R ," An Improved Privacy Policy Inference over the Socially Shared Images with Automated Annotation Process ", / (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 6 (3) , 2015, 3166-3169.

[5] Aishwarya Singh, Bhavesh Mandalkar, Sushmita Singh , Prof. yogesh Pawar, "A Survey on User-Uploaded Images Privacy Policy Prediction Using Classification and Policy Mining",International Journal of Innovative Research in Computer and Communication Engineering .Vol. 3, Issue 9, September 2015.

[6] X. Liu, X. Zhou, Z. Fu, F. Wei, and M. Zhou, "Exacting social events for tweets using a factor graph," in Proc. AAAI Conf. Artif. Intell., 2012, pp. 1692–1698.

[7] A. Cui, M. Zhang, Y. Liu, S. Ma, and K. Zhang, "Discover breaking events with popular hashtags in twitter," in Proc. 21st ACM Int. Conf. Inf. Knowl. Manage., 2012, pp. 1794–1798.

[8] A. Ritter, Mausam, O. Etzioni, and S. Clark, "Open domain event extraction from twitter," in Proc. 18th ACM SIGKDD Int. Conf. Knowledge Discovery Data Mining, 2012, pp. 1104–1112.