Relevancy Classification System for Video Search Results

Ms. Prachi Sharma, Ms. Devika Pathak, Ms. Poonam Rachatwar, Mr. Akshay Vadnere, Ms. Pallavi Khude

Department of Computer Engineering, Dr. D. Y. Patil College of Engineering, Akurdi

Abstract- *Now a day's Internet is becoming one of the basic needs of human being. Day by day different technologies are coming to enhance the experience of user. Our system will also enhance this experience little bit. This system will enhance the video searching experience of the user by providing different categories, dimensions and last but most important, better and most Holocene searching algorithm that is hummingbird algorithm. We are trying to implement one new algorithm, which includes some features of the hummingbird algorithm and also considers user's area of interest to provide more relevant search result. This algorithm searches for results on the basis of user's area of interest. The goal is that matching the meaning do better, rather than matching just a few words.*

Index Terms- Hummingbird algorithm, Relevant, Search engine, Session

I. INTRODUCTION

The main goal of the system is providing most relevant videos to the user. To fulfill this requirement, we will implement Hummingbird algorithm. In this is the gem of the system. By using this algorithm system will provide most relevant search results than existing system. For classification of search results system will use ranking algorithm. It will display results according to ranks given to the videos by users. For the convenience of user, the system will provide different fields like news, music, sports and many other while searching for videos. User can search any of those fields after entering keyword. User can select dimension and extension of videos too while searching. Thus by selecting all settings while searching, user can enjoy his video without doing any settings when it gets played.

Currently there are many websites present for video searching like YouTube, Dailymotion. But these websites don't give relevant result according to user's area of interest. For example, if user want to search for video of coffee band, the existing system will shows results like how to make instant coffee, coffee the greatest addiction ever. But in our system it will ask about field, so user can enter his field to get relevant results.

II. LITERATURE SURVEY

To understand the technical contribution of Relevancy Classification System let us take overview of the related work-

It is possible to build relevant classification system using keyword matching and keyword scoring algorithm. For keyword matching users comment can be used. Videos are classified by applying both algorithms and search results are provided to the user [1].

Online resources present document retrieval systems return long list of ranked documents that users are forced to sift through to find relevant documents. All search engines uses ranking and clustering techniques for presenting the result to the user. But, still it needs improvements. Instead of reducing number of results, it is most important to provide relevant and required search results to the users. Currently Simple ranking algorithm is used by the search engines, which provides scattered results in their search results. To overcome these problem different areas will be provided to the user by the system. For this purpose list of different areas will be provided [2].

Web pages can be efficiently collected using Web information classification system (WICS) and classified into several subject and search result are provided to user. Before extracting the feature phrases of the subject, automatic segmentation of words and frequency statistic will be done [3].

Powerful method for representing videos is decomposing video frames into coherent two dimensional motion layers. Intermediate description is provided by such representation that enables application such as object tracking, video summarization, visualization, video insertion and sprite-based video compression [4].

III. EXISTING SYSTEM

YouTube is a video-hosting website that allows users to upload videos to the site and allowing visitors to search and watch videos anytime they want, conveniently, quickly, and at no cost. A search function allows visitors to search videos by keywords or topics whatever they want. They can be watched in the normal screen mode or full screen mode. One special feature of YouTube that has made it gained such widespread use is its enabling of users to embed videos in their own websites and blogs. YouTube does this by putting up an HTML code for visitors to merely copy and input it into their websites' source code to allow visitors to their own websites watch the videos there and then.

Unlike many other video-hosting websites, YouTube was one of the first to offer videos in streaming media format. With streaming media, viewers do not have to wait to download a large file before watching the video. Instead, the media is sent in a continuous stream and is played as it arrives. Unlike many other video-hosting websites, YouTube was one of the first to offer videos in streaming media format.Thus, this significantly shortens waiting time and as all things go, makes the entire YouTube experience a smooth and enjoyable one.

Currently YouTube is third most visited website after Google and the Facebook. Conventional video searching systems return long list of ranked videos. The login is not required to view any video. But for uploading any video, to do comment, to give ratings login is necessary.
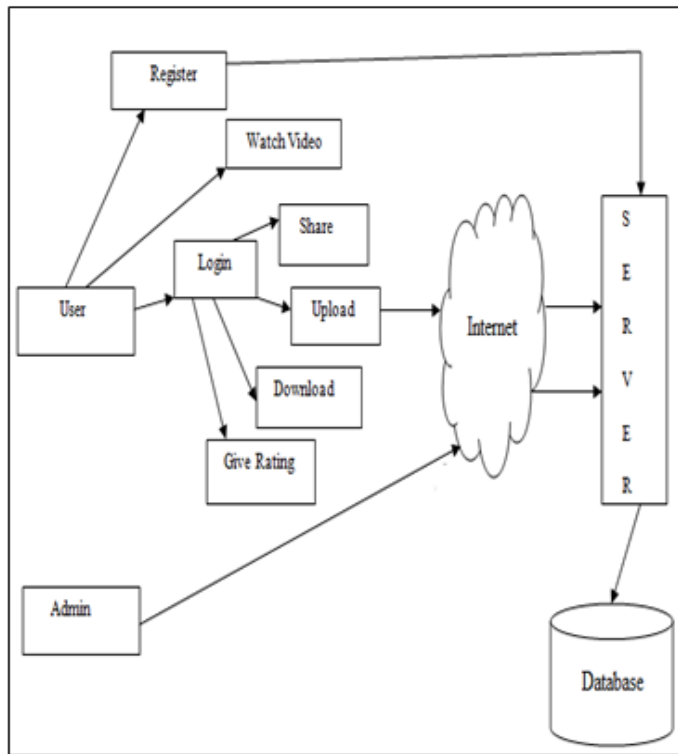
Figure 1: Existing System

IV. PROPOSED SYSTEM

For using this system user must do registration first. At the time of registration different information like name, email id, date of birth and user's area of interest is taken form user. The next time while using this system only login is required. After login system will provide two options for searching the video and for uploading the video. When user wants to search the video, system will provide many options like category, dimensions, keyword facility, most ranked video, previous search history of user etc. When user enters the keyword for searching the video, one drag down list will be provided so that user will get most relevant result. For example user wants to search the video of band journey; the meaning of journey is travelling also. In that case drag down list will show you options like band, travelling. By choosing related option user can get the result that he actually want. User can also give rating to the video. In case if the required video is not present in the database, system will show you default YouTube result. When user wants to upload any video, video will get uploaded first then it will go to the administrator for categorisation purpose. Admin will categorise it according to its dimensions, quality, rate etc.

Analyze and understand all the provided review comments thoroughly. Now make the required amendments in your paper. If you are not confident about any review comment, then don't forget to get clarity about that comment. And in some cases there could be chances where your paper receives number of critical remarks. In that cases don't get disheartened and try to improvise the maximum. To use this system user must do registration first. At the time of registration different information like name, date of birth and user's area of interest is taken from the user. Then next time while using this system only login is required. The session is maintained in this way.
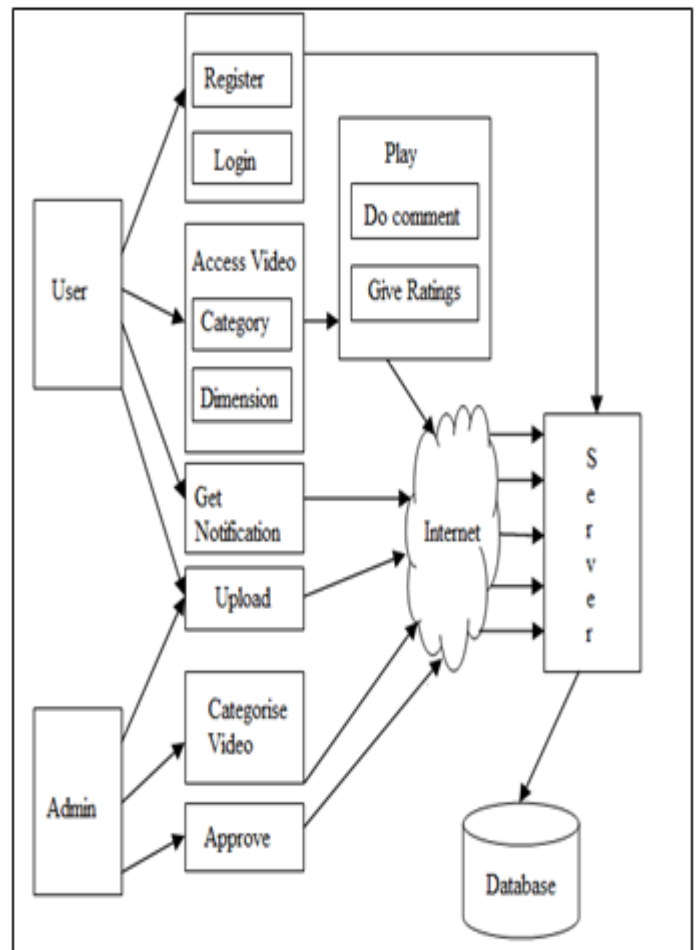

Figure 2: System Architecture of project

The modules of the relevancy classification system for video search result will be:

a] Admin module: This module will be used by Admin only. The admin can upload different videos according to their quality, type, dimensions. He will also approve the request of video approval made by user. Approval is required to neglect the irrelevant search results. Sometimes while uploading any video user can enter wrong information about its field. So if any other user is searching for video of that field, that user will gate irrelevant results. So to avoid this, admin will just check whether that tags are appropriate or not.

b] User Module: This module will be used by all users to view all videos and to upload videos. To access this module user first requires opening an account on this website. The uploaded videos are sending to the admin for approval. User also can share, upload videos as well as can do comments and give rating to the video. When user shares or uploads any video, the notification of that will be send to the user's friends. For this purpose this system requires access to the user's face book friend list. User can decide whether he wants to add all of his facebook friends or some of them to get notification from the system. When user will login, on the home page of the system, he will be able to see search bar, list of suggested videos, list of videos watched by him.

c] Database Module: All data will be stored in this module. It is most important module of the system. It stores all information regarding users, also stores all videos. This module help in transformation of information in terms of videos form Admin to user and vice-versa.

d] All these modules are independent of each other. API's will be used to communicate between database and user interface.

After building the main modules of the system, these modules will be combined together to form an effective system in video searching field.

The user is required to enter the following search criteria for locating a given video:
- Video's name
- Video's field
- Dimensions

The customised implementation steps required for the classification of the most frequently viewed videos are:
1. Searching for relevant videos.
2. Sorting of the videos.
3. Displaying list of relevant result.

1. Searching for relevant videos:

For searching purpose, system will calculate the weight of an each video. When user enters any keyword, system will calculate the weight by using following steps:

- The total weight of ball search results will be 1.
- Out of total weight, 0.3 weights will be given if the video is present in user's area of interest.
- 0.2 weight will be given to the most recent videos.
- 0.5 weight will be given to the videos having more number of views.
- If videos are present in the area of interest of the user and they match to the keyword entered by the user, 0.3 weight will be given to them by system.
- i.e. weigh t= 0.3 depending upon if category exist in the area of interest of the user.
- System will calculate the difference between the present time and the time when video was uploaded.0.2 weight will be given to the most recent videos.
- Now in the next step, weight will be decided according to the number of views to the video. System will calculate it by using following formula:
- Weight = (video views/total views) * 0.5

So the most important aspect for weight calculation is timestamp. The timestamp will be calculated by the system and the half weight has been given to it. User can give ranking to the system. These ranks will be used by the system for weight calculation. 30% of weight has been given to it.

To understand users query more effectively, we are using one feature of the hummingbird algorithm. For example, suppose user enters keyword "Satya Nandella" in his first search, and for second search if user enters "His Education" the system will show results which contain keyword education. As search engine is a machine it doesn't understand human language. It just takes keyword from user and matches it to find the results. To make better communication between searching system and the user we are implementing this feature. In above example suppose user enters keyword "Satya Nandella" in his first search, and for second search if user enters "His Education" the system will take "Satya Nandella" instead of " his". Thus system will give more relevant search results and the communication between user and the search system will be more efficient.

For this purpose system will extract noun from the keyword entered by the user and will replace it with a words like "he, his, she, her, it", to understand users query in the better way. It indexes search result more efficiently. Hummingbird should better focus on the meaning behind the words. Knowing all these meanings may help search engine go beyond just finding pages with matching words. In particular, search engine said that Hummingbird algorithm is paying more attention to each word in a query, ensuring that the whole query-the whole sentence or conversion or meaning-is taken into account, rather than particular words. The goal is that pages matching the meaning do better, rather than pages matching just a few words. So by using this algorithm the list of most relevant videos will be searched for user.

2. Sorting of the videos:

System will sort the videos according to the weight calculated during searching and will display the search results according to the weight in descending order. The weight of the videos will be less than 1. The sum of total weight of the videos will be 1.

It is impossible to have weight 0.99 to 1 to any video. If video is present in user's area of interest, have more number of views and it is more recently uploaded then only it is possible. Similarly it is impossible to have negative weight to any video. Sorted list will be displayed to the user.

User can also click on watch later option to watch the selected video later. User will be able to see list of suggested videos when after logging in.

We are not too bothered about the security aspect of the system as it does not deal with any kind of sensitive data such as banking norms, etc. So we don't need to bother about the security aspect of our system. The login will only be used in case of accessing and uploading of the videos. The data handled in this system is not too vital. There is no need of servers that always be confirmed to run properly.

The important aspect of system is time constrain. This system is real time and hence should be performed. The accountability is not a vital feature in this system and this could be assured if the system is working in its full capability. The system should not be too lagging. It is not possible to develop a system like this without any kind of lag, but the best we can do is to minimize that as much as possible so that we can deliver best performance.

The table is given below. It contains the keyword and the expected result from the system. When user will enter any keyword, which results are supposed to display is given in the table. System will also display the number of results available for the user's query.

Table I: Expected results

Input			Process		Output
Keyword	Category	Dimensions	Searching	Sorting	
Coffee	Music	480*360	Videos in music field.	Top 10 weighted videos of coffee band in music field.	Number of search results found & list in descending order.
Rafael Nadal	Sports	640*360	Videos in sports field.	Top 10 weighted videos of Rafael Nadal in sports field.	Number of search results found & list in descending order.
Ahjhkngv	Music	-	Videos in music field.	No search results.	0 results found.
Rahul	-	-	All videos.	Top 10 weighted videos of Rahul in all fields.	Number of search results found & list in descending order.
Al	Educational	480*360	Videos in educational field containing "al".	Top 10 weighted videos containing "al" in educational field.	Number of search results found & list in descending order.
Coffeee	Music	-	Videos in music field.	No search results.	0 results found.

V. CONCLUSION

The system is still being built, so we can't conclude anything at this moment. But the table of expected result shows the result expected by the system.

FUTURE WORK

After implementation of this project, we are thinking to develop an android app and ios app for this work. Also we are planning to develop the relevancy classification system for normal web search, so that user can get more relevant results. At present we are providing some standard dimensions, but in future we are thinking to provide all dimensions.

ACKNOWLEDGMENT

REFERENCES

[1] Mark Smith, Ray Hashemi, "A Relevancy Classification System for Web Search Results in Mobile Device" in 2011 Englisg International Conference on Information Technology: New Generations. H. Poor, *An Introduction to Signal Detection and Estimation*. New York: Springer-Verlag, 1985, ch. 4.

[2] Rajeev Kumar Mishra, T. V. Prabhakar, "Khojyantra: An Integrated Metasearch Engine with Classification, Clustering and ranking" in 2000 IEEE

[3] Yishan HUANG, Qianping WANG, Jing YANG, "The Design and Implementation of a Subject-oriented Web Information Classification System" in The 9th International Conference on Comput6er Supported Co-Operative Work in Design Proceedings.

[4] H. Tao, "Object tracking with Bayesian estimation of dynamic layer representations," *IEEE Trans. On* Pattern Anal. And Mach Intelli., vol. 24, no. 1, pp.

[5] J. Goldberger and H. Greenspan," Context-based segmentation of image sequences," *IEEE Transactions On Pattern Analysis And Machine Intelligence*, vol. 28, no. 3, pp. 463-468, March 2006.

[6] M. Smith and A. Khotanzad, "Unsupervised object-based video segmentation using color and texture features," *IEEE Southwest Symposium on Image Analysis*, March, 2006.

AUTHORS

First Author – Prachi Sharma, B.E. [Computer Science], Dr. D. Y. Patil College of Engineering, Akurdi and prachibsharma27@gmail.com

Second Author – Devika Pathak, B.E. [Computer Science], Dr. D. Y. Patil College of Engineering, Akurdi and devikapathak155@gmail.com

Third Author – Poonam Rachatwar, B.E. [Computer Science], Dr. D. Y. Patil College of Engineering, Akurdi and poonam40927@gmail.com

Fourth Author – Akshay Vadnere, B.E. [Computer Science], Dr. D. Y. Patil College of Engineering, Akurdi and akshayvadnere1@gmail.com

Project Guide – Pallavi Khude, Lecturer at Dr. D. Y. Patil College of Engineering, Akurdi in the department of computer science, and khude.pallavi@gmail.com