



International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijirccce.com

Vol. 6, Issue 11, November 2018

Content Curation Engine for Youtube Learning Analytics

PROF. Y.K.Sharma¹, Ali Zaidi², Anvay Joshi³, Vishal Jaiswal⁴, Atharva Ashtaputre⁵

Prof, Department of Computer Engineering, Vishwakarma Institute of Information Technology, Savitribai Phule Pune University Pune, Pune, India

Student, Department of Computer Engineering, Vishwakarma Institute of Information Technology, Savitribai Phule Pune University Pune, Pune, India

ABSTRACT: Videos are one of the main components of Web 2.0. Videos are depiction of information in a graphical format. YouTube is one of the main platforms on which videos are viewed on Web 2.0. People share their experiences, knowledge and views with the help of such sites. However, it is not easy to fetch valuable information from the various videos in available time, which normally is very short. In this paper, a new approach of ranking the videos on YouTube based on various factors like user interest, views, likes/dislikes, comments etc is introduced. This new method of content curation will improve the knowledge experience of the user.

I. INTRODUCTION

YouTube channels are like online diaries handled by an individual, maybe a person or even an organization. The videos share knowledge and information across a wide audience. YouTube videos allow one or more individuals to upload about things they want to share with others. YouTube, as a new platform possesses big differences compared to other social media platforms on the aspect of information updating frequency, organization structure, user connection etc, which has an astonishing power of convergence and penetration. People usually create a video as a hobby to share their information and experience on a particular subject. It completely depends on the user what he wants to publish or post. Entries, on the actual YouTube website are ranked based on a multitude of factors.

The exact technical details of the YouTube algorithm are a closely guarded secret. But some speculations can be made by reverse engineering the algorithm. Initially, YouTube only paid attention to the play button as a factor of determining the ranking of a video. But views only showed the number of times a video player loaded, rewarding creators whose videos got a lot of clicks-regardless of how long the viewer watched. So in 2012, YouTube changed their ranking algorithm to reward engagement instead of clicks by dropping views in favor of watch time, a more accurate measure of how engaging users find a video. Today videos are prioritized not by quantity, but by quality. Statistics such as average watch time better determines the success of a video rather than it's view count. Along with the addition to SEO, viewer feedback plays a crucial role in the algorithm. The key thing to remember is that YouTube isn't in the business of judging whether a particular video is good or not. Instead, the YouTube ranking algorithm focuses on how the audience interacts with videos, using artificial intelligence that learns from over 80 billion bits of feedback from the audience daily to understand how to serve the right videos to the right viewers at the right time. In broad terms, that audience feedback includes:

- What they do (and don't) watch
- How much time they spend watching a video (watch time)
- How much time they spend watching videos during each visit (session time)
- Likes, dislikes, and 'not interested' feedback



International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 11, November 2018

A. Watch Time

- When YouTube decided to mothball the view metric, it replaced it with watch time, or how long a viewer watches a video. However, one shouldn't be fooled into thinking that improving your watch time is as simple as creating longer videos. A 30-second video that people watch from beginning to end will rank better than a 10-minute video that only gets watched for a couple of minutes.

II. PREVIOUS WORK

A. Blog Searching and Curating

Based on the gaps of current available systems, a new approach is proposed to improve the knowledge searching experience of the user. The proposed model and consists of four major modules: Search Manager, Curator, Personalized Module, and Rating Engine. Working of the proposed model is similar to a mining activity, which mines the blog posts as per the need of the user, and thus the name of the model Blog Miner. The model extracts the blog posts from various blog websites. The very first part of proposed model is its interface

- Blog Miner Interface, through which all its modules are connected. User can interact via this interface to track blog posts and feed input search query.

1) *Blog Miner Interface*: The very first part of proposed model is its interface - Blog Miner Interface, through which all its modules are connected. User can interact via this interface to track blog posts and feed input search query. To interact with the user, Blog Miner Interface consists of four sub-interfaces: Curator Interface, Login Interface, Keyword rate, comment and share the blog posts results of curator module. User has to login to the blogging site, if he/she wants to access the services. On Login Interface, there are two fields: user name and password. There is one more option of sign up. Search Interface consists of search bars where the user feeds the query in form of keywords. Two search bars are given on the home page; one is for local blog posts search result, and second is for the global blog posts search results. The blog posts search results are shown using Post Visualize. All blog posts are visible to user in reverse chronological order i.e. latest post first.

2) *Login and personalization*: As name shows, the user logs in using this module. After login, user can share, add, comment, rate and modify the blog posts. When user performs any operation on bog post, then his/her name also displays with that blog post. There is an option for Forgot Password and New User. Mail id of user is taken for verification purpose.

As name shows, the user logs in using this module. After login, user can share, add, comment, rate and modify the blog posts. When user performs any operation on bog post, then his/her name also displays with that blog post. There is an option for Forgot Password and New User. Mail id of user is taken for verification purpose.

3) *Search Manager*: This module consists of all the processes of searching the blog posts. Search can be done internally i.e. a local Blog site search, and globally i.e. blogosphere. As user inputs the keyword in the search bar to search blog posts relevant to the keyword, search manager takes input in the form of keyword and then the requisite search process starts. If the user wants to fetch the post results from local blog site then the local search module will be operational. Moreover, the global search will start if user wants to fetch the blog posts from various other blogs from WWW.

After searching, "Result Aggregator" aggregates resultant blog posts. Searching can be done on any type of data available on World Wide Web. The data may be structured, unstructured or semi-structured. Guoliang Li et al. (2008) gave an algorithm that works well for all types of data whether structured, semi-structured, or unstructured. A similar algorithm has been designed and is used for searching the blog posts in Blog Miner.



International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 11, November 2018

Keyword based searching

1. Keyword = localsearch
2. for each post in postlist
3. if (keyword is in posttitle) or (keyword is in posttext)
4. POP(post)
5. enqueue(post)
6. if (queue == empty)
7. return No search result found;
8. for each post in postqueue
9. print(post);
10. dequeue(post);

Searching on global search

1. Keyword = globalsearch
2. Dbtemp = blog(keyword)
3. for each url in dbtemp
4. if (keyword is in urlposttitle) or (keyword is in urlposttext) a. enqueue(post)
5. if (queue == empty)
 - a. return No search result found;
6. for each post in postqueue
7. print(post);
8. dequeue(post);

4) *Curator*: Curator means a method, which curates the contents from various sources and makes clusters of the similar objects i.e. blogs posts. However, here, an automatic content curation is performed. An approach used here is to fetch the posts from various blogs by traversing on their websites via their URLs. There are We get utilities which gives all hyperlinks available on a URL link. On a blogging site, these hyperlinks consist of the links of blog posts. All blog posts fetch via hyperlinks and stored in a temporary database for posts refinement. The refined post result is stored in the blog post database and put forward to the user using the post visualize i.e. Blog Miner Interface. The series of steps required to perform curation are:

Step 1. Check for the blog URLs from the database, which are inserted manually by a user.

Step 2. Process traverses each URL and use we get utilities on them.



International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 11, November 2018

Step 3. These utilities via blog URLs fetch all blog posts and hyperlinks on that page and temporarily stored in a database.

Step 4. The next step is to search top rated blog posts from the temporary database within the specified period of time.

Step 5. Selected blog posts are stored into the blog post database and rest of the contents is dropped.

Step 6. Relevant blog posts are presented to the user through Post Visualizer (Blog Miner Interface).

5) *Rating Engine*: User is allowed to rate the local blog posts. The rating is visible publicly to the users in the form of stars. Based on the star rating, a numeric value is assigned and sorting of blog posts are done in descending order of the numeric value. Two points are given to each star. If rating of a blog post is four stars, it means eight points or eight numeric value is assigned to that post. Data structure used for storing all blog posts is simply a queue. An algorithm is used to sort the blog posts in the descending order. The blog posts are displayed to the user under the tag Top Rated.

Step 1. All blog posts are stored in posts database.

Step 2. When user rates the posts, it is shown in the form of stars.

Step 3. Each star has allotted a numeric value, 2, (5 stars means numeric value 10).

Step 4. As user hits on star, as per the assigned value, the rating is decided.

Step 5. All rated posts are enqueue to the queue. Then based on the numeric value assigned to the posts, sorting is performed.

Step 6. The result is shown under a separate label Top rated posts and the blog posts are shown in decreasing order of their rating or numerical values.

Step 7. If two or more users rate the post, then the average of their numeric values are taken out.

6) *Working of blog miner*: General working of proposed model, Blog Miner, is divided into three parts, first is to search the content, second is to curate it, and finally to present it to the user.

The proposed system works in the manner as discussed below:

Step 1. Enter the keyword in search bar, i.e. local search bar or global search bar, to fetch the relevant blog posts. The working for local and global search is discussed separately in tabular form in Table II.

Step 2. When a user logs in, he/she interacts with blog posts of his interest only. Search Manger searches for the blog posts of user's interest from the blog posts database.

Step 3. Side by side, Curator is on with its work. Curator curates the blog posts automatically. It takes the URLs from the URL Database. Periodically, Curator fetches all blog posts from to those URLs. Only the selected blog posts are stored into the blog post database and rest of the content is dropped.

Step 4. User is permissible to rate the local blog posts by stars and the rating is visible publicly to all the users in the form of stars.



International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijirccce.com

Vol. 6, Issue 11, November 2018

Step 5. All the blog posts are displayed in reverse chronological order, and user can comment, rate, add and share the post information.

III. IMPLEMENTATION

We are going to design an algorithm which will take into consideration parameters such as likes to dislikes ratio, sentiment analysis of comment section of the video, total number of views of the video, number of subscribers etc. Based on these parameters we shall rank the video. This rank will be different compared to the ranking made by YouTube because our approach to ranking the videos is going to be different, for multiple reasons.

Once our algorithm provides a rank to a particular video, then the next step will be crowd sourcing. In this step, we shall provide the video to the users and based on how well the user understood the concepts explained in that particular video, the user will provide a rating to that video.

The final step will be the combination of the rank provided by the algorithm and the ratings provided by the users. After combination of these two metrics, we shall be providing a new rank which will ultimately be the combination of technology and user understanding.

IV. CONCLUSIONS

YouTube comes under the category of important web tools. Nowadays, major part of knowledge and recent activities are shared using YouTube. To enhance the performance of YouTube, a new approach has been discussed in this paper, which will surely improve the information searching and knowledge experience of the user.

REFERENCES

1. Harsh Khatter Brij Mohan Kalra "A New Approach to Blog Information Searching and Curating".
2. Sergiu Chelaru, Claudia Orellana-Rodriguez, and Ismail Sengor Altin-govde "Can Social Features Help Learning to Rank YouTube Videos?" Conference Paper November 2012.
3. Vincent Simonet, Classifying YouTube Channels: a Practical System, May 13–17, 2013.
4. Web Strategies for the Curation and Discovery of Open Educational Resources Vivien Rolfe, vol. 8 issue 4, October–December 2016.
5. Participatory culture on YouTube: a case study of the multichannel network Machinima Bryan Mueller, Bryan Mueller © 2014.
6. Social Learning Analytics Simon Buckingham Shum1 and Rebecca Ferguson, Buckingham Shum, S., & Ferguson, R. (2012). Social Learning Analytics. Educational Technology & Society, 15 (3), 3–26.
7. The Impact of Content Curation for Personal / Informal Learning Ponerulappan. P1, DR. N. hilagavathy, International Journal of Science and Research (IJSR), 23 February 2016.
8. The Five Cs of Digital Curation: Supporting Twenty-First- Century Teaching and Learning Mark E. Deschaine, 2015.
9. DIGITAL CURATION AS LEARNING ACTIVITY Rivki Gadot, Ilya Levin, EDULEARN12 Conference, 2nd-4th July 2012.
10. [Digital Curation as a Core Competency in Current Learning and Literacy: A Higher Education Perspective, Leona M. Ungerer, International Review of Research in Open and Distributed Learning Volume 17, Number 5.
11. Teachers as Curators: Curating Authentic Online Content for Beginning and Intermediate CFL Learners Shenglan Zhang, 2015.