# Analysis of External Links of video using Association Rule Mining

Pooja D. Mahamuni
ME Computer Student
Department of Computer Engineering
JSPM's JSCOE, Hadapsar
Pune, India
*pooja.mh4709@gmail.com*

Prof. H. A. Hingoliwala
Assistant Professor
Department of Computer Engineering
JSPM's JSCOE, Hadapsar
Pune, India
*ali_hyderi@yahoo.com*

***Abstract***— Now a days, to increase the popularity of video sharing site they provide a external link, so that their video or audio content can be embedded into external website. User can copy that URL of that embedded link & post on their own blog or website. In this paper our intension is to increase the popularity & measure the quantification. We provide the measurement & analysis on this external link. With the results collected from two major video sharing site like You tube & Youku we observed that these links have an various impact on popularity. Videos that have been uploaded for ten months in Youku, around 20% of views can come from external links. Next, we analyzed that correlation between internal & external link. Another observation is Youku has comparatively more impact than You tube.

***Keywords***—*Video sharing sites,external links,UGC.*

*****

## I. INTRODUCTION

Historically, most of the media content which was distributed through media organizations that o partitioned users in regional markets & releasing new content. So, popularity of video was global phenomenon because users could not access the same content all over the world.

UGC sites are most popular in recent years. In these sites, people are not only the access the information, but they can also upload their own content. There are various UGC sites like Facebook, Flicker, Twitter & video sharing sites. Among these UGC sites, this paper will focus on video sharing sites, which are nicely represented by both YouTube and Youku. Because of information distributed much faster through UGC sites they are popular enough.

There are various functionality of You tubes like related video links which arrange videos by similar topics. To increase the popularity of video distribution, video sharing sites introduce external links.In YouTube for each video, an embedded links are provided there. The user can copy and paste one of these embedded link URL into anywhere such as their personal webpage, blogs. When people watch video through external links the count goes through You tube & thus popularity increased. Clearly, the external links allow YouTube videos to be embedded in non-YouTube sites to attract views.

Motivation behind this is popularity of UGC sites. YouTube and Youku are classes of UGC sites. These studies, focus on user-to-user, user-to-video or video-to-video relationship within these video sites. Users can obtain an embedded link URL of a video and paste the link to any web pages in other web sites, such as forums, or their blogs and it then the internal links as those maintaining a relationship within the web sites. There are many UGC sites among them we are going to concentrate on video sharing site.

## II. LITERATURE SURVEY

The relationship between popularity of videos & locality of online videos. Geographic locality of interest occurs in online video consumption. and using new measures which quantify their popularity distribution across different geographic regions[1].

Analysis of content duplication, popularity evolution & distribution of UGC video content. We understand the growth of UGC sites & its impact on behavior of user, infrastructure & different probability shapes[2].

For discovering correlation between set of items we use Association Rule Mining techniques.It is motivated by supermarket analysis that is probability of purchasing two items together[3].

It study the latent user interaction by including three component. First by analyzing characteristics of RenRen social graph & compare with other OSN. Second, focusing on latent interaction by describing log reconstruction algorithm which uses clock to merge visitors log. Finally build latent interaction graph from visitors log[4].

Study of web based video sharing services which include user generated video clip to be uploaded and other users allowed to view that clip, rate & comment that video[5].

In this peer to peer structure is generated. Improvement in multimedia content delivery in Youtube[6].

## III. PROPOSED SYSTEM

***Problem definition:-* "Analysis of external links of videos using Association Rule Mining".**

In a proposed method, find & analysis on no. of hits of video from external links with different categories, personalization of user & also find relevant video links with comparative analysis of Apriori algorithm and parallel FP-Growth algorithms.

## IV. ARCHITECTURE OF PROPOSED SYSTEM MODEL

In this diagram we can see the overall picture of the system. There are collected videos. When universal Java Script engine parse the java script pages then it stores the external links. This process is called crawling. It use the information like ages,no.of views of videos. When user upload video he insert like from which category it belongs to, link, name of the video.

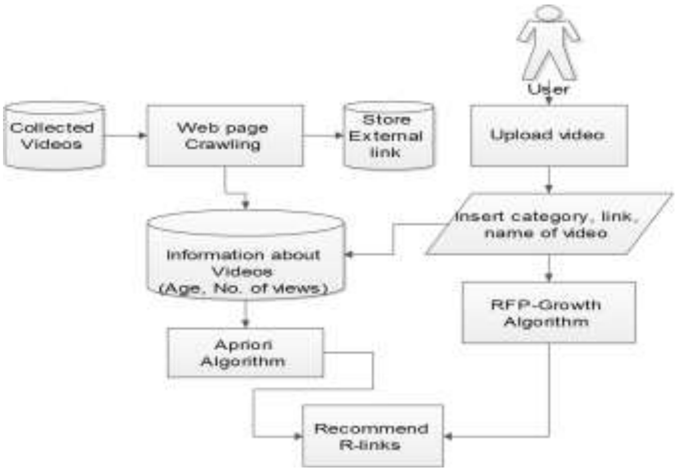


**Fig: System Architecture**

Experimental datasets come from two user generated content VOD sites, YouTube and Youku. YouTube is one of the largest UGC VOD sites in the world and at the time, and it accepts roughly 1.886 billion views every day. Youku is the most popular video site in China, so views comes through external links in Youku is relatively more than You tube.

## V. MATHEMATICAL MODELING

**Mathematical Model**

Let system S

S={U,Vdatabase,C,EL,IL,RFP-Growth,V_Details}

U is no. of users{*U1, U2,U3,….*}

Vdatabase=No.of videos in database={*V1,V2,V3*}

V_Details={Upload_date, no of views, Related IL,EL,C}

Upload_date= to find age of video

EL={External Links of videos}

IL={Internal links of video}

No of views NV={*NV1,NV2…*}

C is category of video={*C1,C2,C3....*}

RFP Growth={*C,N,U,T,Q,H,N*}

H=Header table={*H1,H2*}

Q=node links={*Q1,Q2,....*}

T=Tree generated by link nodes

N=Nodes={*N1,N2,....*}

## VI. IMPLEMENTATION STRATEGY AND EXPERIMENTAL SETUP

We proposed following technique & algorithm to implement this paper.

### A. Association Rule Mining

Association rule mining are one of the major techniques of data mining and it is used to find most common form of local pattern discovery in unsupervised learning systems. It serves as a useful tool for finding correlations between items in large databases.

Algorithms of ARM is used as follows:

*a) **Apriori Algorithm:***

Apriori is an algorithm for finding the frequent itemsets by using candidate generation. It is a level wise complete search algorithm using anti-monotonicity of itemsets, if an itemset is not frequent one, any of its superset is also nota frequent. Let the set of frequent itemsets having size k & Lk and their candidates be Nk It iterates over the following three steps and extracts all the frequent itemsets:

1. Generate Nk+1 , candidates of frequent itemsets of size k +1, from the frequent itemsets of size k.

2. Scan the database and calculate the support of each candidate of frequent itemsets.

3. Add those itemsets that satisfies the minimum support requirement to Lk+1.

It has two steps:

- Join step: Generate Rk+1 , the initial candidates of frequent itemsets of size k+1 by taking the union of the two frequent itemsets of size k, Pk and Qk that have the first k1.
- Prune step: We have to check if all the itemsets of size k in Rk-1 are frequent and generate N by removing those that do not pass this requirement from Rk+1.

*b) **RFP growth algorithm:***

***FP growth***:

It also works like Apriori algorithm but Allows frequent itemsets discovery without candidate itemset generation.It has two main steps:

1) Build a FP Tree- Build a compact data structure called the FP-tree. It Built using 2 passes over the data-set.

Pass 1:

- Find support
- Discard infrequent items.
- Sorting.

Pass 2:

- Reads 1 transaction at a time.
- Fixed order is used
- Pointers are maintained

2) Extract frequent item set from tree directly.

Since FP growth algorithm is Advanced algorithm of Apriori because if the database size is increased Apriori is not efficient. But in this we are using improved FP growth which is called RFP growth to avoid generating intra-property frequent itemsets, and to further boost its efficiency, implement its MapReduce version with additional prune strategy.

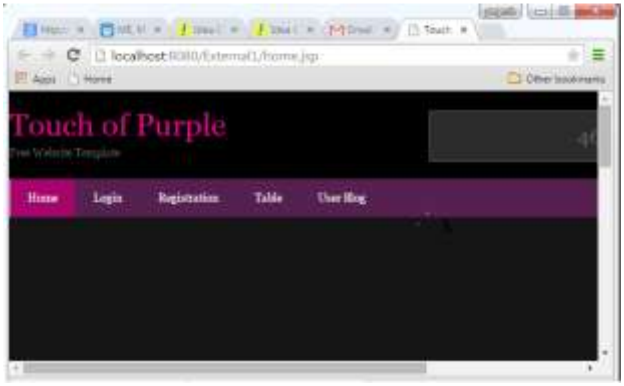
## VII. RESULT ANALYSIS

Software Tools used:

- Operating System: Windows XP/7.
- Language : java
- Frame work : Netbeans 8.0
- Data Base : MYSQL

In this module there are admin & user. Admin upload the video through different catagories.When user watch that uploaded video through his own blog then count will calculated and it displayed in score table. This is shown in following fig.

**Home page:**

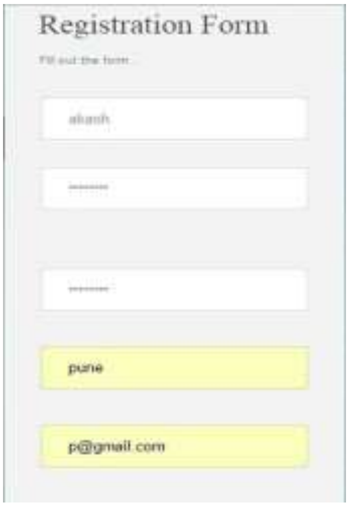


localhost:8080/External1/home.jsp

Touch of Purple

Free Website Template

a) Admin & User registration – First admin register himself by filling information. Then multiple users can do registration.



Registration Form

Fill out the form

akash

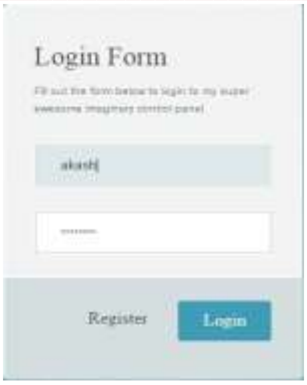
********

********

pune

p@gmail.com

b) Admin & user login – Only one admin can register & login. He able to do operation only when he login himself. Multiple user can register & login.



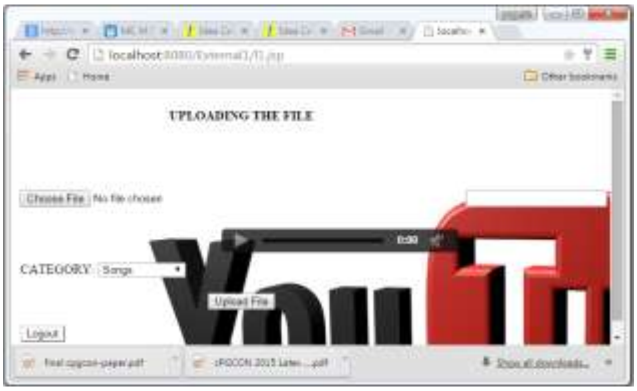
Login Form

Fill out the form below to login to my super awesome imaginary control panel

akash

********

c) Uploading video- Only admin has right to upload video. He choose file & add in respective category & upload it.



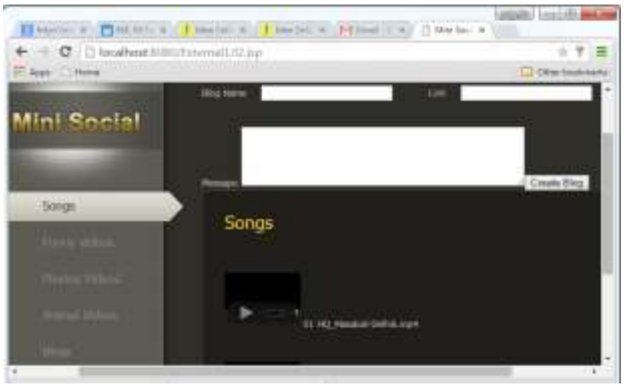
localhost:8080/External1/f1.jsp

UPLOADING THE FILE

Choose File No file chosen

CATEGORY: Songs

d) Blog creation – After logging users can create his blog there. He copy & paste the URL of particular video into his blog. And he can see that video directly from his blog also.



localhost:8080/External1/f2.jsp

Mini Social

Blog Name

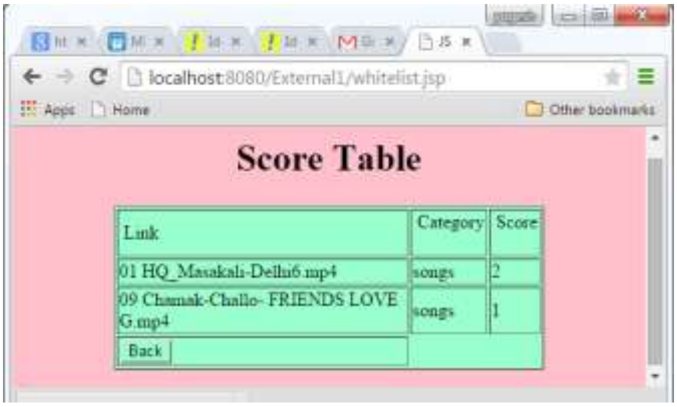
Link

Message

Songs

01_HQ_Masakali-Delhi6.mp4

e) No. of views of video- How many times we see that video through that external link count is recorded.



localhost:8080/External1/whitelist.jsp

Score Table

| Link | Category | Score |
|---|---|---|
| 01 HQ_Masakali-Delhi6.mp4 | songs | 2 |
| 09 Chamak-Challo- FRIENDS LOVE G.mp4 | songs | 1 |
| Back | | |

## VIII. CONCLUSION

The external links plays an important role in the distribution of the videos. The external links have quite different impacts on YouTube and Youku, the correlations of the external links and the internal related video links, personalization of user. With that we are going to find relevant video. For this two algorithms are used Apriori and RFP growth. Both the algorithms selected were able to discover access pattern and user behaviours using support and confidence thresholds accurately. Memory requirement of the Apriori algorithm does not depend on the number of transactions while the memory requirement of the RFP-growth algorithm increases significantly with the growth of the number of transactions. The main drawback of Apriori algorithm is that the candidate set generation is costly, especially if a large number of patterns and/or long patterns exist. From this it can be concluded that RFP is behaves better than Apriori. This study can be extended by classification or clustering algorithm to predict future user requests.

## REFERENCES

[1] A. Brodersen, S. Scellato, and M. Wattenhofer, \YouTube around the world: Geographic popularity of videos," inProc. ACM WWW'11, Lyon, France, Apr. 1620, 2011.

[2] M. Cha, H. Kwak, P. Rodriguez, Y. Ahn, and S. Moon, “I tube, you tube, everybody tubes: Analyzing the world's largest user generated content video system,”inProc. ACM IMC'07, San Diego, CA, Oct. 24{26, 2007

[3] W. Chen, J. Chu, and J. Luan, “Collaborative filtering for orkut communities: Discovery of user latent behavior,” inProc. ACM WWW'09, Madrid, Spain, Apr.20{24, 2009.

[4] J. Jiang, C. Wilson, X. Wang, P. Huang, and W. Sha, “Understanding latent interactions in online social networks,” in Proc. ACM IMC’10, Melbourne, Australia, Nov. 1–3, 2010.

[5] S. Mitra, M. Agrawal, A. Yadav, N. Carlsson, D. Eager, and A. Mahanti, “Characterizing web-based video sharing workload,”ACM Web,vol.5,no.2,May2011.

[6] S.Scellato,C.Mascolo,M.Musolesi,andJ.Crowcroft, “Track glob-ally, deliver locally: Improving content delivery networks by tracking geographic social cascades,” inProc. ACM WWW’11, Hyderabad, India, Mar. 28–Apr. 1, 2011.