DETECTING ORIGINAL AND DUPLICATE YOUTUBE VIDEOS

PRASHANT GUPTA

IPG M.TECH
2010IPG-072
Under the Guidence of
Dr.Joydip Dhar

MTP, 2014

Contents

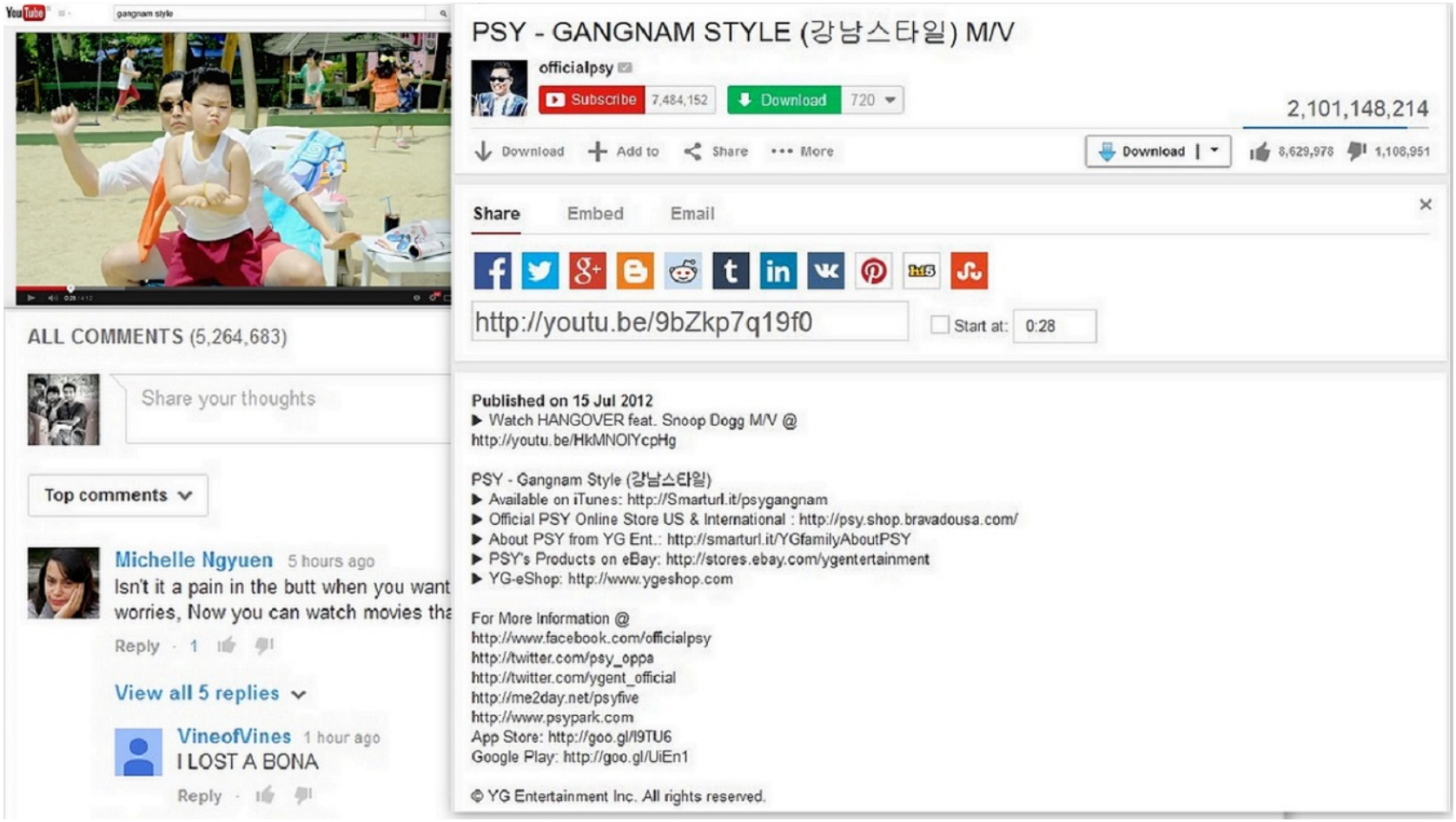
Introduction

- Today, YouTube is the largest user-driven video content provider in the world.
- Allows anyone to upload and watch an unlimited number of videos for free
- YouTube - Alternate TV
- Social Networking on YouTube

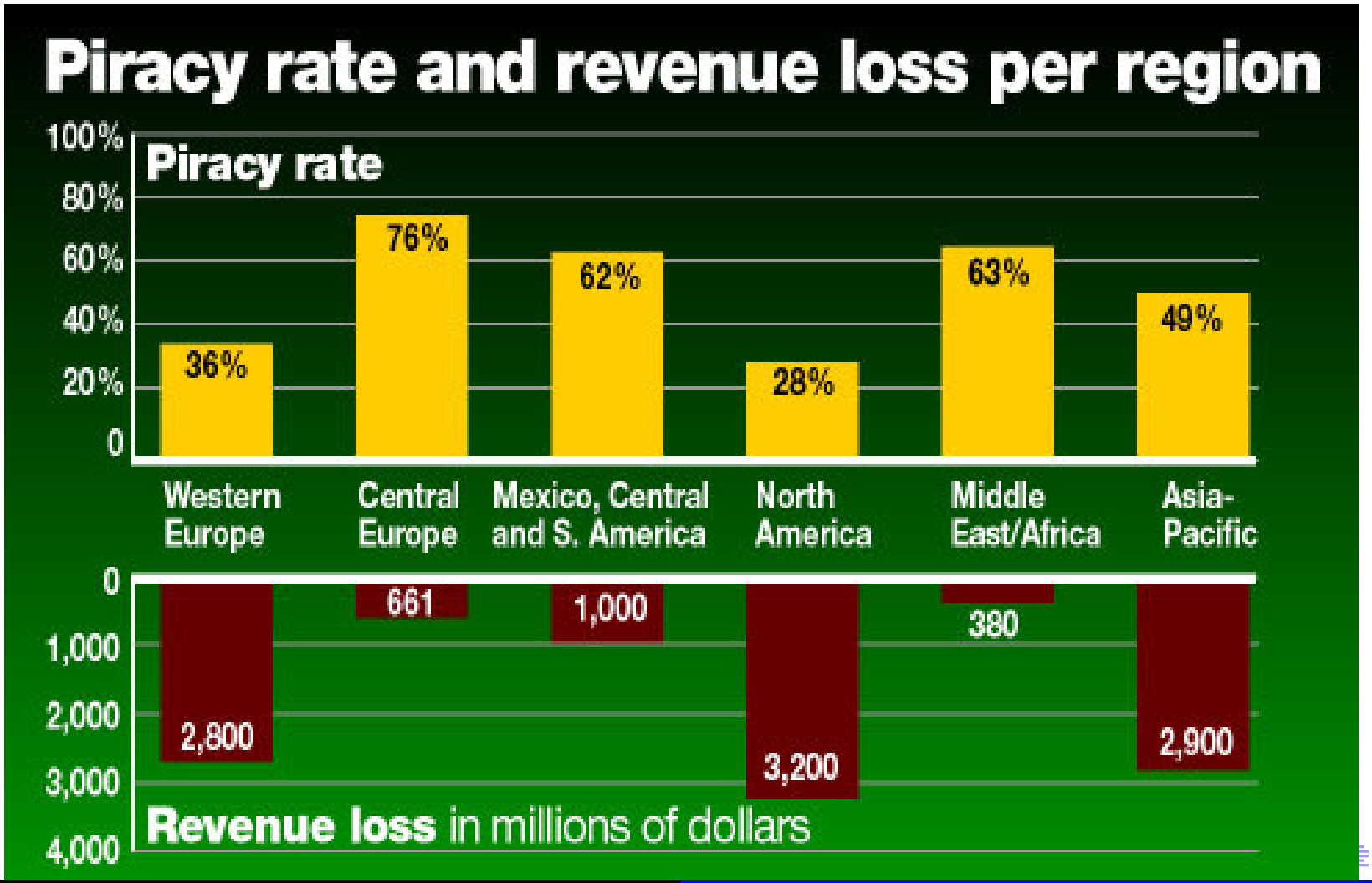
YouTube Statistics

- 500 years of YouTube video are watched everday on facebook
- over 700 YouTube videos are shared on Twitter each minute
- More than 1 billion unique users visit YouTube each month
- Over 6 billion hours of video are watched each month on YouTube that's almost an hour for every person on Earth
- 100 hours of video are uploaded to YouTube every minute
- Tens of thousands of full length movies are available

Snapshot


gangnam style
PSY - GANGNAM STYLE (강남스타일) M/V
officialpsy
Subscribe 7,484,152
Download 720
2,101,148,214
Download
Add to
Share
More
Download
8,629,978
1,108,951
Share
Embed
Email
http://youtu.be/9bZkp7q19f0
Start at: 0:28
ALL COMMENTS (5,264,683)
Share your thoughts
Top comments
Michelle Ngyuen 5 hours ago
Isn't it a pain in the butt when you want
worries, Now you can watch movies tha
Reply · 1
View all 5 replies
VineofVines 1 hour ago
I LOST A BONA
Reply
Published on 15 Jul 2012
▶ Watch HANGOVER feat. Snoop Dogg M/V @
http://youtu.be/HkMNOlYcpHg
PSY - Gangnam Style (강남스타일)
▶ Available on iTunes: http://Smarturl.it/psygangnam
▶ Official PSY Online Store US & International : http://psy.shop.bravadousa.com/
▶ About PSY from YG Ent.: http://smarturl.it/YGfamilyAboutPSY
▶ PSY's Products on eBay: http://stores.ebay.com/ygentertainment
▶ YG-eShop: http://www.ygeshop.com
For More Information @
http://www.facebook.com/officialpsy
http://twitter.com/psy_oppa
http://twitter.com/ygent_official
http://me2day.net/psyfive
http://www.psypark.com
App Store: http://goo.gl/l9TU6
Google Play: http://goo.gl/UiEn1
© YG Entertainment Inc. All rights reserved.

loss of revenue



Motivation

- Online piracy of film and TV network in 2008, cost the Indian film industry $959 million
- Websites are indulging in hosting, streaming providing access to users, infringing its exclusive rights, broadcast and reproduction rights
- Acts of infringement not only cause loss of substantial revenues but will also take away the legitimate revenue to the government through service tax
- In case of YouTube videos infringement, it not only causes loss to the actual uploader of video, but also a loss to YouTube as it has to pay copyright violated video users also.

Literature Review

This section describes the literature survey of the in the area of copyright infringement web video detection, duplicate and near-duplicate video identification. In addition to the application of multimedia and image processing on the video content meta-data based features for the task of copyright infringement detection are also studied in this project.

- youtube uses Content ID as described in [1] to easily identify and manage their content on YouTube.
- youtube also allows us to Submit a copyright infringement notification.
- several discriminatory features are discussed in [2] that can be used to distinguish between the original & the duplicate videos

Literature Review...

- detecting image near-duplicate by stochastic attributed relational graph matching with learning is being discussed in [3].
- a technology to summarize and visualize opinions that are expressed in the form of Web comments is introduced in [4] known as OPINION CLOUD

Literature Review...

- [5] gives an insight to Mine User Comment Activity for Detecting Forum Spammers in YouTube.
- understanding of basic measures for text retrieval is discussed in [6] so that relationship between the set of relevant documents and the set of retrieved documents can be established. Various text retrieval and text indexing techniques are being discussed.

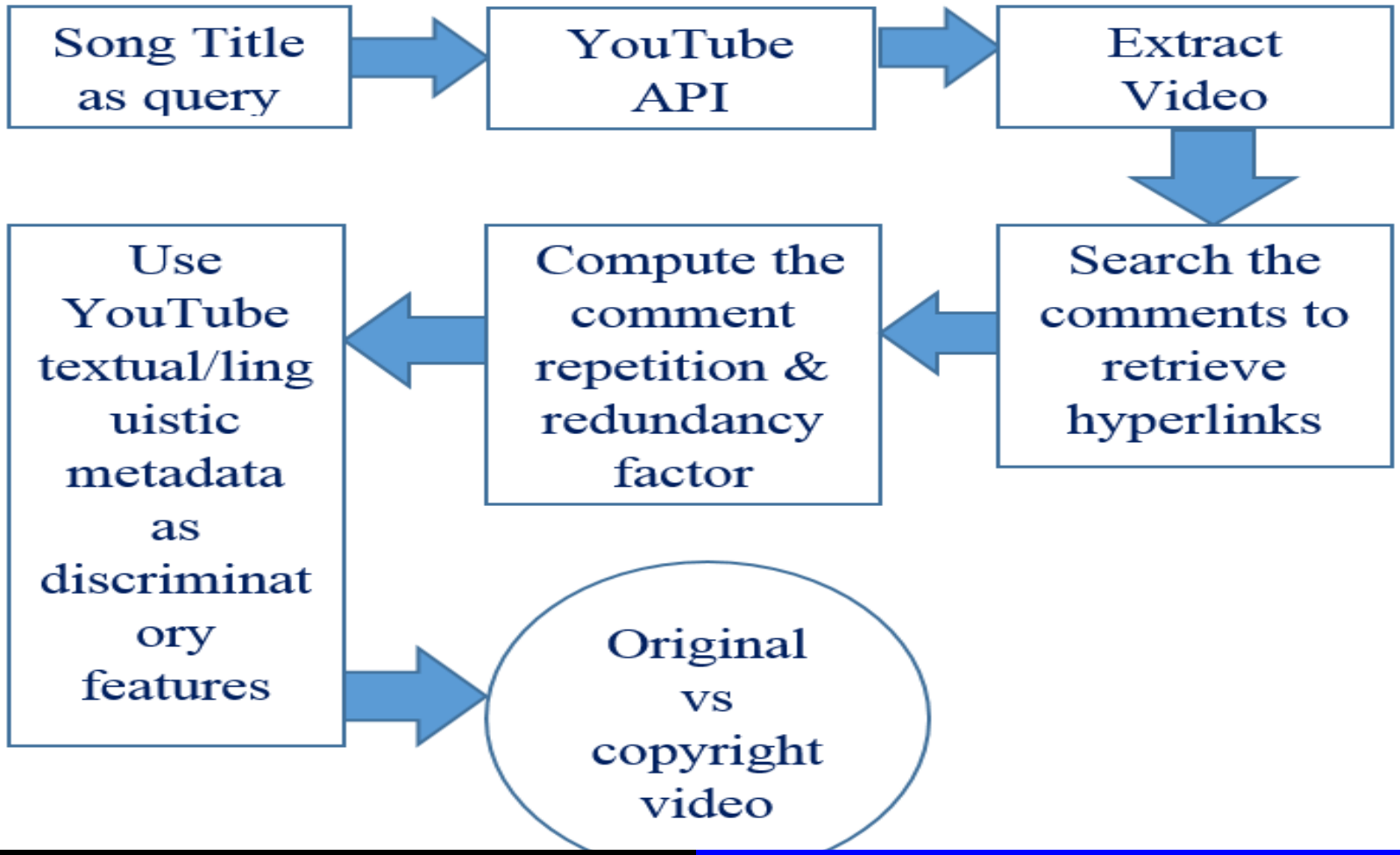
Literature Review Gaps

The proposed literature only explores the metadata of a YouTube video without considering comments on the video. Work done in previous literatures are a limitation to only YouTube Music videos. So, the proposed work is to detect original & duplicate YouTube videos in every form of multimedia content.

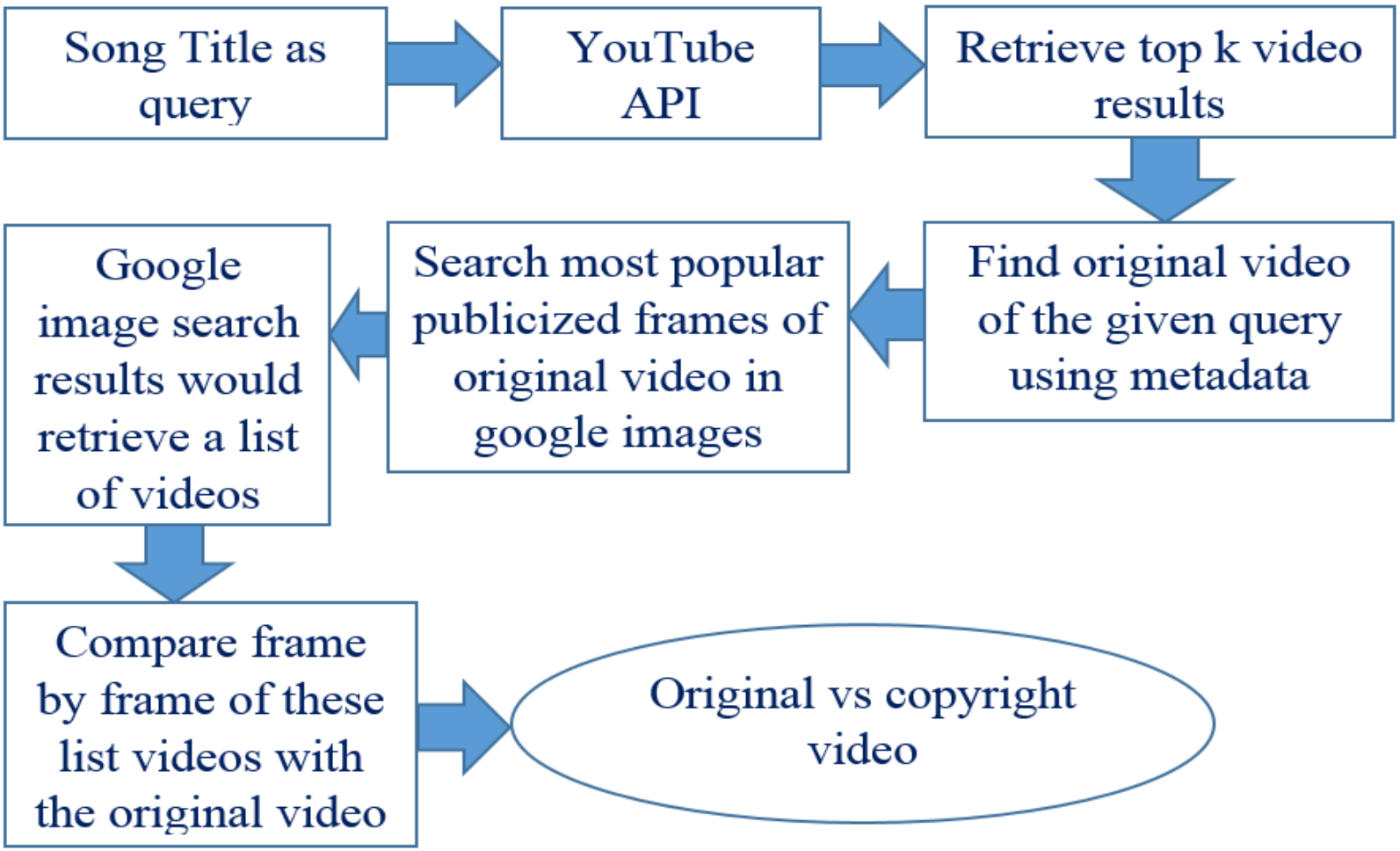
Objective

- To conduct a study on the extent of copyright infringement in YouTube and investigate solutions to detect original and copy-right infringement videos from the search results of a given user query.
- To investigate the effectiveness of contextual features as discriminatory attributes for the task of original and copy-right infringement video detection.

Methodology



Methodology



Methodology

To achieve the objective mentioned, work is divided into four phases.

Phase 1:

- Literature Review
- Devise different ways to detect copyright infringement videos.

Phase 2:

- Study YouTube API
- Information retrieval from Video Metadata

Phase 3:

- devise an appropriate method for near duplicate image detection.
- extract video metadata

Phase 4:

- Integrating above modules
- Comparing & Testing both approaches.

References

- Kim, E. C.: 2007, Youtube: Testing the safe harbors of digital copyright law, S. Cal. Interdisc. LJ 17, 139.
- Agrawal, S. and Sureka, A.: 2013, Copyright infringement detection of music videos on youtube by mining video and uploader meta-data, Big Data Analytics, Springer, pp. 4867.

References

- Zhang, D.-Q. and Chang, S.-F.: 2004, Detecting image near-duplicate by stochastic attributed relational graph matching with learning, Proceedings of the 12th annual ACM international conference on Multimedia, ACM, pp. 877884.
- Potthast, M. and Becker, S.: 2010, Opinion summarization of web comments, Ad- vances in Information Retrieval, Springer, pp. 668669.
- Sureka, A: 2011, Mining user comment activity for detecting forum spammers in youtube, arXiv preprint arXiv:1103.5044 .

References

- Nahm, U. Y. and Mooney, R. J.: 2002, Text mining with information extraction, AAAI 2002 Spring Symposium on Mining Answers from Texts and Knowledge Bases, Vol. 1.