ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

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 10, October 2018

Detecting Media Piracy using AI, ML and Data Mining

Prof. Varsha Dange[1], Kasturi Khedkar[2], Shubhangi More[3], Pavitra Phand[4], Poonam Pawar[5]

Associate Professor, Department of Computer Engineering, Dhole Patil College of Engineering, Kharadi, Pune, India[1]

U.G. Student, Department of Computer Engineering, Dhole Patil College of Engineering, Kharadi, Pune, India[2,3,4,5]

ABSTRACT: Web Technologies are continuously evolving which helps creators in marketing and distribution of media content. Although these technologies benefit the creator and the consumers, it also opens up immense chances of piracy and redistribution. Besides piracy torrent traffic can be a challenge for networks and their management due to flash crowds. People have so many ways to share the content like Social Networking Portals, Free Cloud Spaces and Drives, Email, Chats etc. Detecting and stopping the piracy of content manually is out of the question. We can leverage Artificial Intelligence and Machine Learning to fight piracy by using content monitoring solutions. In this paper, we explain how to use the different Data Mining techniques to search the web and identify piracy threats. These threats must be managed faster and more efficiently. Using different web services and machine learning, the system will produce statistical report of media piracy with various information like; source of piracy, IP Addresses, region, time period etc. Also the system will store or blacklist all the untrusted websites and portals. This database can be used by Private or Government Agencies to get rid of the Piracy.

KEYWORDS:Artificial Intelligence, Information Extraction, Randomized Search, Supervised Learning, Data Cleaning.

I. INTRODUCTION

Artificial Intelligence and Machine Learning to fight piracy by using content monitoring solutions. Data Mining techniques will be used to search the web and identify piracy threats.It will produce statistical report of media Piracy.This database can be used by Private or Government Agencies to get rid of the Pirwork. Web Technologies are continuously evolving which helps creators in marketingand distribution of media contents. Besides piracy torrent traffic can be a challenge for networks and their management due to ash crowds.Detecting and stopping the piracy of content manually is out of the question. Implementing a Media Content Monitoring Solutions using Artificial Intelligence, Machine Learning and Data Mining to fight Media Piracy, it will searchand identify for the pirated content on the Web.

II. RELATED WORK

In the Working system, we have three modules Admin UI layer,Web Crawler and application layer.The Admin UI layer consists of user and Admin registration and from this layer contents will be searched.In the Web Crawler layer the DNS will be fetched and parsed and one by one the contents will be saved in content DB and from there if the contains is found then using AI and ML algorithm URL will be filtered, if the duplicate data is found then from the URL data set,duplicate data is eliminated.In the Application layer a list of websites will be displayed.In the central database,the websites IP,domain,location will be stored and the website will be manually checked,from their websites will be categorized Blacklisted and Whitelisted and accordingly statistics and reports will be generated and send to be Admin.

DOI: 10.15680/IJIRCCE.2018. 0610037

Modules:

A. Admin Module: It consists of mainly Reports.The report is generated fromthe updated database. It includes Black Listed Websites and White ListedWebsites.

B. Black Listed Websites: Black Listed Websites consists of mainly those siteswhich are not legal, and from which we can download the content illegally thosewebsites are displayed in the Black Listed Websites.

C. White Listed Websites: White ListedWebsites consists of mainly those siteswhich are legal, and from which we can download the content legally. So thosewebsites which are legal those are displayed under the White Listed Website.

D. Database: All the White Listed and Black Listed websites are stored in thedatabase.

ProposedSystem Architecture:

The following Fig.1 gives a brief idea about the system architecture:

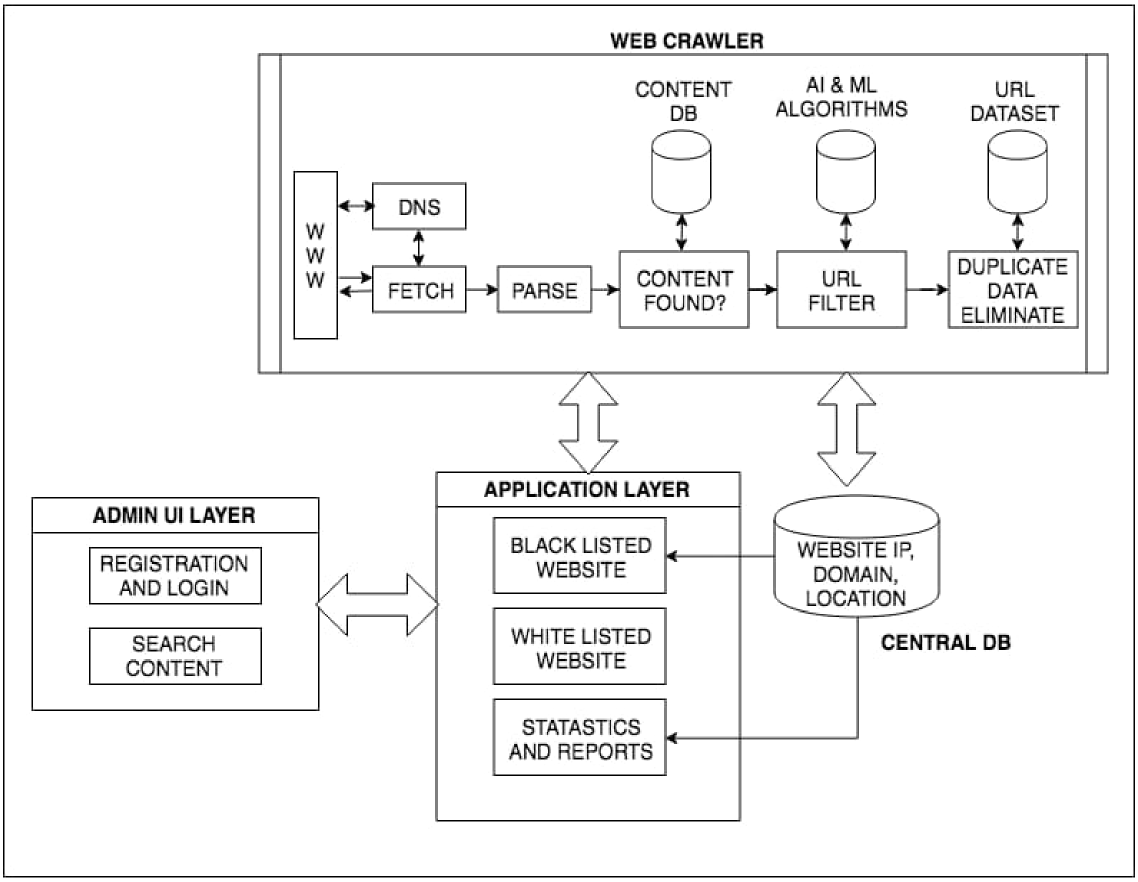

Fig.1.Proposed System Architecture

III.PROPOSED ALGORITHM

A. *Design Considerations:*

LRU Page Replacement Algorithm: The least recently used (LRU)replaces the pagein memory that has been used for the longest period of.This algorithm goes well with the principle of locality.A page that has not been used for a long time is least likely to be referenced in the near future. This algorithm can be implemented by maintaining the backward distance of each page.Whenever a page is referenced its backward distance is set to zero. Backward distance of other pages is incremented by 1. Replace the page in memory that has not to be used for the longest period of time.

B. *Algorithm:*

- **Step 1:** Read initial value i.e., number of frames, length of reference string and reference string.
- **Step 2:** initialize array to -1, indicating that the frames are empty.
- **Step 3:** change array size to 0, indicating it will be used for storing backward distance.
- **Step 4 :** For each page reference i in the reference string, if i not in memory and frame=empty then

  ```
  empty frame=i;
  else if
  i not in memory and frame!=empty then
  longest page distance= i;
  i=0;
  else if
  i in memory then
  i=0;
  else
  i for each page=1;
  ```
- **Step 5:** display result.
- **Step 6:** end.

IV.RESULTS

We are making a System which is Web Application and it will help in detecting Media Piracy on the web.In this System it will search for a Content on the Search Engine and give the output as the list of websites on which the content was found.Then the user will manually check whether the website is a Black listed or White listed website and accordingly create report and send to the Admin from where the Notifications or the Warnings related to the illegal use of the website will be send to the owner of that website and hence reducing the Media Piracy.

V. CONCLUSION AND FUTURE WORK

It will be presented how AI, ML and DM can contribute to fighting the piracy.It will be applicable in the implementation of large scale content monitoring system that track and identify illegal distributed content.Using different web services and machine learning, the system will produce statistical report of media piracy with various information like; source of piracy, IP Addresses, region, time period etc. Also the system will store or blacklist all the untrusted websites and portals. In future we can analyze the previous searched illegal content.We can provide statistical report. In future,we will work on handling large data set.

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

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 10, October 2018

REFERENCES

1. Andri Lareida, Burkhard Stiller,"Bit Torrent Measurement: A Country, Network and Content-Centric Analysis of Video Sharing in BitTorrent", IEEE, 978-1-5386-3416-5/18/31.00 c 2018 IEEE 2017.

2. Tobias HoBfeld,"The Bit Torrent Peer Collector Problem", 978-3-901882-89-0 @2017 1F1P 2017

3. Milosh Stolikj, Dmitri Jarnikov, Andrew Wajs,"Artificial Intelligence For Detecting Media Piracy Digital Object Identifier", 10.5594/JMI.2018.2827181 Date of publication: 22 June 2018.

4. D. Leporini, "Architectures and Protocols Powering Illegal Content Streaming over the Internet", Proc. Int. Broadcasting Convention, p.7, 2015.

5. AKarpathy et al.,"Large-Scale Video Classification with Convolutional Neural Networks", Proc. IEEE Conf. Comput. Vision Patt.Recogn., pp. 17251732, 2014.

6. By Milosh Stolikj, Dmitri Jarnikov, and Andrew Wajs, "Artificial Intelligence for Detecting Media Piracy", in 2018.

7. R. Cuevas, N. Laoutaris, X. Yang, G. Siganos, and P. Rodriguez, "Deep Diving into BitTorrent Locality", IEEE INFOCOM 2011, Shanghai,China, April 2011.

8.M Jain, H. Jegou, and P. Gros. "Asymmetric hamming embedding: taking the best of our bits for large scale image search", In ACM Multimedia, pages 1441–1444, 2011.

9. H Jegou and O. Chum." Negative evidences and co- occurences in image retrieval: The benefit of pca and whitening". In ECCV, pages 774–787, 2012.

10. HJegou, M. Douze, and C. Schmid, "Hamming embedding and weak geometric consistency for large scale image search", In ECCV, 2008.

11. H. Jegou, F. Perronnin, M. Douze, J. S´anchez, P. P´erez, and C. Schmid, "Aggregating local image descriptors into compact codes", IEEE Transactions on Pattern Analysis and Machine Intelligence, 34(9):1704–1716, 20128536

12. B. Cohen, "Incentives Build Robustness in BitTorrent", Workshop on Economics of Peer-to-Peer Systems, Berkeley, CA, USA, June 2003.

13. V. Burger, D. Hock, I. Scholtes, T. Hoßfeld, D. Garcia, M. Seufert, "Social Network Analysis in the Enterprise: Challenges and Opportunities", in Socioinformatics-The Social Impact of Interactions between Humans and IT, Springer, pp. 95–120, 2014.