

International Journal of Innovative Research in Science, Engineering and Technology

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

Visit: www.ijirset.com

Vol. 7, Special Issue 4, April 2018

Process Based Online Contents with Offensive Content Detection

Dinesh B¹, V. P. Sreekanthakumar²

B. Tech, Department of Information Technology, K.C.G. College of Technology, Chennai, India

Associate Professor, Department of Information Technology, K.C.G. College of Technology, Chennai, India

ABSTRACT: Online learning is rapidly replacing traditional, face-to-face lectures in current world. As technology improves, this trend will likely continue and accelerate. This paper is aimed towards providing an open source application for learning concepts through internet. Every user can learn and explore their thoughts via posts and queries. The contents can be searched within the application and also from all over web. Automated content monitoring for offensive contents is one of the main feature of this system. Any contents that are uploaded is first checked for offensive contents, if any such contents are found, the user will be blocked. Only after checking the contents the data is uploaded to the application and stored in the database. This is done using the Naive Bayes Algorithm.

I. INTRODUCTION

Today's application development is focused on the web environment. With spreading of cloud computing the popularity of web services has raised even more. Cloud computing is model for enabling ubiquitous, convenient, on-demand network access to computing resources that can be easily increased or decreased with minimal effort of service provider [5]. The topic of this work is to analyze possibilities of combining the advantages of service-oriented architecture deployable in the cloud environment and process-based composition of services for design and implementation of flexible software as a service (SaaS) applications [1]. As the number of students grows and courses - especially in massive open online courses (MOOC) context - could be nowadays offered for hundreds or even thousands of students simultaneously, there is natural demand for support of educational processes in order to automatize them [1].

This paper is aimed towards providing a process based online learning content application for the use of learners and tech experts. Anybody can share their content or tutorial by posting it into the application and also learn from the application by searching and queries. Content filtering is used to secure the posts and also the queries. Content filtering is done to all types of posts like text [2], image[3], or video[4]. Automated moderation applies machine learning to cost-effectively moderate at scale, without human involvement. Bing search algorithm is used to gather information from all sources of the web [1].

II. ANALYSIS

The process-based services for online learning contents web services and design of the application can deployed in the cloud. Cloud server can be used for scalability and cost effectiveness. With spreading of cloud computing the popularity of web services raises even more. Cloud computing is model for enabling ubiquitous, convenient, on-demand network access to computing resources that can be easily increased or decreased with minimal effort of service provider.

REST [1] is one of the potential web services. They are based on resources that are uniquely identified with Unified Resource Identifiers (URI). Resources can be represented in various formats like XML, JSON3 and others. Operations over those resources are carried by keywords GET, POST, PUT or DELETE. There is no exact description of REST service, so developers should rely on well documented API. In this system, REST API is an interface which is in the form of URL and uses HTTP methods (GET, POST, etc.) for data transmission from and to the application.

International Journal of Innovative Research in Science, Engineering and Technology

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

Visit: www.ijirset.com

Vol. 7, Special Issue 4, April 2018

A. Textual Content Filtering [2]

Naive Bayes Classifiers are a family of simple probabilistic classifiers based on applying Bayes theorem with strong (naive) independence assumptions between the features. In machine learning Naïve Bayes remains a popular (baseline) method for text categorization, the problem of judging documents as belonging to one category or the other (such as spam or legitimate, sports or politics, etc.) with word frequencies as the features. With appropriate pre-processing, it is competitive in this domain with more advanced methods including support vector machines. Naive Bayes classifiers are highly scalable, requiring a number of parameters linear in the number of variables (features/predictors) in a learning problem. Maximum-likelihood training can be done by evaluating a closed-form expression, which takes linear time, rather than by expensive iterative approximation as used for many other types of classifiers. In this system, Naïve Bayes is used for text classification into spam or legitimate. In machine learning, Naive Bayes has got one main advantage i.e. it only requires a small number of training data for classification.

B. Image Content Filtering[3]

Support Vector Machines (SVM) in machine learning are supervised learning models with associated learning algorithms that analyze data used for classification and regression analysis. Given a set of training examples, each marked as belonging to one or the other of two categories, an SVM training algorithm builds a model that assigns new examples to one category or the other, making it a non-probabilistic binary linear classifier. An SVM model is a representation of the examples as points in space, mapped so that the examples of the separate categories are divided by a clear gap that is as wide as possible. New examples are then mapped into that same space and predicted to belong to a category based on which side of the gap they fall. When data are not labeled, supervised learning is not possible, and an unsupervised learning approach is required, which attempts to find natural clustering of the data to groups, and then map new data to these formed groups. The clustering algorithm often used in industrial applications either when data are not labeled or when only some data are labeled as a preprocessing for a classification pass. In this system, SVM is used for classification of images. Experimental results show that SVMs achieve significantly higher search accuracy than traditional query refinement schemes after just three to four rounds of relevance feedback. This is also true of image segmentation systems, including those using a modified version SVM that uses the privileged approach as suggested by Vapnik. Support vectors are the data points nearest to the hyperplane, the points of a data set that, if removed, would alter the position of the dividing hyperplane. Because of this, they can be considered the critical elements of a data set. As a simple example, for a classification task with only two features (like the image above), you can think of a hyperplane as a line that linearly separates and classifies a set of data. Intuitively, the further from the hyperplane our data points lie, the more confident we are that they have been correctly classified. We therefore want our data points to be as far away from the hyperplane as possible, while still being on the correct side of it. So, when new testing data is added, whatever side of the hyperplane it lands will decide the class that we assign to it. The distance between the hyperplane and the nearest data point from either set is known as the margin. The goal is to choose a hyperplane with the greatest possible margin between the hyperplane and any point within the training set, giving a greater chance of new data being classified correctly.

C. Video Content Filtering[4]

Support Vector Machines in machine learning are supervised learning models with associated learning algorithms that analyze data used for classification and regression analysis. The features are extracted from spatial and temporal domains and used for classification of the video segments into two classes based on different SVM classifiers. The two classes are adult and non-offensive video content. The features used are short duration and camera motion from the temporal domain, and skin detection and color histogram from the spatial domain. Two data sets of 7 and 15 hours of video material are used, which achieves an accuracy of 94.44%. Two SVM algorithms are used LibSVM and Sequential Minimal Optimization for training an SVM (SMO). The performances of these two algorithms are compared using Weka.

The goal of the present work is to demonstrate that it is possible to achieve rather high accuracy adult content classification using simple but well-chosen spatial temporal features.

International Journal of Innovative Research in Science, Engineering and Technology

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

Visit: www.ijirset.com

Vol. 7, Special Issue 4, April 2018

In particular we find the combination of short duration and camera motion from the temporal domain, together with skin detection and color histogram representation extracted from the pixel domain lead to good classification results. The classifiers are based on machine learning techniques, explicitly support vector machines (SVM).

III. SYSTEM IMPLEMENTATION

Process based services for online learning content is aimed towards providing an open source application for learning concepts through internet. Every user can use the application to learn and explore their thoughts via posts and queries. One can search all their need for learning in a single application. Once the user registers in the application an account is maintained for that particular user that maintains all the posts and queries made by them. In case the user forgets the password, random password will be sent after verifying the user id or the phone number. In this system user can post their teaching in any format i.e.; text or photo or video. All these posts are maintained in the database. Queries can be raised on the posts to which corresponding answers can be given. The user will be continuously notified on the posts and the queries on the posts. Also, about other posts related to the same domain. Unanswered queries also notified to be answered. The system can be implemented in web to make it accessible for every web users. The contents can be searched within the application and also from all over web which increases the availability of solution to a particular question. Automated content monitoring for offensive contents is one of the main feature of this system. Any contents that are uploaded is first checked for offensive contents, if any such contents

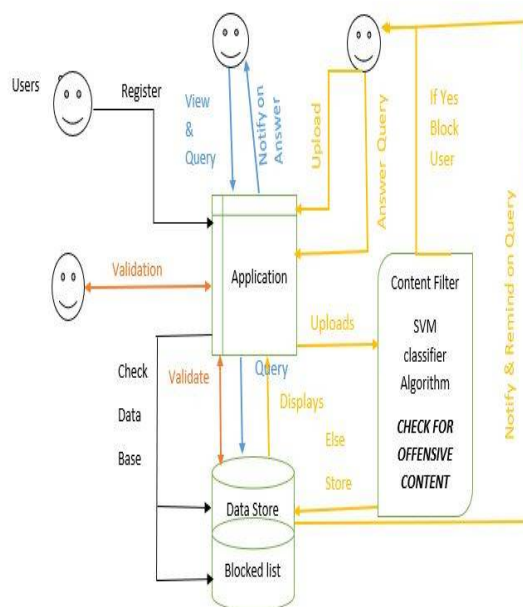


Fig No: 1 SYSTEM DIAGRAM

The proposed system has the following modules:

- i. User profile management

The proposed system collects the user details and the domains of interest during the time of user registration and stores the related data into the database. Authentication and the profile management is done using the application database. The user profile displays the options of change of user name, passwords and domain of interests. The profile screen also displays the list of posts posted by the single user, also the queries are displayed based on the users.

International Journal of Innovative Research in Science, Engineering and Technology

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

Visit: www.ijirset.com

Vol. 7, Special Issue 4, April 2018

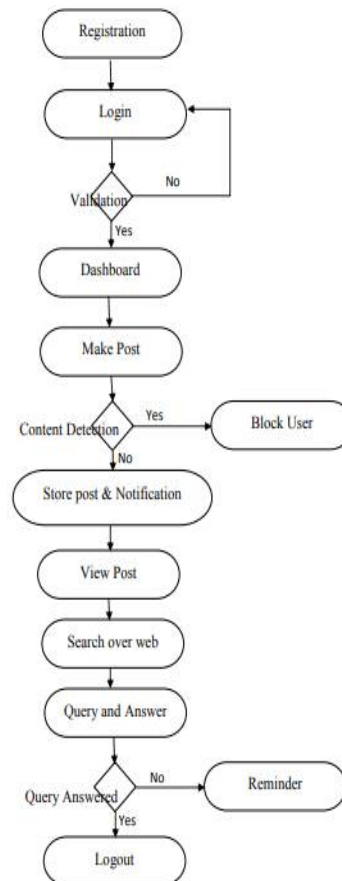


Fig No. 2 FLOW CHART

ii. Post of content & Queries

The screen consists of the posts made by the host users for the viewers, the post may be of any file format, and these may be tutorials that are used to learn through online. It lists the posts in order of user interested domain, with an option to post a tutorial newly. New post option is available to post the tutorial by tagging the domain to which the post belongs to. The viewers have an option of enrich their knowledge by posting queries. These posts and queries are mapped to content filtering and to the databases. They are retrieved from the storage table using RESTful services and the data are visualized in the application.

iii. Content filtering

The content filtering is a module that is hidden between the application and database. This uses the Machine Learning algorithm and SVM classifier to classify the data from the offensive content detection. The classifier splits the input data into two sets of classes and separates them with an enlarged boundary. The algorithm is framed and formatted in the form of API and this API is used in the application to make the system automated.

International Journal of Innovative Research in Science, Engineering and Technology

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

Visit: www.ijirset.com

Vol. 7, Special Issue 4, April 2018

```
{
  "AdultClassificationScore": 0.0324602909386158,
  "IsImageAdultClassified": false,
  "RacyClassificationScore": 0.065064750611782074,
  "IsImageRacyClassified": false,
  "AdvancedInfo": [],
  "Result": false,
  "Status": {
    "Code": 3000,
    "Description": "OK",
    "Exception": null
  },
  "TrackingId": "WU_m2tacsanjeev_1bf77f58-dd85-4db0-93a6-d62f1f7f5"
}
```

Fig No. 2 JSON OUTPUT.

iv. Search

The search operation used in the proposed system is not only responsible for searching the content within the system but also searches the content all over the web for better solution for the search. It features the links of the related contents exactly in the application. Since all the need to search is done in a single user interface.

v. Data base

A database-management system (DBMS) is a computer-software application that interacts with end-users, other applications, and the database itself to capture and analyse data. A general-purpose DBMS allows the definition, creation, querying, update, and administration of databases. The proposed system uses MySQL database for storing and processing. The data are stored in a definite structure to make the processing more efficient.

vi. Notifications & Reminders

Once a new post is posted by a host user on a particular domain the application extracts the data from the application and forward to the database and map these values with the user interested domain and sends notification to the user who are all linked with the domain of post. On the part query a viewers can clarify their doubts on a post by posting queries to a particular the host of the post is also notified on queries and on a particular time period reminder is sent for the un answered queries.

IV. CONCLUSION

In this paper, we have designed and implemented a web application for online learning with content filtering to secure all the posts. This will have a huge impact on the learners and the tech experts who love to learn and teach from anywhere. This single application meets all the needs of a user i.e. one need not waste their time in searching materials and tutorials in different websites. Misusing the application can be controlled by the content filtering algorithm.

REFERENCES

- [1] G. Červeňák, M. Paralič and P. Ragan (2014), "Process Based Composition of Web Services for Support of Flexible and Scalable Educational Application", 12th IEEE International Conference on Emerging eLearning Technologies and Application pp379-384.
- [2] TaekyongNam, Youngsoo Kim (2006), "An Efficient Text Filter for Adult WebDocuments", ICACT, pp 438-440.
- [3] Rainer Lienhart, Rudolf Hauke (2008), "Filtering Adult Image Content with TopicModels", IEEE ICME pp 1472-1475.
- [4] FaouziAlayaCheikh, SuleYildirimYayilgan, Victor M. Torres Ochoa (2012),"Adult video content detection using Machine Learning Techniques", EighthInternational Conference on Signal Image Technology and Internet Based Systems,pp 967-974.
- [5] Indragandhi V, Logesh R, Subramaniaswamy V and Vijayakumar V (2015),"Automated Message Filtering System in Online Social Network", 2nd InternationalSymposium on Big Data and Cloud Computing (ISBCC'15).
- [6]SubramaniaswamyVa, LogeshRb,VijayakumarVc, and IndragandhiVd 2015, "Automated Message Filtering System in Online Social Network " ELSEVIER.
- [7]Victor M. Torres Ochoa, SuleYildirimYayilgan, FaouziAlayaCheikh 2014, "Adult video content detection using Machine Learning Techniques" IEEE.