International Journal of Innovative Research in Science and Engineering
Vol. No.3, Issue 05, May 2017
www.ijirse.com
ISSN (O) 2454-9665
ISSN (P) 2454-0663

SENTIMENTAL ANALYSIS ON E-GOVERNANCE

Diksha Joshi[1], Mayuresh Khalegaonkar[2]

Mohini Lohikpure[3], Preeti Maan[4], Prof. R. A. Deshmukh[5]

[1,2,3,4,5]*Computer Department, Rajarshi Shahu College of Engineering (India)*

ABSTRACT

Sentiment analysis or opinion mining is the computational study of people's opinions, sentiments, attitudes, and emotions expressed in written language. It is one of the most active research areas in natural language processing and text mining in recent years. Its popularity is mainly due to two reasons. First, it has a wide range of applications because opinions are central to almost all human activities and are key influencers of our behaviors. Whenever we need to make a decision, we want to hear others' opinions. Hence sentiment analysis is very necessary.

One of the challenges in delivering e-government services is to design the Web sites to make it easier for citizens to find desired information. However, little work is found to evaluate e-government services in this sense. In addition, current efforts on government Web site design mainly concentrate on Web site features that would enhance its usability, but few of them answers why some Web design is better than others to facilitate citizens information seeking. This project aims to contribute to both aspects: it equips government agencies with a model that can not only evaluate their Web-based e-government services, but also helps them understand why their Web sites succeed or fail to help citizens find needed information. In addition to the model itself, instruments for applying this model are also developed.

Keywords : E-Governance, Extract, Natural Language Processing, Sentiment Analysis

I. INTRODUCTION

One-Stop-Government refers to the integration of public services from a citizens or customer of public services point of view.[2] Online one-stop government allows citizens to have 24 hours access to public services from their home or even on the move. Currently however online one-stop government is at its first only steps. Ideally, online one-stop government requires that all public authorities are interconnected and that the citizen is able to access public services by a single point even if these services are actually provided by different departments or authorities.

In this project GUI will be based on the basis of product tool where admin and user will be the 2 user roles that are admin role and user role .Opinion mining is a rising field of data mining used to extract the knowledge from huge volume of online resources like consumer interpretation, feedback and reviews on any product or topic etc. A lot of work has been conducted to mine opinions in form of document, sentence and feature level sentiment analysis. Sentiment analysis helps in determining the attitude of people regarding various contexts, products etc. This paper

ISSN (O) 2454-9665
ISSN (P) 2454-0663

basically discusses the sentiments of the citizens regarding e-governance projects. The E-Governance plans benefit at wider level from villages to cities. We have discussed how opinion mining techniques help in recovery of information and relationships from textual data sources, thereby support policy makers in discovering relations between policies and citizens. We present here, methodology of using opinion mining for e-governance decision support. We have tried to find out the reviews of different people regarding the upcoming Schemes provided by government. Our concern was to know whether the E-Governance is acceptable by the mass or not, and tried to know their expectations in this field. Admin will be the member of the government team where he/she will be able to add the Schemes and the particular Policies for the same. If the normal user logs in, then he/she will be able to view all the schemes and the particular policies. If the users like the policies that have been added by the admin then user will be also able to comment on it. Further the comments will get stored into text file and the text file will be stored into Hadoop for storage purpose. If the new user logs in for the first time then the most positive commented policy or the scheme is showed first as a recommendation for the same. The Comments text will be classified into the basis "POSITIVE", "NEGATIVE " and "NEUTRAL".

1.1 Proposed System

To design a portal for Government Schemes[1] that implies to develop an online application for citizens through which citizens can know different schemes that are provided by government and they can apply for those on same portal submitting required documents. The citizens can register on this portal and get the confirmation and notifications. Citizens will be able to comment on Policies and Schemes. Whenever User checks any Policies or Schemes her/she will first check Reviews regarding those. So this proposed system provides facility to user to comment on any Schemes and Policies added by Admin.

In Proposed System we are going to add schemes like:

1. Rojgar Yojna.
2. Mamta Yojna.
3. KrushiVikas Yojna.

We are providing single Government Portal which uses Big Data that is HADOOP, which is used to store large number of comments about Schemes and Policies. HADOOP uses NLP for Sentiment Analysis. Due to Sentiment Analysis New User (First time login) will be able to see Positive comments about Policies and Schemes. System will be able to store large number of Data because of HADOOP.

1.2 Technologies Used

1.2.1 HADOOP

Hadoop is an open-source framework that allows to store and process big data in a distributed environment across clusters of computers using simple programming models. It is designed to scale up from single servers to thousands of machines, each offering local computation and storage. Hadoop is an Apache Open Source framework written in java that allows distributed processing of large datasets across clusters of computers using simple programming

ISSN (O) 2454-9665
ISSN (P) 2454-0663

models. A Hadoop frame-worked application works in an environment that provides distributed storage and computation across clusters of computers. Hadoop is designed to scale up from single server to thousands of machines, each offering local computation and storage.

1.2.2 HDFS Storage

HDFS holds very large amount of data and provides easier access. To store such huge data, the files are stored across multiple machines. These files are stored in redundant fashion to rescue the system from possible data losses in case of failure. HDFS also makes applications available to parallel processing.

Features of HDFS :

It is suitable for the distributed storage and processing.

Hadoop provides a command interface to interact with HDFS.

The built-in servers of name node and data node help users to easily check the status of cluster.

Streaming access to file system data.

HDFS provides file permissions and authentication.

1.2.3 Sentiment Analysis Using NLP

Sentiment analysis also known as opinion mining refers to the use of natural language processing, text analysis and computational linguistics to identify and extract subjective information in source materials. Sentiment analysis is widely applied to reviews and social media for a variety of applications, ranging from marketing to customer service.[3]

Generally speaking, sentiment analysis aims to determine the attitude of a speaker or a writer with respect to some topic or the overall contextual polarity of a document. The attitude may be his or her judgment or evaluation, affective state, or the intended emotional communication (that is to say, the emotional effect the author wishes to have on the reader).

Sentiment is not analyzed via artificial intelligence, as some people may be tempted to think. Rather, it is analyzed via a systematic process that involves the use of a sentiment lexicon. This lexicon assigns a degree of positivity or negativity to a word by itself that is then used to give meaning to the entirety of the article. This is a way of analyzing sentiment, then, by considering a type of inherent positivity or negativity of each word that would be used by someone to talk about your business or products. For example, 'happy' would be deemed a positive word, as well as 'like' and 'love'. At the opposite end of the spectrum we can see words like 'hate', 'dislike', etc.

1.3. Sentiment Analysis

Sentiment analysis is a task that involves information extraction from customer feedback and other authentic sources like survey agencies. As the word suggests it includes detecting sentiments of any individual from the text that is writes in digital format. There are a wide array of applications of this concept. This concept became centre of attention since industry got revolutionized with the change in paradigm of Seller's Market to Buyer's Market in order to capture market share.

1.3.1 Classification

Classification is a stage in sentiment analysis that can described as a process in which we predict qualitative

ISSN (O) 2454-9665
ISSN (P) 2454-0663

response, or in this case we classify the document into its polarity. Predicting a qualitative response of a document can be referred to as classifying the document since it involves since it involves assigning an observation to a category or class. There are many possible classification techniques, or classifiers that one might use for to predict the qualitative response or class of a document. In sentiment analysis some widely used classification techniques are as follows:

Naive Bayes Classifier

Max Entropy Classifier

Boosted Trees Classifier

Random Forest Classifier

1.4 ALGORITHM

Step 1: Login with two user roles.

Step 1.1: Login with Admin.

Step 1.2: Login with User.

Step 2: Add Schemes and Government Policies when logged in by Admin.

Step 3: call RC function

Step 2.1: Get U (Schemes and Policies) as Input to WC.

Step 2.2 : for i=0 to MAX

//MAX = Maximum no of policies

Step2.3 : visit i (E-Government site) when logged in by normal user .

Step 2.4 : go to step 2.2 till MAX

Step 2.5 : Output as CP is able to view all policies and schemes.

Step 4: call to IE Function

Step 4.1 : Get CP (Comment) as Input .

Step 4.2 :Call Function NLP

Step 4.3: Process NLP as Removing Stop Word.

Step 4.4 : Get Relevant Information of positive, negative and neutral comments

Step5: Display Result in the graphical form.

Step 6: Stop.

III. LITERATURE SURVEY

1.Sentiment Crawling: Extremist Content collection through a sentiment analysis guided web-crawler

With the constant growth of extremist networks and extremist content, unguided methods of data collection and analysis are becoming less and less feasible. There is simply not enough funding or manpower available to allow law enforcement agencies to be able to manually sort through these large amounts of data. As such guided data collection has grown more and more vital. The methods in this study were designed aid in remedying this problem

ISSN (O) 2454-9665
ISSN (P) 2454-0663

by expediting the process of collection and simultaneously collecting specific types of data with set parameters involving extremism.

This study sought to examine a method to classify and interpret extremist content by combining sentiment analysis with the TENE web-crawler as well as classification software. By dividing specific websites into one of four different classes it was focused on determining if sentiment software could effectively apply sentiment to data obtained from the crawler in a way that would allow for the creation of a decision tree, specifically a decision tree that could effectively differentiate between the four classes. After collecting several thousand pages using the web-crawler, 34 nouns were extracted from the various crawled datasets, and the sentiment around those same nouns calculated for all datasets. WEKA's J48 algorithm was then applied to the resulting sentiment data to create a decision tree and to show the percentage of correctly identified cases present in the sample. The 80% correctly analyzed pages from the WEKA analysis shows a promising future for this approach and more specifically the 92% correctly identified pro-extremist pages show that there is a future in identifying extremist content using this approach. The secondary analysis of the pages outside of the analyzed set shows a lower rate of accuracy, although it is still clear that the approach has produced significant results.

2. Concept-Level Sentiment Analysis: A World Wide Web Conference 2014 Tutorial

The WWW'14 tutorial on Concept-Level Sentiment Analysis aims to provide its participants means to efficiently design models, techniques, tools, and services for concept-level sentiment analysis and their commercial realizations. The

tutorial draws on insights resulting from the recent IEEE Intelligent Systems special issues on Concept-Level Opinion and Sentiment Analysis1 and the IEEE CIM special issue on Computational Intelligence for Natural Language Processing

The tutorial includes a hands-on session to illustrate how to build a concept-level opinion-mining engine step-by-step, from semantic parsing to concept-level reasoning.

Sentiment extraction aims to extract and group aspect and opinion words from online reviews. Previous works usually extract aspect and opinion words by leveraging association between a single pair of aspect and opinion word, but the structure of aspect and opinion word clusters has not been fully exploited.

In this paper, we investigate the aspect-opinion association structure, and propose a _rst clustering, then extracting" unsupervised model to leverage properties of the structure for sentiment extraction. For the clustering purpose, we formalize a novel concept syntactic distribution consistency as soft constraint in the framework of posterior regularization; for the extraction purpose, we extract aspect and opinion words based on cluster-cluster association. In comparison to traditional word-word association, we show that cluster-cluster association is a much stronger signal to distinguish aspect (opinion) words from non-aspect (non-opinion)words. Extensive experiments demonstrate the effectiveness

of the proposed approach and the advantages against state-of-the-art baselines.

ISSN (O) 2454-9665
ISSN (P) 2454-0663

3. An Evaluation of Machine Translation for Multilingual Sentence-level Sentiment Analysis

Sentiment analysis has become a key tool for several social media applications, including analysis of user's opinions about products and services, support to politics during campaigns and even for market trending. There are multiple existing sentiment analysis methods that explore different techniques, usually relying on lexical resources or learning approaches. Despite the large interest

on this theme and amount of research efforts in the field, almost all existing methods are designed to work with only English content. Most existing strategies in specific languages consist of adapting existing lexical resources, without presenting proper validations and basic baseline comparisons. In this paper, we take a different step into this field. We focus on evaluating existing efforts proposed to do language specific sentiment analysis. To do it, we evaluated twenty-one methods for sentence-level sentiment analysis proposed for English, comparing them with two language-specific methods. Based on nine language-specific datasets, we provide an extensive quantitative analysis of existing multi-language approaches. Our main result suggests that simply translating the input text on a specific language to English and then using one of the existing English

methods can be better than the existing language specific efforts evaluated. We also rank those implementations comparing their prediction performance and identifying the methods that acquired the best results using machine translation across different languages. As a final contribution to the research community, we

release our codes and datasets. We hope our effort can help sentiment analysis to become English independent.

4. Data Intensive Review Mining for Sentiment Classification across Heterogeneous Domains

Many factors contribute to complicate sentiment classification in user review mining. Linguistic nuances, first of all, such as the quality of lexicon and peculiarities in syntax, make classical text-mining approaches hardly applicable (from among the 23,628 terms used in one Amazon corpus, 12.030 terms appear once, mostly due to typos and grammar errors, and 15,202 appear only twice). In such a sparse scenario, drawing separating surfaces can be tricky for a sophisticated classifier. Another issue arises at the document level: the patterns lying closest to the separating surfaces draw class boundaries in SVM-related models; thus classification is ultimately ruled by the most uncertain reviews. Robust approaches, such as kernel logistic or the kNN method, seem preferable. Cognitive issues arise from the need for an established model of the text-to-sentiment mapping; this supports holistic approaches, that integrate lexical and semantic information in a flexible manner.

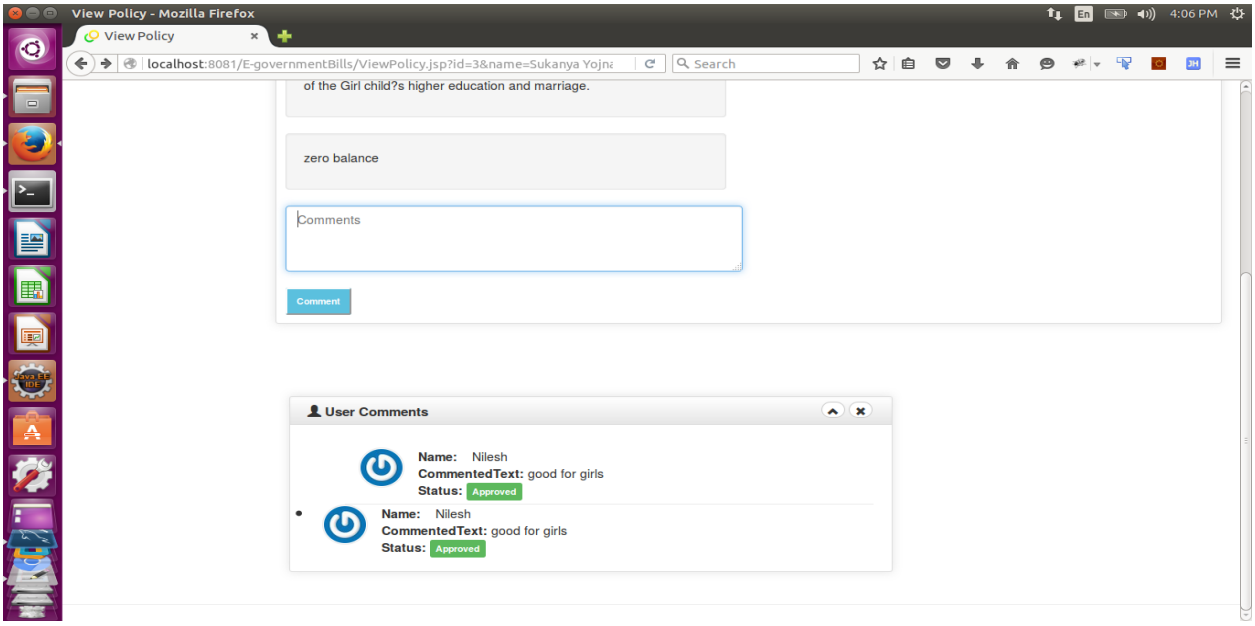
IV. RELATED WORKS

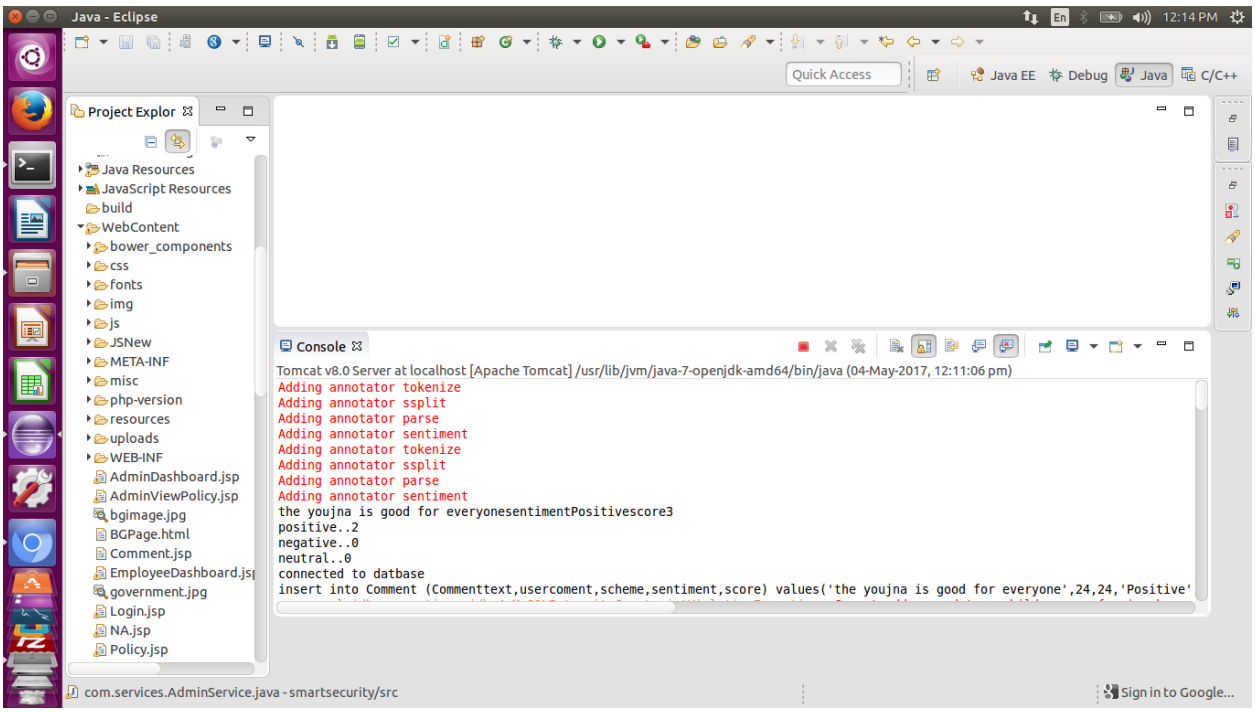
According to the European Commission the issue at stake is not that Member States should produce more Information, but that the information which is already available to the public should be clearer and more accessible to potential users. The e-Europe 2002 Action Plan suggests: e Government could transform old public organization and provide faster, more responsive services. It can increase efficiency, cut costs, increase transparency and speed up standard administrative processes for citizens and business similar project was implemented in us on small scale

ISSN (O) 2454-9665
ISSN (P) 2454-0663

basis. Which was not that successful. Also following UK countries like France also tried the similar concept. In India also official government websites are present .But only information about schemes are present not all the forms and detail procedure.

V. RESULT





ISSN (O) 2454-9665
ISSN (P) 2454-0663

VI. FIGURES

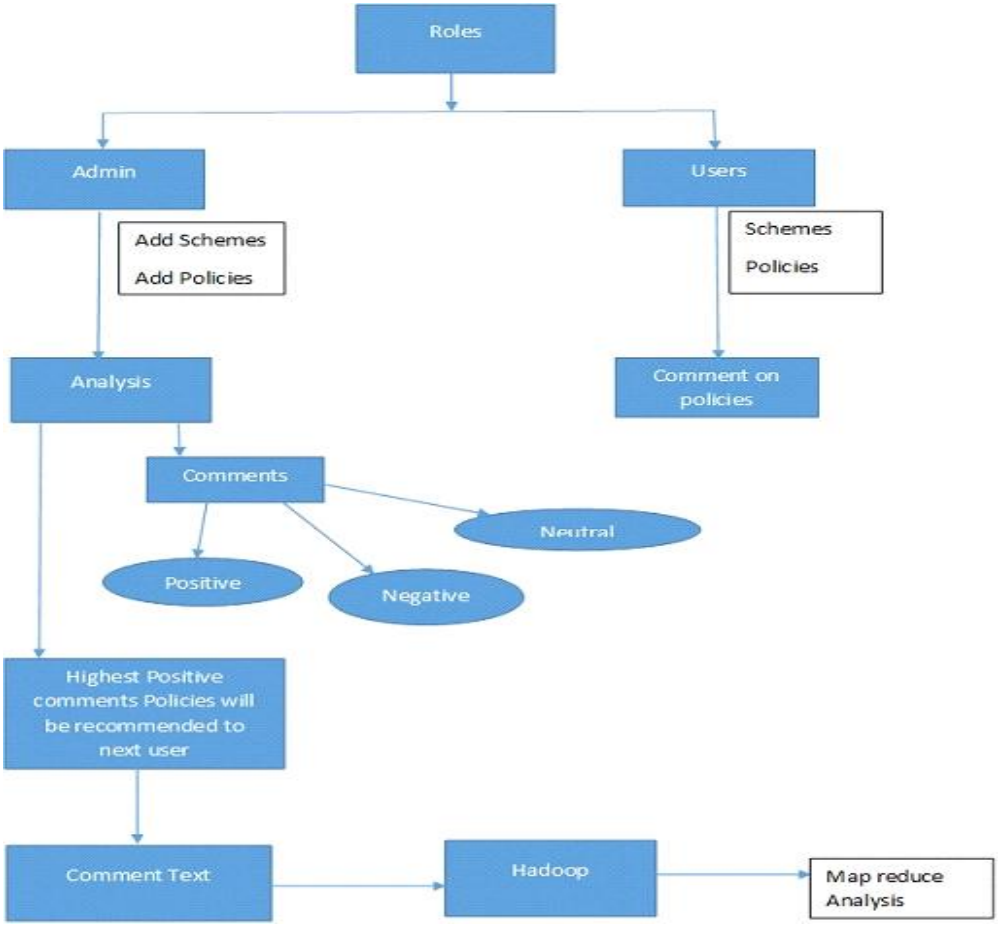


Fig: System Working / Architecture

VII. CONCLUSION

Opinion mining is a rising field of data mining used to extract the knowledge from huge volume of online resources like consumer interpretation, feedback and reviews on any product or topic etc. A lot of work has been conducted to mine opinions in form of document, sentence and feature level sentiment analysis. Sentiment analysis helps in determining the attitude of people regarding various contexts, products etc. This paper basically discusses the sentiments of the citizens regarding e-governance projects. The E-Governance plans benefit at wider level from villages to cities. We have discussed how opinion mining techniques help in recovery of information and relationships from textual data sources, thereby support policy makers in discovering relations between policies and citizens. We present here, methodology of using opinion mining for e-governance decision support. We have tried to find out the reviews of different people regarding the upcoming Schemes provided by government. Our concern was to know whether the E-Governance is acceptable by the mass or not, and tried to know their expectations in this field.

ISSN (O) 2454-9665
ISSN (P) 2454-0663

VIII. ACKNOWLEDGEMENTS

We owe a great many thanks to a great many people who helped and supported us during our project work.

Our deepest thanks to the Guide of the project **Prof. R.A.Deshmukh** for guiding and correcting various documents with attention and care. She has taken pain to go through the project and make necessary correction as and when needed.

We would like to thank our HOD **Prof. S. V. Keddar** for providing us with a platform on which we could conduct research extensively on a topic of our choice.

We express our thanks to the Principal **Dr.R K.Jain**, for extending his support.

We would also thank our College and our faculty members without whom this project would have been a distant reality. We also extend our heartfelt thanks to our family and well-wishers.

REFERENCES

Papers:

[1] *Guidelines for Indian Government Websites*
An Integral Part of Central Secretariat manual of office procedure.
http://darpg.gov.in/sites/default/files/Guidelines_for_Government_websites.pdf

[2] *E-Governance using Sentiment analysi*s
http://ijariie.com/AdminUploadPdf/E_Governance_using_Sentiment_Analysis_ijariie1830.pdf

[3] *Sentiment Crawling: Extremist Content collection through a sentiment analysis guided web-crawler*
Joseph Mei, Richard Frank* International CyberCrime Research Center ,Simon Fraser University, School of Criminology , Burnaby, Canada {jmmei, rfrank}@sfu

[4] *Concept-Level Sentiment Analysis: A World Wide Web Conference 2014 Tutorial* Erik Cambria Temasek Laboratories , National University of Singapore, cambria@nus.edu.sg

[5] *Sentiment Extraction by Leveraging Aspect-Opinion Association Structure* Li Zhao, Minlie Huang, Jiashen Sun*, Hengliang Luo*, Xiankai Yang, Xiaoyan Zhu State Key Lab. of Intelligent Technology and Systems Tsinghua National Laboratory for Information Science and Technology ,Dept. of Computer Science and Technology, Tsinghua University, Beijing 100084, PR China

Books:

[6] *Data Mining: Concepts and Techniques*: Elsevier; Third edition (2007)

[7] *Natural Language Understanding*- Allen, James, Allen 1995: Natural Language Understanding – Contents , University of Rochester, New York / Ontario / Wokingham, U.K. / Sydney / Singapore / Tokyo / Madrid / San Juan 1995,The Benjamin/Cummings Publishing Company, Inc. , Redwood City, California • Menlo Park, California Reading, Massachusetts, New York • Don Mills, Ontario • Wokingham, U.K. • Amsterdam • Bonn • Sydney • Singapore • Tokyo • Madrid • San Juan , ISBN: 0-8053-0334-0