FEEDBACK PORTAL SYSTEM FOR UNIVERSITY USING NLP AND DATA ANALYTICS

A PROJECT REPORT

Submitted by

ANINDITA GUHA
HARSH PATEL
NITIN LALWANI
VATSAL MISTRY

In fulfilment for the award of the degree

of

BACHELOR OF TECHNOLOGY

in

COMPUTER AND SCIENCE DEPARTMENT

Under the Guidance of
Prof Jaydeep Viradiya

PARUL UNIVERSITY, VADODARA

October, 2018

PARUL UNIVERSITY

CERTIFICATE

This is to certify that Project-I -Subject code 031054016 of 7th Semester entitled "FEEDBACK PORTAL FOR UNIVERSITY USING NLP AND DATA ANALYTICS" of Group No. PUCSE_41 has been successfully completed by

ANINDITA GUHA - 150303105055
HARSH PATEL - 150303105127
VATSAL MISTRY - 150303105090
NITIN LALWANI - 150303105071

under my guidance in partial fulfillment of the Bachelor of Technology (B. TECH) in Computer and Science Engineering of Parul University in Academic Year 2018-2019.

Project Guide
Prof. Jaydeep Viradiya

Project Coordinator
Prof. Jaydeep Viradiya

Head of Department,
CSE
Prof. Harshal Shah

External Examiner

ACKNOWLEDGEMENT

Behind any major work undertaken by an individual there lies the contribution of the people who helped him to cross all the hurdles to achieve his goal.

It gives me the immense pleasure to express my sense of sincere gratitude towards my respected guide **Prof. Jaydeep Viradiya** (Asst. Prof, CSE), for his persistent, outstanding, invaluable co-operation and guidance. It is my achievement to be guided under him. He is a constant source of encouragement and momentum that any intricacy becomes simple. I gained a lot of invaluable guidance and prompt suggestions from him during entire project work. I will be indebted of him forever and I take pride to work under him. I also express my deep sense of regards and thanks to **Prof. Harshal Shah** (Asst. Prof. Head of CSE Department)., and Head of CSE Engineering Department. I feel very privileged to have had their precious advices, guidance and leadership.

Last but not the least, my humble thanks to the Almighty God.

Place: Vadodara

Anindita Guha
Harsh Patel
Vatsal Mistry
Nitin Lalwani

Date: 13/10/2018

150303105055
150303105127
150303105090
150303105071

ABSTRACT

Feedback system has proven to be one of the greatest assets for any organization from time immemorial. Feedbacks from customers or employee does not only consist of the positive response it also incorporates certain negative experiences. Both these responses help in improving the stance of any institute.

There has been a lot of experiments in the process of conducting feedbacks in the organizations. It all began with taking feedbacks in a written format, now it leaped further as people moved on to less cumbersome online format.

Our system is designed in a way that it automatically takes feedback from the users in their leisure time, periodically. The interaction is highly human friendly yet concise to the points. The system will be able to detect the semantic and emotion behind the feedbacks and segregate accordingly. The design will reduce the manpower required behind segregation of feedbacks and thus can focus more on working on the negative aspect.

TABLE OF CONTENTS

1. INTRODUCTION

Feedback plays a very important role in any system. Feedback system has been changing periodically with technological advancements. From verbal in person feedbacks to paper written feedbacks to google forms.

World today is surrounded by data. Every data can be considered as structured or unstructured.

Semantic analysis majorly deals with detecting and extracting semantics or meaning from a statement. We can get hidden insights. With these insights we can further predict the cause and nature of the statements.

PROBLEM DEFINITION

1.1 PROBLEM STATEMENT

To devise a system where a chat-bot will conduct feedback sessions from the students in a very amicable and human friendly way. This system will highly incorporate natural language processing, data analytics and machine learning.

1.2 SCOPE

1. Better productivity
2. Less time consuming
3. Friendly conversation
4. Fully automated system

1.3 OBJECTIVE

To reduce the human efforts in segregating the feedbacks according to the departments as well as nature.

This system will conduct feedback sessions in a human friendly way and hence the conversation will be light as well as more information regarding issues can be extracted from the students.

2. LITERATURE SURVEY

2.1 Research Paper on Analysis of student learning experience by mining social media data (IJESC); Akanksha A. Pande, S.A Kinariwala, Volume 7 Issue No. 5

Many issues like depression, suicide, anger, anxiety are increasing among students. These issues are necessary to seek out and analyze, but students never discuss their issues with anyone. Today Social media is very popular medium where individuals share their feeling and opinion. Students also terribly active on social sites like Facebook and Twitter. Their unceremonious discussion on social media (e.g. Twitter, Facebook) illuminates light on their educational experiences—vote, sentiment, opinions, feelings, and concerns about the learning process. Data from such environments can supply valuable information which is helpful knowledge to understand student learning experiences. Analyzing such data can be challenging. This Paper emphasized on student's twitter posts to learn problems in student life as well as positive things occurred in their educational life. First conducted a qualitative analysis on sample tweets related to student's college life. Students face issues such as heavy work load of study, lack of social engagement, and sleep deprivation, employment issue, etc. In this paper "positive things" happen in student's life is also taken in to consideration. To classify tweets reflecting student's problem multi label classification algorithms is implemented. Naïve Bayes and Linear Support Vector Machine Learning algorithms are used.

This paper keeps track and extracts sentiments form student's online media posts. It deals with learning about student's experiences and opinion. Since our project majorly deals with student's feedbacks and experiences, this paper can help us to extract them efficiently.

2.2 Research paper on Automatic Detection of Political Opinion in Tweets;
(IRJET) Diana Maynard, Adam Funk, Apr -2017

In this paper, they discuss a variety of issues related to opinion mining from micro posts, and the challenges they impose on an NLP system, along with an example application we have developed to determine political leanings from a set of pre-election tweets. While there are a number of sentiment analysis tools available which summarize positive, negative and neutral tweets about a given keyword or topic, these tools generally produce poor results, and operate in a fairly simplistic way, using only the presence of certain positive and negative adjectives as indicators, or simple learning techniques which do not work well on short micro posts. On the other hand, intelligent tools which work well on movie and customer reviews cannot be used on micro posts due to their brevity and lack of context. Our methods make use of a variety of sophisticated NLP techniques in order to extract more meaningful and higher quality opinions, and incorporate extra-linguistic contextual information. This paper discusses variety of issues related to opinion mining from micro posts and the challenges they impose on an NLP system. This will seem useful to us since the technique incorporated in this paper can extract meaning or opinion from micro posts, since we are dealing with a lot of different type of students, it is a possibility that we may encounter with such micro posts or feedbacks without any contextual information.

2.3 A comprehensive study of text mining approach; Abhishek Kaushik, Sudhanshu Nathani (IJCSNS), Vol. 16, Issue 2, February 2016

Text mining or knowledge discovery is that sub process of data mining, which is widely being used to discover hidden patterns and significant information from the huge amount of unstructured written material. This kind of data cannot be used until or unless specific information or pattern is discovered. For this text mining uses techniques of different fields like machine learning, visualization, case-based reasoning, text analysis, database technology statistics, knowledge management, natural language processing and information retrieval. Third paper is a comprehensive study which showcases various premises and applications of text mining approach. The data cannot be used until or unless specific information or pattern is discovered. This paper describes the need of text mining as well as the areas where it can be explicitly used.

2.4 Techniques of Semantic Analysis for Natural Language Processing- A Detailed Survey; (IJARCCE) Rajani S, M. Hanumanthappa, Volume 5 Issue 2, October 2016.

Semantic analysis is an important part of natural language processing system. It determines the meaning of given sentence and represents that meaning in an appropriate form. Semantics, as a part of linguistics, aims to study the meaning in language.

In this paper, survey is done on semantic analysis and explores different works that have been done in semantic analysis by different researchers.

In the examination, two important research fields are noticed, one of the popular statistical model called as LSA model and another active research area called as ontology which represents a set of primitives of domain of knowledge. In the analysis, it is noted that, LSA is used in automated evaluation against human evaluation and also used for extracting semantic information from textual information. Ontology technique is used to extract structure information from unstructured data, retrieving the information from database and in the semantic web applications. This paper provides an overview of the most widely used techniques for semantic analysis detection. LSA technique is used in automated evaluation against human evaluation and further used to extract semantic information from textual information. Ontology technique is used to extract semantic from unstructured data. Since we have to deal with a lot of people and unformatted textual feedbacks, it is important for us to make use of both these techniques in order to extract meanings from unstructured statements.

2.5 Understanding sentiment of people from news articles; Tomohiro Fukhara, Hiroshi Nakagawa, Toyoaki Nishida (ICWSM)

Temporal sentiment analysis that analyzes temporal trends of sentiments and topics from a text archive that has timestamps is proposed. The method accepts texts with timestamp such as Weblog and news articles, and produces two kinds of graphs, i.e., (1) topic graph that shows temporal change of topics associated with a sentiment, and (2) sentiment graph that shows temporal change of sentiments associated with a topic. Sample results obtained by applying the method to news articles are described.

This paper is all about visualisation. After scrutinizing every other aspect of semantic analysis, and gathering all the segregates information from the feedbacks it is necessary for the institute to plot out a graph which showcases the growth as well as downfall of different changes in the institute. Hence, our final step drawing an accurate visualisation can be achieved by the method/ algorithm mentioned in this paper.

2.6 Research Paper on *Sentiment Analysis Algorithms and Applications - A survey* (Ain Shams Engineering Journal); Walaa Medhat, Ahmed Hassan, Hoda Korashy, *(2014) 5, 1093-1113*

Sentiment Analysis (SA) is an ongoing field of research in text mining field. The survey paper tackles a comprehensive overview of the last update in this field.

Sentiment Analysis (SA) or Opinion Mining (OM) is the computational study of people's opinions, attitudes and emotions toward an entity. The entity can represent individuals, events or topics.

2.7 Research paper on Sentiment Analysis on Twitter Data; (International Journal of Computer Applications) Vishal A. Kharde, S. S. Sonawane, Vol 139-No.11, April 2016

The paper mainly focuses mainly on sentiment analysis of twitter data which is helpful to analyze the information in the tweets where opinions are highly unstructured, heterogeneous and are either positive or negative, or neutral in some cases. In this paper, we provide a survey and a comparative analysis of existing techniques for opinion mining like machine learning and lexicon-based approaches, together with evaluation metrics. Using various machine learning algorithms like Naive Bayes, Max Entropy, and Support Vector Machine, we provide research on twitter data streams. We have also discussed general challenges and applications of Sentiment Analysis on Twitter.

Sentiment analysis can be defined as a process that automates mining of attitudes, opinions, views and emotions from text, speech, tweets and database sources through Natural Language Processing (NLP). Sentiment analysis involves classifying opinions in text into categories like "positive" or "negative" or "neutral". It's also referred as subjectivity analysis, opinion.

The words opinion, sentiment, view and belief are used interchangeably but there are differences between them.

- Opinion: A conclusion open to dispute (because different experts have different opinions)
- View: subjective opinion
- Belief: deliberate acceptance and intellectual assent
- Sentiment: opinion representing one’s feelings

The authors have used the following approach:

- Pre-processing of Data.
- Feature Extraction.
- Training.
- Classification.

The authors conclude that Machine Learning methods such as SVM and Naïve Bayes have highest accuracy and can be regarded as baseline learning methods while lexicon-based methods are very effective only in some cases which require few efforts in human-labelled document. Also, cleaner the data more accurate the results.

2.8 Sentiment Classification on Customer feedback data noisy data, large feature vectors and role of linguistics analysis; Michael Gamon (Microsoft Research).

The paper mainly demonstrates that it is possible to perform automatic sentiment classification in the very noisy domain of customer feedback data. We show that by using large feature vectors in combination with feature reduction, we can train linear support vector machines that achieve high classification accuracy on data that present classification challenges even for a human annotator.

The authors have used the following approach:

- What is the feedback about?

Authors use Text Mining Tools to answer this question.

- Is the feedback positive or negative?

Automatic Sentiment Classification is used to answer this question.

The authors conclude that by using large initial feature vectors combined with feature reduction can be used for very noisy domain of customer feedback and perform sentiment classification.

2.9 **Online Feedback Analysis using Sentiment Analyser;** (IJETTCS) Abhishek Redekar, Sanket Metkar, Harshada Shirole, Gayatri Pawar, Prof. Ms. S.S Pophale, VOLUME 6, ISSUE 2, MARCH-APRIL 2017.

Through Online Feedback Analysis Using Sentiment Analyzer paper, the authors are showing their work done in implementation of a system of online feedback collection based on sentiment analysis. The paper is much of a review paper than a research paper. It more says on what they have done in implementing the system. The idea of paper revolves around the concept of Sentiment Analysis. The system used earlier for feedback collection and analysis had main importance on object type questions/answers, but the authors have importance on both objective as well as subjective/descriptive type of questions/answers in their project and upon which sentiment analysis will be done in order to get more accurate feedbacks.

The sentiment analyzer is build using Machine Learning Algorithm. The algorithm model is trained using positive and negative words of English language. Feedbacks will be processed through the algorithm to get the results.

The system is a secured system. The identity of students giving feedback is not disclosed to anyone including the admin. The project is beneficial for college. With the help of this, more accurate system college will be able to work in more effective way knowing what students' actually want and what they face problems in.

2.10 Online Student Feedback Analysis System with Sentiment Analysis; Divyansh Shrivastava, Shubham Kesarwani, Amol K. Kadam, Aarushi Chhibber, Naveenkumar Jayakumar (IJIRSET) VOLUME 6, ISSUE 5, MAY 2017.

The Online Student Feedback Analysis System (OSFAS) by authors, is more of a review paper than a research paper. It describes the project done by the authors namely OSFAS for education institute. The OSFAS is a web-based system which collects the feedback from every individual student and provides an automatic generation of a collective feedback which has been taken from the students. It is a management information analysis system for educational institutes to manage student feedback data. The system provides facility to generate reports automatically. Responses are gathered and analyzed on behalf of human power. The main purpose according to the authors is quality enhancement and save time as well as decrease human efforts.

Sentiment analysis is an approach of using natural language processing and text analysis for extracting and identifying the sentiment of a text into a positive, negative or neutral category. Two algorithms are used:

1.) General Sentiment Analysis Algorithm.

2.) Multi-use Sentiment Analysis Algorithm based on STANFORD CoreNLP Toolkit.

The main motive to design he system is to reduce time and save efforts of the faculties from maintaining huge amount of records where feedback is concerned.

Techniques and tried to include all these in our project to make it better and more effective.

2.11: Information Lifecycle Modelling Framework for Construction Project Lifecycle Management; Hitoshi ISAHARA

Information in construction project has always been dispersed and scattered in every independent phase. It is always a challenge to integrate project lifecycle information in construction system management to improve productivity. Based on comparison of information lifecycle and construction project lifecycle, this paper presents an information lifecycle modeling to manage construction projects through lifecycles. An innovative framework of information lifecycle modeling for construction project management is suggested for all participants who can access construction lifecycle information. The information in the construction projects is acquired and shared in central database and traditional information loss in construction projects is eliminated.

Information lifecycle management (ILM) is a comprehensive approach to manage the flow of an information system's data and associated metadata from creation until the time when it becomes obsolete and is deleted. ILM involves all aspects of dealing with data, starting with user practices, rather than just automating storage procedures. Information Lifecycle Management is comprised of the policies, processes, practices, and tools used to align the business value of information with the most appropriate and cost effective IT infrastructure from the time information is conceived through its final disposition. Information is aligned with business requirements through management policies and service levels associated with applications, metadata, and data.

2.12 An Overview of Publications on Artificial Intelligence Research: A Quantitative Analysis on Recent Papers; Saiyan Cheng

The study on artificial intelligence (AI) is highly interdisciplinary, which involves increasing number of researchers from different academic fields. In order to provide an overview for the researchers on recent publications in related fields, the conducted a statistic analysis using bibliometric methods. Using the data source from Web of Science (ISI), we have studied the articles published in the SCI and SSCI journals on the subject of AI between the year of 2000 and 2011. The data were analysed from six aspects, including article distribution by years, journals, languages, countries/regions, research fields, and authors. With the fast development of computer science and information technology, Artificial Intelligence (AI) has become an important discipline that may have great impact on both academic researches and practical applications. AI makes machines to imitate human thinking and behaviour. AI has been developed into a new and comprehensive discipline, which involves computer science, cybernetics, information theory, neurophysiology, psychology, linguistics and other many subjects Since the term “AI” was first reported in 1959, researches in this field have stimulated the generation of many new research subjects, such as problem solving, logical reasoning and theorem proving, natural language understanding, gaming, automatic programming, expert system, learning and robotics. The computer science is the major research field for AI study. In this subject, there were over 2682 papers published in the given period, which is 51.75 percent of all the papers. The papers were published on 596 journals, corresponding to 32.27 percent of the journals. It is shown that the majority of AI research members are computing scientists, and the computer science journal is the major source for AI papers. The bibliometric study on AI related publications have revealed several important issues. It is found that AI study is highly interdisciplinary, which involves very diverse fields. This is evident by the wide range of journals that published AI researches. In the same time, the contribution to the field is rather concentrated. More than 70 % papers come from 11 counties. We have identified the core journals, and core authors on this area which may provide a guideline for those interested in the AI research field.

2.13 Artificial Intelligence; Huang Ling-fang

AI has always been on the pioneering end of computer science. This talk outline presents an overview of AI, by summarizing its historical development, family, as well as some representative projects and some Expert system of AI.

- Knowledge and Interference
- Dendral
- Mycin
- The CYC project

AI is a new science of researching theories, methods and technologies in simulating or developing thinking process of human beings. We can simulate or develop human's brain by computers, that is how to use computer technology more effectively and how to use computer to design and construct computer systems as same intelligent as human's or more intelligent than human's. AI research follows two distinct, and to some extent competing, methods, the symbolic (or "top-down") approach, and the connectionist (or "bottom-up") approach. The top-down approach seeks to replicate intelligence by analyzing cognition independent of the biological structure of the brain, in terms of the processing of symbols—whence the symbolic label. The bottom-up approach, on the other hand, involves creating artificial neural networks in imitation of the brain's structure—whence the connectionist label. To illustrate the difference between these approaches, consider the task of building a system, equipped with an optical scanner, that recognizes the letters of the alphabet. Bottom-up approach typically involves training an artificial neural network by presenting letters to it one by one, gradually improving performance by "tuning" the network.(Tuning adjusts the responsiveness of different neural pathways to different stimuli.) In contrast, a top-down approach typically involves writing a computer program that compares each letter with geometric descriptions. Simply put, neural activities are the basis of the bottom-up approach, while symbolic descriptions are the basis of the top-down approach.

2.14 Constructing Test Cases Using Natural Language Processing; Prof. Ahlam Ansari, Ansari Sadaf Fatima, Mirza Baig Shagufta, Shaikh Tehreem

Testing of product is performing to discover or detect the errors and defects in the developed system. But testing is usually time consuming especially when complex projects are canvass. Testing of a product lead off with generation of test cases. The Test case generations are based on three parts coding, design and specification. The Specification based testing deals with generation of test cases from the functional requirements. The proposed system deals with automatic generation of test cases from functional requirement using Natural Language Processing (NLP). The proposed system constructs test cases based on keywords in context from functional requirement of Software Requirement Specification document. The proposed system is beneficial as it can automatically analyze the functional requirement from Software Requirement Specification in order to extract test cases for testing. The goal of the proposed system is to reduce effort and time consumed by software tester to test the product.

2.15 Understanding Causal Feedback Using the Strategic Planning System (SPS); Michell Smith, Peter B Riggs, Edward Freeman.

The Strategic Planning System (SPS) is a causal modeling tool that gives planners a way to express underlying causal relationships and feedback loops in a strategic plan and determine the effects and side-effects of different strategic alternatives. Corporate strategic planning is setting a direction for a corporation in the present based upon uncertain information about the future; SPS provides an environment where situations can be defined, and plans refined, abstracted, analyzed, critiqued, and then further refined through an explicit mutually understandable fork. The tools available for planners tend to be traditional spreadsheets or simulations such as Stella? While these tools provide quantitative analyses and results, they offer no qualitative explanations. SPS is a tool in which the underlying structure of a model is elucidated through qualitative explanations and incorporates "Systems Dynamics" concepts (non-linear systems with feedback). Numerical tools are then ked afterward to further refine the models. Overall, strategic planners have found SPS to be a valuable tool. They have been impressed with the system and have used and are using SPS in a number of planning. The strategic plans of a corporation describe the long-range direction of the corporation. Typically, these plans include both textual and analytic descriptions of where the corporation views itself in the marketplace of today, where it views itself in the marketplace of tomorrow, the actions or strategies that will be followed to affect change, and the reasons why those actions are appropriate. The plan tries to lay out the best course for achieving the goals of the corporation based on an understanding of the current state of the business, actions that could be taken by the business, and any external or uncontrollable influences that could be imposed.

Strategic planners have the unenviable task of trying to produce an understandable, coherent, and optimal plan in an uncertain business future. Planners rely on complex, generally qualitative, mental models of the world, and how to get from today to tomorrow. They generally try to figure out the "right" actions to take and then derive the financial impacts of the actions.

2.16 **Boosting Applied to Word Sense Disambiguation;** Gerard Escudero, Llu´ıs M`arquez, and German Rigau

Resolving the ambiguity of words is a central problem for language understanding applications and their associated tasks. Till date no highly accurate WSD system is built. The most successful current line of research is the corpus–based approach in which statistical or Machine Learning (ML) algorithms have been applied to learn statistical models or classifiers from corpora in order to perform WSD. Naive bayes and Exemplar based algorithms are present currently and supervised methods are also present but it contains high computation cost. Boosting algorithm is proposed in the paper. The main idea of boosting algorithms is to combine many simple and moderately accurate hypotheses (called weak classifiers) into a single, highly accurate classifier for the task at hand. The weak classifiers are trained sequentially, each of them is trained on the examples which were most difficult to classify by the preceding weak classifiers. Purpose of boosting algorithm is to find classification rule by combining weak hypothesis. weak learner is a procedure which iterates and provides weak hypothesis each time. Weak hypothesis gets combined into single rule called combined hypothesis. The magnitude of prediction is interpreted as measure of confidence in the prediction. WSD is not a multi–label classification problem since a unique sense is expected for each word in context. Experiment is performed comparing boosting algorithm with EB and MB and results showed that boosting algorithm is better for rules over 100 and significantly gets better for increase in rules.

2.17 Combined Optimization of Feature Selection and Algorithm Parameters in Machine Learning of Language; Walter Daelemans, V´eronique Hoste, Fien De Meulder, and Bart Naudts

Methodology currently used in Word sense disambiguation may not be reliable and using genetic algorithms for WSG increases reliability and accuracy.

Supervised machine learning methods are investigated intensively in empirical computation linguistics because they potentially have a number of advantages compared to standard statistical approaches. Machine learning techniques used in NLP for empirical learning have outperform the hand craft methods. Many factors play role in outcome of ML experiment are data, information sources, representation and algorithm parameter settings. Joint optimization can lead to significantly higher generalization accuracies.

GA succeeds in finding solutions that are significantly better than the default solutions and the best solutions obtained by a heuristic combined feature selection and algorithm parameter optimization approach optimization can be used for obtaining higher predictive accuracy.

2.18 Empirical Learning of Natural Language Processing Tasks; Walter Daelemans, Antal van den Bosch,Ton Weijters

Research paper shows that using empirical learning techniques will be best suited for processing

NLP tasks.

Some of the reasons why NLP should focus on ML techniques are

o Complexity of tasks

o Real world applications

o Availability of large datasets

Corpora which has been used for research purposes which can be used for empirical learning include-

o CELEX

o Pen Treebank II

o WordNet

All linguistic problems can be described as a mapping of one of two types of classifications

o Disambiguation

o Segmentation

lazy learning, information encountered in training is not abstracted, whereas in greedy learning

information is abstracted by restricting and removing redundant or unimportant information.

3 approaches to greedy learning are –

o Decision tree learning

o Artificial Neural networks

o Inductive logic programming

simple lazy-learning algorithms, extended with feature weighting and probabilistic decision rules,

consistently obtain the best generalization accuracy on a large collection of linguistic tasks.

greedy learning techniques, such as ILP and Rule Induction, induce structures which may add the understanding of the domain, and indeed sometimes generate new linguistic descriptions of the domain. Third, the learning techniques described are well-suited for integrating different information sources.

2.19 **Detecting Inflection Patterns in Natural Language by Minimization of Morphological Model**; Alexander Gelbukh, Mikhail Alexandrov, Sang-Yong Han

Stemming algorithm is to map the words having the same base meaning but differing in grammatical forms, to the same letter string that can be used to identify the word independently of its morphological form. Manual construction of dictionary and rules is complicated and tedious so an attractive alternative is automatic learning of the necessary models from the texts themselves.

Languages Spoken in the world can be classified as:

o Inflective languages

o Agglutinative languages

o Isolating languages

o Intra reflective languages

o Incorporating languages

Main previous Approaches to stemming are:

o Dictionary based

o Rule based

o Statistical based

Proposed unsupervised algorithm for stemming reduces data and priori information and removes those words where S+E occurring once or those words which are rare. This Algorithm currently ignores some phenomena’s and when used for longer time may provide good results and can be used for further natural
language researches.

2.20 Syntactic Structural Kernels for Natural Language Interfaces to Databases; Alessandra Giordani, Alessandro Moschitti

The automatic conversion or translation of natural language questions into SQL queries would allow for the design of effective and useful database system from a user viewpoint.
In this paper, the authors have approached the problem by deriving a shared semantic between natural language and programming language by automatically learning a model based the syntactical representation of the training examples. In their experiment, they consider pairs of natural language questions and SQL queries as training examples.
They designed innovative combinations between different kernels for structured data applied to pairs of objects, that to the best of our knowledge, represent a novel approach to describe relational semantics between NL and SQL languages. The approach taken up seems viable to mine semantic relations between natural language and SQL.

3. RESEARCH METHODOLOGY

After going through detailed literature survey and analysing the past survey, we have amalgamated all the techniques and tried to include all these in our project to make it better and more effective. We will be applying the techniques such as ontology technique in order to extract information from unstructured data. The approach used in the paper automatic detection of political opinion in tweets is helpful for us since we are dealing with students who have the tendency to frame answers in a form of micro-posts or hashtags. To extract meaningful information from the micro-post it is necessary for us to use the methodology used in order to take relevant data sets and semantic from the feedback processed. The method of visualization is a critical aspect in our project. Since we are dealing with a lot of data it is necessary for us to keep a track record and maintain a graph in order to predict some further changes in the near future. Therefore, the visualisation survey and methods used in the paper "understanding sentiment of people from news articles" is quite crucial. The use and plotting graph for sentiment and topical is necessary for us.

4. Conclusion

After meticulous analysis and research, we have picked few techniques and methods along with some additional concept in order to improve our system. We have scrutinized every aspect of the research paper we read, we further plan to incept these concepts in our project. Our main motive is to reduce the manpower effort needed for segregating the forms in today's world according to their nature, department as well as the solution that can be imposed.

Feedback has seen a lot of changes in the nature of their conduct, and shall leap further into seeing some more advancements in this field. Our world has been turning into an AI system and thus devising a system which will be enforced in getting involved in friendly conversation with students and thus extracting critical information in form of question and answers.

5. References

[1]. Xin Chen, Mihaela Vorvoreanu, and Krishna Madhavan, "Mining social media data for understanding students' learning experiences", IEEE Transaction, 2014.

[2]. Mariam Adedoyin-Olowe,, Mohamed Medhat Gaber and Frederic Stahl," A Survey of Data Mining Techniques for Social Network Analysis", School of Computing Science and Digital Media, Robert GordonUniversity Aberdeen, AB10 7QB, UK

[3]. E. Goffman, The Presentation of Self in Everyday Life. Lightning Source Inc., 1959.

[4]. J.M. DiMicco and D.R. Millen, "Identity Management: Multiple Presentations of Self in Facebook," Proc. the Int'l ACM Conf. Supporting Group Work, pp. 383-386, 2007.

[5]. M. Vorvoreanu and Q. Clark, "Managing Identity Across Social Networks," Proc. Poster Session at the ACM Conf. Computer Supported Cooperative Work, 2010.

[6]. B. Sriram, D. Fuhry, E. Demir, H. Ferhatosmanoglu, and M. Demirbas, "Short Text Classification in Twitter to Improve Information Filtering," Proc. 33rd Int'l ACM SIGIR Conf. Research and Development in Information Retrieval, pp. 841- 842, 2010.

[7]. Bodong Chen, Xin Chen, Wanli Xing, "Twitter Archeology of Learning Analytics and Knowledge Conferences" Proceedings of fifth international conference on lerning analytic and knowledge, pp. 340-349,2015.

[8]. A. Go, R. Bhayani, and L. Huang, "Twitter Sentiment Classification Using Distant Supervision," CS224N Project Report, Stanford pp. 1-12, 2009.

[9]. Hsin-Ying Wu, Kuan-Liang Liu and Charles Trappey " The Theory On User Feedback Analysis."

[10]. Suleyman Cetintas, Luo Si, Hans Peter Aagard, Kyle Bowen, and Mariheida Cordova-Sanchez," Microblogging in a Classroom: Classifying Students' Relevant and Irrelevant Questions in a Microblogging-Supported Classroom," IEEE Transactions on Learning Technologies, Vol. 4, No. 4, October- December 2011.

[12].W. Zhao, J. Jiang, J. Weng, J. He, E.P. Lim, H. Yan, and X. Li,"Comparing Twitter and Traditional Media Using Topic Models," Proc. 33rd European Conf. Advances in Information Retrieval, pp. 338- 349, 2011.

[15]. Andreas Hotho," A Brief Survey of Text Mining" KDE Group University of Kassel May 13, 2005.