



# **COLLEGE OF ENGINEERING, PUNE**

(An Autonomous Institute of Government of Maharashtra.)

## **Department of Computer Engineering and Information Technology**

### *PROJECT ABSTRACTS*

#### **BTECH INFORMATION TECHNOLOGY 2016-17**

---

#### **SDDC AUTOMATION SUITE FOR DEPLOYMENT AND CERTIFICATION**

Data Centres have undergone transformation from the traditional on-premises setups to the current cloud based ones. The advent of cloud computing has exponentially reduced the data centre infrastructure costs for organizations and made scalability an easy affair. However data centre management has been a tedious task and continues to remain so even after the advent of cloud. Cloud providers provide wide variety of services which involves management of whole computational resources at the data center side. This involves network, storage, security, computation management. Virtual environments are reducing work related to configuration of physical hardware and giving more power to control every component through software; but understanding complex systems developed by some other party and then deploying them into data center still makes data centre management a complex task. By using the concepts of Software Defined Data Centre (SDDC), Software Defined Networking (SDN) and virtual environment interfacing the Software Defined Data Centre Suite will help data center administrators to solve the problems associated with the on-premises as well as cloud based network administrating data centres. This will reduce the human efforts, time as well as financial costs of data centre management exponentially.

#### **AUTOMATED QUIZ GENERATOR**

Automated Quiz Generator (AQG) is an extension of the factual question generation system implemented by Michael Heilman, which is generic and therefore applicable to any given domain of discourse in natural language. The extensions mainly include the ability to make MCQs out of generated questions and ranking questions by interestingness of the sentence from which the respective question was generated. Besides, it has functionality to extract interesting trivia from Wikipedia articles of important entities in the input text. Being domain independent, this project relies on DBpedia - a database of structured content extracted from Wikipedia, the largest general reference work on the Internet.

An important educational purpose this system serves is gauging a learner's reading comprehension ability by generating cross-checking questions. It can also aid teachers in setting multiple choice/plain factual questions for a quiz. From the economic point of view, interesting trivia generation serves

towards user engagement.

## **BANK CHAT BOT**

Banks play an important role in every country's economic development. In day-to-day life, everybody needs banks. But most of the people, especially the first-timers, struggle to know various procedures and processes required to get their work done at the bank and avail of its different products and services. Currently banks have their own web-sites, mobile applications and facilities like internet banking, mobile banking but sometimes, these sources can be a bit overwhelming for most of the users who are either not well versed with technology or in some cases where the information is too scattered to search for easily. There are different types of platforms provided by different banks but people are facing problems accessing them (different GUIs, too much navigation). Although Customer Care centers are available, there are lot of wait times and redirection in some cases, leaving the customer with no choice but to experience considerable delays getting a simple informational query resolved. People have queries about various bank policies, loans, fixed deposits. This results in unnecessary crowd in banks for enquiry. Banks also face problems solving repeated queries of customers. This is time consuming and banking staff gets frustrated. Manpower and money gets wasted for separate enquiry counter.

We intend to provide a chat bot interface for customers which could be available on the web and on any handheld devices. Customers can interact with mentioning their queries in plain English and the chat bot can resolve their queries with appropriate response in return. Chat bots are intelligent systems that understand user's natural language queries and respond accordingly. It is more like a virtual assistant, people feel like they are talking with real person. Proposed chat bot application is easily accessible to customer thereby solving redundant queries anywhere anytime. The system would help replicate the customer service experience with one difference that the customer would be interacting with a bot instead of a real person and yet get the queries attended and resolved. As there will be fast response for enquiry, this will be time saving for both bank and customers. The proposed system would be a stepping stone in having in place an intelligent query handling program which could in next phases not just respond but self-learn to improve itself thereby increasing not just the quality of customer service but also reducing human load, increase in productivity and of course increasing number of satisfied customers.

## **STATISTICAL ANALYSIS OF SPORTS DATA FOR PLAYER EVALUATION**

Sports and statistics go hand in hand, but except Baseball other sports have been deprived of good statistical analysis methods. This project aims at devising a method for statistically analyzing the huge amount of data, the sports events generate and help the managers/coaches to find out the best fit of

players and how much to bid for a particular player from the team cap. This project is aimed at assimilating the best features of the previous approaches and losing the drawbacks. We aim to create a tool which can be used for the statistical analysis of the information available for any sport with minor customizations for evaluation of various aspects. We aim to devise an approach for statistical analysis of sports data, focused on Cricket and particularly IPL data which can be later extended to other sports with minor customizations.

**Collaboration Information Management System** This project "Collaboration information management system" gives a simple interface for the purpose of maintaining of information or records with regards to the students and the staff member of the institute or an organization. Achieving this objective using a manual system is very difficult. This problem can be solved or handled using this project. This project provides facilities like online registering the user and creating profile for user, attendance system, fee transactions, records generation etc. The main module is the biometric finger print recognition system used for giving authentication to the user to view the profile as per the role of the user in that system. The finger print recognition system increases the security of the system.

## **BUG TRIAGE MANAGEMENT**

Software industry spends more time and cost to handle software bugs. Object of bug triage is to assign appropriate developer to upcoming bug-report. This work is done manually hence it is time consuming as well as not efficient. It is not assured that bug will be assigned to correct developer as it contains lots of redundancy. We need automatic bug triage management system. Our first task will be to reduce data redundancy and improve the quality of bug data. This will be achieved by feature and instance selection. Attributes of the bugs will be extracted from the historical bug data. Second task will be mapping of these bugs to correct developer.

## **APPLICATION OF THE HIERARCHICAL TEMPORAL MEMORY THEORY FOR DOCUMENT CATEGORIZATION**

The current work intends to study the performance of the Hierarchical Temporal Memory (HTM) theory for automated categorization of text as well as documents. HTM is a biologically inspired theory based on the working principles of the human neocortex. Traditionally, document categorization has been done using various machine learning techniques, mainly, Support Vector Machines (SVM), KNN Classification, Naive Bayes Classification and Decision Trees. The current study intends to provide an alternative framework for document categorization using the principles of the Spatial Pooler learning algorithm in the HTM Theory. As HTM is modelled after the mammalian brain, its input should also be in accordance with the input format received by the brain. Latent Semantic Indexing (LSI) technique is used for

extracting the top features from the input and converting them into binary format. The Spatial Pooler algorithm converts the binary input into sparse patterns with similar input text having overlapping spatial patterns making it easy for classifying the patterns into pre-defined categories. Our results prove that HTM theory, although is in its early stages, performs well for categorizing documents.

## **FRAUD DETECTION APPLICATION ON CLOUD**

We aim on building a cloud-native client-server application for Fraud Detection, a major security concern for financial institution such as MasterCard. For a particular transaction, this application provides a score which indicates the probability of the transaction being a fraudulent one. The application is deployed onto Pivotal Cloud Foundry (PCF), where the ability of the application to expand or shrink as per the load is demonstrated.

## **CAREER BUILDER**

We are creating an efficient, user friendly, simple platform, which will help user for overall development. This software application has information about various fields not only related to study but also different field like music, sports, politics etc. In this application we are providing platform to the user to choose his or her field of interest which is helpful for his/her career. Now-a-days most of the people are using smart phones. So, we are using the power of mobile phones to distribute information among the people. Android app is the efficient way as mobile user spend 86 percent of their time on app and only 14 percent on website.

## **MEETING ROOM BOOKING SYSTEM**

Our project addresses the necessity of role based access in the allocation of resources. Meeting Room Booking System (MRBS) is an existing resource allocation system, which is used to grant or reject meeting room booking requests. However, the current versions of MRBS have only 2 access levels, namely: user and admin. This does not sufficiently define access permissions. Our implementation of MRBS eliminates these limitations, by properly defining access rights in the form of roles and their capabilities. We have made MRBS implementable for any generic resources, not just meeting rooms, by allowing the creation of user-defined roles by selecting specific capabilities. Thus, we have implemented a fine-grained access control via Role-Based Access Control (RBAC).

## **SQL INJECTION**

As the popularity of the web increases and web applications become tools of everyday use, the

role of web security has been gaining importance as well. In the last few years, web-based attacks have shown a significant increase. SQL Injection is the most common and widely known attack on the web server. But still, data breaches due to SQL injection is continuously increasing from year to year.

Our system automatically analyses web sites with the aim of finding exploitable SQL injection. It is able to find many potentially vulnerable web sites.

### **STATISTICAL PART OF SPEECH TAGGER USING HIDDEN MARKOV MODEL**

People normally switch between multiple languages during the conversation. This results in Code-mixing. It is common in places, where people are primarily bilingual in nature. People rather than using Unicode in writing, prefer to use phonetic typing. It causes difficulties in analysis of such text. CodeMixedPosTagger attempts to find Part-of-Speech tag for such text, which is the first and important step in the analysis of such text. It is a statistical part-of-speech tagger, based on Hidden Markov model. It uses machine learning approach to find the most probable grammatical tag of each word in the input text. As it is based purely on the statistical approach, it is generic in nature in terms of language. Thus, can be used to tag both pure language (single language) as well as bilingual text.

### **PROFILE SHOWCASE IN MIDASFUSION**

The issue of security is very paramount to any organization, especially when it comes to dealing with the original certificates of an individual. Therefore, we intend to build a module 'Showcase' for the product called 'MidasFusion' of MidasBlue Inc.

MidasFusion is a platform for Industry Academia collaboration to improve match of new hires, innovation, research jobs and Mentorships. Show- case, i.e. the module in consideration is a vitrine as well as a safe house for all the documents and certificates that the user will upload for his profile. The user will have full access to his/her showcase and also the user will decide who can see it's showcase and who cannot. In order to view to any user's (mentor, student, organization etc.) showcase, one needs to send a request to another and the receiver has the option to accept or reject the request. Once the request has been granted, one can view the document in the browser. Neither can he download it nor can he take a screenshot, and that is how the issue if security has been taken care of. Also, the showcase has been categorized as per convenience.

### **SPORTSLIVE - A PLATFORM FOR LIVE SCORING**

'SportsLive' is a Android Base Mobile application that provides a platform for live scoring of

many sports. Sport Live application is controlled by admin panel side, which gives control to admin to update and remove score. Also admin can add two opposite teams and broadcast a message to all user's who have installed 'SportsLive' App (user side). In user side of 'SportsLive', user can see live and instant update of score which is added by Admin. User can check all the modules viz. Ongoing matches, Next matches, Todays Schedule, Live Score, Winner Board, Emergency Broadcast. SportsLive will provide a dynamic platform for providing live scoring of all sports arranged by colleges, schools, local communities, institutes, society and companies. A dynamic platform means sport arranging committees can configure live sport app according to their needs. They can add, create, remove and or update sports events according to their choice. At the same time SportsLive app can be utilized by many sports arranging committees like colleges, schools, local communities, institutes, society and companies for live scoring and users can see the updated scores according to their choices. SportsLive app will not only provide a live scoring platform but also provide a dynamic sports event management app.

## **WEB APPLICATION ATTACK DETECTION SYSTEM**

The utilization of web applications by the clients over the Internet is expanding step by step. Analyzing the increasing demand, many organizations are providing their services through web applications. Common web applications include webmail, online retail sales, online auctions, wikis, instant messaging services and many other functions.[3] Number of clients and web applications are in wrinking ceaselessly. In addition, the pernicious movement is moreover expanding with a similar rate. Therefore, there is a need to secure web servers and applications. Web servers and applications are experiencing Flooding attacks from network to the application layer. Numerous arrangements are accessible for flooding attack at the system and transport layer. However, very few arrangements or solutions are accessible for flooding attacks at application layer because of complexity in detecting the behavior of attack.

Firewall is a system security framework that screens and controls the approaching or incoming and active system movement. It mostly work between Physical to Network layer and little bit of peeking into the Transport layer to figure out source and destination port number. Even thought Firewall is used, Web applications are insecured due to various type of Attacks. So, Fire- wall are not effective for security of Web Applications.

IDS -Intrusion Detection System, developing an Detection System to find out Intrusion an Effective way to find out Attacks on Web Application. Here we have designed and developed Ontology(Greek Meaning: Onto means "existence or being" and logy means "science or study" i.e study of being ) which comes under Artificial Intelligence Domain, which finds Attacks and Behaviour. For testing out our system we are using CSIC 2010 Dataset and also measured the

performance.

## **PRODUCT RATING USING OPINION MINING**

The advent of technology has changed the way we work today and has made our lives so comfortable. With the development of Internet, the idea of shopping has also changed. A new trend of online shopping has begun and more and more people are joining in because of the advantages it provides like more variety, lower prices and less wastage of time. The industries too have completely revolutionized their processes and business strategies and many companies today are trying to get their businesses online. Amazon is one such company which has become hugely successful in the domain of online shopping.

Amazon.com is one of the largest electronic commerce website in the world which allows users to purchase different products belonging to various categories. The customers can also rate the products and submit reviews on each one of them. The reviews allow the first-time buyers to understand the quality of the products and decide whether to make a purchase or not. These reviews result in unstructured big data which can be analyzed to understand the purchase patterns and recommend the products.

Nowadays, a lot of research is being carried out on big data which is generated online. Different approaches for analysis of the product reviews obtained on the Amazon website have been developed to provide rating to the products but there are some issues that have not been addressed yet like detecting fake reviews and increasing the speed of big data analysis. So the main objective of our project is to provide a faster implementation for analysis of only genuine Amazon reviews after removing fake reviews from the dataset and provide a more accurate rating to the products.

## **APPLICATION OF THE HIERARCHICAL TEMPORAL MEMORY THEORY FOR DOCUMENT CATEGORIZATION**

The current work intends to study the performance of the Hierarchical Temporal Memory(HTM) theory for automated categorization of text as well as documents. HTM is a biologically inspired theory based on the working principles of the human neocortex. Traditionally, document categorization has been done using various machine learning techniques, mainly, Support Vector Machines (SVM), KNN Classification, Naive Bayes Classification and Decision Trees. The current study intends to provide an alternative framework for document categorization using the principles of the Spatial Pooler learning algorithm in the HTM Theory. As HTM is modelled after the mammalian brain, its input should also be in accordance with the input format received by the brain. Latent Semantic Indexing (LSI)

technique is used for extracting the top features from the input and converting them into binary format. The Spatial Pooler algorithm converts the binary input into sparse patterns with similar input text having overlapping spatial patterns making it easy for classifying the patterns into pre-defined categories. Our results prove that HTM theory, although is in its early stages, performs well for categorizing documents.

### **PERFORMANCE PARAMETER EXTRACTION TOOL FOR BIG DATA PLATFORMS**

Many big data applications have critical performance requirement for making application more interactive and to make efficient use of resources in the big data cluster. There are numerous performance parameters in big data platforms. Optimum tuning of these parameters is required to get best performance. System administrators should be able to understand impact of different parameters on their platform and should tune it accordingly. In this paper we are proposing a performance parameter extraction tool for big data platform which will extract all the different types of performance metrics which are important to the performance of big data platforms. This tool can be used as a module for integration with performance monitoring and performance analysis tool.

### **CONVOLUTIONAL NEURAL NETWORKS FOR TEXT CATEGORIZATION WITH LATENT SEMANTIC ANALYSIS**

The recent emphasis on intelligent systems has increased the focus on categorization techniques as it is an important step in information retrieval and natural language processing. The text categorization is largely achieved using machine learning techniques. In most of the approaches one-hot encoding or pre-trained word embedding such as word2vec or glove vectors are used. This study explores the feature vectors based encoding using Latent Semantic Analysis (LSA) technique along with the Convolutional Neural Network (CNN) being used as a classifier. It was found that applying LSA followed by CNN for text classification offers better accuracy than the conventional methods of CNN with other encodings such as one-hot. This research, thus, highlights the importance of Latent Semantic Analysis technique coupled with convolutional neural networks for text classification.

### **DETECTION OF FORGERY IN ART PAINTINGS USING MACHINE LEARNING**

This project aims at the identification of art through the use of machine learning. In today's world, computers are advancing faster than ever and more and more applications of machine learning have been explored, like computer vision. We have identified this problem statement as relevant and challenging. Currently detection of forgery in museums is done by examining the



painting in detail by an art expert. Ongoing research on automated art identification is limited. Through machine learning, we aim to identify if two paintings are painted by the same person. We believe that our proposed solution for detecting forgery in art paintings hold interesting applications for curators and art historians, and for connoisseurs and art lovers. Through our project, the similarity between the different artists can be found out, with the characteristics and style of the paintings identified through machine learning.

## **AUTOMATION TOOL**

Software Quality Engineers manually check for bugs, logical errors and flaws in the source code of the product before it goes out to the market. The product in consideration apart from being logically correct, should also conform to the parent company's standards and policies. Hence, here our challenge is to effectively automate the testing process which is often prone to human errors and provide a user friendly tool to perform the various checks involved during the release management process.

Our aim is to create standalone utilities to automate the various tests involved throughout the testing cycle, eminently increasing the reliability of the tests which are prone to human errors otherwise. Once all the utilities have been created, we propose to create an interface which will integrate all the tools and make them accessible collectively.

## **DATA SECURITY SYSTEM**

Security of the computer files and folders have been one of the most important issues. Important data present on the machine can be accessed by the people who know the machine's password. The project focuses on making an application that can be used to secure the folders using new authentication system. The new system uses most common mobile features such as Bluetooth, fingerprint scanner to identify the owner of the files. The user will be able to select the level of security and the level of security can be managed by changing the algorithm and key size.

The software gives another layer of security, i.e., by password also by biometric sensors, also it gives a secured environment to manage data.

## **USER ACTIVITY ANALYSIS WITH ZEITGEIST**

Users everywhere are using computers everyday, and as the Linux Operating System is growing in market share more and more linux machines are showing up as servers, hosts, mainframes and app-servers. A normal PC users biggest challenge is today's convenient world of one app for every requirement and terabytes of storage is searching through folder after subfolder of files,

hidden files and related files to find the file he modified just yesterday. Organizing his/her files is tedious and tiresome, But as one can think about it, aren't there better ways to organize ones files?

For a system administrator, sometimes the most important responsibility is to monitor system performance and to verify the causes for system or application crashes, in pursuit of this responsibility, he usually carefully checks application logs for errors or crash reports. But poring over thousands of lines of logs is tedious. What if there were a way to monitor all the relevant actions taking place in our machine?

In today's world, there are several cyber-criminals who use Linux every- where and cyber-crimes being executed with linux systems as the weapons, facilitators or targets. There is a larger need for forensic programs that will allow a cyber-security-forensic expert to determine what a suspect was doing on his or her personal computer at a specific time and date, as well as which user was operating what programs.

A contemporary solution to the for-mentioned problem is necessary to aid the user, administrator and forensic specialist. Our project is a proposed solution addressed to a plethora of problems all lying in the fields of Software Engineering, User Activity Analysis and Computer Organization.

### **A NOVEL APPROACH TO SOLVE THE CHALLENGES IN SENTIMENT ANALYSIS**

With the increased use of Internet nowadays for sharing opinions, views and sentiments about products, services, people and organizations; social networking sites are gaining popularity. Twitter, one of the largest social networking sites is used by many people to share their opinions, life events, views about various topics. Our project performs sentiment analysis on these tweets which are highly unstructured and heterogeneous using various machine learning algorithms like Multinomial Naive Bayes, SVM, Random Forrest, Decision Tree. There are various challenges involved in sentiment analysis of these tweets. Our project focuses on solving some of these challenges like sarcasm, negation, slang word and emojis handling.

### **COGNITIVE LEARNING ASSISTANT**

In today's world, the scientific community is undertaking great endeavors to make machines think more human like. Machines in general aren't that great at analyzing the patters that exist in the real world, but today's computers possess great prowess that can be handy in solving real life problems. Thus, the oncoming of new technologies like Machine Learning and Natural Language Processing completely changing the scenario, providing the best of both the applications : an aptitude to find patterns and analyses subjects possessed by humans along

with the speed and accuracy of the new age computers. Recent developments have given birth to a new field, Cognitive Computing. Cognitive Computing refers to those systems that have the ability to reason with some motive, able to scale themselves and go in par with humans natural way of thinking. Cognitive Computing systems are not programmed they are prepared and trained to detect, foresee, gather and in a way to also think, utilizing computerized reasoning and machine learning calculations that are exposed to enormous data sets. Cognitive Computing will allow humans to interact with the computers at a more natural level. It will allow the computers to behave and think like humans; it is the technology that sci-fi has always fantasized about. We plan to innovate in this nascent field by developing a prototype for a system that learn and understand what it's user knows, and will aid the user by helping him to build newer concepts using the ones he already is aware of. The system then dynamically keeps learning about the user's knowledge base and adapts accordingly. The system will be initially optimized with the age, education and some extra factors to predict the user's knowledge base. The system will use examples from the known knowledge base to answer the user's queries. This will be great solution to aid people of all diversities and capabilities in educating themselves.

## **DYNAMIC PRICE PREDICTION FOR E-COMMERCE USING NEURAL NETWORK**

Prediction of real world entity is very crucial. But due to increasing technology various models have built which can overcome the challenges in the field of finance and economics. Nowadays Neural networks is the trending field in various areas. We have implemented the project on Dynamic Price Prediction using Artificial Neural Network by giving input to it. Dynamic Price Prediction is a type of pricing strategy in which the price of the product is predicted on the basis of current demand of it in market. This strategy is used to gain maximum profit. This technique can be used in various areas like airlines, hotels, train tickets, movie tickets etc. We have applied this technique for E-commerce site. In this project, we aim to give maximum profit to seller for that purpose we have used brnn algorithm. We observed that our built model predicts prices as demand changes. Price Prediction is done on the half hourly basis. We kept threshold values for prices beyond and above that which price should not go.

## **DOCUMENT CLASSIFICATION USING SUPPORT VECTOR MACHINES AND LATENT SEMANTIC ANALYSIS**

Document Classification has become a key technique for handling and organizing of vast documents. There are various text classifiers available with varied features and varied

algorithms. Every classifier aims to gain maximum accuracy and fast results. In this project, we aim to achieve better accuracy using multiple text classification algorithms. After study of the results of different algorithms on different datasets, we observed that with increase in size of dataset i.e when number of documents increase, the number of words increase and there by analyzing takes more time and lesser accuracy. We aim to study the changes in time required for the classification and the accuracy by decreasing the number of features. After reviewing some classifiers, we found that our data was best classified by Support Vector Machine (SVM). SVM offer better speed and accuracy over latest document classification techniques and provide better solution to many document classification problems. SVM are fully automated i.e. there is no need of any parameter tuning. So we are using SVM for classification. While for dimensionality reduction, we are using Latent Semantic Analysis (LSA).

### **AUTHENTICATION USING A SMARTPHONE**

There is only single method of authentication prevalent these days which are password-based authentication. But password base authentication have many flaws. In this project, we are trying to build an authentication system which will help users to authenticate into various different websites using their smartphones by just two clicks. Since users increasingly owns a smartphone today, our system will connect the browser with our smartphone and browser will then retrieve necessary data from smartphone without asking for any password authentication on corresponding websites.

Here, we are proposing a system which will help us to sign in to any websites using any browser. This system is based on Android application. While using this system, we do not need to type username and password again and again. This system will therefore let websites to migrate to better authentication scheme irrespective of each other.

However, achieving this is somewhat difficult. Migration of services will not be possible easily if not much users are using the new system. In a similar way, users will not make use of such authentication scheme which is acceptable at very few services. In addition to this, when we enable an android application to sign in to a website browsed over computer is also non-trivial.

### **DATA ORGANIZATION AND VISUALIZATION TOOL**

A lot of data was being collected from the COEP Satellite which was to be distributed to various subsystems of the satellite team. The data being collected in raw file format could not be used without converting it into suitable formats. The aim of the project is to convert all the data in raw files into relational database and to create a tool for analysis and visualization of data. The aim is to create the tool with a minimal learning curve. The tools must also be created

in a way that small functionalities can easily be added to the tool.