COLLEGE OF ENGINEERING, PUNE

(An Autonomous Institute of Government of Maharashtra)

Department of Computer Engineering and Information Technology

PROJECT ABSTRACTS

BTECH COMPUTER ENGINEERING 2016-17

VEHICULAR DETECTION AND ANALYSIS

Computer vision is an area of study, which incorporates different techniques for capturing, processing and analyzing images. It generally deals with complex data from the real world find derives some useful information from them. In today's world, this becomes an intriguing and a tough task to identify the different vehicles from an image or video feed. Image processing is an important part of Computer Science. The main focus of the project is to identify different types of vehicles from a video feed. Types of the vehicles will be cars, motorcycles and rickshaws. A video of traffic will be supplied to the software which is taken in a predefined way meaning the angle at which the video is taken is fixed. The types of vehicles from the video will be identified. The number of vehicles of each type will be given by the application of a vehicle counter. Also the graph of the vehicle count is generated on demand. Registered users will also get the mail containing the details of the traffic analysis.

POLYGLOT PERSISTENCE AS ASERVICE

There has been an increase in the amount and heterogeneity of the data being used by organizations today . Because of this using a single database solution for the complete dataset has become inefficient and using different types of databases are necessary to improve the performance. This paper describes a polyglot persistence framework that exposes a simple REST API for application developers to integrate with various database technologies (RDBMS, NOSQL, Search, Cache, etc). The term polyglot persistence can be defined as the use of different database solutions within a single application. These different database solutions are used to address functional and non- functional needs of the application, using the optimal solution based on the application requirement instead of a single solution with one size fits all approach. The framework supports connection pooling and sharding and thereby provides higher performance and throughput. A complex enterprise application uses different kinds of data, and usually integrates information from different sources. This project is all about simplifying the work that an engineer has to do to integrate with multiple data sources.

PORTRAIT - A FREE SYSTEMS MANAGEMENT SOFTWARE

Many-a-times we see that in a computer lab the are multiple systems and the person in charge or the lab admin needs to run the same task (like install ting a particular software or getting some statistical data from the systems) on multiple systems. This is very cumbersome. So in this case there is a huge need of some systems management software which would help ease the task. Our project is about creating a free and open source systems management software. This software would not only let the user do the normal tasks that a systems management soft- ware but also have provisions for a remote desktop view function which would let the user to directly view the desktop of another system.

PHIS-STORM

Phishing URLs have very little relationship among few part of URL which is registered and remaining part of it. It defines new concept of intra-URL relatedness and its evaluation done by using features which are derived from words that forms URL links based upon query data from Google, Yahoo and Bing. The classification depend on machine learning for detection of phishing URL defined by the features. Now-A-days one of the most dangerous threats to cybercrime activities is 'Phishing'. Surveys have been conducted total loss due to phishing are in several Billions. To determine actual financial losses caused by phishing is very difficult. To counter such problems we have developed a rating system to protect end users from phishing content called Phish Storm. Phish-Storm is automated and gives URL phishing rating in real time. The relation between the words that form URL and state the difference between the legitimate and phishing links or content of URL can be implemented by using Phish-Storm. In Phish-Storm computation technique which gives relation between the various parts of the URL for detecting phishing URL in real time. It takes data from Google, Yahoo and Bing for operation. Due to this range of Phish-Storm is fairly large. PhishStorm can be applied for various levels of network for the prevention to phishing. In Phish-Storm detection system depends on the lexical analysis of the URL. These methods basically rely on the relation between the parts of composed URL. The word relatedness show relationship between words used in URL of registered Domain. In last few years, many methods were designed for countering dangerous phishing techniques and are focusing on real time implementation of phishing problems. By correlating data which is used recently to counter phishing is one way to achieve solution.

BOOK RECOMMENDER CHATBOT

Today, chatbots are changing the way world communicates. They are getting popular day

by day. According to recent study, more than 3 billion users are active on top four messaging apps. Chatbots are being used in various domains. The purpose of this project titled "Book Recommender ChatBot" is to develop the chatbot in learning or educational domain. Whenever we want to learn something we have to search a lot on that topic. We have to find good reference material for that. It requires surfing over the Internet. Many times after searching a lot, we realize that we have not got the material what we expected or we are not cleared with that concept. Instead of doing all these things, it would be better if a chatbot either recommends you a good reference book to learn particular topic or it will teach you that topic through some series of question and answers. It will check your understanding through some tests. We hope that this project will definitely help those who want to learn different things using proper reference material and change the way the world learns.

NETWORK RECOMMENDATION BASED ON ROUTE PREDICTION AND MOBILE TOWER LOCALISATION

Poor signal reception in mobile phones is a major issue that many people face today. With an increasing number of people getting access to mobile phones every year, it is vital to improve the quality of services for all. The reception of every network provider is not uniform at all locations. Mobile Virtual Network Operators (MVNOs), like Google's Project Fi, have come up with a solution to switch to the best network available in the area, to reduce the problems of bad network connectivity. Switching to the strongest network ensures that that the user will have the best possible connection strength at all times. For this, knowing which mobile network to switch to is essential. Moreover, knowing what location the user might go to in the near future beforehand can help ensure a seamless transition to the best network in that location. In this project, we propose a mobile network switching system to predict a user's route, estimate the locations of mobile towers through trilateration and clustering, and find and recommend the strongest network in the user's predicted location beforehand. This will solve the problem of weak connectivity and delays in transitioning from one network to another.

GNUKhata

Applications of Computer Science are ubiquitous. It has touched almost all the application areas. Along with the proprietary software’s used in diversified application areas, alternative open source software’s have been increasingly made available. In spite of this, there was a dearth of open source software for accounting and inventory management in financial sector in the last decade. Identifying this need, GNUKhata was started in 2009. It started as accounting software, serving the Chartered Accountants to help them in daily

bookkeeping. Since last year, GNUKhata has been actively developed to also include inventory management system. Here, our work mainly focuses on understanding the underlying framework of the currently existing software and contributing to its further development. The milestone for GNUKhata version 4.0, was to provide a full fledged inventory management system which includes features like category, products, godown, purchase order, delivery note, invoice and some reports. Our aim is to work on category, products, godown, delivery note, invoice and generate some reports like unbilled deliveries report, log report, etc. which will facilitate the completion of milestone 4.0.

IOT BASED GARBAGE MONITORING SYSTEM

This project IOT Garbage monitoring system is useful to make the cities
clean, smart and good. This technique observes the dust bins and informs the extent of garbage collected at intervals the rubbish bins via website. The level of garbage is being calculated by ultrasonic sensors that square measure compared to the bin's height. The system makes use of Alf and Vegard's architecture family micro controller, liquid crystal display monitor, wireless native space network equipment wireless fidelity for inflicting information and a buzzer. A 12 Volt device is used to power the system. The liquid crystal monitor is supposed to look at the progress of the extent of garbage collected at intervals the bins. Where as an internet page is utilized to indicate the results of garbage watching system to user exploitation it. web content ofers a pictorial browse of the dirt bins and entails the rubbish collected in modification order to purpose out the extent of garbage collected. The digital display screen shows the standing of the rubbish level. The system alert the system by noisy on the buzzer once the extent of garbage collected crosses the set limit. so this method is helpful to stay town clean by informing the rubbish levels of the bins by providing graphical image of the bins via associate website.

DATA MIGRATION FROM RELATIONAL TO NOSQL DATABASES A CASE OF MYSQL TO BAYESDB

NoSQL databases have become increasingly important with the advent of big data, corresponding scalability issues and inefficient handling of data by relational databases. But many existing systems which are using relational database management systems (RDBMS) face the problem of migrating the current data to NoSQL databases. The proposed work lies in the domain of data migration and focuses on solving the problem of data migration from a relational database to a scientific NoSQL database. The objective is to carry out data migration from MySQL as the source database and BayesDB, a scientific NoSQL database, as the target database. While designing a methodology for denor- malization, three separate approaches have been considered for optimization, static priority list, greedy approach and

dynamic programming approach. Based on both time and space complex- ity considerations, dynamic programming has been chosen as the best approach for generating an optimal join order for denormalization. A solution to automate the process of migrating data of any generic schema from MySQL to BayesDB has been proposed and implemented. During data migration, several challenges posed by MySQL have been overcome while implementing the proposed algorithm. An additional feature of query translation from SQL to Bayesian Query Language (BQL) to validate the denormalization process has also been developed. Furthermore, experimental statistics have been used to analyze the efficiency and accuracy of the methodology. The query translation interface has been used to compute the efficiency of execution of BQL queries as compared to SQL queries.

SENTIMENT ANALYSIS OF SOCIAL MEDIA AND WEB DATA USING MACHINE LEARNING

The aim of the project is to develop a system that will be a generic machine learning tool which has been depicted through development of sentiment analysis system comprising polarity calculation, domain identification, language identification, spam/ham identification, clustering of data obtained by response to user query on natural language text extracted from social media and web data. All the processing on the data is done using machine learning algorithm. This data will be firstly normalized using various pre-processing techniques and then further undergo natural language processing. It will also perform tasks like multi word extraction and named entity recognition, Sarcasm identification, analysis of Contextual tweets for cumulative sentiments of a given base tweet, long distance dependency, ranking of documents. The data will be stored and indexed through Apache Solr.It will then be accessed by user query on a graphical user interface. Clustering of data will take place due information retrieval. The entire output is depicted through comprehensive visualization of sentiment analysis of social and web media where the efficiency is determined by inherent Machine Learning algorithms and training of the Learning systems.

HIGH FREQUENCY TRADING USING DEEP LEARNING

High-frequency trading (HFT) is an algorithmic trading platform that leverages the power of advanced computing machinery to execute large number of trades in a very small duration of time. To bring this to perspective, the duration for which a stock is held in London Stock Exchange is roughly two seconds and by every passing day, the hold period is gradually decreasing. To maintain such high trading frequency, it is a necessity to leverage the computing power of machines and to train these machines to take trading decisions. Accurately predicting the price fluctuations in stock market is a huge economical advantage. The aforementioned task is generally achieved by analyzing the company, this is called as fundamental analysis. Another method, which is undergoing a lot of research work recently, is to create a predictive algorithmic model using machine learning. To train machines to

take trading decisions in a high frequency trading scenario, the latter method needs to be adopted.Deep Learning algorithms have come up as one of the recent breakthroughs in machine learning techniques. These models are a great match when it comes to making sense of large amounts of random and unpredictable data. A huge database of highly granular stock market data will be used to train the model, to make short term predictions ahead of time. Specifically this research will be using Deep Neural Networks to tackle the complexity and randomness involved in the prediction of stock prices. A lot of corporations and individuals who have been successful in the market have always been guided by the policy of buying low, selling high. By combining the stock prediction and various trading strategies, a complete platform suitable for algorithmic traders can be created. There are a plethora of strategies to choose from, broadly categorized in mean-reversion and momentum based. This research uses strategies appropriate for HFT. Few of the strategies chosen by the HFT traders across the globe are, Pair Trading, Volume Weighted Average Price, Time-Weighted Average Price, Crossover Moving Averages, etc. In this research, a model which will consider the pre- diction of the stock price and apply a trading strategy on that prediction to realize simple positions either bullish or bearish will be created. And doing so in a very short span of time and trading large volumes, a lot of risk free profit can be reaped. During the past decade, the stock market has won one of the many races to being truly unpredictable, and because of this very trait it has become a true classic in the field of artificial intelligence; however it still needs some attention in order to attain a near perfect solution. And given the huge history of use of AI in algorithmic trading, Deep Learning and HFT would become a powerful combination which would increase the speed and efficiency of current HFT techniques.

AUTOMATED EMAILS CATEGORIZATION

As the world entered the digital age, the rise of people using electronic mail has surged rapidly. While in the past, it was used as a medium to communicate important information regarding business; it has paved its way into our everyday lives. Email clients such as G-mail and outlook are vastly used by people for both personal and professional use. However, with the boom of information available to us on the Internet, our email inboxes are cluttered with emails of myriad content ranging from important bank statements to constant product promotions making it difficult for us to sort through the inbox. Our inbox remains unorganized in spite of Gmail's addition of Primary, social and promotions folders. Our proposed model aims to alleviate this difficulty by labelling each email with an appropriate topic and grouping them together under a single folder. We improve on the existing system where you have to manage your inbox folders manually or specify rules explicitly. The aim

is to use unsupervised learning techniques to obtain our objective. For this purpose, we have selected the Enron corpus dataset to build and test our model. We have used the elbow method in collaboration with the topic modelling algorithm Latent Dirichlet Allocation to perform this operation.

GRAPH BASED INFORMATION EXTRACTION AND RETRIEVAL SYSTEM USING NLP

Natural Language textual data in digital age in the WWW and social media is unstructured and the lack of structure makes it difficult to retrieve. Text analytics is most important in the age of Big Data and data is ever growing. There is a need for an automated retrieval process with proper Information Extraction to structure the data for relevant and fast text analytics. The first big challenge is the conversion of unstructured textual data to structured data. Our project aims at developing a graph based information extraction and retrieval system. The project involves two main phases: 1) XML/English text to graph database creation and 2) Querying and visualization. In the first phase, we build a graph of entities and relationships from given input XML and unstructured English sentences. We have identified patterns in sentences to extract relationships between entities. In the second phase, we query the graph database with natural language queries to retrieve results. The query results are visualized with a graphical depiction of entities and their relations so as to give better and quicker insight into the information.

ACADEMIC ADMINISTRATION MODULE FOR EDUCATIONAL INSTITUTES

Gone are the days when students used to write their assignments on paper and come to college to submit them or even stand in queues for registration. Today is the era of technology where everything is done at the click of a button. Academic institutes need software which is robust, efficient and fast, which handles all the tasks previously done manually. An open source software capable of doing all these tasks dosen't exist. This project is development of such kind of a software. Some of the basic features of this software are provision for storing all the information related to a student, teacher and all other staff members, being able to view the grade reports, teachers can enter student marks and attendance. Also course related information like the score, feedback etc are stored in this software. This software can be used by multiple different users like students, teachers and all the other staf members from the academic institution. Different users have different roles and permissions. For example a person assigned the role of a teacher will have the permission to upload a new course while a student role would not .A user's role can also be changed (eg after geting promotion) which will also change her permissions. To manage all these permissions, the software should have a robust access management software.

IOT BASED GARBAGE MONITORING SYSTEM

This project IOT Garbage monitoring system is useful to make the cities clean, smart and good. This technique observes the dust bins and informs the extent of garbage collected at intervals the rubbish bins via website. The level of garbage is being calculated by ultrasonic sensors that square measure compared to the bin's height. The system makes use of Alf and Vegard's architecture family micro controller, liquid crystal display monitor, wireless native space network equipment wireless fidelity for inflicting information and a buzzer. A 12 Volt device is used to power the system. The liquid crystal monitor is supposed to look at the progress of the extent of garbage collected at intervals the bins. Where as an internet page is utilized to indicate the results of garbage watching system to user exploitation it. web content ofers a pictorial browse of the dirt bins and entails the rubbish collected in modification order to purpose out the extent of garbage collected. The digital display screen shows the standing of the rubbish level. The system alert the system by noisy on the buzzer once the extent of garbage collected crosses the set limit. so this method is helpful to stay town clean by informing the rub- bish levels of the bins by providing graphical image of the bins via associate website.

SCALABLE END-TO-END IOT PLATFORM FOR THE DEPLOYMENT AND MANAGEMENT OF THE INTERNET OF THINGS

The project is an end to end IoT platform ready to meet the growing amount of data and security threats. The platform enables users to deploy complete IoT projects, with minimal coding involved. The system deployment is completely automated, easily scalable and has easy integration support for any 3rd party APIs. The platform has real time data analytics running to analyze the data streams. This platform is a big data system, and it supports infinite horizontal scalability as it is based on clustered distributed backend providing best in class latency, throughput, and response parameters. Connect-X : the platform developed has a graphical interface where users can design their IoT systems using drag-drop/drop-down lists based tools. The platform will then auto-generate codes for hardware, IoT hubs and cloud; and deploy them on the servers. The key feature for this platform is that its every component ensures scalability and security. It supports every IoT data communication protocols available (MQTT, CoAP, DDS, HTTP2.0, XMPP). The real time data analytics provides easy interfacing with 3rd party databases and services thus enabling users to monitor their IoT devices and trigger actions they want when certain conditions are set. Light-weight, secure and reliable M2M communication support is an inbuilt feature of the platform. This platform enables users to easily create web-dashboards, android applications

to monitor their IoT enabled devices. Through the web-UI, users can decide Rules for the data streaming to the cloud. These rules will enable real time analysis of data streaming to perform actions as decided by the users. Connect-X enables a code-less deployment of IoT systems. The users need not code any part of the backend, and every component is auto-configured to scale up as per incoming data traffic. IoT platforms are an upcoming technology concept with the current market leaders having launched their platforms not more than 5 months ago. The proposed platform architecture makes it better than all existing platforms like AWS IoT, Azure IoT, IBM Watson IoT, etc. As the demonstration of the capabilities of the platform, the project includes a home- automation use case based on Connect-X platform. Users can monitor the state of electronic devices in their home and control them via an android application. We also implemented a contextual personal assistant (Chat bot) to control various devices at home. Users can automate the home applications by commanding the chat-bot to set time and conditions based triggers like Start AC on 8am or Start AC when temperature is greater than 40 degree all via voice commands. Thus the project is a complete end to end implementation of an IoT Platform using Open Source Technologies allowing the users of the Platform to deploy IoT Solutions easily.

BIDDING MODULES FOR OPENCART E-COMMERCE

Our project aims at building auction process modules for web-applications built using Open-cart Framework. Firstly, with the changing lifestyle and constant desire to upgrade to a higher standard of living people usually want to purchase new products such as cars and homes. However, this leads to the need to sell their old properties and earn a value on that. Secondly, in case of non-payment of loans on properties the banks repossess the goods for which loan is offered. These goods are non-value to the banks and need to be auctioned to enable the banks to recover their loss. In both these cases, the customer's i.e people and banks trying to buy/sell properties end up paying a hefty amount of brokerage. Our auction portal enables a direct end to end interaction between the two parties willing to buy and sell the products thereby eliminating the broker chains in between. Also online auctioning via our portal makes it easy and convenient for people from different cities across India to be able to bid on a particular product rather than local live auctions.

WEB DOCUMENT CLASSIFICATION BASED ON ONTOLOGY

In this project we propose a methodology to make domain ontology of a given website. In this project we get a website and parse through its contents. After parsing, the output is then processed upon by the statistical algorithm, i.e. TF-IDF. This statistical analysis will give us the required information about the domain which can be used to build ontology. The ontology thus created will represent the proper taxonomy of classes and their interrelationships. This ontology will help in proper

classification of a website and ability to give relevant output when complex queries are given. The system will then make use of the ontology to classify it in comparison with the ontology's in the database.

VISUALIZATION OF HIERARCHICAL DATA

This paper hopes to tackle various problems associated with the visualization of any kind of hierarchical data. It provides a background on data visualization methods, hierarchical data and mentions various existing visualization techniques. It also discusses in brief the merits, shortcomings and compatibilities of these techniques in relation with hierarchical data. We aim to assimilate appropriate and usable visualization techniques to create a new technique compatible with almost all forms of hierarchical data and present that data in a convenient and user friendly manner.

OPTIMIZATION OF PERFORMANCE PARAMETERS ON HADOOP

The aim of the project is to fine-tune parameters of Hadoop in order to achieve a configuration which gives best performance of the system the user is working on. We work on tuning each individual parameter from a series of ranked parameters and seeing its effect on the other parameters. We perform a wide range of analysis on performance parameters which can be used to deploy cloud services to customers by service providers.

STUDY AND ANALYSIS OF OPTIMUM GRAPH PARTITIONING FOR ARCFLAG BASED DIJKSTRA SPEEDUP

Determining the shortest path from one node to another in a graph is probably the most popular question in graph theory. If the graph is non-negatively weighted, Dijkstras algorithm is the clas- sic algorithm used to answer this question. For the application of dealing with huge numbers of shortest-path queries in static graphs, we consider an algorithm, which uses pre-processed data to decrease the search space for each shortest-path request. One of such algorithms is Arc-flag acceleration. It keeps one flag vector per arc which tells about reachable nodes through this arc. To improve efficiency, this method partitions the graph and, for each edge, the pre-processing considers the relevant regions which have the shortest path over this edge. So, it is interesting to know the search space optimal partition for a graph. Here, our work will mainly focus on studying the problem of finding Arc-flag optimum partition for graphs. Since, finding such partitions for an arbitrary graph instance is NP-Complete, we focus on restricted graph classes. Our work mainly deals with theoretical analysis of such optimum partition for some undirected trees. Those undirected trees carry some sort of inherent symmetry within them. Finally, we have done some experimentation to support our claims.

ANALYSIS OF FEATURE SELECTION ALGORITHMS ON HADOOP PERFORMANCE PARAMETERS

This paper aims to analyze some of the most popular Feature Selection algorithms using Hadoop parameters. Apache Hadoop is used as a Big-Data platform for Big-data applications. In order to analyze the performance of these applications, one needs to get hold of the most relevant Performance parameters which are responsible for performance efficiency of Hadoop. As the Hadoop job dataset is unsupervised, we have proposed a work-flow for Fea- ture Selection on unsupervised dataset. We have used Clustering (k-means), ReliefF, IRelief, Random Forest and Support Vector Machine(SVM). Also, a better combination of Ranker and classifier is proved using the results. ANOVA and chi-square methods are also used as a statistical support for our findings. In the end we have been able to shortlist 26 parameters.

PREDICTING THE LIKELIHOOD OF A RECOMMENDED ADVERTISEMENT BEING CLICKED BY A USER USING FEATURE ENGINEERING AND OTHER MACHINE LEARNING TECHNIQUES

Most of the popular websites have sections of their pages dedicated to recommend content that may interest their viewers. The task of recommending content on a publisher's web page is done by Content Recommendation Platforms (CRPs) like Outbrain, Taboola. The revenue generated is in terms of cost per click of advertisement and is shared by CRPs and the publisher. To maximize revenue, it is necessary to maximize the clicks on advertisements. CRPs use sophisticated algorithms to decide which advertisement should be shown on which publisher site and to which users.The project aims at developing a system to predict likelihood of an ad being clicked that have been recommended to a particular user. The system will have as input meta information of the current web page, user and of the advertisements shown on the webpage. The project focusses on exploring various machine learning techniques and extracting significant features to maximize the prediction accuracy. This helps in increasing revenue for CRP and enriching user experience on the internet by showing content that appeals to the users.

DOORBELL ALERT SYSTEM FOR THE ELDELY AND DEAF

Deaf and old age people are faced with everyday challenges in identifying various sounds like baby crying, doorbell ringing, fire alarms, phone ringing, etc. The objective of the project is to design and implement a reliable, effective and efficient system for such people who stay alone in their house, to notify them when the doorbell rings. When the visitor presses the doorbell, captured image of the visitor is stored on the server with the specific

date and time. These details are also sent to the registered device via an Android application. The doorbell also triggers the vibrating wearable device which notifies the person that someone is at the door. The captured image with date and time can be retrieved later using the Android application. A user- friendly interface is provided for the same. This system is easy to use and also enhances the security of the user, making him less vulnerable to unwanted visitors and burglars.

ENGLISH LANGUAGE UNSTRUCTURED DISCOURSE LINKING TOOL COMPRISING LEXICON BUILDING AND ONTOLOGY CREATION

Natural Language text is not bound by a fixed structure. The usage of language keeps evolving with time. For a machine to understand the language, the challenge lies in resolving the ambiguities and capturing innovativeness. Due to its unstructured nature, discourse linking, required for understanding and generating text through machine is a challenging task. Also, dealing with sentences varied in nature and changing them into generic structure is an additional challenge. Considering these difficulties, we still believe that an effort to create a tool which can link the discourse structures in any unstructured natural text is a big step in NLP. In this project proposal, we aim to create a discourse linking tool for English language. This tool will be based on specially created lexicon suited for discourse linking purpose, which also needs to be developed. Currently available lexicons such as dictionaries, WordNet, etc. are not suited for their lacking of context information and pragmatic properties. We need a cus- tomized lexicon with contextual and pragmatic properties. An ontology will be developed for political news domain as a part of this project. The lexicon built will be based on this ontology and inherent relations of the ontology will be used for discourse linking purpose.

N-QUERY - A NATURAL LANGUAGE STATEMENT TO SQL QUERY GENERATOR

Today, virtually every relational database management system(RDBMS) uses Structured Query Language (SQL) for querying and maintaining the database. Users accessing relational databases need to learn SQL and build queries in the right syntax for retrieving the data. It becomes a big hurdle for all those who are not technically knowledgeable in this domain to write the queries in SQL. It would be very convenient if the relational database system can be queried using natural language like English. In this research, an intelligent system is designed between the user and the database system which accepts natural language input and then converts it into an SQL query. The research focuses on incorporating complex queries along with simple queries irrespective of the database. The system accommodates aggregate func- tions, multiple conditions in WHERE clause,

advance clauses like ORDER BY, GROUP BY and HAVING. The system handles single sentence natural language inputs, which are with respect to the underlying database. The research currently concentrates on MySQL database system. The natural language statement goes through various stages of Natural Language Processing like morphological, lexical, syntactic and semantic analysis resulting in SQL query formation.

AN APPROACH TOWARDS SECURITY OF WEB APPLICATIONS

Security is protection and prevention of any attack that happens on the system. In today's world the use of computer applications and Internet is very huge so to keep your information safe is big challenge. Security of web applications is very important research topic in recent years. This paper deals with detection of various types of cyber attacks. We used an ontology based detection method to detect attacks on the systems. According to current research more than 75 percent attacks are on HTTP protocol and more than 85 percent of the applications are vulnerable to these attacks, so the importance of intrusion detection systems increases. Intrusion Detection Systems (IDSs) are useful tools for identification of various attacks in the respective network and protect the respective systems without the modification of the end user software. So here in this project we have proposed and implemented an Intrusion Detection System basedon ontology.

AUGMENTING E-COMMERCE PRODUCT RECOMMENDATIONS BY ANALYZING USER PERSONALITY

It is a common observation today that most of the websites we visit on a daily basis, are a result of certain curated content based on our online activities especially related to our use of Google or social media platforms. This kind of personalization has become imperative for e-commerce web- sites, allowing them to convert browsers (visitors) into buyers by enticing them to increase their cart spend. The e-commerce industry predominantly uses various prediction models for product recommendations which play a crucial role in exposing customers to new products, based on their online behaviour. These are items that have been frequently viewed, considered, or purchased with the one which the customer is currently considering. Harnessing the power of data science has undoubtedly helped e-businesses to improve customer satisfaction and keep customers more engaged with their e-store. Can we better this e-personalization model? Our main objective with this project is to give the e-commerce product recommendation model a new dimension, by incorporating the personality of the customer in providing product suggestions. Contemporary personality psychologists believe that there are five basic facets of personality, often referred to as the Big 5 personality traits. Certain studies have explored the possibility of automatic recognition of these personality traits by mining conversational and textual data through

Linguistic Inquiry and Word Count (LIWC) analysis module. Considering the existence of an e-commerce chat application, we utilize insights from the aforementioned studies to train a recognition model based on historical conversational data that the application would gather over time from its users. This model can then be used to satisfy our objective by equipping us with the tool to automatically recognize a customers personality based on his conversational pattern.

GNUKHATA FREE AND OPEN SOURCE SOFTWARE CONTRIBUTION

GNUKhata is a free and flexible software for Accounting and Inventory Management, being developed at Digital Freedom Foundation. It is a Free and Open Source Software that supports a wide range of applications in every field of economic activity, including factory or farm-based production, point of sales accounting and inventory and work in the service sector. While it serves such traditional accounting requirements, its special feature is that it can support emerging sectors of the economy who are being required to keep audited accounts, such as self-help groups, craft producers and micro-finance groups. Our aim in this project is to contribute in the inventory module of GNUKhata. We have added new features and modules for milestone 3.2.5 and version 4.0 of GNUKhata. This includes purchase and sales order modules. Also new tab feature is added.