



COLLEGE OF ENGINEERING, PUNE

(An Autonomous Institute of Government of Maharashtra)

Department of Computer Engineering and Information Technology

PROJECT ABSTRACTS

BTECH COMPUTER ENGINEERING 2015-16

ENHANCING COMPILER TO IMPROVE PERFORMANCE OF PROGRAMS RUNNING ON GPU

Our project aims at enhancing the NVIDIA compiler by adding features to it so that it can in turn improve the performance of the programs running on GPU compiled by it. We have designed two different approaches for this purpose. The first assesses the problems and solutions for adding the feature of Inline Assembly to the compiler. The second one compares the use of Data Flow Analysis approach for implementing optimizations with the current ad-hoc approach. We have developed a Generic Data Flow Analysis Framework for this purpose. We tested the correctness and performance of the framework using Convert to Boolean Optimization. The effects of this optimization on GPU shaders were also tested using various Nvidia test suites.

A NETWORK INTRUSION DETECTION SYSTEM

As the usage of internet which is a global system of interconnected networks is wide, the network intrusion detection, that is, information assurance, data integrity, network security has become a dynamic and major research area. There are many network intrusion detection systems and methods developed. We provide a structured and exhaustive overview of different aspects of the network intrusion detection methods and systems (NIDS) so that anomaly can be analysed in a systematic way and can make our system more secure from the various attacks that is, intrusions and can assure the flow of the information which is the most important service of the networking systems. We are also providing the information about the various anomalies and categorizing them into different domains and compare them and discussing the different methods and tools that can be used effectively to develop such a network intrusion detection systems (NIDS).

CROSS PLATFORM (RDBMS TO NoSQL) VALIDATION TOOL

Data - structured, semi-structured and unstructured is growing at an exponential rate. Most organizations are dependent on this huge volume of data to make real time decisions. To provide accurate information to these decision making models, it is crucial that the data is of superior quality. Since most organizations are migrating from relational databases to NoSQL databases, it is vital to validate the data after the migration process. Considering the different

structures of the two databases, the process of validation is a formidable task. This report talks about an approach to validate data between cross platform databases using de-normalized schema structures.

ANDROID APPLICATION FOR FRIENDS AND FAMILY TRACKING

In today's fast moving life services based on location has very much importance in everyone's life. Purpose to develop the system is to keep track of your family members and close persons location using smart phone (Android Phones). All over the world in every one in five people in the world has Smartphone, with its attractive features.

Any location or the place can be identified in terms of the geographical coordinates the geographical coordinates of the places latitude and longitude. The location of the user can be tracked by any technique either by implementing GPS device or by using one of the services by Google i.e. Google Geo-location API, Google places API, Google maps API. Recently Google launched new API services based on Google location API where we need not explicitly choose either GPS or network location provider explicitly, as Fused Location Provider automatically chooses the underlying technology and gives best location as per need. Services tries to triangulate your location by looking at the cell towers and Wi-Fi access points around.

System holding facility to work as client-server based application. Clients are the users having android phone with family friend tracker app and the server connected to clients and central database through internet. Server keeps track of all users and log the information in central database. As soon as the clients has installed application, server starts finding the current location of user using Google location API and store it in into the database. If the internet is of then log of last received location information of provider's device is maintained. Also in emergency case there is facility name as SOS message which will be send to saved no of SOS system with the location.

SOFTREC :QUESTION ANSWER BASED PRODUCT RECOMMENDER SYSTEM

SoftRec is a Question Answer based Software Recommender mobile application. The goal of this project is to help users get recommendation about the software that user has asked for. This application is based on the concepts of NLP and intention mining. There isn't any application that recommends software available as of now. Thus our main focus is to develop a system that processes data that is unstructured and is obtained from the question answer database that will be built through question answer portal and recommends required software among all existing software. This is always beneficial than search engines like Google or Bing as Question answer provides selected and narrowed down results.

REMOTE LAB MONITORING SYSTEM

In most computer labs Instructor have to be present physically to monitor

students progress. If there are large number of students, it becomes tiring job for an instructor to go to each commuter to solve doubts of students. To solve this problem we are designing framework that will help an instructor to remotely monitor the lab, facilitate better grading and evaluation methods. The system provides user friendly interface between teacher and student for better doubt resolution.

FIT-O-METER: A MOBILE HEALTH ASSISTANT

In today's stagnant and stressful life, people tend to ignore taking care of their personal health and fitness. Hence, there is a need to develop a system which assists the people in maintaining a healthy lifestyle by keeping a track of their daily routine. This report presents an android application (Fit-O-Meter) which at the very initial stage, it takes in the parameters like height, weight, age, gender, lifestyle and stores in the database on the server. It uses recently available advances in the smart phones such as accelerometer, digital compass and gyroscope to record daily activities like walking, running, riding a bicycle or travelling in a vehicle and determine calories burned during these activities. Based on an individual's parameters like height, weight, age, sleep, daily calories need and daily calories burnt, it recommends a daily diet plan along with health tips. This recommendation is done using user based collaborative filtering recommendation algorithm.

VOICE ACTIVATED HOME AUTOMATION

Home automation is a growing industry. This is developed to provide help for the old and disabled people. Controlling home appliances is the working of home automation. In home automation, a wireless communication link must be provided to the remote user. Main task of this system is, make a system who will control household electronic things remotely via voice command. This project explains the design and structure of wireless home automation on the basis of voice command which has been built and implemented. This system controls all the electrical appliances through voice commands in an office or house.

AUTOMATED RANKED SENTIMENT ANALYSIS OF SPEECH

This project involves the identification of prosodic features, their applicability to prosodic segregation and clustering for classification. The following prosodic features are used; Mel Frequency Cepstral Coefficients (MFCC), first order rate of change of the MFCCs (Velocity), second order rate of change of the MFCCs (acceleration), Pitch, and Spectral Entropy of the input speech data.

First, the gender of the speaker, and the emotional state of the input data is probabilistically determined based on the pitch. Then the probability of their emotional state is calculated once again using MFCCs, Velocity of MFCCs, Acceleration of MFCCs, and Spectral Entropy. The resultant probabilities from the two methods are then combined, to arrive at a final set of

probabilities for the speaker's gender, and emotional state.

ENTERPRISE SEARCH APPLICATION

Different legal documents in organization are in scanned format like Master Service Agreement (MSA) and Standard of Work (SOF) between client and any organization. These documents have to sign up by client and organizations employee (HR/ CEO/ Any Employee). For that purpose these documents are double scanned and sign up by required authority. Other than these digital documents many documents are there in organization in structured as well as unstructured format. We have to design an application which will not only make these double scanned documents searchable but also help search through structured as well as unstructured data. For that we have to extract the text from digital documents using Optical Character Recognition (OCR) and search through these documents using Solr.

COEP ASPYRA: PRIVATE CLOUD PLATFORM FOR COEP STUDENTS

Cloud computing has revolutionised the world of computer and information technology. Today a number of computer solutions are being provided using cloud technologies. Many popular desktop applications such as Microsoft Office and Adobe Photoshop have been moved to the cloud. Deploying applications on a cloud platform has a number of advantages such as scalability, high availability and better performance while cloud storage ensures fast, easy and universal access to important data.

The aim of our project was to design and deploy a private cloud platform on the COEP server. We surveyed and compared a number of technologies that could be used for designing the platform. The final design of the cloud platform makes use of container virtualization technology as the virtualization environment. Open source packages like Docker Engine, Docker Machine and Compose form core of the cloud platform. The platform is called COEP Aspyra. A general user can access Aspyra using it's web interface. COEP Aspyra provides a number of services like application deployment, storage, dynamic code execution and a linux learning portal.

SECURITY IN MANET USING DETECTION ENGINE

In today's world many systems are becoming more and more sophisticated with the growing need of the time. Simple systems are changing drastically into more useful and complicated systems. Computer networks saw a major change from wired to wireless networks and quick wireless infrastructure became a new trend. With the growth of wireless networks emerged an infrastructure less network MANET. They are used in many different places including places of natural disaster where it is difficult to set up an infrastructure and access medium for communication. MANETs have wide range of applications which makes them a popular choice.

Any node in a MANET is capable of forwarding packets from another node i.e. they act as both routers and host. Besides certain events like node joining or leaving a network, merging or fragmentation of network need to handled effectively in MANETs. Hence, communication in MANET requires different routing protocols than the other networks.

All these characteristics of MANETs make them susceptible to much different kind of attacks and malicious activities. Any intermediate node that handles a packet can alter or read the packets. Hence security is a major concern in MANETs. In the light of this fact, many different ideas for detecting malicious activities in a MANET are proposed. An effective algorithm is the one which can identify maximum possible malicious activities with a minimum possible overhead and battery utilization. Here we give a short survey of proposed solutions and attacks in MANETs. We also provide an alternative solution for providing security in MANET using a Detection Engine.

STOCK MARKET ANALYSIS USING TWITTER

Social Media Analytics using twitter to obtain huge amount of historical unstructured data generated by different Twitter users and trending hash tags. The data downloading is supported by Python libraries like Tweepy and Twitter API. The whole data is stored and maintained in MongoDB using Stream Listener in the form of a JSON object.

This data will basically be used in performing sentimental analysis on the tweets. This will help in analysing the stocks of a particular company and in studying different trends such as Word Cloud, Sentiment Analysis, Emotion Analysis, and Average Sentiment Value on hourly basis resulting in average value for the day. This analysis can be somewhat compared with actual variation in the NSE market with help of historic data obtained from NSE. The app is also useful for investors who want to study the market trends of any company and make their decision of investing appropriately.

SENTIMENT ANALYSIS ON TWITTER DATA

In Sentiment Analysis we do the computational study of people's sentiment, opinions, attitudes and emotion expressed in written language. It is one of the most research area in recent times as the social sites are getting more and more popular these days. Whenever we want to take decisions, we want to hear others opinions, starting from movie to buying gadgets. So many companies does sentiment analysis from various sites. We have chosen twitter as a platform because in tweet we can express large information in small character and it has global reach. It really helps companies to revisit their product strategy and make important changes in their product. Thus it really helps market and consumers for making economical business decisions.

ONLINE RECRUITMENT PROCESS

As we know, recruitment process is very time consuming and tedious. Lots of students submit their resume to a company. Afterwards, they shortlist students. Take their technical test and various interviews. This project is aimed at developing web-based recruitment process where all activities regarding recruitment process will be held online. From submitting resumes to getting offer letter, all things will occur online.

The process is divided into six steps. First a candidate will register for the site. Then he will upload his personal information and required documents for the job which he will apply. Company will verify his information and documents in verification process. Selected students will give online test. Then shortlisted students will choose their time for interview according to the slots given. Further wards online interviews will take place and finally selected students will get offer letter.

GENERIC BINARY CLASSIFIER TOOL FOR CLASSIFICATION OF PATIENTS SUFFERING FROM BRAIN DISORDERS IN R

We created a multipurpose binary classifier package in R known as Brain- oread. The package works as a standalone tool which is trained using two types of datasets. First dataset classifies the prediction data as Schizophrenic or Healthy Control. Second dataset classifies the prediction data as Demented or Non-demented (Healthy Control).

The package consists of models built using Machine Learning algorithms from three packages namely Caret from R, Weka from Java and Scikit Learn from Python. On loading the package in RStudio, user can supply the initial training data along with labels. This training data is split in 70:30 ratios. Each model is trained using 70% data and predicted over 30% data. It is repeated for 100 iterations. This procedure is repeated for all the models in the three packages and the models with highest accuracy are sorted in decreasing order of their accuracies. The user can then select the most accurate model and train it for 100% dataset which can be further used for prediction of new data.

STRENGTHENING E-BUSINESS USING WEB DATA MINING

The world of Internet today is bringing a massive revolution in the Business marketing strategies. Nowadays, normal traditional Business are now transforming into e-business. E-business can be defined as selling and buying the goods and also working together with business partners, performing electronic transactions, servicing the customers and conducting e-learning within organization. In e-commerce websites different kind of services and products can be sold, advertised and introduced. Through it, you can reach a majority of people without being concerned about time and distance. Organizations have started making use of strategies to make their web-pages better in terms of performance than their competitors. This increasing business operations through internet have resulted in unstructured data found on net. Thus, making

extraction of information from such a huge data, a very tiring and time consuming task. Data mining concentrates on the browsing pattern in the historical data and try to find patterns within data, not identifiable by simple analysis. E-business companies can improve sales on product quality by using this. With the help of defined algorithms, it provides users a chance to find main elements of business. Web mining is the data mining technique discovers information from web documents and services for obtaining useful information. Web mining is used for large records from a database. For web mining the data of corporate information is public and rarely requires access rights. It can process unstructured and semi- structured data from web pages. Thus, in our project we make use of Web Mining to give e-business a data mining solution for discovering hidden aspect of business and web data. This helps the companies to make cognitive business planning and make their customer better relationship.

SENTIMENT ANALYSIS OF TWEETS

Social media and micro blog instruments are progressively utilized by people to express their sentiments and feelings in the structure of tweets. These tweets convey opinions about different topics ranging in variety of fields. We intend to develop an automatic sentiment (positive, negative or neutral) extractor from a tweet.

Our project aim is to build a library/system which can detect the sentiment of the tweets of some drugs. For this, we recognize at such of favourite micro blog called Twitter and build ideal for classifying tweets into positive, negative and neutral sentiment. In this project, we use machine learning algorithms to predict the sentiment of Tweets. We appoint Twitter messages as input data exist, as they extend a absolutely large, divergent and willingly available jointly of emotions. This is exceptionally valuable since it permits input to be totalled without manual intercession. Utilizing this analyzer:

1. Advertisers may personalize ads based on the sentiment of a user towards certain products.
2. Buyers can use sentiment evaluation to analyze products or services before making a purchase. E.g mobile phones, laptops etc.
3. Organizations/companies can make use this analyzer to gather important feedback about critical problems in newly launched product.

Our results suggest that hash-tags and other conventional markers of tweets are useful features for sentiment and emotion classification. In future work, we will explore more moods of tweets.

STOCK MARKET ANALYSIS AND PREDICTION

Our project will analyse previous stock value of companies in IT sector(Microsoft Corp., Oracle Corp., Symantec Corp., Cognizant ,Google Inc.) with help of certain parameters that affects stock values like Close, Open, High, Low, Volume, Bullish and Bearish intensity for the company over a span of time.

We will implement this data in data mining algorithms and extract patterns in the way stock

markets behave and respond to external stimuli. We will consider the Bullish and Bearish data of Twitter users and its resemblance with the up and down changes in the closing values of stocks. This will help us to predict future values of that stock.

These predictions will be useful for investors to invest in stock market. Our software will allow users to set an alert price for particular company. As prices reach to higher or lower the set price, user will receives an alert. So users don't need to keep a constant track of the company's stock price.

WEB BASED E-LEARNING ANALYTICS

We have developed a web based platform which uses different analysis models and learner produced data to retrieve information. There is a need for accurate and meaningful analysis of data. Our system provides an environment in which a user can solve questions, read blogs, and participate in discussions and access study material. Based on the user interaction and using analytical models, our system gives related output. We have included pie charts, stacked bar graphs, simple time series visualization and wide range of data analytics. The analysis is about demand/popularity of any subject or area, result analysis for number of users and usage analysis within particular period. We have also included functionality to run R programs remotely on the platform.

ROAD TRAFFIC PREDICTION AND CONGESTION CONTROL USING NEURAL NETWORK

In past decade, the problem of traffic has become severe due to industrialization especially in big cities. Hence, the urban population has to invest much valuable time during travelling. Increased road traffic results in more number of road accidents and more consumption of fuel, thus wasting energy.

Dynamic traffic flow and static traffic signal is major problem which results in congestion of traffic. Hence for solving this issue, this project proposes prediction of road traffic using artificial neural networks which will ultimately control congestion and results in the smoothening of overall traffic. Artificial Neural Network possess power to learn from past and predict the future. Neurons - the back-bone of the neural network which are trained with real-time data (time-based data) based on which it predicts future traffic volume. This project proposes the significance of Jordan sequential network for prediction of future values, depending upon the current value and aggregate past values and also guarantees prediction of traffic flow with accuracy of about 92-98% using Jordan sequential Network.

PUNE CITY BUS TRANSPORTATION

Transportation is playing an important role in development process of every nation and has

became an important aspect in our life for mobility. Nowadays we can see that excessive use of private vehicles has caused traffic problems in big cities. Public transportation, while may not be as enjoyable as traveling with our personal vehicle. But public transportation definitely allows you to relax, read or nap during travel time instead of fighting and feeling the road rage. It is undeniable that public transportation have some weakness points. One of weakness is lack of information about routes, available buses, timings of City Transportation. Despite of this Public transportation brings a lot of benefits for individual and society. The extensive use of public transportation will automatically reduce traffic problems. There will be reduction in consumption of fossile fuels and environmental pollution also. We believe benefits overweigh drawbacks. So our aim of project is to encourage people to use public transport instead of their private vehicles. By providing them an application which enable them to get easy access to information about routes, available buses, timings of Pune City Bus Transportation , Monthly pass registration and pass renovation.

TIMETABLE MANAGEMENT SOFTWARE

Time table generation is a time consuming problem faced by many educational institutes. It belongs to the class of combinatorial optimization problems. We propose a semi-automated approach for solving this heavily constrained problem for educational institutes like College of Engineering Pune (CoEP). It will allow the users to make time table as per his/her choice while ensuring all constraints are satisfied and there are no conflicts. There are some existing solutions for this problem which are fully automated but difficult to use. We propose a much simpler approach for solving it. We aim to develop a desktop application using Object Oriented Programming paradigm with a user friendly interface.

PREDICTING RETWEETERS IN TWITTER

While a lot of work has been done in the fields related to predicting the number of retweets or the virality of a tweet, not many works focus on the followers who are most probable to retweet a given tweet. In this project, we aim to create a system for predicting the followers of a tweeter who are most likely to retweet a given tweet. A study of essential features was done in order to determine their importance in retweet prediction. A training set was then designed as input for the Support Vector Machine (SVM) Rank algorithm. Within this learning-to-rank framework, the various feature families considered are interests' similarity, retweet history and followers' status. The Machine Learning Algorithm learns based on the training set, and given a tweet as an input (test set), the Machine Learning Algorithm identifies the list of most probable 10 retweeters.

E-MAIL VIRUS DETECTION: BY NETWORK TRAFFIC MONITORING AND VIRUSTOTAL API

As E-mail viruses can cause substantial damage to our system and network, so it has become high requirement benchmark to stop the spread of such mails. The purpose of our project is to detect and to stop the spread of email viruses. We are implementing our Virus detection process in two steps. In first step we are using the Virus Total API to detect the virus. In this step we detect the viruses whose behaviour is already know and stored in one of Anti-virus application. And in second step we are using the tree structure algorithm to detect the new viruses whose behaviour is not known to anyone.

SMART CITY ANALYTICS

A Smart City uses digital technologies to improve urban areas, to reduce costs of services, resource consumption and to interact more effectively and actively with its citizens. Smart city technology is developing under these areas: transport, electricity, education, healthcare, water and waste. Smart City is a very important initiative by Govt. of India. PMC is also actively working on collecting citizens thoughts through various forums on this topic.

Our project is mainly based on collecting data related to Smart City. Then perform analytics on collected data and generate some meaningful insights from it. Mainly task is to perform Text and Sentiment Analytics on collected data

SOFTWARE DEFINED DATA CENTER COMPOSER

Most of the data centres today are software defined. In SDDC (Software Defined Data Centre) the entire infrastructure is virtualized and delivered as a service. Core components of SDDC are Network virtualization, Server virtualization and Storage virtualization. The architecture of SDDC (server, software defined network, software defined storage) is defined by the applications running on it. This project talks about Software Defined Data Centre Composer which allows visualizing the architecture of SDDC (i.e. interconnection between compute, storage, network and security components). This architecture is realized over the virtualization platform that runs over the actual hardware. The advantages of SDDC Composer are effortless designing and fast deployment. Templates for popular network architectures are provided which facilitates rapid deployment of services.

AUDIO STREAMING PIPELINE OPTIMIZATION AND MULTIMEDIA VALIDATION FRAMEWORK

Nvidia GameStream is a technology that allows streaming of high definition interactive applications or multimedia content to a portable device such as Nvidia Shield or tablet. The audio pipeline involves components like audio encoder and decoder, audio capture libraries, FEC, UDP protocol sink and jitter buffers and the flow of control varies on the client and the

server side. As part of SW-GPU Audio team, this project involves optimizations and additions to the existing audio framework.

1. Tool for capturing, converting GameStream network captures with packet reordering and logging.
2. Adding functionality to existing network streaming framework to measure audio bandwidth.
3. A generic Google Chrome extension that can communicate bidirectionally outside the sandbox to background services and drivers.
4. Automating tests for Nvidia CoPlay.
5. Test feasibility of a GPU driven, geometry based, audio ray tracing algorithm.

The project will help NVIDIA Corporation to add more functionality and immersion to the Nvidia audio framework and optimize and increase the fault tolerance in company internal tools.