

Department of Computer Engineering & Information Technology

ABSTRACTS OF PROJECTS FROM B. TECH. (Computer Engineering)

OpenNJ - Stay notified, optimally

OpenNJ, (where NJ stands for Notification Junction and Open indicates that it is open source), is an extensible framework that will interface with multiple applications and multiple notification providers, abbreviated as NP(s). It is available on github[13]. OpenNJ is similar to an M x N multiplexer, there being M applications and N Notification Providers. OpenNJ primarily aims at solving what is called the doorbell analogy. In a house, it does not need to be checked every few minutes whether there is someone at the door. Similarly, CPU cycles can be saved by monitoring NPs intelligently, and by preventing polling. OpenNJ has to be extensible, in that more applications and more NPs should be easy to add without having to recompile OpenNJ. OpenNJ needs to be robust so that an Application should be notified in the case that the NP it is seeking notifications from, fails. An Application needs to register with OpenNJ, informing it about the notifications it needs.

OpenNJ has two interfaces -

1. Application - OpenNJ Interface
2. OpenNJ - NPs Interface

It is intended that OpenNJ be extensible, robust and modular. A prime feature of OpenNJ is that it is an aggregation of NPs. OpenNJ will abstract the notification acquisition process, making it simple, organised and efficient.

Implementing Security in Message Passing Interface using One-Time Pad

Distributed Computing system is a cluster of computers connected on a network which coordinate their actions via passing messages. Since the cluster can also be accessed publicly, the security of the messages becomes crucial. In this paper we propose a highly efficient and secure implementation of message passing interface (MPI) in distributed environment. This implementation, which is named as VAN-MPICH2, integrates security measures to ensure data confidentiality using One-Time Pad (OTP) encryption technique. Since the proposed encryption implementation decreases the security overhead substantially, VAN-MPICH2 manages to provide confidentiality with very insignificant decrease in performance.

DISTRIBUTED FACE RECOGNITION USING HADOOP

This paper introduces a scalable and flexible solution for face recognition using distributed Hadoop cluster. With the increase of use of CCTV for security purposes there is a need of efficient solution to utilize the vast data generated by them .In case of an investigation it is ideally required that the evidences are evaluated as fast as possible. Usually a week's recording of a CCTV footage is approximately 100 GB. Hence in this proposed solution we use avconv, OpenCV and Hadoop MapReduce for fast distributed video processing. We have experimented and analyzed the performance of facial recognition on different models and distributed configuration. We found Hadoop as an effective framework for video processing to recognize face.

DEVELOPMENT OF AN HCE BASED APPLICATION FOR EVENT TICKETING

We aid in designing an application on the Android phone that will carry out an NFC transaction using HCE for event ticketing. NFC or Near Field Communication, is one of the most promising new technologies to be incorporated in Smart Phones. NFC allows communication between a point of transaction such as a terminal, and the users phone, enabling it to be used for proximity transactions such as Metro Transit, Credit Card payments, etc. It has a great role in digital commerce. The latest version of Android KitKat 4.4 and above, has now made it possible to implement NFC transactions without a hardware secure element. This is realizable through a mode called HCE or host card emulation. In HCE, the device application directly communicates with the NFC terminal. The objective of the project is to develop an android application showcasing HCE and Terminal and/or server emulation to use with the above application The various entities involved are:

- 1) Server
- 2) Client (Mobile)
- 3) Terminal
- 4) VAS (Value Added Services) Lite

The Server contains the information about all the tickets categorized by: entertainment, match, concert, play etc. When the client syncs on start-up with the server, it sends information about the tickets on mobile. The client then performs the booking of tickets and saves it locally. At the same time this transaction information through server is shared with the terminal which maintains it in its own database. On redeeming the ticket, Vas Lite comes into picture. The terminal reads data

through vas lite and compares it with the information stored in its own database and accordingly approves/denies the ticket. The mobile then retrieves the status of the ticket from VAS lite.

CASCKET: A Binary protocol based C client-driver for Apache Cassandra

Few can deny the diverse applications of NoSQL Database technology today. The last decade has witnessed an exponential rise in not only data generation but also in the manipulation and management of the same. NoSQL Technologies have engendered a new platform for large enterprises to come up and have provided them a wide vista for data manipulation. Moreover, existing technologies have been augmented with the unique features that this technology offers. Apache Cassandra is one of the leading NoSQL Databases today. The rise of Cassandra has also given rise to corresponding high level client drivers. There are several drivers present today for interfacing Cassandra with languages such as PHP, Java, C++, etc. However, to the best of our knowledge, no such driver for C exists. This report proposes such a driver, Cascket that allows direct communication of Cassandra over C without the existence of a middle-ware.

BUILDING A PLATFORM TO PROVIDE SECURITY TO THE VIRTUAL MACHINES ON A SAME HOST

VMware NSX is the network virtualization and security platform for the software-defined data center. NSX brings virtualization to your existing networking and security constructs and transforms networking and security operations and economics. Running multiple virtual machines on hypervisors (host machines) significantly reduces operational costs by employing cpu and memory consolidation. Similar consolidation for security is provided by VMware NSX Guest introspection and Network Introspection solutions. NSX allows security vendors to register and deploy third party appliances as a security services that enable security for guest virtual machines without requiring to have components running inside the virtual machines. One instance of an appliance is deployed on every host. This project demonstrates the ability to create a security appliance, register it with NSX, deploy the appliance on all hosts that are to be protected using this appliance and invoke the EPSec, NetX API's provided by NSX to partners. This reference solution will be used by VMware partners (like Symantec, McAfee, Trend Micro etc.) for integrating with NSX and building their security solution.

Image Analytics for Classifying Data Sets of Large Class

Applications analytics are growing for effective results. The techniques of image processing, extraction, compression, and recognition are established in applications and are in use. The presence of large data set of images on the web is increasing exponentially. However, there is a need for an automated online methodology for processing the relevant sets of images. In this paper, we propose a mechanism that retrieves images from a dataset of large class and arrive at a cluster for deciding behavioural trends across four well-known locations on the globe. The proposed system is implemented using distributed cloud technology and a crawler script to show interesting results.

SUPPLY CHAIN ANALYTICS

Management of data in an organization is a critical factor for its excellence. Managing supply chain is a complex task. Companies, firms, banks etc. are generating various kinds of data. The problem of understanding what the data is representing is quite common and one easy solution to that is to deduct useful information, patterns, connections that are relevant. One cannot just rely on practices to survive against its competitors. There is a need of accurate or rather more meaningful analysis of data. Descriptive analytic is used to deduct the insights. We therefore present a simple web platform that can help the above entities to understand the data visually so that generated data makes more sense and focus on whats more important. We have used d3.js library to generate interactive visualizations. Interactive visualizations include stacked bar graphs, pie charts, simple time series visualization, multi time series graph, bipartite graph and bubble graph to visualize wide range of data sets.

Data Insight file system interface for Hadoop

MapReduce is programming model that analyzes large data sets in a parallel and distributed environment[2,7]. It works on principle of splitting the input data set into multiple chunks that are processed concurrently. MapReduce contains Map function that processes data sets and generates key/value pairs while Reduce function combines key/value pairs with same key. Both the input and the output of the MapReduce jobs are stored in HDFS or in Local File System[6]. However carrying out the same functions on remote data set brings new issues like need of manually uploading data to HDFS. We propose a method to carry out MapReduce job on remote data (data available on machine without hadoop installation) through custom file system implementation and TCP/IP

socket programming.

COEP Placement Guidance Portal

In Final year of Engineering every student has queries about the placements and as senior are not present there at that instant to solve their doubts and to help them. A portal should be there so that pass out students can post their experience about company and can discuss with Junior to help them out. Detailed information of Placed student will be provided so if individual wish to contact them personally he can. The online sites of study material of placement will also be provided for their betterment . This will help the student to prepare in a better way for placement. We are also providing student profile to companies, so that they can select students according to their interest(eg. Pointer). Most interviews make students feel like the interviewer is putting them under a microscope. In this portal, we suggest student put themselves under a microscope following an interview to evaluate their own performance. If they find there are some areas where they fell short, improve them prior to their next interview using this portal. Doing this after each interview will improve their performance over time.Our goal is to prepare students to understand the interviewing process. We work to help students know interviewing strategies and structure so they can be ready to showcase everything that makes them strong applicants.

Rapid 3Dimensional CAD product designs using Automatic region detection in 3D space

The project deals with the Automatic detection of non-overlapping closed region in 3D and 2D curves. The 3D curves are represented using CAD/CAM software which are used in modeling a 3D CAD product. A 3D product like Shoes, jewellery etc. requires random curves (graph) drawing on their 3D solid model with the help of CAD software (like Powershape). While modeling a 3D product user has to trace all the edges of loop manually. Now to ease this process the algorithm automatically detect the non overlapping loop in 2D which is mapped from 3D curves designed on solid model.The non-overlapping loops from given design get automatically detected by just mouse hovering on it. For rapidness in designing a product, apart from automatic region detection there is provision of different loop arithmetic such as loop merging (addition),loop subtraction which helps user to simplify the process of modeling.

Remote Desktop Application Using RFB Protocol

In today's world, science and technology have made one's life very easy. Tools and devices have been invented to do work in a faster and simpler way. With the help of technology we can remotely access these devices. There are already many applications available in the market to control and monitor these devices. A user can access any computer, with a server running on it, by using our PocketPC application. It is an Android app to control a target PC. The requirement to do so is a VNC server must be running on the target PC and the user who has the control should know the IP address of it. Of course both the devices should be connected via Internet or Wi-Fi. A user can have full access of the target PC. PocketPC surrounds the client and server application in which, the server application has been implemented in Java and the client application is implemented in Android. The reason behind why Java or Android is that both of them are open source platforms which allow working on new ideas with a set of open standards. We can perform operations like sharing a file, editing and saving it, start and stop applications, troubleshoot the problems on the target PC, shut-down the target PC and much more. An IT administrator can remotely control any computer connected in the network, to troubleshoot and solve the problems in a faster way. As the number of smartphone users and mobile applications are growing rapidly, this could be the future of computer technology.

ASK - A – FRIEND

Owing to the vast scale of globalization in today's world, consumers everywhere can choose from a plethora of services and products. Making this choice may become an overwhelming task sometimes. Social media tries to simplify this with the advent of various applications such as Zomato, Just Dial, etc. Reviews on such platforms come from total strangers, which may not always be entirely reliable. However, for more crucial services - such as an orthopedist, or a tuition class, people would prefer advice from their own friends, which would be more trustworthy. Presently, there is no such application that provides consolidated information on various services provided from the user's friends themselves. This project aims at providing a scalable and reliable web application that will integrate the information given by the user's social circle thus enabling the user to make informed choices about various services. The web application will intelligently analyze and display the data retrieved from the users. The application makes the users login through their Facebook accounts so as to retrieve their friend list. The users are then provided with a variety of services (bookstores, tuition classes, General Physician, gyms etc.) to choose from. The database of

each of these services are populated by the users i.e. the users add their respective providers for each of these services themselves. Facebook friends of the user, using the app, also do the same. On querying for a particular service, the users can make an informative decision judging by the service providers of their friends displayed by the app.

TRAINING AND PLACEMENT SYSTEM AUTOMATION

The Training and Placement Automation System is mainly designed for automating all the activities of the training and placement cell of any Institute. This system provides opportunities to students to use collective intelligence in order to improve the placement percentage. It simplifies the job of student, training and placement system authority and company representatives. It allows students, companies and TPO to create their individual accounts and view all the available options respective to them. Students can upload their resume, submit the academic information like C.G.P.A., backlogs etc. as well as personal details. They can view the placement schedule and check out the company profile as well as the discussion forum. TPO can verify the company and student details. TPO can give the dates to the company. TPO can also communicate all the important documents like offer letter to the students. Company representative can upload the documents like company details, Job profile, short lists etc. and can set the minimum eligibility criteria. It can also communicate with the TPO for scheduling details.

Data Mining on Telecom Dataset

The aim of our project is to apply various data mining and machine learning algorithms on telecom dataset. Machine learning algorithms use input dataset to develop a model and then this model can be used to make predictions. We have used tournament dataset for developing models. This dataset is freely available on Fuqua University's website. The performance of models is determined by comparing the accuracy which is calculated by checking the predicted output and the actual output. Sensitivity and Specificity are other two measures for comparing the models. Due to application of prediction, companies will surely earn profit. We developed the prediction models on one dataset i.e. Training dataset. These models are tested on data which contains the instances, recorded 3 months after the training dataset. The result suggested that the model gives expected results, so these models can be used for future improvements. The models are developed in three stages:

1. Filtering Data
2. Attribute Selection

3. Using algorithms for making predictions

We have used three algorithms for developing models: Decision tree(J48), Logistic Regression(Logistic) and Support Vector Machine (SMO). SMO predict fairly. But Logistic and Decision Tree developed more accurate models.

A Self-configurable New Generation Children Tracking System based on Android Mobile Terminals

This report proposes a new generation child tracking solution for android platform. It enables the user to closely monitor their child's cellular phone activity in a detailed and comprehensive manner. Apart from monitoring the call logs and the messages sent and received, this solution also gives the option of setting up a geofence and a system to trigger an alarm if their ward goes outside the aforementioned geofence. Details of the current and previous locations can be accessed by using the GPS functionality. Various parameters like the call logs, the messages exchanged and their realtime GPS positions are regularly updated in the server database. This information can later be accessed by the appropriate parties. Access is only granted after the user provides the correct username/password combination.

Life @ one Touch

Life@onetouch is website which incorporate most probably the information about the availability of blood, location and path to reach the location using GIS technology This project is aimed to developing an application which will help to check the availability, location of the blood bank in case of emergency. This project also shows the nearest path from our current location to the blood bank where the required blood is available. This project is developed with simple GUI so that at the time of emergency the end user would be comfortably interact with the application and get fast availability of blood. Any authenticated blood bank owner can register his blood bank on the system and also provides information about available stock i.e blood group, quantity, purity. In addition to that any person who wants to donate the blood can register himself on the website. once registered, his details will be sent to the registered blood banks. So when there is any blood donation camp the blood bank owner will manually contact the person. This project has been planned to be having a centralised database which will be common for both the website and an android application. The database is also synchronise. Those user who dont have a notebook and the ethernet connection can also check the availability of the blood with their smartphone having an android platform. Overall,

our project provides reliability and ease of use to the end users.

REPLICATION PLANNING TOOL

The issue of data protection and recovery is of supreme importance in any organization. With the arrival of virtualization in the world of technology, organizations now are only a click away from making copies of their data. However the process of copy or what we call 'Replication' comes with a cost, usually dependent on the speed at which copying needs to be done. This is called the 'Bandwidth'. Therefore we intend to make a tool that recommends a client an optimal bandwidth for their replication process. This tool is designed for the IBM V7000 Storwize systems that have virtualized RAID architecture. The problem without using our analytical tool is that, clients will have to calculate the bandwidth by analyzing their i/o. Also, without knowing the working of V7000 products, it is difficult to write a proper logic for the same. Clients usually don't know how much i/o is being done on their storage. They might over-provision the bandwidth and incur more cost for replication. Otherwise they might under-provision the bandwidth and put their data at risk during a disaster since the data will be transferred slowly now. In this project, we have kept looking from the customers perspective about the minimal information a client might require to adjust the parameters of the replication process properly.

VeracityNXT : Multimedia Validation Framework Video Encoding Test Development

VeracityNXT is a Multimedia Validation and Regression Testing framework developed by NVIDIA. As part of the SW-GPU Video team, the project aims to deploy tests related to video encoding on VeracityNXT. These tests will run on a nightly basis or on a per-commit basis for regression testing and code sanity. Understanding the VeracityNXT test development infrastructure, making necessary code level changes and deploying the tests are the tasks that fall under the purview of the project. The tests are deployed using NVIDIA's DVS (Driver Validation System) ecosystem which manages their compilation and execution. The tests are executed on VRL machines with various OS-GPU configurations as per the requirement of each test.

The various tests deployed are,

1. H.264 and H.265 encoding
2. AModel tests (NVIDIA's software implementation of the GPU)
3. Encode Feature tests (H.264/HEVC encoding on DX9/DX10/DX11/CUDA platforms)
4. Encode Performance tests (Calculating FPS encoding speed)

The project will help NVIDIA detect bugs and performance issues in their code.

Streaming of Opus Encoded Audio Data

Since its initial release in 2012, opus audio codec has proved itself to be an extremely good mode of being a carrier for audio data in all types over the Internet as compared to the already prevailing audio codecs. It has provided developers a tool which they can use without having to worry about format compatibilities across platforms and license problems. Being a open source project it has enabled developers to create softwares that provide a very high quality of user experience with unmatched audio qualities. It also has proved to be very useful in creating applications that work in a narrow b/w environment, where previously such communications would not have been allowed by the Internet connection. In spite of having so many perks, this codec is still new and even after being declared as a mandatory to have support for codec for WebRTC applications, very few applications actually support this codec or have incorporated support for it are less. This project is aimed at understanding and exploiting the perks of Opus codec and to create an application that will support the streaming of an audio stream which is encoded in opus format to a server that can distribute this stream to any client that wishes to listen to it.

CHRONIC DISEASE MANAGEMENT APP

India online health is apotal providing connected health to patients our projects aim is to improve health of people , especially those with chronic andlifestyle disorders. Wearables are an exciting recent technological development. So far they have been used more for as are creational tool,as a fitness tracker. But they can also be used for more serious purposes of saving life and also ensuring a better quality of life. This project aims to do so.We are developing a application considering following aims. We are going to develop an application that will be very useful to doctors as well as patients. We are developing this application in Sencha. Sencha touch makes the application platform independant i. e. you can run application on any phone without changing the application code. There are two sides of application. One is doctor's side, another is patient's side. Based on the profile of the user, doctor will generate instructions for patients. For eg; a fit user will have fitness program design, where as user having some diesease like diabetes will have a different instructions generated. The business rules for this will be obtained from renowned specialist. Once the instructions are generated, the data is stored in database. On user side user will be able to see the instructions by doctor.He can enter tasks completed by him. Based on the activities and health

parameter user will gain points. Doctor can see the progress of user. It will be very easy for doctor to keep record and access patients information. Doctor can do survey about health of people very easily.

WIDS Using Reputation Based Approach

Growth in IT sector has led to increase in speed of Internet. With the increase in internet users there is a tremendous increase in network traffic. Increase in network traffic has also increased the threat of intrusion over the network. For the ever growing internet speed there is a necessity to modify the intrusion detection system in order to detect the intrusion threats all over the network. As a result, to meet these needs of Intrusion detection system we can use the combination of packet based and flow based systems. Reduction of false alarm is also a matter of concern in intrusion detection. Wireless Technologies within the most simple sense empower a number of systems in order to communicate jointly, without utilizing any physical medium including cables. That is the reason wireless circle technology is the quickest developing portion of communication organization sector. But unsecured wifi circle importantly 'Open top door' of respective network in order to attacker. Possibly he could open yours current shared gadgets n facts, access the current mail, use internet and take data regarding future process and has taken providing those should crack rest of your respective systems. Thus intrusion means an steps done purposefully or inadvertently, which threaten the safety, Integrity, Availability and also Confidentiality from the network program

PEER TO PEER FILE SHARING

Twas brillig and the slithy toves, Did gyre and gimball in the wabe. All mimsy were the borogoves, And the mome-raths outgrabe. While conceding that India were the favourites to lift the World Cup on Saturday, Sri Lanka captain Kumar Sangakkara refused to term his own band of men as the underdogs in the mega event's summit clash at the Wankhede Stadium. "They (India) are a very good side and they have always been the favourites to win this tournament. They've got to the finals and everyone will be looking for them to keep going," he told a media conference on the eve of the match.

A Secure Intrusion detection system against DDOS attack in Wireless Mobile Ad-Hoc Network

MANET will be a bunch of mobile nodes create having each and every a radio transmitter along with a radio of which connect mutually by using bidirectional wireless backlinks generally right away as well as ultimately. Drastically help liberty standard liberty as well as scalability, wireless systems usually are usually desired as the first day time on the development. On account of the elevated technologies as well as reduced costs, wireless systems have accumulated much more possibilities above experiencing ignited systems formerly number of age range. Mainly because within the rising wish with regards to wireless specifics as well as facts answers, providing faster as well as trustworthy mobile gain access to is growing a vital matter. Presently, not merely cellular phones, as well as laptop computers as well as PDAs are widely used by means of men and women of their professional as well as exclusive life. The item are widely-used on their own for the most part that may be their own application ordinarily do not come together. Often, all the same, several cell phones kind a quickly fixed, non permanent elliptical whenever they continue better. This excellent allows us to talk about specifics by means of files, offering sales pitches whether my business is going around as well as inside of a obtaining.

BOOKS++

While reading a book, the best friend we have is our imagination. We create characters, places, movements, feels in our brains to fill in the gaps left by the words. But the words provide us enough raw materials to construct the rest. Yet, we do like books which are nicely illustrated. They provide us with some additional material to build our worlds. Some books like Shakespeare's plays, books by Charles Dickens and ancient classics are annotated with information about places, characters, pictures and so on to make the reading more interesting. What if we could take web pages and augment it with interesting information? About places, people, maps, etc? What if we could augment the reading experience with pertinent emotional content such as images, paintings, poetry and music? What if the content were personalized to the tastes of the reader? What if the content was diverse enough to keep the reader engaged? One could always do this manually. But then where is the fun in that? Our page enhancer system would help to overcome these problems. Using Books++, we create a tool for automatically augmenting the web page content. It would analyze the text of the webpage, and get relevant and varied content from the Internet to augment it. It would also select the content according to its relevance to the text. So this eventually enriches user's

reading experience for the travel destination with interesting information and providing emotional content of the place. Thus this tool successfully keeps the user engaged by providing quality content selected from across the Internet.

MEDICAL DIAGNOSIS USING CLASSIFICATION ALGORITHM

This is a graphical user interface web based application which is used to diagnose health condition of user. It is a prediction model which uses machine learning classification algorithm support vector machine to process the provided observation in the datasets. Classification is done on all possible conditions provided in statistical data in dataset. With every new observation, SVM training takes place and results or accuracy of results change accordingly. The user can himself use the application to find his medical health condition by providing his symptoms to the prediction model. The authorized person like doctor can add new data entry to the dataset after each diagnosis. Looking at the results the application provides best doctors and hospitals for current health condition.

AUTOMATIC QUESTION GENERATION FROM STRUCTURED DATA

Given the emergence of competitive examinations such as GMAT and CAT, there is a need for questions to be generated for fields such as Data Interpretation, which entails analyzing given input data tables to answer related questions. Today, such a task is performed manually utilizing extra time and resources. To simplify this task, we propose an Automatic Question Generation system to generate questions from structured data. This project explains an approach which takes data table input and generates questions of varying types. The data is preprocessed and categorized according to their class (numeric and nonnumeric). Tuple formation and Template selection helps provide variety in and increase quantity of questions generated. We have built a custom Tagger, which takes the data entry and assigns it an entity, such that it can be generalized to a particular entity name. For uncategorized or unrecognized entities, options to add or modify entities is also provided. Peers and colleagues were consulted for system efficiency analysis and statistical interpretation. Preliminary experimental work carried out using this template system shows promising results.

AGRI-QAS: Question Answering System for Agriculture Domain

AGRI-QAS is a real time Question answering system for Agriculture domain. We focus on the need for a robust domain specific question answering system targeting agriculture domain. The goal of

the project is to help farmers get information and resolve their queries related to agriculture and thereby improving agriculture literacy. The system is based on the principles of natural language processing and information retrieval. Most of the currently available information retrieval tools do not support runtime answer retrieval and return ranked list of documents instead of precise answers. Thus we focus on developing a system which processes unstructured data and returns actual answer for FACTOID questions such as which, what, who, where. For example, which diseases affect the wheat crop?, what are the prevalent diseases in North-America region? Etc.

CONTENT MANAGER

The project aims at building an application for selecting inbox messages based on certain criterions which are nothing but entities extracted using Gate tool. In daily life and in organizations etc people text messages to their friends, colleagues, partners, etc. to discuss about their daily conversation. But these messages are not taken care of by the people as they are informal. our application extracts entities from messages at a particular location, time and date. these entities along with location discovers interesting association and correlation relationship with the help of association rule mining. This module help to fetch and analyse inbox messages with the help of gate. Additionally an Inference Model has been developed based on entities extracted by using Gate tool and relations extracted using association rule mining which will help in various decision making processes such as behaviour of person at a particular location related to food, movies and shopping domain. Given a text message, what the two parties are communicating at a particular time will help to decide their behaviour based on different locations. This is the main motive for building this Inference Module. This module mainly extract entities and relations related to various domains like the food, movies, shopping etc. and these are considered as the base factor while getting the inference from a message.

Commodity trading assistance using twitter sentimental analysis

Twitter is a extensively used micro blogging site where account holders to display their opinion write messages which are called as tweets. These tweets convey opinions about different topics ranging in variety of fields. We intend to develop an automatic sentiment (positive or neutral or negative) extractor from a tweet. This is very useful because it allows feedback to be aggregated without manual intervention. Using this analyzer 1) Common users (typically consumers) can use this sentiment(opinion/review) analysis to research products or services before buying them. E.g.

Mobile phones, laptops and such electronic equipment. 2) Marketers can use this to research public opinion OAS for classifying the sentiment of Twitter messages using distant supervision. 3) Organizations can also use this analyzer to gather important feedback about critical problems in newly released products. E.g. Brand Management (Nike, Adidas). We have written a python script in such a way that it can be used for downloading the required amount of tweets of any user existing on twitter provided the user name is correctly known. We specifically aim at downloading the tweets related to finance such as tweets related to launch of a new electronic product like Samsung SIII, Iphone6 etc and also of various commercial institutes such as financial banks ICICI,Barclay, HDFC. This type of data constitute of training set and this training set is fed to machine learning algorithm (Naive Bias) which gives a detailed report about the sentiment of the training data fed. Before feeding the training data to the machine learning algorithm the raw tweets need to be processed by removing all the noise such as exclamations and special characters then stop words need to be filtered 2out so that machine learning algorithm can be set ready to extract all the informative features from the training data set.

TEXTUAL DATA INFERENCE MODEL

Our project is in the domain of Information Extraction. In this project we intend to infer the meaning of the text. This is a quite complex process which involves several phases as we are processing a natural language that is often ambiguous. Getting rid of the ambiguity is one of the main hurdle. So there are various steps to achieve this. In our project first we try and find out different entities, this is done using Gate tool (A open source tool). Then we resolve the co references. Then we try to obtain the relation between the recognized entities and then finally we go for semantic resolution which gives us the final result.