**Extended Eye**

### BITS ZG628T: Dissertation

by

D BALA VENKATA MANIKANTA

2015HT13078

# Dissertation work carried out at

## Societe Generale Global Solution Centre Private Limited & Bangalore

****

**BIRLA INSTITUTE OF TECHNOLOGY & SCIENCE**

**PILANI (RAJASTHAN)**

November 2017

**Extended Eye**

### BITS ZG628T: Dissertation

by

D BALA VENKATA MANIKANTA

2015HT13078

# Dissertation work carried out at

## Societe Generale Global Solution C

## entre Private Limited & Bangalore

Submitted in partial fulfillment of M.Tech. Software Systems degree programme

Under the Supervision of

Bharat Babaso Mane & Technical Manager

Societe Generale Global Solution Centre Private Limited & Bangalore

****

**BIRLA INSTITUTE OF TECHNOLOGY & SCIENCE**

**PILANI (RAJASTHAN)**

November, 2017

#### CERTIFICATE

This is to certify that the Dissertation entitled Extended Eye and submitted by D Bala Venkata Manikanta having ID-No. 2015HT13078 for the partial fulfillment of the requirements of M.Tech. Software Systems degree of BITS, embodies the bonafide work done by him/her under my supervision.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Signature of the Supervisor

Place : \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Date : \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Name, Designation & Organization &Location

**Birla Institute of Technology & Science, Pilani**

**Work-Integrated Learning Programmes Division**

**First Semester 2017-2018**

**BITS ZG628T: Dissertation**

**ABSTRACT**

**BITS ID No. : 2015HT13078**

**NAME OF THE STUDENT : D BALA VENKATA MANIKANTA**

**EMAIL ADDRESS : manikanta.donthu@gmail.com**

**STUDENT’S EMPLOYING : Societe Generale Global Solution Centre**

**ORGANIZATION & LOCATION Private Limited & Bangalore**

**SUPERVISOR’S NAME : Bharat Babaso Mane**

**SUPERVISOR’S EMPLOYING : Societe Generale Global Solution Centre**

**ORGANIZATION & LOCATION Private Limited & Bangalore**

**SUPERVISOR’S EMAIL ADDRESS: bharat.mane@gmail.com**

**DISSERTATION TITLE : Extended Eye**

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**ABSTRACT** :

Extended Eye mainly to track the shipping dealers and companies to detect

fruad and risk management. Activities which are going to illegal like some delears

are black listed and some sanctions on countries etc.

To detect the shipping companies fraud we have to search different websites.

Extended Eye will simplify the process of detecting the fraud and risk management.

Activities are legal as per the UN / US / operating countries regulations. Today the

customer has to visit many websites to collect information and embargos and fraudulent deals. Involving various shipping agencies, dealers and few blocked countries to calculate the risk.

The document which given by customer will consists the dealer names, shipping

corporation names, type of ships, ship id, vessel id, shipping route etc. Using Natural Language Processing Techniques like Parts of Speech will search this keywords and Lucene Search engine Technology Apache solr will use to Search these keywords in web sites and come up with small user interface page where we can display fraud related data.

Broad Academic Area of Work: **Artificial Intelligence**

**Key words** **: Java, PDF Reader, NLP technique Parts of Speech, Apache Solr, HTML**

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Signature of the Supervisor**

**Signature of the Student**

**Name: \_\_\_\_\_\_\_\_\_\_\_\_\_\_ Name: \_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Date: Date:**

**Place: Place:**

**Acknowledgements**

I would like to express my sincere gratitude towards my mentor **Bharat Babaso Mane** for examining my progress in splite of his busy schedule.

I would like to thank **Posham Sathyanarayana** for his guidance, efforts and patience. His valuable comments have been immensely helpful in enhancing the quality of work.

I would like to thank my **Colleagues** for their support and co-operation during the course of my project.

I am also thankful to **Societe Generale Global Solutions Center** for giving me this opportunity and also making available all the resources required for this dissertation.

I am also thankful to **BITS**, Pilani and the Training department for all their efforts in organizing this course.

Finally, I take this opportunity to extend my earnest gratitude and respect to Teaching &Non-teaching staffs of the department, the library staff and all my friends, who have directly or indirectly supported me during the period of my project work.

**Table of Contents**

**Intruduction**

**Introduction to vessel trackers**

**Problem**

**Purpose of the project**

**Requirement Analysys**

**Requirement Analysys**

**Funcational Analysys**

**Non Functional Analysis**

**Project Model with Design**

**proces overwin**

**Workflow diagram**

**Low Level Desgign**

**Project Structure**

**Data desing**

**System requirements**

**Hardware requirements**

**Software Requiremnt**

**Implementation**

**Platform used**

**Language selectin**

**Java**

**Apache solr**

**NLP parts of speech**

**html**

**Testing**

**System testing**

**Unit testing**

**Accept testing**

1. **RESULTS AND ANALYSYS**

**Input Data**

**Result After convert PDF to text**

**Result for NLP Technique Parts of Speech**

**Final Result in UI**

**List of Figures**

**List of Tables**

**CHAPTER 1**

**INTRODUCTION**

* 1. **INTRODUCTION TO VESSEL TRACKERS**

Now a day’s in world shipping goods or cargo is used widely by sea because of it is very cheap compare to air transport. This transport is inside country called domestic transport or one country to other country called international transport. So once these vessels are travelling with military or commerce or passengers and soon on purpose. Once vessel is travelling they may take illegal goods, because of that dealers and shipping companies are black listed.

In World Wide Web, there are so many vessel tracker and vessel finder web sites. In that many dealers and shipping companies are involved internationally. There is different type of ship tracking like Live Ship Tracking, Ship Vessel Tracker and Tracking containers. For each cargo ship vessel, there is Name with Country name, Image and International Maritime Organization number and Maritime Mobile Service Identity will be existing.

Present vessel tracking services increase in online and All are using automatic systems. Information overload and confusion with information has becoming obstructions. For cargo shipping tracking where dealers and shipping companies are involved internationally, to know their activities are legal as per the UN / US / operating countries regulations. The important aspects in this to track the vessel and analysis of activities of vessel.

Today the customer must visit many websites to collect information and embargos and fraudulent deals. Involving various shipping agencies, dealers and few blocked countries to calculate the risk.

**1.2 PROBLEM WITH VESSEL TRACKERS**

There are so many companies and dealers are present. Over million users from multiple sectors will use the automatic identification system and they will track the ship location and destination but it is difficult to track the ship is going in correct track or container is taking any illegal goods or not. Sometimes the ship is black listed also and some there is no registration.

Present online services or website are failed to get the illegal activities because of that customer must visit many websites to collect information and embargos and fraudulent deals. Involving various shipping agencies, dealers and few blocked countries to calculate the risk. This entire data is dynamic and static document / list will not help and in the volatile environment.

* 1. **PURPOSE OF THE PROJECT**

Customers must be careful about the sanctions and black listed companies to avoid the fraudulent deals. Extended Eye Application will process the physical document to get the deal information, which contains the dealer info, shipping info, vessel details and route of the shipping and countries to be halted and extract all the nouns to associated in the document provided.

Use Natural Language Processing (NLP) techniques will extract the nouns and search in various regulator websites, informational sites, to better know the deals and their activities to know the risk accordingly and take the corrective steps to minimize the risk of fraudulent deals, reputation loss and monetary loss.

After searching the available information, parse the HTML content and made available for decision making and show the risk. Over the period, application will gather all the data and will be able to suggest the risk based on its past experience.

**CHAPTER 2**

**REQUIREMENT ANALYSIS**

* 1. **REQUIREMENT ANALYSIS**

Requirement analysis is the very import part of the project. Requirement analysis mainly involved in determine the user expectations. Extended Eye mainly concentrate on the risk and fraud management and illegal activities of dealers and shipping agencies. After process, the document and searching the data in websites finally in user interface will show the final result.

The tasks involve getting to know the functional requirements which means what the system is required to do. Thus, a good understanding of the application requirements is needed in order to determine the specific features that should be implemented. Once the requirements are known then we can provide solution for so far problems that have occurred in existing system. This contains all the duties that go into investigation, scoping and explanation of new or changed system.

The main intension of the application is to detect fraud and risk in shipping companies and vessel dealers. At present, there are so many companies and vessels are black listed because of their sanctions and activities.

**Table 1**: **Present black listed companies and Vessels.**

|  |  |
| --- | --- |
| **Blacklisted Entity** | **Blacklisted By** |
| BEIJING LION HEART INTERNATIONAL TRADING | Japan |
| CHINA NATIONAL COMMERCIAL NEW TONE | United States |
| COSAILING BUSINESS TRADING | Japan |
| CHINA SHIPBUILDING TRADING | United States |
| DANDONG RICH EARTH TRADING CO | United States |
| EASTNINE INTERNATIONAL TRADING | United States |
| GUANGYUFA INDUSTRIAL CO., LTD | Egypt |
| IMAIZUMI TEA MANUFACTURING & TRADING | United States |
| KINGLEAD INTERNATIONAL TRADING | Japan |
| LEADER (HONG KONG) INTERNATIONAL TRAD | United Kingdom |
| MINGZHENG INTERNATIONAL TRADING | United States, Japan |
| NINGBO JIAHE TRADING CO | United States, Japan |
| POLESTAR TRADING COMPANY, LTD. | United States |
| SHANGHAI NORTH TRANSWAY INTERNATIONAL | United States |
| SOUTH INDUSTRIES SCIENCE AND TECHNOLOGY | United States |

**3.2 FUNCTIONAL REQUIREMENTS**

A functional requirement will explain the behavior of the software and in built third parties will give some specific functionalities for the inputs. Based on the issue functional requirement will change.

Based on previous results of those trackers will provide some recommendations. According into the Extended Eye Functional requirement will split below tasks. Most of the functional requirement will determine using Use cases or BDD scenarios.

Getting the customer data, which consists dealer names, shipping corporation names, type of ships and ship id, vessel id and shipping route etc.

Convert the PDF or Excel or CSV any format of customer data and Search the nouns and verbs in that document.

Search those data in existing government or any third-party websites and come up some user interface because of that user not able to check all website to conform those dealers are black listed or some sanctions are involved or not.

Functional design will discuss through the use case diagrams.

Use case will capture the list of events or actions regarding the application overview. Client will act as the actor and detail business case will capture the using this use case.

Dictionary

Parts of Speech Engine

Parser Engine

Diagram 1. use case diagram

**3.3 NON-FUNCTIONAL REQUIREMENTS**

Nonfunctional requirement which will focus on operating system. There are so many nonfunctional requirements.

Usability which will be analyses the how efficient and searching is formed. Scalability we can parse the more document or we can search more web sites rather than two or three websites. Adaptability this project user friendly and we will get any other information.

**CHAPTER 3**

**PROJECT MODEL WITH DESIGN**

**3.1 Process Overview**

Project Model and Design is very key to application which tells how much heavy or light is our project. Normally we will draw the work flow diagram to tell the process of the application.

In this process, there are four steps involved. Convert the customer data into the text format and find the dealer names, shipping corporation names, type of ships, ship id, vessel id, shipping route etc. an Search these data in the existing websites and finally come with user interface. so that customer can easily track the fraud and risk involved in that particulars companies and below Workflow diagram will tell about the application s process.

**3.2** **Process flow Diagram-Overview**

Customer Data Collection

Parts of Speech Library (NLP)

Convert the Data into Plain Text

User Interface

Apache Solr(Lucene Search Engine)

Diagram 2. Process work flow Diagram

**3.3 Low Level Design**

Low Level design is the step by step process application. In this process, we will discuss about Project structure, software architecture, data design and implementation code etc.

**3.3.1 Project Structure**

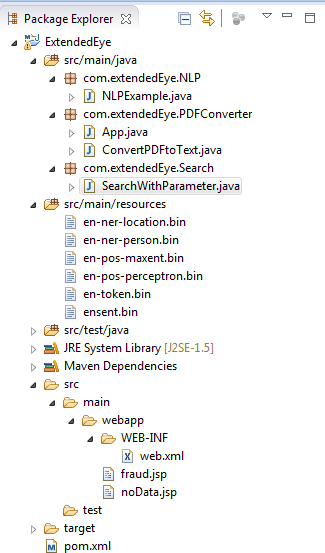
****

Diagram 3: Project Structure of the Project.

**3.3.2 Architecture** Database

Oracle

Apache Solr

Open NLP

IText PDF Parser

Tomcat Server

Data Base Server

HTML

Browser

Client

Service Layer

Client Layer

Diagram 4: Architecture of Project.

Here There are three layers.

* + - 1. Client Layer
      2. Server Layer
      3. Database Layer

Client Layer: Web Browsers in which web application will be rendered and be used to interact with the server-side web application.

Service Layer: All the business logic is going to be implemented in the service layer.

Database Server: Database server node will have the database instance the Application is going to use.

Finally Tomcat where we will deploy for local development.

**3.3.1 Design phase**

Low Level Design will tell actual logic of each every component. we will use the class diagrams to define the application.

Class diagram will tell the structure of classes, attributes and operations of relations of the objects.pasted-image.tiff

Diagram 5: Class diagram of Application.

Here In this class diagram parts of speech will extract Map parts of speech to object model components nouns usually map to classes, objects, or attributes and verbs usually map to operations or associations.

Activity Diagram which will tell the dynamic action of the application. It will tell in each activity Business functionality.

Import the customer Data

Parse the customer Data into Plain Text

NO

Is Parse Successful

Y

Yes

Search the nouns and alphanumeric words

NO

Found the Nouns

sful

*Yes*

Yes

Browse in the Third parties

NO

Found the fraud data

sful

No

Yes

Prepare HTML Page

Diagram 6: Activity Diagram of Application

**CHAPTER 4**

**SYSTEM REQUIREMENT & IMPLEMENTATION**

Software requirement condition is a key file, which outlines the groundwork of the item headway practice. It records the necessities of a structure and additionally has a depiction of its genuine segment. System requirement will have software and hardware requirements.

Another key principle for development is technology used and system requirements, for our business requirement which technology is suitable and what is the performance time and memory consumed it very important for the application.

**4.1 HARDWARE REQUIREMENTS**

The hardware we are utilizing for this project is as follows

* + Processor: Intel core I5
  + RAM : 8GB
  + Hard Disk: 500GB

**4.2 SOFTWARE REQUIREMENTS**

The software we are using in this project is as follows

* + Operating System : Windows 7
  + Technologies used :
    1. Programming language: Java
    2. Client side: HTML
    3. Server Side: Apache solr (A Lucien Search Engine), Natural Language Process (Parts of Speech), Apache PDF Reader.
  + JDK : Version 1.7
  + Eclipse : Luna

**4.3 IMPLEMENTATION**

Project development heart is implementation. In this world, there is no one will be without heart the same way without implementations there is no project.

Implementation means writing the program is to work the software with requirement given by the customers. In order to write the program, we have to select the platform and technology. The platform and technology we use to understand the each and every user. So, let us know the appropriate platform and technologies used in our project.

**4.4 PLATFORM USED**

Platform indicates the operating system which we are used. There are so many operating systems in market. But so many people are using Windows because it is user friendly and windows XP, windows 7 and windows 8 and windows 10. But in our project, we will window 7.

**4.5 LAGUAGE SELECTION**

Any technology we have to write in some language. At present, there so many languages in market but we can technology is any language. But in our project, we will use “JAVA” language and there are certain features in java and it will use so many people.

**4.5.1 JAVA**

Java is a object oriented programming language and each and every element will class and those classed we will access through objects. Object is a instance of class and it has state and behavior. In Simple Way Class is logical component and object is physical component.

Please find the below diagram for some of the feature of java.

Distributed

Secured

Simple

Portable

Interpreted

Object Oriented

Interpreted

Dynamic

High Performance

High Performance

High Performance

High Performance

Platform Independent

Platform Independent

Robust

Object Architectural Neutral

Platform Independent

nted

Architectural Neutral Performance

Multi threading

ural Neutral

Diagram 7: Features of java

**4.6 TECHNOLOGY USED**

To Develop the application here are Technology components.

**4.6.1 ITEXT**

iText is open source library to create or manipulate the PDF documents in java. In this iText we will use the SimpleTextExtractionStrategy

This renderer keeps track of the current Y position of each string. If it detects that the y position has changed, it inserts a line break into the output. If the PDF renders text in a non-top-to-bottom fashion, this will result in the text not being a true representation of how it appears in the PDF.

This renderer also uses a simple strategy based on the font metrics to determine if a blank space should be inserted into the output.

**4.6.2 NATURAL LANGUAGE PROCESSING**

Natural Language Process is small component of the Artificial intelligence and it has the ability to understood the human speech and it is based on machine learning.

if you have lot of data that is Structured or unstructured and we want some data extract from that then These Natural Language Processing Techniques will be helpful.

There are several tools involved in Natural Language Processing.

1. Parts of Speech

2. Name Entity Recognizer

3. Open Information Extraction

4. Coreference resolution system.

**4.6.2.1 Parts of Speech**

In Our Project, we will mainly Parts of Speech library. Parts of Speech is a piece of software that reads the text from language and split each word into the Noun, verbe, adjective, etc.

In Parts of speech classification techniques.

1. Supervised Parts of speech

2. Unsupervised Parts of Speech

Supervised Parts of speech: It is a Structured collection of text it will learn based on already existing data.

Unsupervised Parts of speech: It does not require any exist learning experience based on assumption it will give output.

Both the speech will use Rule based Decision making model. It is tree structure and recursively partition the data.

Please find the native mapping of the Parts of speech.

Diagram 8: Parts of Speech

**4.6.3 APACHE SOLR**

**A**pache Solr will use the Apache Lucene Search engine. A Lucene is high performance, Full featured text search engine written in java.

There are four main operation we will use in this Project. Indexing which will convert document into machine under stable language and query which will form the query to retrieve the data and mapping will map the query to the document and Ranking is last will give the Rank for the searched documents.

Please find the application Solr search architecture.

Index

Query Parser

Request Handler

Send the Request Send Response

Lucene Sever

Response Writer

Rest API

Diagram 9: Apache solr

**4.6.4 HTML**

As Everyone know HTML standards for HyperTextMarkupLanguage and it will use for display data and published by world wide web. Latest version of the HTML is HTML5 which will support all cross-platform sites and model web browsers.

HTML5 has added new tags and attributes those are not exist in the previous version of HTML and it will support some API where it will include in specifications.

**4.7 DEVELOPMENT**

There are four main steps involved in Development process.

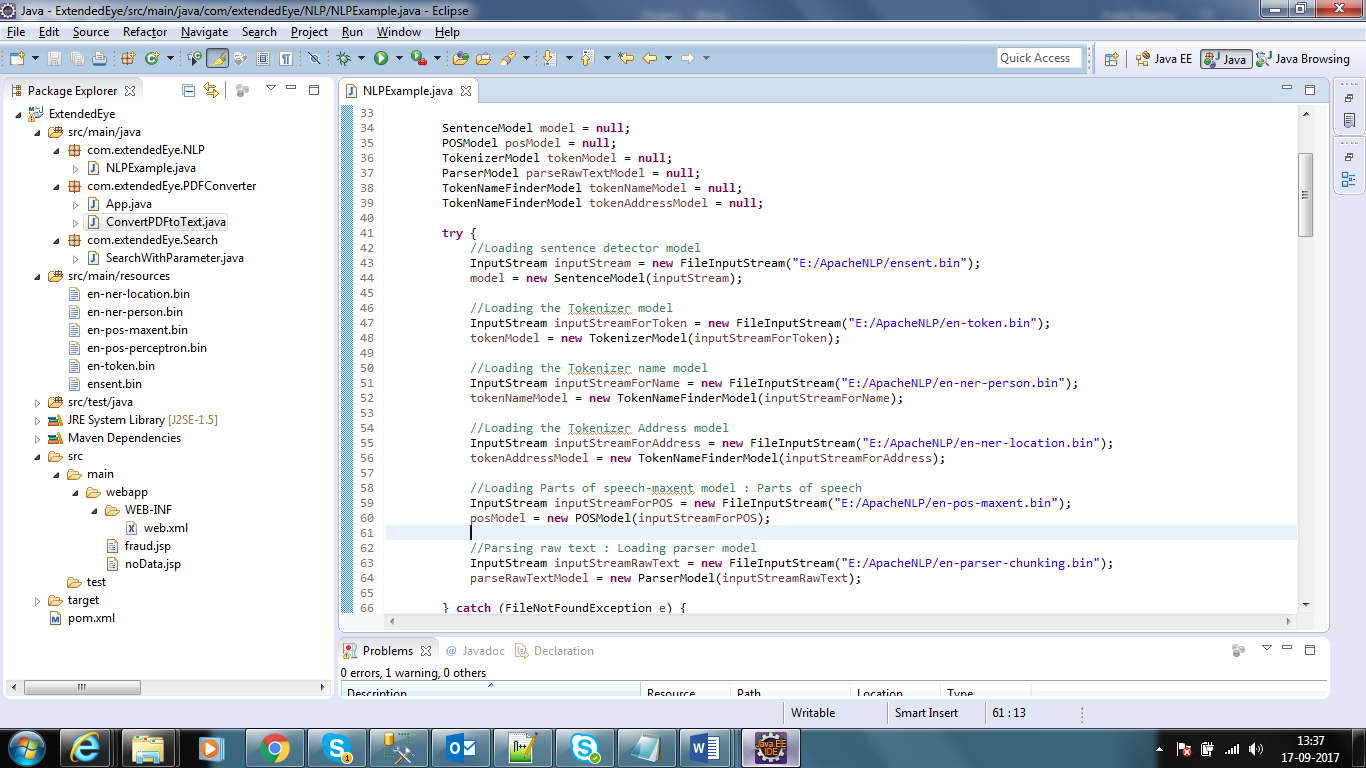
Step1:

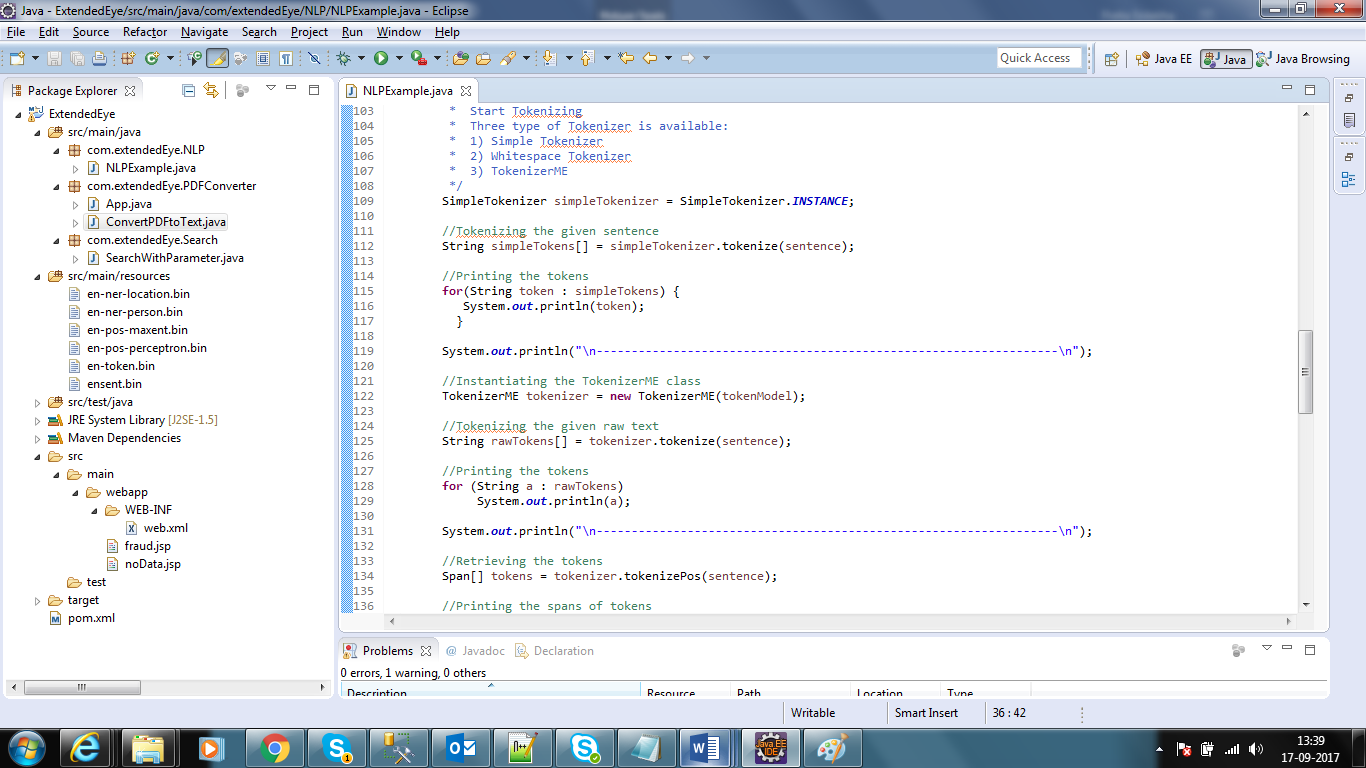
Get the Customer data in PDF format which will contains the key details of the dealers and shipping agencies. Using iText PDF Converter we will convert the PDF to plain Text.

**Code: **

Step2:

This main key implementation and it will extract the all nouns, verbs from the text irrespective of structure or unstructured data using Parts of Speech the library(NLP).





Some of the common parts of speech tags which we have used in this project.

Table 2: Parts of speech tag which we are using in project.

|  |  |
| --- | --- |
| **Tag** | **Description** |
| CC | Coordinate Conjunction |
| JJ | Adjective |
| MD | Model |
| NN | Noun, singular |
| NNS | Noun plural |
| NNPS | Pronoun |
| RB | adverb |
| VB | Verb base form |

Step3:

After getting the all nouns and verbs, which will contain ship or dealer name and some other information using apache solr we will search the in websites.

Continue….

Step4:

Finally, we will come with fraud details if those are existing in websites

Continue….

**CHAPTER 5**

**TESTING**

Testing will tell about quality of the software. In Testing, there are several types. The commonly used Testing are Unit Testing and Acceptance Testing.

Testing will play very significant role in application. It will discover maximum number of bugs or faults on developed modules. Testing is dependent on the object of the requirement.

All the object model must be tested completely with respective to functional and interaction of modules. System Testing will include the memory testing and application performance testing. How much memory is consumed and with how many seconds page is loaded, is client is satisfied with those measurement or not. Those type testing is done is in this testing.

**5.1 ACCEPTANCE TESTING**

In Testing for each and every feature there will be a acceptance criteria given by the business and this is the main for the client so that business can decide this feature will be delivered or not.

In our Application, there are four main features. Please find the below Acceptance criteria for each feature.

**5.2 UNIT TESTING**

In Unit Testing we will test each and every module separately. More faults will detect on this phase it may be functional or technical issues and it may require some sample data to test also. These phase is cycle process and it will do until all faults are rectified from application. In this Unit Testing, there are Black box testing which will run the whole system and Functional testing it will test each feature behavior.

**Test Case 1: Unit Testing of Module 1:**

**Table 7.1 Convert PDF to Text Unit Test Case**

|  |  |
| --- | --- |
| **Name of the Test** | Test Case for Converting PDF to Text |
| **Test Description** | A Test to verify convert Pdf to Text and that text will contain the details dealers and shipping agencies. |
| **Sample Input** | PDF file |
| **Expected Output** | plain text |
| **Actual Result** | As Expected. |
| **Remarks** | Pass. |

**Test Case 2: Unit Testing of Module 2:**

**Table 7.2 Finding Nouns and alphanumeric words Test Case**

|  |  |
| --- | --- |
| **Name of the Test** | Test Case for Finding Nouns and alphanumeric words. |
| **Test Description** | A Test to verify or found the nouns and alphanumeric in data |
| **Sample Input** | text file |
| **Expected Output** | Text file words with nouns |
| **Actual Result** | As Expected. |
| **Remarks** | Pass. |

**CHAPTER 7**

**RESULTS AND ANALYSYS**

The subsequent snapshots characterize the outcomes or yields that we will attain after orderly implementation of the considerable number of modules of the framework.

**SUMMARY**

**BIBLIOGRAPHY / REFERENCES**