SMART RESUME
BI TOOL TO SHORTLIST CVs FOR A JOB VACANCY

Madhushika K.A

(IT15097798)

Bachelor of Science Special (Honours) Degree in SE

Department of Software Engineering

Sri Lanka Institute of Information Technology
Sri Lanka

September 2018

SMART RESUME
BI TOOL TO SHORTLIST CVs FOR A JOB VACANCY

Madhushika K.A

(IT15097798)

Final Report

The dissertation was submitted in partial fulfillment of the requirements for the B.Sc. Special (Honors) Degree in Information Technology.

Department of Software Engineering

Sri Lanka Institute Of Information Technology
Sri Lanka

September 2018

DECLARATION

"I declare that the project would involve material prepared by the Group members and that it would not fully or partially incorporate any material prepared by other persons for a fee or free of charge or that it would include material previously submitted by a candidate for a Degree or Diploma in any other University or Institute of Higher Learning and that, to the best of our knowledge and belief, it would not incorporate any material previously published or written by another person in relation to another project except with prior written approval from the supervisor and/or the coordinator of such project and that such unauthorized reproductions will construe offences punishable under the SLIIT Regulations.

I can aware, that if we are found guilty for the above-mentioned offences or any project related plagiarism, the SLIIT has right to suspend the project at any time and or to suspend us from the examination and or from the Institution for a minimum period of one year".

Signature of the Candidate: ……

(Madhushika K.A - IT15097798)

2nd September 2018

The above candidates have carried out research for the B.Sc. dissertation under my supervision.

Signature of the Supervisor: ……

(Mr. Lakmal Rupasinghe)

2nd September 2018

ABSTRACT

Today in Sri Lanka, most industries follow up one traditional process in hiring new employees. The normal process includes, advertising the vacancy, calling Curriculum Vitaes (CV), short listing them by referring the CVs and interviewing the shortlisted candidates. Having the right set of CVs is vital since a CV is the representation of the qualifications of an applicant. Also, when it comes to an emergency project, the employer should be able to hire the best employee set within a minimum time period. In this case help of a third party CV storage which already has a collection of related CVs, and has the ability to generate the list of most qualified applicants among them, would be helpful. The submitted CVs should be read properly and check several attributes such as skills, experiences and some personal information in order to select the best. It is much time consuming for a human to read and draw a mind image about the applicant. There is a shortage of tools that support in selecting the best qualified set of employees to an employer. Smart Resume is a business intelligence tool for the IT sector, which analyze and classify operational data with classification algorithms to present complex and competitive information to decision makers, in order to dynamically fulfill the business needs. It is built to satisfy the task of generating the list of most suitable candidates. In this paper, we present a combination of desktop and web application that facilitates the task of automating the selection of the most suitable and qualified candidates depending on the attributes given by the user like Age, Gender, Work Experience, Soft skills and Education Qualifications. An ETL (Extract, Transform and Load) tool will analyze the data and prediction models will be designed in order to generate the optimal solutions for candidates. Depending on the relationship of the attributes will dynamically visualize the most suitable candidate on the Dashboard.

ACKNOWLEDGEMENT

The work described in this research paper was carried out as our 4th year research project for the subject Comprehensive Design Analysis Project. The completed final project is the result of combining all the hard work of the group members and the encouragement, support and guidance given by many others. Therefore, it is our duty to express our gratitude to all who gave us the support to complete this major task.

We are deeply indebted to our supervisor Mr. Lakmal Rupasinghe, Lecturer of Sri Lanka Institute of Information Technology whose suggestions, constant encouragement and support in the development of this research, particularly for the many stimulating and instructive discussions. We are also extremely grateful to Mr. Jayantha Amararachchi, Senior Lecturer/ Head-SLIIT Centre for Research who gave and confirmed the permission to carry out this research and for all the encouragement and guidance given.

We also wish to thank all our colleagues and friends for all their help, support, interest and valuable advice. Finally, we would like to thank all others whose names are not listed particularly but have given their support in many ways and encouraged us to make this a success.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

ETL	Extract, Transform, Load
SRS	System Requirement Specification
BI	Business Intelligence
WBS	Work Breakdown Structure
GUI	Graphical User Interface
SQL	Structured Query Language
CV	Curriculum vitae
HR	Human Resources
MySQL	Relational Database Management System
CSV	Comma Separated Values
DB	Data Base
DW	Data Warehouse

Table 1 – List of abbreviations

1. Introduction

1.1. Background Context

Smart Resume BI Tool - Data in business is very useful only if data is analyzed properly which will aid in strategic business decision making. In the IT industry, workforce management is critical since there is a huge workflow in the industry. Hiring individuals make a big impact on the productivity of the company. To hire best-qualified employees there must be a system, as it is time-consuming to read and study all the curriculum vitae. Standardly curriculum vitae have 2-3 pages sometimes it may have more than 3 pages. So it may take a long time to go through all the details on it. And also, there is a probability of missing some CVs, as there is no proper method to store them. Therefore, we must find solutions to automatically analyze data, classify and summarize it along with discovering and characterizing trends and flagging anomalies in order to ease the decision-making process effectively in a company. Smart Resume targets at developing a set of tools, technologies and programmed products that are used to collect, integrate and make data available for better, faster decision making.

There are various attributes help to identify the best candidates for a given vacancy. An ETL (Extract, Transform, Load) tool will extract relevant attributes from the Source Files, then transform data and loading data. These are critical steps of an ETL tool when considering drawing Data for a Business Intelligence systems. In our System mainly focus on flat files format like CSV (Comma Separated Values) to extract data. These extracted data and cleaned data will be stored in the data warehouse for further analyzes such as making decisions generating reports etc. The Accuracy of the Business Intelligence tool mostly relies on an ETL Tool.

1.2. Literature Surveys

Prior to the proposal of our research project we conducted a Literature survey on the existing platforms with similar functionalities and technologies. Some of the prominent researches are reviewed here.

1. Data migration from a product to a data warehouse using ETL tool

The Objective of this research is to migrate Historical Data of Risk Management Product to a Data Warehouse. [1] And Get Previous Day Reports for Further Analysis and Creating Reports. And They Use Large No of Data Such as Risk Management data, Ledger Data and Financial Application Data.

In Methodology, they considered about migration Data from ETL tool instead of PL/SQL based transformation, Manually Transformation when Mapping Data. In Mapping Data- First, they identify the Source Tables then what are the Entities That They want to Map into the Oracle Database. They Use Agile Methodology For their Approach in Research.

2. A New Tool for ETL Process

The Objective of this research is to create a Model to active ETL tool. Use SQL Queries to Approach the Process for Mapping Data. [2]

SQL Queries use for all construction Process like data extraction, cleaning Procedures, direct Storage and for front-end information delivery. Try to reduce the complexity of the process by applying SQL queries.

3. The Research & Application of Business Intelligence System in Retail Industry

The objective of this research paper addressing the important of a BI tool in retail industry [5] including ETL tool. Because of business decision makings processes complicated traditional database system is not support for analytical and data intelligence processing. so it come up with a solution which includes ETL, data warehouse and data mining in order to implement BI tool which support in retail industry[5].

1.3. Research Gap

Even though there are existing proposed products in the market area, they don't address most of the problems that the proposed system is going to address. The following table shows a comparison of features between the existing products or applications and the proposed solution.

Features	**Oracle BI**	**Birst**	**Jobscan**	**Smart Recruit**	**Smart Resume**
BOT(Automated)					✓
ETL Tool					✓
Optimal Solution			✓		✓
Feasible Solution					✓
All the user does not have to interact with the system	✓	✓	✓	✓	✓
Visualization of data in an abstract way		✓	✓	✓	✓
Predictive Analytics	✓	✓	✓		✓

Table 2 - Comparison with existing System

1.4. Research Problem

The world is a chain of businesses. As the businesses get bigger day by day, the complexity and the competition are highly increasing. New trends are being incorporated into business ecosystems. New technologies are evolving rapidly and it has significant influence in the organizational processes. So day by day small to large all companies have to update themselves in terms of resources, manpower, and infrastructures in order to maintain a competent and business system.

It is a known and documented fact that information-driven culture is needed in a company to meet customer needs in today's world. To achieve this, they need a tool or service to discover and prioritize business challenges across their organization with these new one's assessment methods and bring information closer to them so they can make a smarter decision.

When hiring new employees, a company will have to spend much time, effort and cost on finding suitable candidates among thousands of the educated and qualified ones. As of the recruiting process using in the industry nowadays, companies have to spend a huge cost and time on selecting the perfect ones for the vacant position.

1. Advertise the vacancy
2. Call Curriculum Vitae of the interested candidates.
3. Shortlisting the applied candidate list by referring their Curriculum Vitaes.
4. Interview the shortlisted candidates and recruit the most suitable ones for the position.

But practically, it takes a lot of time and effort for a human to judge an employee's skill and talent just by reading their Curriculum Vitaes. Normally, a CV should contain 2 to 3 pages and all the relevant qualifications should be listed there. Because, according to normal policy, the time dedicated to reading one CV is 6 to 7 seconds. The reader should be able to grab the relevant information within that time period.

But, practically, there may be well qualified, talented candidates, who have a large skill set and a CV extended from 7 to 8 pages since it has to hold each and every qualification they achieved. Sometimes, the required qualifications for the specific position they applied, would be included in the last pages of the CV. In this kind of scenario, the reader would miss the important skills or points because they cannot waste much time on one CV. It is much time consuming for a human reader to read one CV end to end. And also, the most qualified candidates would not be called to the interview just because their CV is too long or not well formatted. It is a huge disadvantage not only to the candidate but also to the company, since the company may lose the best employee to their vacancy.

On the other hand, there may be hundreds of applicants for a vacancy of a large IT industry. In such a scenario, it is very hard and time-consuming to download each and every CV and read them one by one in order to shortlist in human hands and send them to relevant companies who need suitable candidates. And also they need to save those candidates details in their databases for further uses such that another company also want the same set of candidates.

There is no any similar tool to satisfy the exact issue, but there are some similar commercially available BI tools but below are the drawbacks.

1. They are far too expensive, which are being developed by big vendors and often target the big clients.
2. Small and medium scale companies cannot afford a big cost or a time to find the suitable candidates because many of the employees have to engage in multi-tasks inside the company.
3. Cost for advertising for each candidate is very high.

Considering all the above facts, Most Companies have understood the need of a business intelligence tool to achieve goals through business intelligence concepts [3]. We can come to a point that there is a real need of cost-effective Business Intelligent (BI) tools that can cater the need of recruiting the best employees to a company by sending a set of selected candidates for relevant companies who require them. Therefore, the purpose of this research is to address such mentioned issues.

1.5. Research Objectives

Any action that must remain in the way of success requires sound strategies and management planning. A business intelligence tool can be very appropriate to achieve this goal. The Proposed ETL is a research project component focused on several objectives. With the completion of the project the developed ETL tool is supposed to fulfill these research objectives. The main objectives of the ETL tool component are mentioned as follows.

1.5.1 Main Objectives

Introduce an intelligent system to select the CVs based on the characteristics/attributes. Depending on that given characteristics an optimal or feasible CVs will be filtered. The optimal CV selection is the one with the lowest probability of unfavorable outcomes an optimal solution is a theoretically proven solution. But this might not be a correct logical solution and we may have to come up with a feasible team. Hence the tool has the option of providing the most feasible (possible and practical) solution as well. The **Smart Resume** enables users to select the most optimal or feasible CVs according to the given attributes for the job vacancy of IT company or industry. The current market of the BI tools has a very complex interface, which requires professional knowledge to perform tasks. Smart Resume will not require technical knowledge or professional expertise to interact and it will be developed in a simple way with fewer controls to increase adaptability and comfort for the user.

1.5.2 Specific Objectives

- Download the CVs in the automatically
- To read the downloaded CVs
- Classify the data into relevant columns
- Save the classified data in CSV format
- Mapping CSV data into data warehouse by Extracting most suitable attributes, then cleaning the extracted data into more analytical manner by removing redundant data omitting null values and apply appropriate values. Then map the cleaned and transformed data into the data warehouse created using MySQL for further Processing and analysis.

  - ✓ Identify the Attributes with the use of Relevant data
  - ✓ Extract the Relevant Data from CSV
  - ✓ Cleaning Unorganized, Redundant Data (Data Cleansing)
  - ✓ Transform Cleaned Data into Different Format

- ✓ Creating data warehouse using MySQL
- ✓ Load (Map) Cleaned and Transformed Data into the data warehouse

- Build a solution to select the most optimal and feasible candidate list.
  - ✓ The main objective of this research part is to generate the most suitable candidate list in a faster and more accurate way considering attributes given by the user, related to the IT sector. The prediction is expected to be highly accurate and the final decision is sent to create a graphical view.

After building the Predictive Model, the predicted details will be stored in a MySQL database. Using that data will find out that that result is the optimal or feasible CV selection according to the given attributes. ROC and classification Metrics will be used for those things. Also, the predicted result will be representing as Graphs, charts, hierarchies, and tables. It will provide a better user interaction through the interactive dashboard

Centralized Error Handling:

ETL tools often include generic error handling routines which are invoked regardless of where the error originated from or how it is raised.

Performance

Provide high performance with accurate results with cleansed data

In the data Cleansing section, the errors found can be fixed based on a pre-defined set of metadata rules. Here a distinction needs to be made between completely or partly rejecting the record and enabling a manual correction of the issue or by fixing the data through for example completing the record, correcting the inaccurate data fields, adjusting the data formatting etc.

2. Methodology

This section contains detailed information about the tools and techniques used to make an ETL (Extract Transform Load). Including the way the software is implemented, how much data and functions it needs and how they have been extracted. And also includes time frames and schedules that are necessary for some purposes. In addition to this, the research areas identified to carry out this component have been explicitly explained in this section.

To carry out this research project, follow the Agile Methodology. Prior to the start of the process, a research was made of the identity of the major problems faced by IT industries that when selecting suitable candidates. After reviewing the field problems, we have made a literature review for BI (Business Intelligence) tools available and what are the drawbacks of them. Then we come up with a most suitable solution for creating Business Intelligence tool by addressing major business problems in order to select most suitable candidates for a job vacancy.

The feasibility study has been proved that, the system is technically, operatively and financially feasible, because it is built on open source technology and there is no limitations and dependencies. Then we have a focus on the demands of gathering functional and non-functional requirements

During design phase, we developed the high-level architecture design inorder to incorporate the gathered funtional and non-functional requirements of

- Desktop Application
  This includes core components of Smart Resume which are automated ETL tool. These components will be displayed using controls in the main user interface.
- Web Application
  Use for Visualization of Data.

2.1. Requirement gathering

Requirement gathering was done as a team as it was a common information for the entire system. There are so many software designed to be necessary for the business analysis. Unfortunately, these tools are very expensive because they are mainly focused on high level organization. It's true that medium and small-sized company cannot benefit greatly from these BI tools that exist in the market. Our primary objective is to develop business intelligence tool that can select most suitable candidates for IT industry vacancies even for a small company afford these business intelligence tools.

Smart Resume BI Tool mainly focused on IT industry as its initial step.

The information gathering was difficult, since we had to gather secure data (Candidates related data) from IT Industry giants who supply candidates for companies which need candidates for their vacancies on time. The accuracy of the system depends on the amount of data gathered.

We gathered data and find existing business tools, what are the drawbacks, what are the technologies used, and finally come up with selected attributes list by brainstorming with team members and Supervisor.

When it comes to Predictive models and the results of each vacancy, manipulation of data is more important. The data may contain different formats with different impurity levels. That's why we have to use most efficient methods for extraction and data transformation. We cannot miss any data as it’s affected to the prediction model and analyze. So we come up with highly reliable ETL tool for Data warehousing.

2.2. Implementation

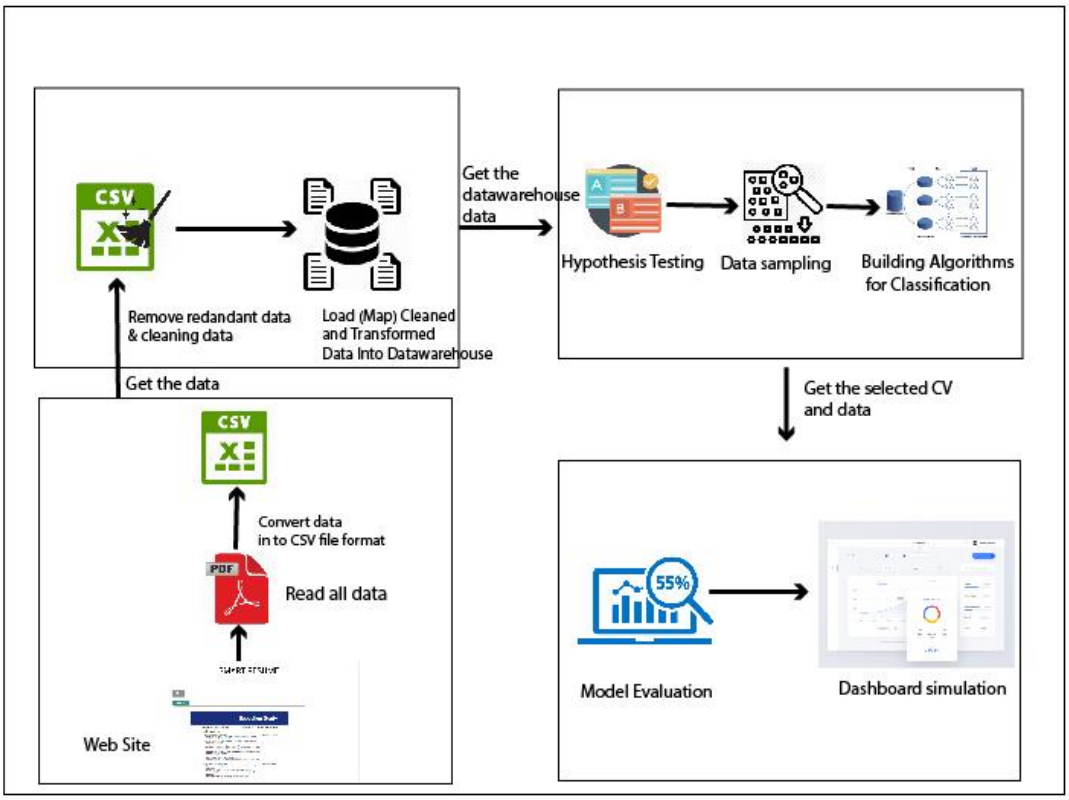


Figure 1 - High Level Architecture

According to Figure 1 there are three major components in Smart Resume

1. Automated ETL (Extract, Transform and Load) Tool
2. Prediction Model and Predictive Model Evaluation
3. Interactive Dashboard for Graphical representation

This documentation explains the process followed to develop the ETL tool

ETL (Extract, Transform and Load) Tool

Data in a business can be very useful if and only if analyzed properly [4] which leads to take strategic business decisions. It is acceptable that the data can be in various formats and locations. Thus **Smart Resume** should be able to extract data automatically in whatever form it is available.

The user can select only required attributes to analyses the data hence **Smart Resume** will only consume the selected attributes and precede for predictive analysis. Furthermore **Smart Resume** will automatically detects odd data and remove it from processing. So that it will result in high- accuracy and best quality output. We only focused on Flat files as Data Source.

Basically ETL consists of 3 parts – Extract, Transform, and Load.

Extract includes 2 sub-parts, "identification of the data sources" and "extraction" of selected data automatically. In most of the organizations, there can be different data sources related to the system, stored in different locations in different formats such as CSV, XML, SQL, and JSON etc. So the system should be able to identify all these data sources and process them. After identification, necessary attributes has to be extracted and stored in a suitable way for local processing.

After that, data needs to be transformed to a more organized format. This is done in two steps, "data cleansing" and "data transformation". Data cleansing is performed by detecting and removing and/or correcting a database's dirty data (i.e., data that is incorrect, out-of-date, redundant, incomplete, or formatted incorrectly). Data imputation techniques will be applied in order to cope up with missing data values and remove duplicate data so as to fit it to the models [6]. Then in the transformation step, data may need to be merged, aggregated, enriched, summarized, or filtered depending on the nature of the integration scenario.

Finally the cleansed and transformed data are loaded and stored in data warehouse for further processing.

Implementing ETL Tool

In the implementation stage, as first step, Focused on what are the technologies use in order to make ETL Tool

NetBeans 8.2 IDE was mainly used to develop the GUI of the ETL tool. Python together with JAVA was used to implement the ETL process. The main reason for choosing python is, it is comparatively fast in data mining. Database management of Smart Resume ETL was done using MySQL.

Following are the programming languages and technologies used in implementation stage of ETL tool.

- Java with JDK 1.8 and Swing for interface development of ETL Tool.
- Python 3.5

During the first phase, the attributes which are related to IT industry was collected through existing literature surveys and questionnaires from IT companies. This will give a clear idea about the suitable attributes that system want to focus in order to make decisions. These attributes were very helpful for next phases.

The Data Warehouse generation occurs via 3 major steps in the ETL tool as follows.

Extraction

As in the figure 2, the Extract step includes the data extraction from flat files like CSV (Comma Separated Value). The core target of the extraction is to retrieve all the relevant data from the data source with the minimal resources as possible and display the extracted data in table as figure 3.

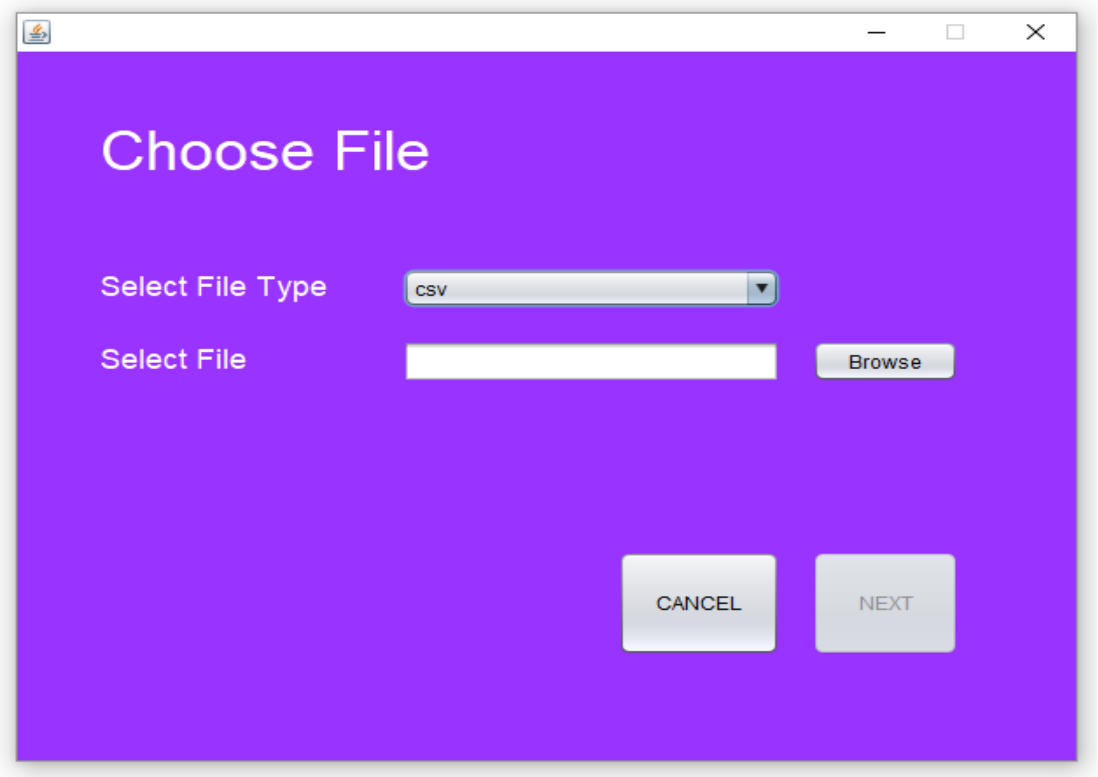


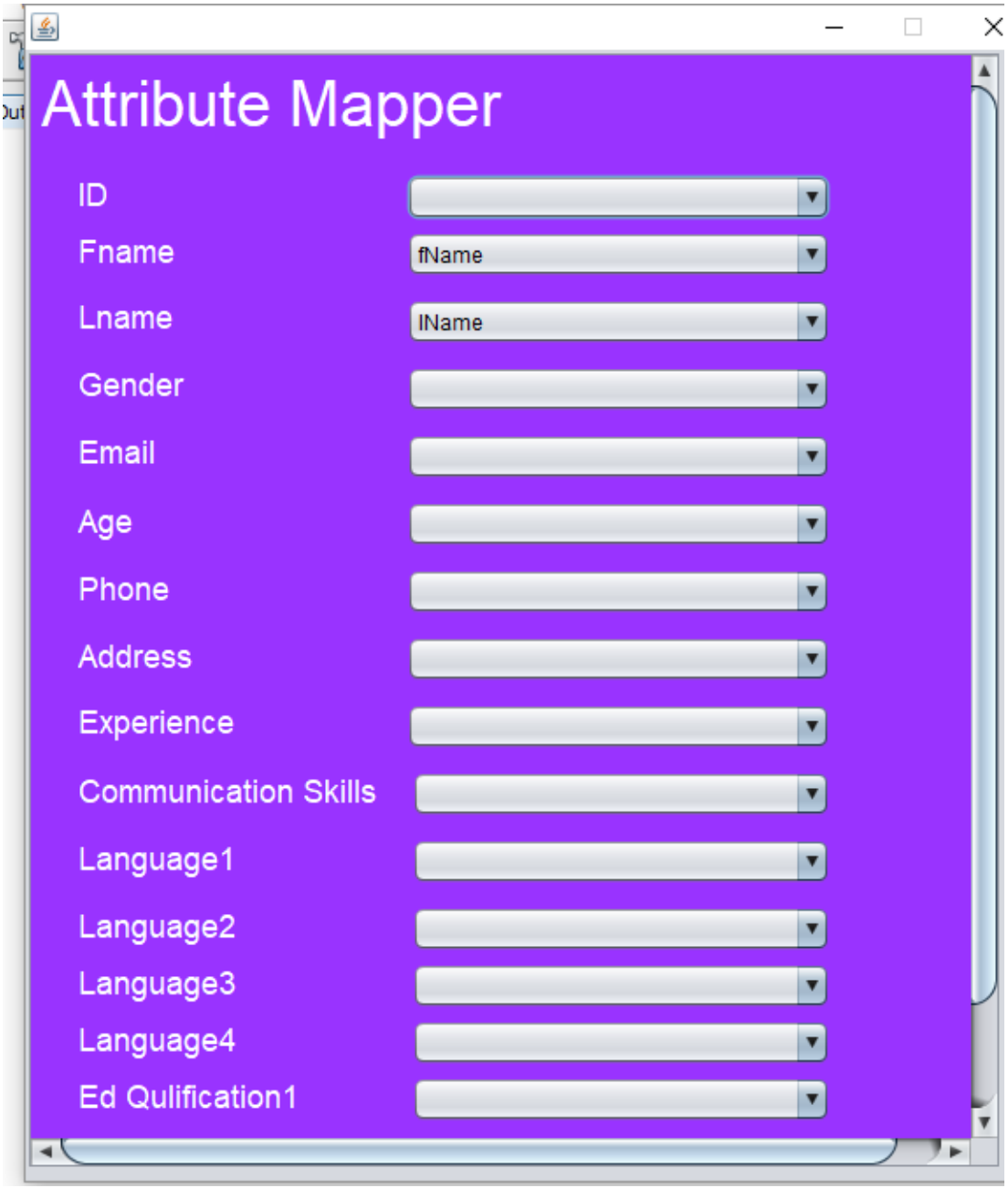
Figure 2 Extraction Interface

Table 3: Candidates Attribute List

Attributes of Train dataset	Attributes of Prediction dataset
1. Age 2. Address 3. Gender 4. Civil Status 5. Email 6. Educational Level 7. Previous Experience 8. Communication Skill 9. Languages	1. Age 2. Address 3. Gender 4. Civil Status 5. Email 6. Educational Level 7. Previous Experience 8. Communication Skill 9. Languages

The extraction has been done in separate thread so it doesn't affect the performance, response time of the source system. The size of the extracted data varies from hundreds of kilobytes up to gigabytes. This depends on the company size.

Figure 3 Attribute Mapping interface



Next map the attributes which are mentioned above with the column names as in figure 3. Attribute mapping part is 90% automated and attributes which are not mapped correctly needs to be mapped accordingly. Library JAVACSV was used to accomplish the mapping process.

Transformation

Transformation is one of the most vital as it certifies the quality of the data in the data warehouse. Transformation is performed according to the following base rules, such as:

- Remove rows which contain null values for given field.
- Validate numerical fields (remove rows which contain characters other than numbers for given field).
- Remove duplicate values.
- Fill null value columns with average value

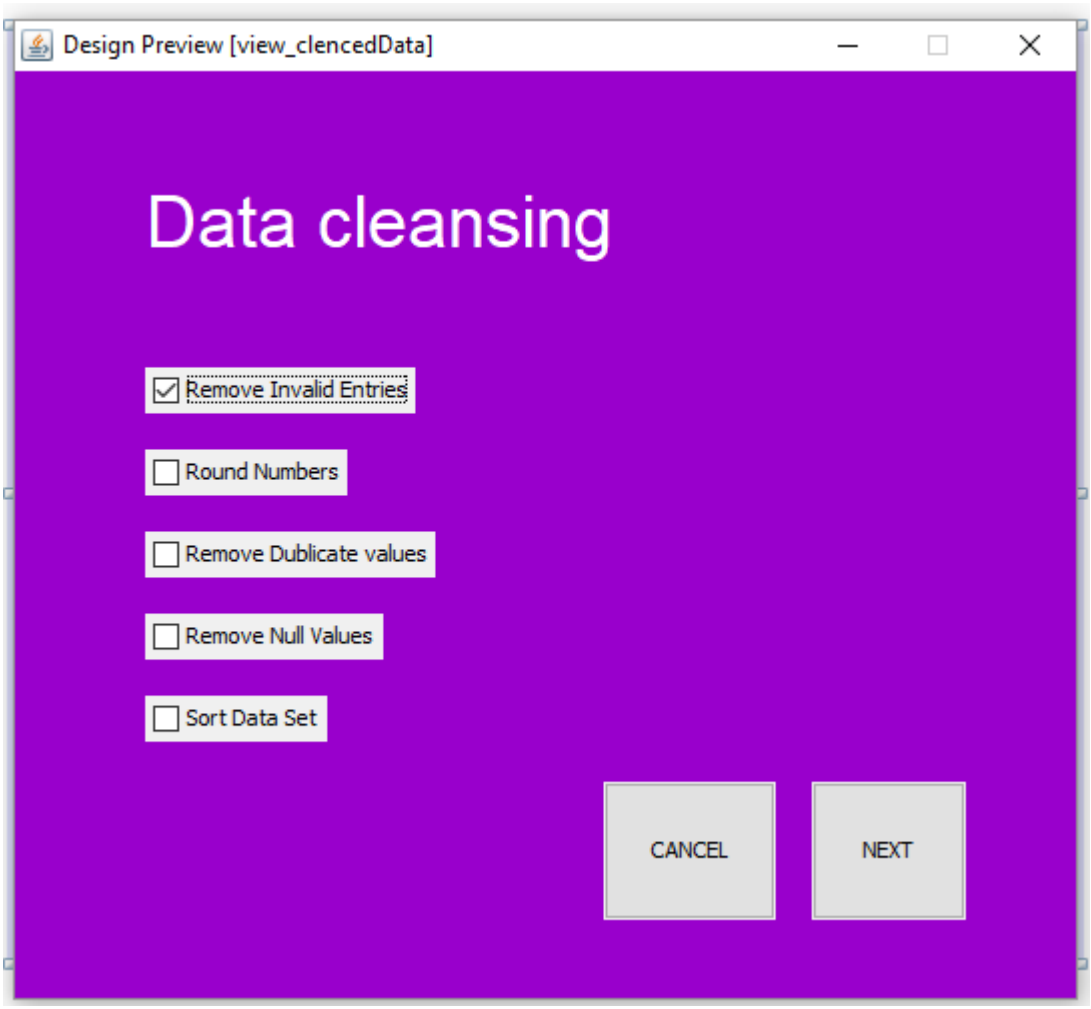


Figure 4 Transformation Interface

Load

Loading phase focuses on creating a data warehouse. Data warehouse has created using MySQL.

2.3. Tools and Techniques

Tools

- Notepad++
- MySQL Workbench
- Netbeans 8.2
- PyCharm
- Microsoft Excel
- Inscape
- Photoshop

Technologies:

- Python 3.5
- MySQL
- Java
- Jython Library

2.4. Testing

In the testing phase of the ETL tool I have followed below steps in order to maintain the product quality.

- Unit Testing – Each interface of the ETL tool was tested by me produced a defects free unit of coding. Following are the test cases I have followed in my Unit Testing phase.

Table 4: Test Case 1

Test case ID	TC1
Test case Description	Validate empty fields of Login Interface
Pre-Condition	Login Interface of the Smart Resume is loaded
Test Procedure	Press Login button without typing anything in Username and Password fields.
Test Input	Username : <Blank> Password : <Blank>
Expected Output	Error should be displayed saying empty fields are detected
Actual Output	Display an error message saying “Empty fields detected. Please fill up all the fields”

Table 5: Test Case 2

Test case ID	TC2
Test case Description	Validate invalid credentials
Pre-Condition	Login Interface of the People Clues is loaded
Test Procedure	Type username and password Press Login button
Test Input	Username : abc Password : 123

Expected Output	Error should be displayed saying invalid credentials are detected.
Actual Output	Display an error message saying “Incorrect Login Credentials”

Table 6: Test Case 3

Test case ID	TC3
Test case Description	Check for valid login credentials
Pre-Condition	Login Interface of the Smart Resume is loaded
Test Procedure	Type username and password Press Login button
Test Input	Username : ashi Password : 1234
Expected Output	Data Extraction Interface should be displayed.
Actual Output	People Clues home page is displayed.

Table 7: Test Case 4

Test case ID	TC4
Test case Description	Validate empty field of Extraction Interface
Pre-Condition	Have valid login credentials and Extraction Interface of the Smart Resume is loaded
Test Procedure	Press Extract button without selecting any source files

Test Input	File Type : CSV File Source : <Blank>
Expected Output	Error should be displayed saying empty source file is detected
Actual Output	Display an error message saying “Please select a source file.”

Table 8: Test Case 5

Test case ID	TC5
Test case Description	Validate empty field of Extraction Interface
Pre-Condition	Have valid login credentials and Extraction Interface of the Smart Resume is loaded
Test Procedure	Press Extract button without selecting any source files
Test Input	File Type : CSV File Source : <Blank>
Expected Output	Error should be displayed saying empty source file is detected
Actual Output	Display an error message saying “Please select a source file.”

Table 9: Test Case 6

Test case ID	TC6
Test case Description	Change of source file type in Extraction Interface

Pre-Condition	Have valid login credentials and Extraction Interface of the Smart Resume is loaded
Test Procedure	Select another source file type from the dropdown
Test Input	File Type : CSV
Expected Output	Warning should be displayed saying change of source file
Actual Output	Display a warning message saying “Would you like to remove the selected file?”

Table 10: Test Case 7

Test case ID	TC7
Test case Description	Validate valid file in Extraction Interface
Pre-Condition	Have valid login credentials and Extraction Interface of the Smart Resume is loaded
Test Procedure	Select a valid source file and click on Extract Button
Test Input	File Type : CSV Source File : \etl\datasets\candidate.csv
Expected Output	People Clues Data View page should be displayed.
Actual Output	People Clues Data View page is displayed.

Table 11: Test Case 8

Test case ID	TC8
Test case Description	Validate valid file in Extraction Interface

Pre-Condition	Have valid login credentials and Extraction Interface of the Smart Resume is loaded
Test Procedure	Select a valid source file and click on Extract Button
Test Input	File Type : CSV Source File : \etl\datasets\candidate.csv
Expected Output	Smart Resume Data View page should be displayed.
Actual Output	Smart Resume Data View page is displayed.

Table 12: Test Case 9

Test case ID	TC9
Test case Description	Check Map Attribute button of Data View Interface is working
Pre-Condition	Have valid login credentials and Data View Interface of the Smart Resume is loaded
Test Procedure	Select all the columns needed and click on Map Attributes button Click on Done button
Test Input	Dataset type : Train Dataset
Expected Output	Smart Resume Attribute Mapper page should be displayed.
Actual Output	Smart Resume Attribute Mapper page is displayed.

Table 13: Test Case 10

Test case ID	TC10
Test case Description	Check Next button of Attribute Mapper Interface is working
Pre-Condition	Have valid login credentials and Attribute Mapper Interface of the Smart Resume is loaded
Test Procedure	Check all the attributes are mapped correctly Click on Next button
Test Input	-
Expected Output	Smart Resume Data Cleansing page should be displayed.
Actual Output	Smart Resume Data Cleansing page is displayed.

Table 14: Test Case 11

Test case ID	TC11
Test case Description	Check Finish button of Data Cleansing Interface is working
Pre-Condition	Have valid login credentials and Add Rule for Attributes Interface of the Smart Resume is loaded
Test Procedure	After selecting data cleansing type click on Finish button
Test Input	Check Cleansing type

Expected Output	Data Insertion Successfully message should be displayed
Actual Output	Data Insertion Successfully message should be displayed

- Integration Testing – In this testing level I integrated each interface of the ETL tool and tested.
- System Testing – In this testing level we have integrated the ETL tool, prediction tool and dashboard and tested as a whole system.

2.5. Research Findings

Final Smart Resume system is consists with ETL Component, predictive model and Dashboard features. The results are shown in this section.

For ETL component, Candidate selection attributes were needed. Following are the attributes which were identified mainly from research papers.

1. Age
2. Address
3. Gender
4. Civil Status
5. Email
6. Educational Level
7. Previous Experience
8. Communication Skill
9. Languages

To implement ETL, I have found PETL which is a python library that is helpful in building ETL tool.

3. Results and Discussion

3.1. Results

Gathering data was the hardest and the most important part of this research because in ETL process I have used attributes which use to predict suitable candidates. These candidates' data were secured in those company and it is not accessible by outsiders. Secure data of candidates in the company which is not given to the outsiders of the company. After spent lot of times and gathering information by reading research papers also finally find the set of most suitable attributes use to implement ETL Tool. Smart Resume has the following ETL tool interfaces.

For the extraction part I have implemented Data extraction interface as Figure 2, in order to select source files such as CSV (comma separated values). Extraction interface is very user friendly and simple. This extraction process is implemented mostly for CSV files to extract Data.

After the source file selection, user can select the necessary columns as needed for data warehouse creation. Then the user navigate to the data mapping interface where the raw data column names are mapped according to column names defined in the data warehouse.

In order to map these columns with the data source which is user has selected interface in Figure 3 can be used.

After data mapping is done data needs to be transformed. In order to have a better cleansing for data I have provided a way to check the rules that need to be cleansed.

After all these steps done extracted, transformed and cleansed data loaded into Data warehouse as the result.

3.2. Discussion

With hard working through the year, we have completed the Smart Resume BI tool for selecting CVs for IT Industry. The research was successful because of the help of our supervisor Mr. Lakmal Rupasinghe and my team members who supported to achieve this goal.

Candidates CVs data is the most important data set in this system. To gather these data it was very hardest part of the project as these Candidates data are not given to outsiders as they are very important to the company. Spending lots of time we gathered Data sets and select the attributes which are needed to store in data warehouse for further analysis. These attributes were gathered by reading many research papers, brainstorming and studying related BI tools.

As ETL tool is a core of our system. We need to identify which technologies used for better performance and libraries to create most powerful and useful ETL tool for the system.

Due to the time constraint, we wouldn't able to finish some feature which is useful for future researchers and it will discuss in the next chapter.

4. Conclusion

Today for most companies like IT, receive huge number of CVs for a vacancy as there are lot of graduates coming out from a university within a year. The quality of the company depends with the capabilities of the recruiters. Therefore getting the best qualified people from incoming applications is very difficult for them. Currently the selections of CVs is done manually. As the huge number of CVs are summarized by manually, qualified CVs can be skipped by CV errors and human errors. That is where Smart Resume come in to play by reducing all the difficult tasks of selecting optimal set of candidates in place of referring CVs.

Smart Resume contains a Desktop application with ETL (Extract, Transform, Load) Tool to transform data into a meaningful order and Smart Resume has a Dashboard as

a Web application in order to predict best candidates for a given job opportunity by analyzing the CV data.

Smart Resume has been developed to choose the suitable Candidate for IT Industry.

As the initial step Smart Resume analyze and extract the professional skills, personal skills, personal details and etc. from the CV according to the job vacancy. Finally, Smart Resume display the results in an attractive proper manner in a dashboard with simple graphs and charts which shows the user friendliness of the system

In next step our team will focusing on creating ETL tool which extract data from any file types like word file, json files, access files and sql files.

5. References

[1] C. Shrinivasan, “Data migration from a product to a data warehouse using ETL tool,” *Proc. Eur. Conf. Softw. Maint. Reengineering, CSMR*, pp. 63–65, 2011.

[2] Z. Chen and T. Zhao, “A new tool for ETL process,” *Proc. 2012 Int. Conf. Image Anal. Signal Process. IASP 2012*, pp. 269–273, 2012.

[3] T. Gang, C. Kai, and S. Bei, “The research & application of Business Intelligence system in retail industry,” *Autom. Logist. 2008. ICAL 2008. IEEE Int. Conf.*, no. September, pp. 87–91, 2008.

[4] R. Samarasinghe, G. Perera, N. Perera, P. Senaratna, and L. Samarasingha, “People Clues : Business Intelligence Tool for Team Dynamics,” pp. 179–184, 2017

[5] T. Gang, C. Kai, and S. Bei, “The research & application of Business Intelligence system in retail industry,” *Autom. Logist. 2008. ICAL 2008. IEEE Int. Conf.*, no. September, pp. 87–91, 2008.

[6] Agustín-Blas, Luis E.,Salcedo-Sanz, Sancho, Ortiz-García, Emilio G.,Portilla-Figueras, Antonio,Pérez-Bellido, Ángel M.,Jiménez-Fernández, Silvia, “Team formation based on group technology: A hybrid grouping genetic algorithm approach Computers & Operations Research, vol. 38, issue. 2, p. 484-495, 2011. [Online serial]. Available: sciencedirect, http://www.sciencedirect.com/science/article/pii/S0305054810001371

[Accessed March 06, 2016].

6. Appendices

Work Breakdown Structure:

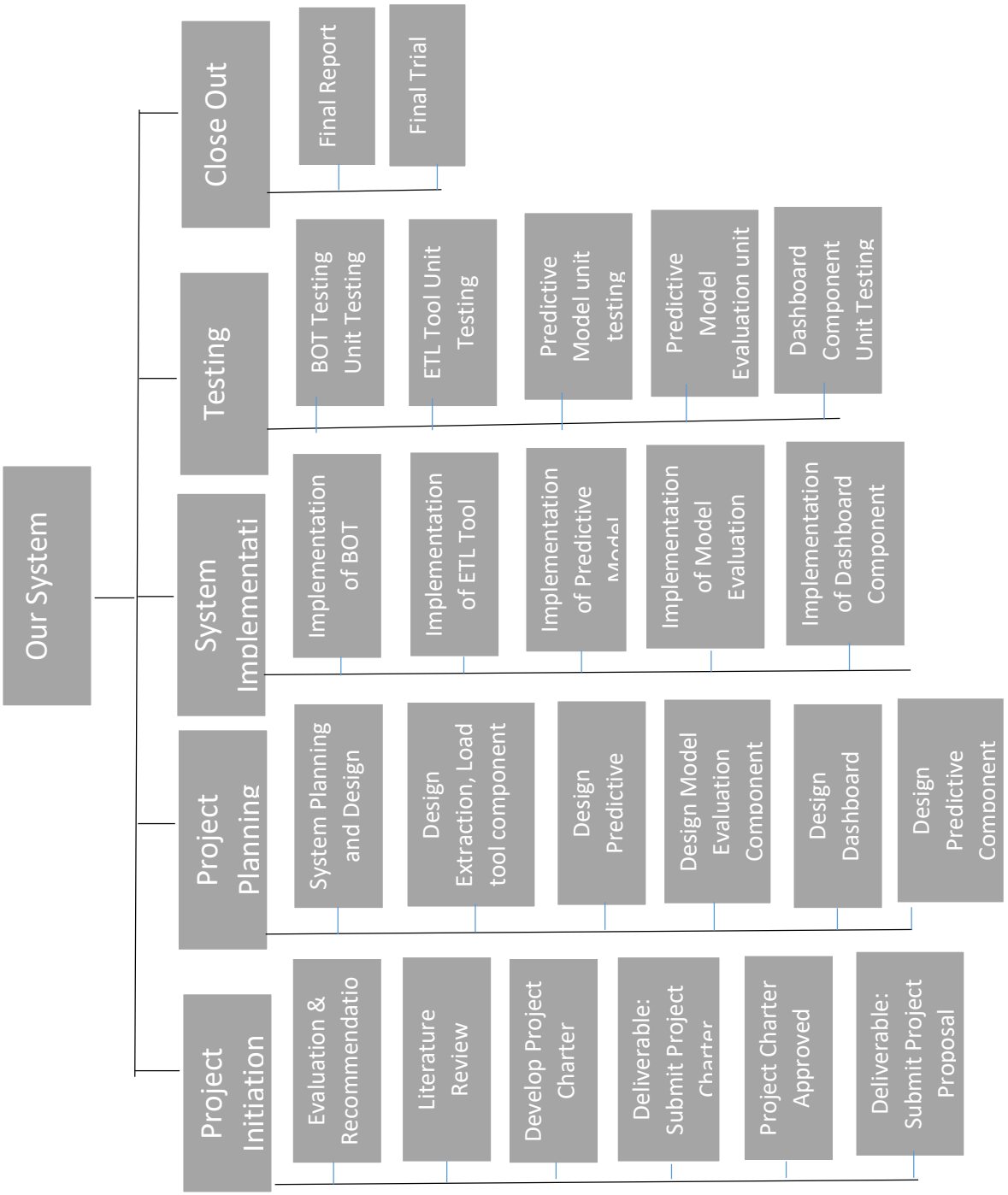


Figure 5 Work Breakdown Structure