Proceedings of Second Shri Chhatrapati Shivaji Maharaj QIP Conference on Engineering Innovations
Organized by Shri Chhatrapati Shivaji Maharaj College of Engineering, Ahmednagar
In Association with Novateur Publications JournalNX-ISSN No: 2581-4230
February, 22nd and 23rd, 2019

Linkedin Profile Extractor

Ms.Samriddhi Shantilal Kothari
Department of Computer Engineering
SNJB's KBJ College of Engineering
Chandwad, India
samkothari1810@gmail.com

Mr.Jayesh Prakashchandra Bafna
Department of Computer Engineering
SNJB's KBJ College of Engineering
Chandwad, India
jayeshbafna619@gmail.com

Mrs.Dipali Prashant Pawar
Assistant Professor Department of Computer Engineering
SNJB's KBJ College of Engineering
Chandwad,India
Pawar.dpcoe@snjb.org

Abstract— **Today, most of the professionals may require the past working history of their employees, colleagues at one glance so they will approach towards the professional social site where they can get all the professional past history of the person. Linkedin network is the best of them but because it contains huge data which makes this process of manual information extraction from the whole data structure complicated. The Linkedin Data structure contains data which is not in structured format. Linkedin profile extraction is the process of extracting user required information from Linkedin website. From the word Linkedin profile extraction, we mean the extraction of profile data that is present in the Linkedin data structure . Learning the information or patterns or features present in that Linkedin data structure, one can easily extract the required profile data in efficient way and in special manner. Linkedin Extractor is the way to do that.**

Keywords:-Extractor, Profile, Past History, Pattern.

I. INTRODUCTION

The information present on the internet is very large and it is not possible to manually take the information from any website to perform information extraction. The web data is un-
structured that means it contains noise or unwanted data. Many data mining techniques are available today to perform analysis on the web data. Data mining means gaining information from the huge amount of data. In our project the data is studied from multiple views and taking the summaries from that useful data which is used for extracting desired profile. This mined data is used by higher authorities to filter out the profile of their passed out batch of their institute, reduce manual work of entering data again and again. Also it will reduce the work of calling or texting them regarding their current working profile.

II. TECHNIQUE AND PLATFORM

A. *Text Mining*

Text Mining is also known as Text Data Mining. The purpose is too unstructured information, extract meaningful numeric indices from the text. Thus, make the information contained in the text accessible to the various algorithms. Text mining is an exercise to gain knowledge from stores of language text.The root working of the software is based on the Text Mining.It is the first basic step followed by the Extractor for processing of further actions/operations.
Typically falls into one of two categories

- Analysis of text: To analyze the certain pattern and provide the specific results. E.g. Which of the profile has the institute name as "SNJB"?
- Retrieval: To extract the analyzed pattern to perform certain operation. E.g. Retrieval of the profile data to map it into the table?

B. *Salesforce*

Salesforce is the primary enterprise offering within the Salesforce platform. It provides companies with an interface for case management and task management, and a system for
automatically routing and escalating important events. The Salesforce customer portal provides customers the ability to track their own cases, includes a social networking plug-in that enables the user to join the conversation about their company on social networking websites, provides analytical tools and other services including email alert, Google search, and access to customers' entitlement and contracts.

- Integration:

User Interface Integration is one great way to surface various applications inside Salesforce with little redesign of each individual app. It provides your users a single point of entry into multiple applications.

- App Manager:

Salesforce application manager, is provided for development of user defined business application.

- Report and dashboard generator:

Using this functionality of Salesforce ,we can generate report of desired profile in main format of xls sheet. Also we can display the reports in the form of pie chart and bar chart. Dashboard is used to provide the overview of the report generated.

III. RELATED WORK

A. *Literature Survey*

- ARPPA

Crawling and mining professional profiles from LinkedIn are challenging problems. e. Moreover, manual analysis is

an exhaustive and prohibitive task often demanding the use of data mining techniques for fast, less expensive and more effective analytical processing. In this article, we introduce ARPPA, a novel approach to discover professional profile patterns from LinkedIn by using association rules mining. ARPPA is an acronym for "Association Rules for Professional Profile Analysis"

- LinkedIn Lead Extractor

LinkedIn Lead Extractor is a desktop application which allows you to extract data fromLinkedIn at an exceptionally fast rate. It automatically extracts available business name,email address, phone number, Yahoo messenger id, Skype Id, Google Talk ID, etc. Lead Extractor searches your targeted customers based on your search keywords. Its drawback is such that,it does not support mobile accessibility.Also customization is not possible.It is on premise cloud based.It is not open source.

- eXtractorONE

eXtractorONE is capable of decomposing html pages into a tabular model in SQL Server.Youcan _lter on any element if you just need to extract data from web tables or any other component.The advanced web connector in eXtractorONE can parametrize URLs to have them populated with a sequence to grab all pagesSOAP Web Services can be used to distribute data via the web. eXtractorONE can connect to any web service just by entering the WSDL URL. The parameters can be populated by values in the database of your company. I.e.: upload the VAT numbers of your complete supplier table to the VIES web service and retrieve the validity of the VAT number including the registered location. All data will be stored in readable tables. IT departments also tend to open data via WSDL web services for other departments. As the data is distributed in XML format eXtractorONE is the right choice. Its drawback is such,that it cannot provide the extracted data in proper format.The data which is provided by extraction is not much reliable. Also the API for data is provided by entering the API URI with token.Asloit only provide scrapped data in structured format in excel sheet.

B. Equations(TNR- 10, Italic,Bold)

- Precision

$$precision = \frac{|\{Relevant\} \cap \{Retrieved\}|}{|\{Retrieved\}|}$$

- Recall

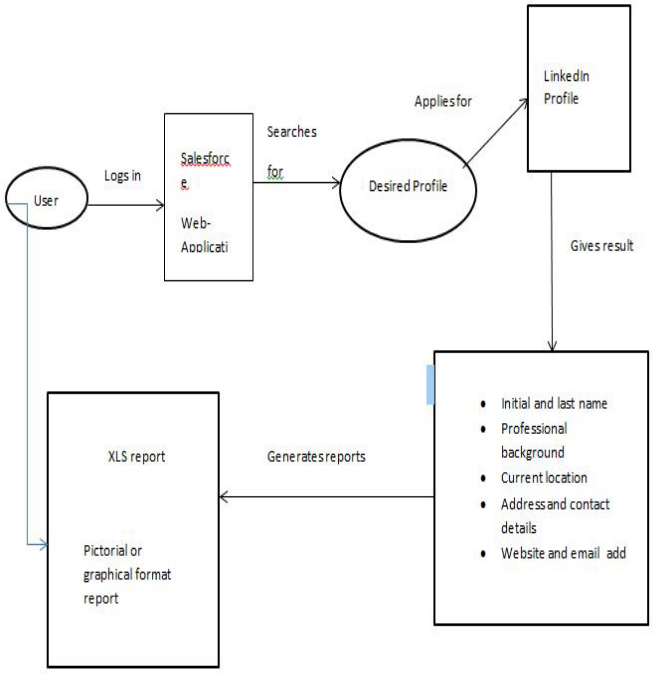
$$recall = \frac{|\{Relevant\} \cap \{Retrieved\}|}{|\{Relevant\}|}$$

- F-Score

$$F\_score = \frac{recall \times precision}{(recall + precision)/2}$$

C. Proposed System

Following is the concept to be implemented under the proposed system to be developed In the proposed system we are going to integrate the Linkedin API Base Along with the Sales Cloud provided by Salesforce.In Our System the profile data will be get analysed according to partitioned basis Like Graduation History,WorkingHistory,Current Professional Background,CurrentLocation,Initials & Last Names,Websites and Email Adress,Address and Contact Details,etc. All this Data will be gathered under the Sales Cloud where the Data will go through Extraction,Transform And Load Process to Generate the Result According to User Applied Filters,Finally It will Represent the Result in the format of XLS reports, Bar Charts,and Pie Diagram.User can also get his generated result on his Login Dashboard.



IV. SOFTWARE REQUIREMENT SPECIFICATION

A. Functional Requirements:

- The system will provide the data in xls format which will contain whole information of profile by exporting it.
- It will also provide the data in the form of pictorial representation in the form of graphs and charts.

B. ***External Interface Requirement:***

- User Interface: One window will appear for entering required profile name. Also require any laptop or PC with internet connection.
- Software Interface: Salesforce platform.*:*

C. ***Non-Functional Requirements***

- Performance requirement

Extractor should provide the Correct Extracted Data and also Work according to Time Efficient Way.

- Safety Requirement

Safety Against the Web Threats should be taken under considerations As this project works on Web Based Background

- Security Requirement

While LinkedIn is valuable for building your professional presence, it's important to be
conscious of your individual privacy and security when using the network.
1.The authentication for extraction is provided to authorised person.
2.It is not possible to edit the data of profile which is in xls sheet

D ***System Requirement***

- Database Requirement

Database is the most important key for the project.DataBase is provided by Sales Cloud itself and will be maintained by sales cloud itself.DataBase is required to Store the Data gained through the LinkedIn Api.So Sales Cloud Data Base is the required DB for the Project.

- Hardware Requirement

Ram-1 GB, HDD-100GB, Processor-Quad core

- Software Requirement

Salesforce studio as a platform for accessing their own cloud. Salesforce is the world's no1 Customer Relationship Management (CRM) platform. Our cloud-based applications for sales, service, marketing, and more don't require IT experts to set up or manage – simply log in and start connecting to customers in a whole new way. Treats all their customers like they're their only customer with Salesforce CRM. Understand their needs, solve their problems, and identify opportunities to help by managing their information and interactions with our company on a single platform that's always accessible from anywhere.

V. TABLE

Parameters	ARPPA	Linkedin Lead Extractor	Extractor One
Purpose	Scrapping Data	Scrapping Data	Scrapping Data
Efficiency	Less	Medium	Good
Accuracy	Less	Moderate	Less
Speed	Moderate	Less	Good
Data_type	Structured	Structured	Structured
Security	Less	Moderate	Moderate
Filters	No	Yes	Yes

Fig. 1. Literature Survey

Acknowledgment

We would like to acknowledge all the people who have been of the help and assisted me throughout my project work. First of all we would like to thank our respected guide
Prof.D.P.Pawar , Professor in Department of Computer Engineering for introducing me throughout features needed. The time-to-time guidance, encouragement, and valuable suggestions received from him are unforgettable in my life. This work would not have been possible without the enthusiastic response, insight, and new ideas from him. We are also grateful to all the faculty members of SNJB's College of Engineering for their support and cooperation. I would like to thank my lovely parents for time-to-time support and encouragement and valuable suggestions, and thank our friends for their valuable support and encouragement. The acknowledgement would be incomplete without mention of the blessing of the Almighty, which helped me in keeping high moral during most difficult period.

References

[1]. Paula R. C. Silva, Wladmir C. Brando "ARPPA: Mining Professional Profiles from LinkedIn Using Association Rules." The Seventh International Conference on Information, Process, and Knowledge Management

[2]. Klaus Zechner "A Literature Survey On Information Extraction And Text Summarization." April 14 1997

[3]. "Integration Pattern And Practices." By SalesforceEdition.Version 44.0 Winter 19'

[4]. Miss. N. V. KamanwarM.Tech., Dept. of Information TechnologyY.C.C.E., Hingna Road, WanadongariNagpur- 441110, India. Prof. S. G. Kale Dept. of Information Technology Y.C.C.E., Hingna Road, Wanadongari Nagpur- 441110, India "Web Data Extraction Techniques: A Review."

[5]. Data Mining -Volinsky-2011 -Columbia University "Text Mining."