ISSN No:-2456-2165

Resume Refinement System using Semantic Analysis

Ruchita Haresh Makwana
IT Department,
K J Somaiya Institute of Engineering & IT, Sion,
Mumbai, India.

Aishwarya Mohan Medhekar
IT Department,
K J Somaiya Institute of Engineering & IT, Sion,
Mumbai, India.

Bhavika Vishanji Mrug
IT Department,
K J Somaiya Institute of Engineering & IT, Sion,
Mumbai, India.

Abstract:- The Resume Refinement System focuses on online recruitment process, where the recruiter uses this system to hire the suitable candidate. Recruiter post the job-post, where the candidates upload their resumes and based on semantic analysis the web application will provide the list of shortlisted candidates to the recruiters. Since web documents have different formats and contents; it is necessary for various documents to use standards to normalize their modeling in order to facilitate retrieval task. The model must take into consideration, both the syntactic structure, and the semantic content of the documents. Resume is the document that summaries our education, skills, accomplishments, and experience. Job seekers submit their Resume via the web. Therefore, in their recruitment process, companies are requiring systems for extraction and analysis of information from Resume: identifying specific patterns, which meet with certain profile. To extract the essential component of Resumes and to relate them with recruiter's requirements needs first, a study of their most significant elements and a better understanding of the resume feature. This work focuses on resume analysis.

Keywords:-Online recruitment, semantic analysis, neural network, missing background knowledge.

I. INTRODUCTION

Few years back recruitment process was done manually. HR posts the job requirements on a particular site where candidates upload their resumes. Then the HR will download the resumes and will do the keyword matching manually ignoring the semantics of the job post and the resume contents due to which large number of gained results are irrelevant. To overcome this we have proposed this system where it will do semantic analysis and will give appropriate results.

An online recruitment system that first employs Natural Language Processing (NLP) tools to find and extract a list of candidate concepts from both job posts and candidates resumes. Next, existing semantic resources (also referred to as ontologies) are cooperatively incorporated to analyze the list of candidate concepts at the semantics level. When a concept is not recognized by the used semantic resources, statistical-based concept-relatedness techniques are then used to address this issue.

The scope of our online recruitment system: The Resume Refinement System will help the recruiter to find the appropriate candidate for the job post. It is the time saving process for the recruitment. Since previously all this process was done manually the chances of right candidate was minimum. By using this system the chances of selecting right candidate is maximum. Besides using programming languages to select candidate, we are going to use educational background, years of experience etc. The scope of this project is to get right candidate for the job post, with minimum time requirement and to get efficient results.

II. LITERATURE SURVEY

A. An Automatic Online Recruitment System Based on Exploiting Multiple Semantic Resources and Concept-Relatedness Measures.

Recruitment is considered among the most challenging functions for job portals and human resource (HR) departments. This is because employers often receive a huge number of resumes – some of which are uploaded as unstructured documents in different formats such as .pdf, doc, and .rtf, while others are uploaded according to specific forms prepared by employers – that are difficult to manually process and analyze. Recently, many companies have shifted to automatic online recruitment systems in an attempt to reduce the cost, time, and efforts required for screening out applicants and matching candidate resumes to their relevant job posts. Several techniques/approaches have been employed by online recruitment systems. Examples of these techniques are Boolean Retrieval, models based on Relevance Feedback , Analytic Hierarchy Process, Semantics- based techniques, and Natural Language Processing (NLP) and Machine learning based approaches. Although these techniques achieve good matching results, they are still limited by the obstacles.

In order to avoid the obstacles they have proposed an automatic online recruitment system that exploits multiple semantic resources in an attempt to highlight and capture the semantic aspects of both job posts and candidate resumes. The proposed system employs NLP pre-processing techniques to identify and extract lists of candidate concepts from job posts and resumes. In addition, it utilizes statistical concept-relatedness measures (extracted from Hiring solved Dataset) to enrich and expand the lists of candidate concepts with entities

i.e. concepts that were not initially recognized by the employed semantic resources.

B. Matching Sem Online recruitment system based on multiple semantic resources

They have proposed an online recruitment system where they use a combination of multiple semantic resources and statistical based techniques. The aim behind using multiple semantic resources is that they are capable of representing several domains and derive the semantic aspects of resumes and job posts. When some of the identified concepts (from the resumes of job posts) are not found in the employed semantic resources, then, the statistical-based semantic relatedness measures are used in an attempt to find the relation between those missing concepts i.e. from the semantic resources and those that are defined in them.

C. Toward the Next Generation of Recruitment Tools: An Online Social Network-based Job Recommender System

The rapid growth of social networks in recent years has developed a new business: the trade of social networks users data. These social networks data are becoming important for many companies around the world and are often used to determine social networks users interests for items in order to propose or advertise items to them. Recommender systems help users deal with data overload by recommending to them items that they would like. There has been a lot of work done on designing recommender systems during the last two decades. Amazon.com and Netflix are two popular applications of recommender systems.

Here they have proposed an online social network-based recommender system where it extracts the users interests for job and recommends them accordingly. For doing so they have considered the users interaction data like comments, likes, publications etc and jobs descriptions to predict users interests for jobs. This is particularly done for Facebook and LinkedIn Users.

III. COMPARISON BETWEEN EXISTING & PROPOSED SYSTEM

A. Existing System

In this system they have used various modules which are as follows. First they have used Concept Identification and Extraction to create concept list from both resume and job post. Here they have used td-idf and features list to do. Then using feature list words they have retrieved words from semantic resources like word net and yago2 and using those words they have constructed a semantic network. This network is formed by connecting the words by various types of semantic relations obtained from semantic resources. Also they have used Missing Background Knowledge Handler to enrich the constructed semantic networks.

B. Proposed System

Here we will create a web-application where resumes and job post is given as a input to the system. Since web documents have different formats and contents it is necessary to preprocess in order to facilitate retrieval task. To do so we use NLP pre-processing Techniques which are as follows. First we will convert the resumes into a particular format i.e txt format. Then we will perform tokenization like sentence tokenization and word tokenization. Next we will remove the stop words like 'the' , 'an' , 'a' etc. Then we will do tf-idf, speech tagging, named entity recognition. Next we will create semantic networks of job post and resumes by using Word net, Dbpedia ontologies and the algorithm used will be Jaro distance for finding the similarity between both the networks. Finally recruiter will get the resumes of suitable candidates.

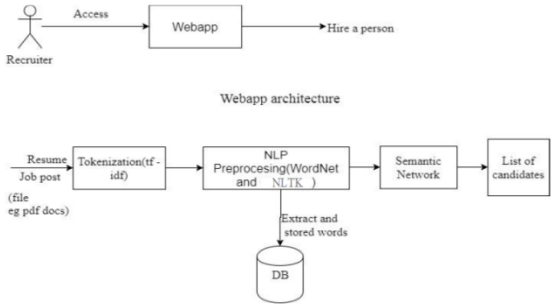

Fig. 1:- Proposed System

Existing System	Proposed System
In the existing system, there is exact keyword matching.	In the proposed system we use semantic analysis for arranging words to show relationship among them.
Here they use only keyword matching for refinement process.	Here we are considering additional years of experience, missing background knowledge for refinement process.
Specific document types were only supported.	Different document types such as pdf, docx are supported where we convert it into standardized format like txt.
Techniques used by existing system are WordNet, Yago2.	Techniques used in proposed system are NLP pre processing,WordNet,NLTK, DBpedia

Table 1: Existing System v/s Proposed System

ISSN No:-2456-2165

IV. METHODOLOGY

Proposed Online Resume Recruitment System is based on semantic analysis where the system focuses on online recruitment process, where the recruiter uses this system to hire the suitable candidate. Recruiter post the job-post and accordingly the candidates upload their resumes and based on semantic analysis the web application will provide the list of shortlisted candidates to the recruiters

➢ *Natural Language Toolkit*

- The Natural Language Toolkit, or more commonly NLTK, is a suite of libraries and programs for symbolic and statistical natural language processing (NLP) for English written in the Python programming language. It was developed by Steven Bird and Edward Loper in the Department of Computer and Information Science at the University of Pennsylvania. NLTK includes graphical demonstrations and sample data. It is accompanied by a book that explains the underlying concepts behind the language processing tasks supported by the toolkit, plus a cookbook.
- NLTK is intended to support research and teaching in NLP or closely related areas, including empirical linguistics, cognitive science, artificial intelligence, information retrieval, and machine learning. NLTK has been used successfully as a teaching tool, as an individual study tool, and as a platform for prototyping and building research systems. NLTK supports classification, tokenization, stemming, tagging, parsing, and semantic reasoning functionalities.[4]

➢ *DBpedia*

- DBpedia is a crowd-sourced community effort to extract structured information from Wikipedia and make this information available on the Web. DBpedia allows you to ask sophisticated queries against Wikipedia, and to link the different data sets on the Web to Wikipedia data. Knowledge bases are playing an increasingly important role in enhancing the intelligence of Web and enterprise search and in supporting information integration. Today, most knowledge bases cover only specific domains, are created by relatively small groups of knowledge engineers, and are very cost intensive to keep up-to-date as domains change. At the same time, Wikipedia has grown into one of the central knowledge sources of mankind, maintained by thousands of contributors.[5]
- The DBpedia project leverages this gigantic source of knowledge by extracting structured information from Wikipedia and by making this information accessible on the Web under the terms of the Creative Commons Attribution-Share Alike 3.0 License and the GNU Free Documentation License.[5]

V. ALGORITHM

The Jaro distance is a measure of similarity between two strings R(Resume) and J(Job Post).This function is useful to create semantic networks of job post and resumes. The higher the Jaro distance for two strings is, the more similar the strings are. The score is normalized such that 0 equates to no similarity and 1 is an exact match.

- *Input: R and J*

Output: Measure of similarity between the strings R and J based on the set of correspondences Semantic Network(SN).

```
1: int similarity;
2: result ();
3: for i=0;i<R.length;i++
4: for j=0;j<J.length;j++
5: result= Jaro(R[i],J[j])
6: if(result<m) then
7: add (R[i],J[j]) to SN
8: similarity++;
9: end if
10: end for
11: end for
12:return result
```

VI. CONCLUSION

The proposed system is

- Reliable, fast and scalable approach.
- Selects the appropriate candidate for the job post.
- To increase the efficiency of result as compared to existing system.
- To make the recruitment process easier.
- Minimizes the time of the selection process.

REFERENCE

[1] An Automatic Online Recruitment System Based on Exploiting Multiple Semantic Resources and Concept-Relatedness Measures, 2015 IEEE.

[2] Matching Sem: Online recruitment system based on multiple semantic resources, IEEE

[3] Toward the Next Generation of Recruitment Tools: An Online Social Network-based Job Recommender System, 2013 IEEE.

[4] Natural Language Toolkit. Available: http://www.nltk.org/

[5] DBpedia. Available: http://wiki.dbpedia.org/