**RESUME PARSER**

**1. INTRODUCTION**

These days, recruiters and human resources professionals are flooded with applications and searching resumes often becomes a highly challenging task for them. Sifting through thousands of resumes for some of the job vacancies is something not as easy as it seems.

Therefore, to ensure the act of resume search becomes easier through the stacks of applications received on a daily basis, organizations have long resorted to Applicant Tracking Systems (ATS), and of late, they are focusing on more advanced HR technology to make the process of recruitment and hiring simpler.

* 1. **Background**

Resume searches have been made much simpler with the help of a technology named resume parser and the process is termed as resume parsing.

So, what is resume parsing?

Resume parsing is the name given to the technology that allows an employer to process resumes submitted online by extracting data from it and then converting those into appropriately structured information. To put it simply, parsing allows recruiters to search resumes in an intelligent manner and thereby focus on the right candidate.

How does resume parsing benefit recruiters?

Resume parsing technology helps create recruiters and HR professionals, a more convenient resume application as well as a screening process. It allows recruiters to organize resume data electronically by implementing semantic search technology in the background.

Key measures of a good resume parser

Needless to say, to make resume searches simpler and easier, there are some steps or features, which distinguishes a good resume parser from others. Some of the most advanced resume parsers possess the ability to even extract references, candidates’ interests/hobbies, desired location and compensation, demographics and other related information, just by implementing the process of resume searching.

* 1. **Objective:**

The specific objectives of this project include

* Extract Current Title of candidate profile.
* Extract Current Company of candidate profile.
* Extract Current Location of candidate profile.
* Extract Contact Number and Email ID of candidate profile.
  1. **Project Justification:**

Human resources professionals are flooded with applications and searching resumes often becomes a highly challenging task for them. So using Resume parsing technology helps create recruiters and HR professionals, a more convenient resume application as well as a screening process. It allows recruiters to organize resume data electronically by implementing semantic search technology in the background.

**2. SURVEY OF TECHNOLOGIES**

* 1. **Software requirements:**

**MYSQL DBMS**

MYSQL, the most popular Open Source SQL database management system, is developed. Distributed and supported by Oracle Corporation.

A database is a structured collection of data. It may be anything from a simple shopping list to picture gallery or the vast amounts of information in a corporate network. To add, access, and process data stored in a computer database, you need a database management such as MYSQL Server. Since computers are very good at handling large amounts of data, database management systems play a central role in computing, as standalone utilities, or as parts of other applications.

A relational database stores data in separate tables rather than putting all the data in one big storeroom. The database structures are organized into physical files optimized for speed. The logical model, with objects such as databases, tables, views, rows, and columns, offers a flexible programming environment. You set up rules governing the relationships between different data fields, such as one-to-one, one-to-many, unique, required or optional, and “pointers” between different tables.

The SQL part of “MySQL” stands for “Structured Query Language”. SQL is the most common standardized language used to access databases. Depending on your programming environment, you might enter SQL directly (for example, to generate reports), embed SQL statements into code written in another language, or use a language-specific API that hides the SQL syntax.

**Visual Studio 2015**

The Visual Studio 2015 Preview included many new features that enhanced the way developers work with everything from the web and desktop to mobile apps. Several features have had the spotlight, such as gesture support in the editor, C++ enhancements and the new Android emulator.

C# is a modern, general-purpose, object-oriented programming language developed by Microsoft and approved by European Computer Manufacturers Association (ECMA) and International Standards Organization (ISO).

**C# Coding.**

C# is a modern, general-purpose, object-oriented programming language developed by Microsoft and approved by European Computer Manufacturers Association (ECMA) and International Standards Organization (ISO).

**Testing** – Unit testing.

**Web browsers:** Mozilla Firefox, Google Chrome, Opera and Internet Explorer.

* 1. **Hardware Requirements**

**Microsoft Windows XP Professional SP3/Vista SP1/Windows 7/Windows 10 Professional:**

**Processor:** 800MHz Intel Pentium III or equivalent

**Memory:** 512 MB

**Disk Space:** 750 MB of free disk space

**3 PLANNING AND SCHEDULING**

The project is planned to complete in three months’ timeline, covering all the stages in the chosen conventional model. In our case, we have chosen the Waterfall model. We have estimated two weeks’ time for identifying the requirement and analysis for Resume parser. The design of the project and identifying the entity and their relations will be done in next two weeks. Fully implementation of the project and coding will need a month. Testing for the project will require two weeks and fixing the issues found / maintenance will be done in the last two weeks.

**Schedule:**

|  |  |
| --- | --- |
| **Task** | **No of Days / Weeks** |
| Requirement Analysis | 2 weeks |
| Designing | 2 weeks |
| Coding | 4 weeks |
| Testing | 2 weeks |
| Maintenance | 2 weeks |
| Documentation | 3 days |

**Implementation:**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **TASK** | **Week 1-2** | **Week 3-4** | **Week 5-6** | **Week 7-8** | **Week 9-10** | **Week 11-12** |
| **Requirement Gathering** |  |  |  |  |  |  |
| **Designing** |  |  |  |  |  |  |
| **Coding** |  |  |  |  |  |  |
| **Testing** |  |  |  |  |  |  |
| **Maintenance** |  |  |  |  |  |  |

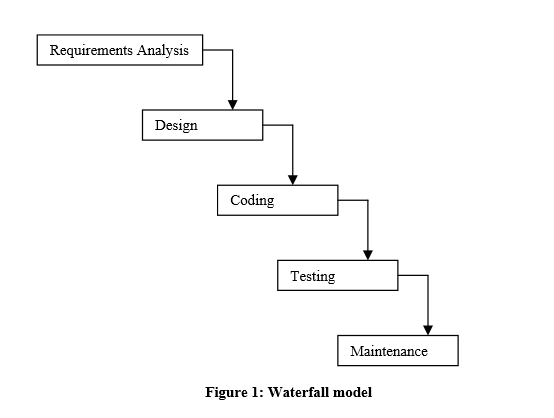
* 1. **Scope of Solution**
* In today’s recruiting world social media recruiting and online recruitment have become the order of the day. Web sources like social networking sites, job boards, blogs, professional networking sites, wiki, blogs, forums, search engines, etc. have now become the hunting grounds for recruiters to look for resumes. However, the challenge here is how to parse resumes and transfer the candidate information into the database in a more structured manner.
* Surveys reveal that recruiters spend most of their valuable recruiting time to extract resumes and transfer applicant’s details to the database. It is one of the reasons that smart recruiters invest in Resume Parsing software to make the most of their opportunities..

**4. CONCEPTUAL MODEL**

We are approaching Waterfall model to implement Resume parser.

**Waterfall** **Model:**

It is the simplest, oldest and most widely used process model. In this model, each phase of the life cycle is completed before the start of a new phase. It is actually the first engineering approach of software development. Below figure depicts the Waterfall Model.



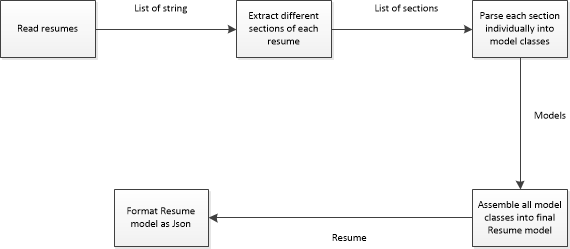
The function of various phases is discussed in software process technology. The waterfall model provides a systematic and sequential approach to software development and is better than the build and fix approach. But, in this model, complete requirement should be available at the time of commencement of the project, but in actual practice, the requirements keep originating during different phases. The waterfall model can accommodate new requirements only in maintenance phase. Moreover, it does not incorporate any kind of risk assessment. In the waterfall model, a working model of software is not available. Thus, there is no way of judging the problems between different phases. A slight modification of the waterfall model is a model with feedback. Once software is developed and is operational, then the feedback to various phases may be provided.

**5. DATA FLOW DIAGRAMS**

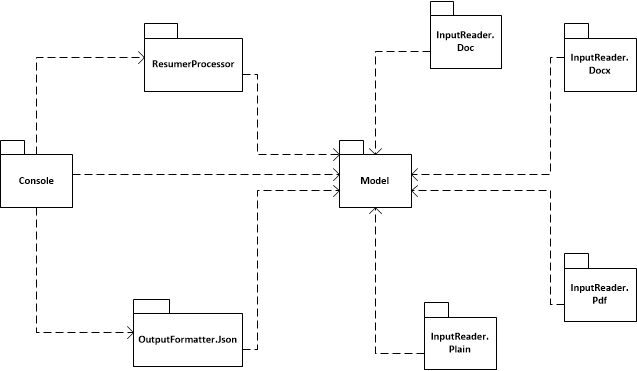
Data flow diagram is the starting point of the design phase that functionally decomposes the requirements specification. A DFD consists of a series of bubbles joined by lines. The bubbles represent data transformation and the lines represent data flows in the system. A DFD describes what data flow rather than how they are processed, so it does not require hardware, software and data structure.

A data flow diagram (DFD) is a graphical representation of the flow of data through an information system. DFDs can also be used for the visualization of data processing. A data flow diagram is a significant modeling technique for analyzing and constructing information processes. DFD literally means an illustration that explains the course or movement of information in a process. DFD illustrates this flow of information in a process based on the inputs and outputs. A DFD can be referred to as a process model.

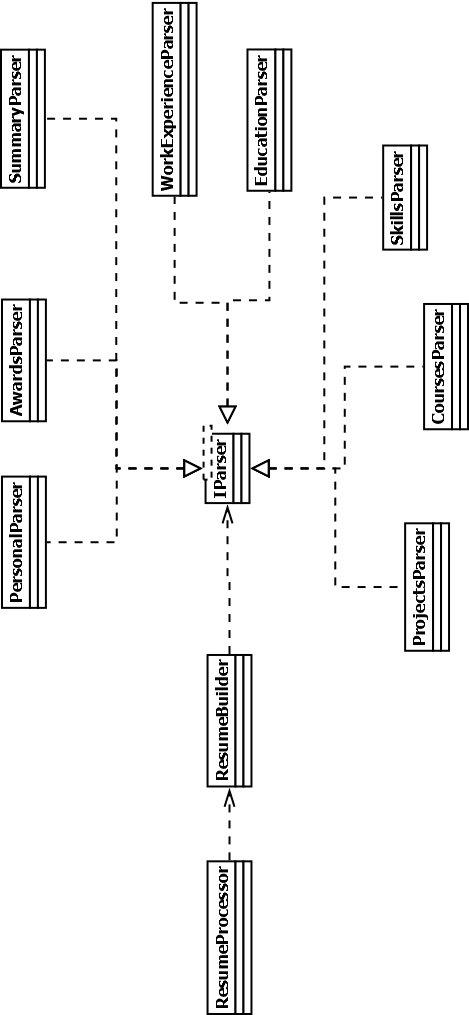
**Level 0**



**Level 1**



**6. ACTIVITY DIAGRAM**



**7. DATABASE TABLES**

This project uses below tables

1. Candidate
2. Experience
3. Education

**Candidate table:**

|  |  |  |
| --- | --- | --- |
| **Field Name** | **Data Type** | **Description** |
| CandidateID | Integer | Resume ID (Primary key) |
| ResumeName | Varchar | Name of Resume. |
| CandidateName | Varchar | Parsed Name of Candidate |
| CurrentCompany | Varchar | Parsed Current company |
| CurrentTitle | Varchar | Parsed Current Title |
| Current Location | Varchar | Parsed Current Location |

**Experience table:**

|  |  |  |
| --- | --- | --- |
| **Field Name** | **Data Type** | **Description** |
| ExpericneID | Integer | ExpericneID (Primary key) |
| CandidateID\_FK | Integer | CandidateID from Candidate Table. This is Foreign Key Relation. |
| Order | Integer | Order of Experience. |
| Title | Varchar | Title of Candidate in present position. |
| Company | Varchar | Company of Candidate in present position. |
| Location | Varchar | Location of Candidate in present position. |
| StartDate | Varchar | Start Date of Candidate in present position. |
| EndDate | Varchar | End Date of Candidate in present position. |

**Education table:**

|  |  |  |
| --- | --- | --- |
| **Field Name** | **Data Type** | **Description** |
| EducationID | Integer | EducationID (Primary key) |
| CandidateID\_FK | Integer | CandidateID from Candidate Table. This is Foreign Key Relation. |
| Order | Integer | Order of Education. |
| Degree | Varchar | Degree of Candidate in present Institute. |
| College | Varchar | Name of the Institute. |
| Location | Varchar | Location of Institute. |
| StartDate | Varchar | Start Date of Candidate in that Institute. |
| EndDate | Varchar | End Date of Candidate in that Institute. |

**8. MODULES OF THE PROJECT**

**8.1 Resume Parser App:**

This App takes resume as Input and provides parsed output in JSON format.

**8.2 Resume Processor:**

This module extracts plain text content of the provided resume which is in .docx, .doc, .pdf formats.

**8.3** **Resume Processor Business Logic:**

This module extracts data at each section level like personal section which contains Name, EmailId , Phone number and Location.

Experience Section in which parser extracts all experiences of candidate.

Education section in which parser extracts all educational background.

**8.4 Resume Processor DAL:**

This module helps to interact with database and business logic.

All the extracted entities will be saved in database through DAL (Data Access Layer).

**9. FUTURE SCOPE**

The project is one of the emerging products in industry. We can get the better quality using NLP and Artificial Intelligence. Deep learning the core technology to be used to get best quality. We have train with millions of resumes for machine learning model to get best parsed entities.

**10. BIBLIOGRAPHY AND REFERENCES**

**Books:**

1. Head First PHP & MySQL – by Lynn Beighley & Michael Morrison
2. Code complete: A practical handbook of programming and software construction
3. Software Engineering by Rajib Mall

**References:**

[www.codeproject.com](http://www.codeproject.com)