



Instant Resume Evaluation Engine

Jinesh Dhruv (jad6566@g.rit.edu)
Advisor: Prof. Christopher Homan



Introduction

- The job seeker prefers having the resume in PDF format as it can be opened on any operating system, formatting does not get messed up and it provides a better appearance
- Existing resume parsing system still has trouble parsing and extracting the content from the PDF due to the different layout-structures

Contribution

- Created a hybrid content-based & segmentation-based technique for resume parsing
- Developed a tool where user can upload the resume for information extraction

Tools and Technology

- Software & Python packages used: Meteor, MongoDB, Python 2.7, pdfminer, NLTK, Fuzzy

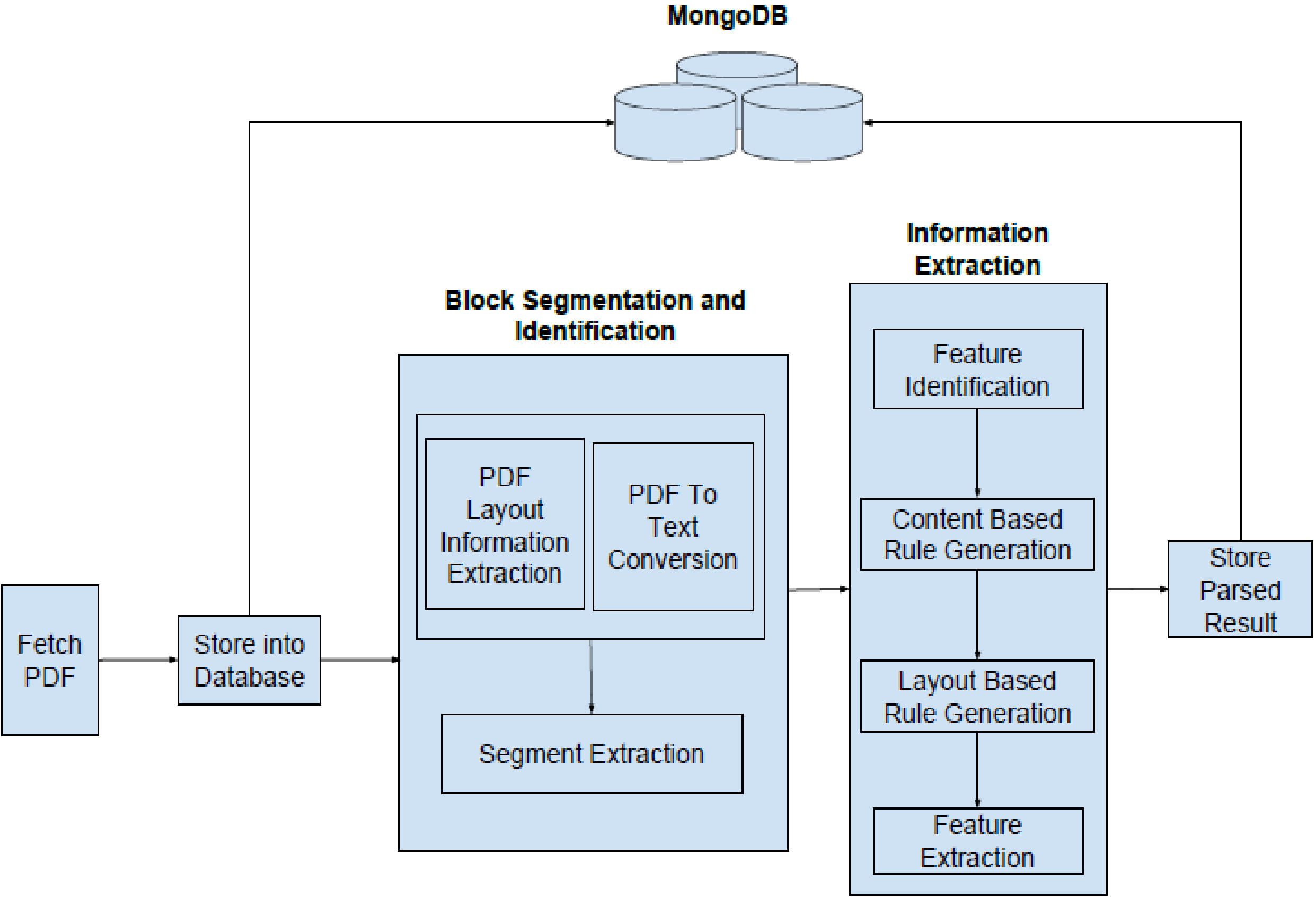
Corpus Word-List

Domain List	Definition
Segment keyword list	List of all possible segment titles like education, qualification, university, project, personal project, academic project, etc.
University list	A university list in the US
US location list	A list of US Zip Codes, State and City
Degree list	A list of aliases & abbreviations of Masters, M.S., MS, Bachelor's, etc.
Major list	A list of majors in United States
Company list	A list of all company in US

Hierarchical Layout

JINESH DHURV	Personal	Lakshmi Ravi
EDUCATION: M.S. Computer Science Rochester Institute of Technology Rochester, New York, USA Coursework: Algorithms, Distributed Systems, Machine Learning, Web Services, Big Data Analytics, Intelligent Systems B.E. Computer Engineering University of Mumbai Mumbai, India Coursework: Data Structures, Software Architecture, System Programming & Compiler Construction, System Security Expected: Dec 2017 GPA: 3.5 Aug 10-May 14 in: https://www.linkedin.com/in/jineshdhruv/ https://github.com/jineshdhruv/	Education University, Major, Degree, GPA, Year	EXPERIENCE AMAZON SOFTWARE DEVELOPMENT ENGINEER, TEST May 2014 - June 2015 Kindle Devices Chennai, India Designed & developed tools to calibrate system parameters, fluidity & multi-tasking capabilities of Kindle devices under dynamic scenarios. Responsibilities: A/B result analysis, Monthly On-call for performance tests. AMAZON SOFTWARE DEVELOPMENT ENGINEER INTERN June 2015 - Aug 2015 Fulfillment by Amazon Seattle, WA Designed & developed workflow based service execution (using AWS-SWF). Scaled the system from single host to distributed environment. Implemented & deployed failure handler, which reduced failure rate by 60%.
WORK EXPERIENCE: Data Science Intern Cause + Effect Strategy & Marketing (CESM) Rochester, New York May 17-Aug 17 Developed machine learning model to predict propensity to vote in election, which maximized relevant Ads retargeting by 25% Engineered & tested the pipeline to migrate data from relational data warehouse to big data platform to tackle scalability issues Created interactive dashboards on Science by preparing quantitative and qualitative data to assist end users' decision-making Data Strategy Intern Viacom New York City, New York Sep 16-Dec 16 Built the Data Warehouse infrastructure, and executed OLAP queries to analyze the performance of post campaign data Displayed KPIs for benchmarking successful posts using predictive modeling techniques, which increased successful posts by 35% Accomplished data-driven audience insights, campaign optimization and effectiveness measurement across integrated marketing Software Developer Intern Adventure Rochester, New York May 16-Aug 16 Led restructuring of the Ad-Builder tool to build platform independent seamless web application for clients Boosted Angular JS modules to tackle memory leaks & scalability issues that increased Ad-Builder tool responsiveness by 20% Implemented & deployed failure handler, which reduced failure rate by 40% Software Developer Imagina Analytics Mumbai, India Jun 14-Jun 15 Food Panda Built information retrieval and recommendation system for global master dish database across 42 countries with User Interface Implemented data mining techniques to classify the unstructured data & formulate association rules, which earned more \$30k AdPersonia Digital Marketing Integrated Android SDK with additional functionalities to collect & track user's in-app activity for profile building Constructed backend infrastructure with a team of 3 for Real-Time Bidding using ad exchanges & supply-side platforms	Work Company, Position, Year, Description	DISCOVER LAB ANDROID APP DEVELOPER Dec 2015 - May 2016 Roc-Reader RIT, Rochester, NY Application displays digital content for magazine pages in augmented reality. RESEARCH DOCUMENT & PATTERN RECOGNITION LAB RESEARCH MEMBER Aug 2016 - Present Machine Learning RIT, Rochester, NY Improved handwritten math symbols recognition in formulas, using contextual information from visual density histograms. Identify the layout structure of expressions by building a graph representation, visual features for layout classification and extracting final tree using Maximum Spanning Tree Algorithm. (Best Poster award) NANO COMPUTING LAB RESEARCH ASSISTANT Aug 2016 - Present Neural Networks RIT, Rochester, NY Spatial pooler (SP), an adaptive learning tool, based on brain inspired computing. Formulated & implemented, mathematical approach, to formalize SP.
TECHNICAL SKILLS: Programming Languages: Java, Python, R, SQL, C++, C# Web Technology: Angular JS, JavaScript, Node.js, HTML5, SOAP, REST Big Data ETL: MapReduce, RapidMiner, Pentaho, Weka Database & Data Visualization: MySQL, Neo4j, MongoDB, Siasse, Tableau Technology Expertise: Distributed System, Machine Learning, Data Mining, Parallel Computing, Artificial Intelligence Tools / Technology: IntelliJ IDEA, PyCharm, Eclipse, Visual Studio, Android Studio, NetBeans, JIRA, AWS, Git, Bitbucket, MySQL, Workbench	Skill Technical, Non-Technical	SKILLS ★★★★ Java • Python • Pattern Recognition Machine Learning • AWS-Simple Workflow • Likelihood • Neo4j Graph Mining ★★★★ C++ • R • JSP • Shell • Neural Networks • MySQL • Oracle • Spring MongoDB ★★★★ Android • Distributed Systems Hypothesis Testing • W/Write
PROJECTS: Instant Resume Evaluation Engine (Python, Meteor, JavaScript, MongoDB) Created a hybrid content-based & segmentation-based technique for resume parsing with unrivaled level of accuracy & efficiency Provided resume feedback about skills, vocabulary & third-party interpretation, to help job seeker for creating compelling resume Intelligent Parallel File Processing System (Java, Shell Script) Developed personalized MapReduce across the distributed network for real-time file processing based on its size Tracked processors & allocated work based on its past performance results thereby increasing file processing speed by 45% Book Insight Retrieval Android Application (Java, Android Studio) Formulated supervised machine learning algorithm for book reviews feature extraction, ranking, outlier detection & normalization Enriched the application with new user-review to train the model with new corpus for better insights API Search Engine (Java, Node.js, MongoDB, Java, RapidMiner, JavaScript) Developed a web-based query system to parse, store, manage & analyze all the APIs & Mashups listed in Programmable Web Improved the accuracy of keyword based API / Mashup discovery by 15% using machine learning algorithms User Social Profile-based Web Service Discovery (Java, Neo4j) Created a distributed system for web service discovery based on the user's social ego-centric networks Achieved satisfactory web service discovery results by considering the user's social profile and its past invocation history Text Translation & Recognition from Image (Java, JavaScript, iQuery, HTML5) Built an application that can take an image, recognize & translate text in the image & query the browser for text search results Used Microsoft Optical Character Recognition & Transcribe API for text extraction & translation from the image	Project Title, Description, Year	PROJECTS PATTERN RECOGNITION - HAND WRITTEN FORMULA Segmentation - Find potential Symbols candidates For K-linear strokes, predict chances of forming a potential symbol using geometric features, shape content features & find connected components. Classification - Label Potential symbols with a class name Using visual features and orientation based histograms classify symbols. Accuracy symbol labels - 90.0% (second best rate in class) Parsing - Find spatial relations between symbols Build expression tree by considering relations with right-most symbol. GRAPH MINING - PEOPLE TRAVEL PATTERNS Identify People's travel pattern in Four-square dataset, using frequent sub-graph mining algorithms. PUBLICATIONS A Maximum Likelihood Approach to Adaptive Learning of the Spatial Pooler, at International Joint Conference on Neural Networks, Jun-2017.
LEADERSHIP & HONORS: Completed Advanced Certification in Big Data Analytics from Rochester Institute of Technology Awarded Graduate Scholarship in Computer Science at Rochester Institute of Technology (AY 2016-17)	Other Awards, Publications, Thesis	

Workflow



Implementation

Block Segmentation and Identification

- Stored document layout information
- Converted PDF to text
- Search segment keywords using fuzzy logic
- Created bounding box near the keyword
- Stored each sentence into respective segment

Feature Extraction

- Created segment specific feature list
- Searched main feature in the respective segment
 - Used keyword-based technique
 - Used layout information like font-size, font-name and position

Segmentation Result

Segment Detection	Precision	Recall	F1-score
Personal	97%	97%	97%
Education	76%	72%	74%
Work Experience	90%	75%	81%
Project	53%	47%	50%
Skills	75%	70%	72%
Other	41%	40%	40%

Feature Extraction Result

Attributes	Precision	Recall	F1-score
Name	92%	90%	91%
Email	92%	90%	91%
Phone	92%	90%	91%
Profile Link	81%	77%	78%
Address	65%	57%	60%
University	88%	75%	80%
Major	70%	65%	67%
Date	55%	35%	42%
Degree	65%	60%	62%
GPA	60%	40%	48%

Conclusion

- The proposed system produces better results for resume parsing due to the use of content-based and layout-based techniques