Data Intake Report

Name: NLP::Resume Extraction
Report date: 17th of May 2023
Internship Batch: LISUM20
Version:1.0
Data intake by: TeamNLP
Data intake reviewer: Nilsu Bozan
Data storage location: <https://github.com/nilsubozan/Resume-Extraction.git>

Tabular data details:

Total number of observations	200
Total number of files	1
Total number of features	2
Base format of the file	.json
Size of the data	1.1 MB

NLP INTERN AT DATA GLACIER

Project: NLP::Resume Extraction

BatchCode: LISUM20
Group Name: TeamNLP

Name	Email	Country	College/Company	Specialization
Nilsu Bozan	bozannilsu@gmail.com	Turkey	Binghamton University/ Istanbul Technical University	NLP
Nishchay Vaid	Nishchay89@gmail.com	USA	Rutgers University	NLP
Anish Mitra	anishmitra9666@gmail.com	USA	Montana State University	NLP
Sukriti Macker	sm11017@nyu.edu	USA	NYU Steinhardt School of Culture	NLP

Resume extraction

Problem description

Named entity recognition is a natural language processing system that involves extracting and identifying the important information from the text while discarding superfluous information. It is an information extraction subtype. It takes unstructured text and breaks it down into person names, organizations, locations, medical codes, time expressions, quantities, monetary values, percentages, etc.

Implementing named entity recognition algorithms, we hope to reduce these difficulties involved in applying for a job and thus streamline the process in the human resources industry. We have further context on how to improve the human resources application industry below.

Business Understanding

One of the industries where named entity recognition is most important is parsing resumes with very different formats and inputting the important information into the company database. This is something that sites with widespread usage such as workday do not do accurately and is very frustrating to both job seekers and human resources professionals and also the problem that we are trying to ameliorate for them.