Resume Extraction

Group Name: Overloading

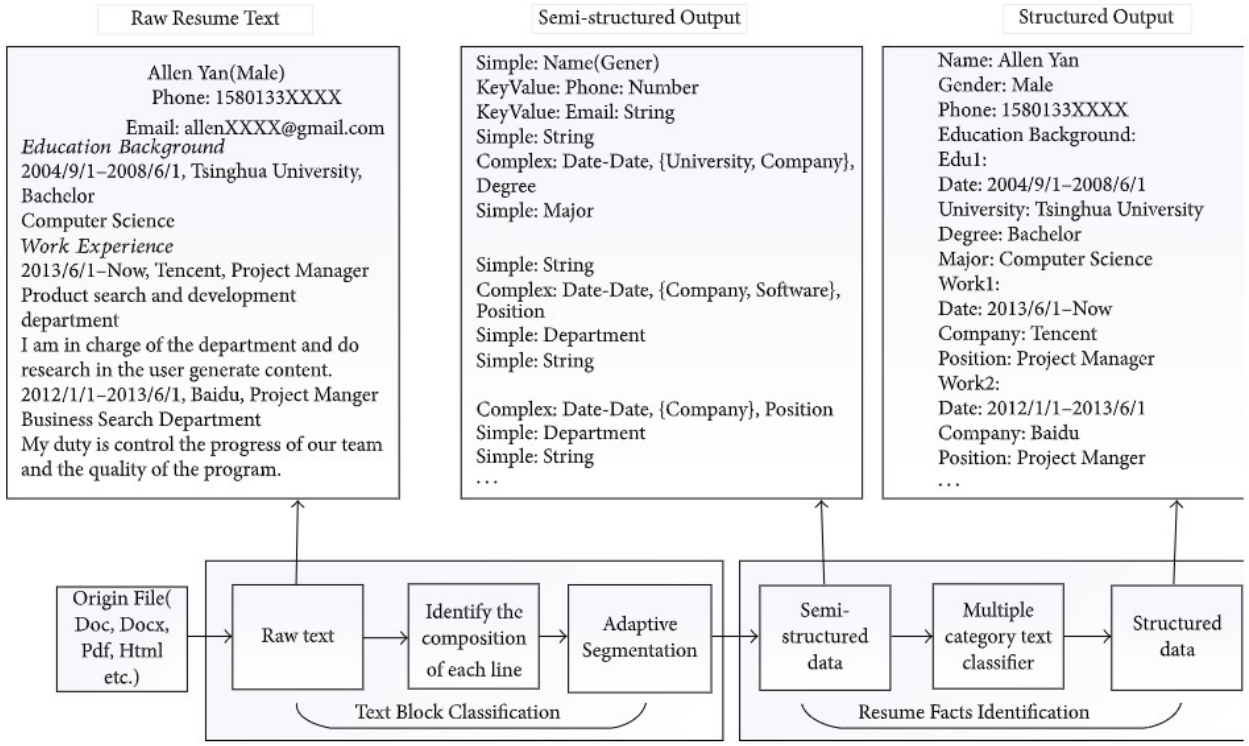
Arda Cem Özmen , ardaozmen3@gmail.com, Turkey, Istanbul Bilgi University, NLP
Buse Küçükçoban , buse_kucukcoban@hotmail.com, Turkey, Istanbul Technical University, NLP

Github Repo link: https://github.com/ardaozmen/Resume_Extraction

.

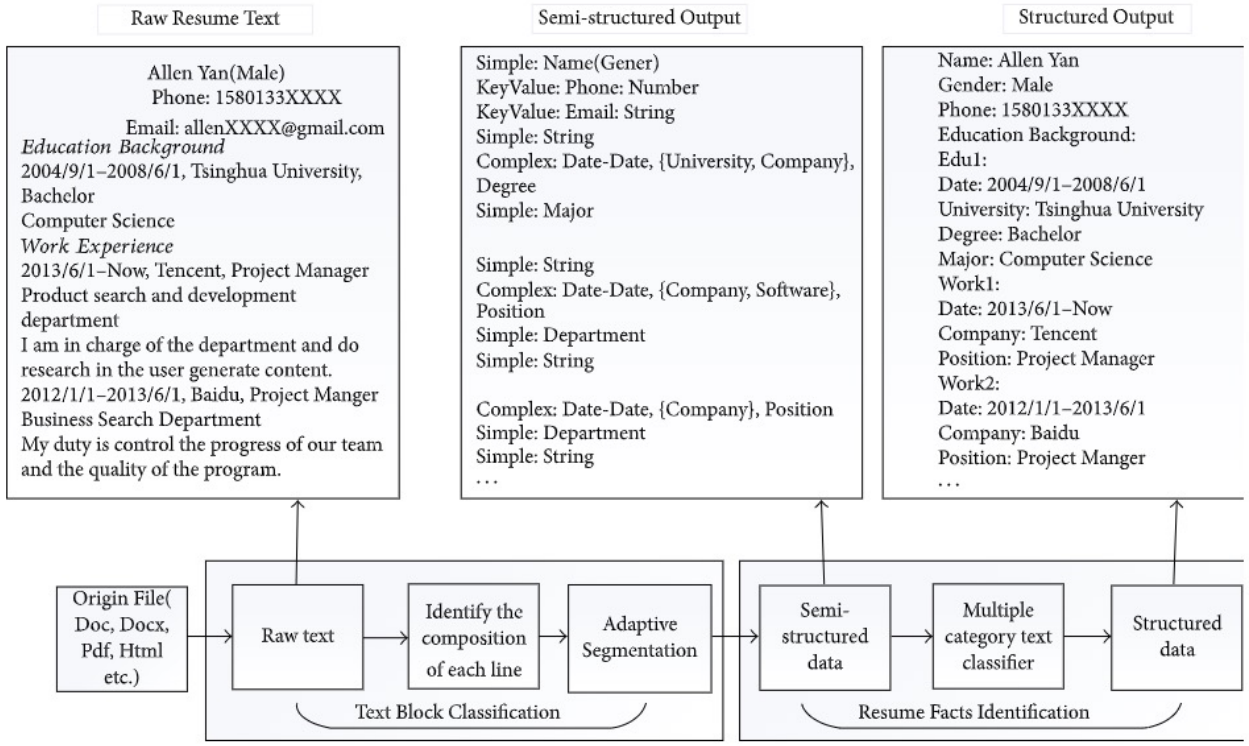
Problem Description

Resumes contain surfeit information that is not relevant for the HR/authority, and they have to manually process the resumes to shortlist the promising candidates for them. And, thus making the shortlisting task a herculean task for HR.



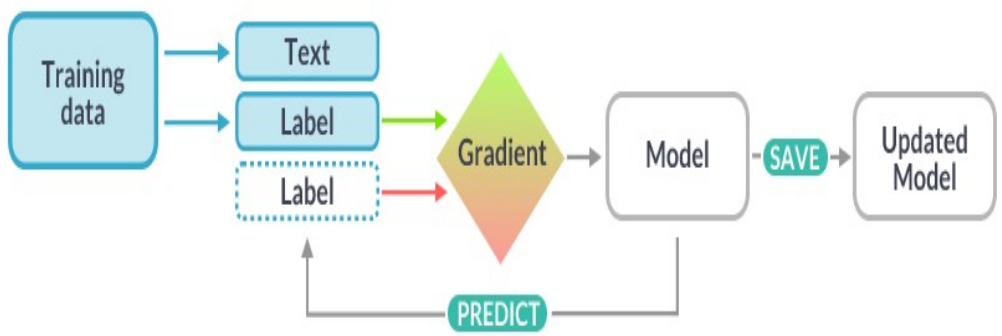
Problem Description

By making use of the NER(Named Entity Recognition) model of NLP this problem can be solved by finding and classifying the entities that are present in each resume into predefined classes such as person name, college name, academics information, relevant experiences, skill set, etc.



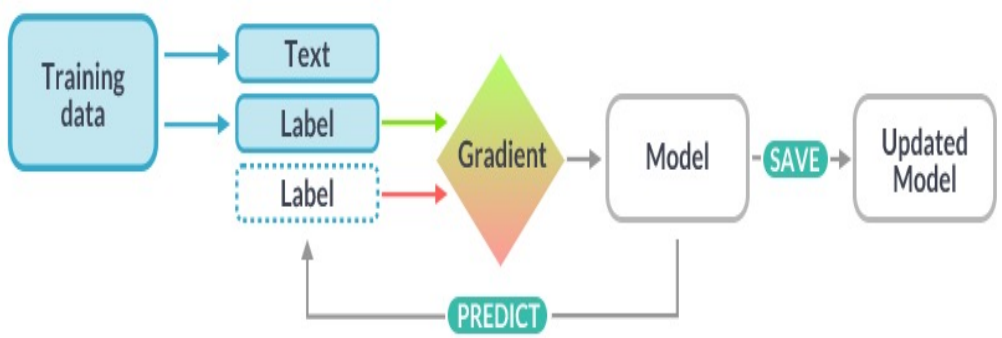
Approach

We used Spacy library, it has made advanced Natural Language Processing (NLP) much simpler in Python. Spacy provides an exceptionally efficient statistical system for named entity recognition in Python, which can assign labels to groups of tokens which are contiguous.



Approach

It provides a default model which can recognize a wide range of named or numerical entities, which include company-name, location, organization, product-name, etc to name a few. Apart from these default entities, Spacy enables the addition of arbitrary classes to the entity-recognition model, by training the model to update it with newer trained examples.



Exploratory Data Analysis

Data Glacier
Your Deep Learning Partner

We got an average result with the available data.
Increasing the number of data allows us to get more serious results.

	content	annotation	Companies worked at	Skills	Graduation Year	College Name	Degree	Designation	Email Address	Location	Name	Years of Experienc
63	Sivaganesh Selvakumar\nDevOps Consultant with ...	[{'label': ['Skills'], 'points': [{'start': 23...	[{'start': 1915, 'end': 1930, 'text': 'Infosys...	[{'start': 2349, 'end': 2641, 'text': 'TECHNIC...	NaN	[{'start': 2257, 'end': 2290, 'text': 'Saranat...	[{'start': 1990, 'end': 2043, 'text': 'Bachelo...	[{'start': 1895, 'end': 1912, 'text': 'Technol...	NaN	[{'start': 1934, 'end': 1940, 'text': 'Chennai...	[{'start': 0, 'end': 20, 'text': 'Sivaganesh S...	NaN
59	Sharan Adla\n- Email me on Indeed: indeed.com/...	[{'label': ['College Name'], 'points': [{'star...	NaN	NaN	NaN	[{'start': 2421, 'end': 2449, 'text': 'Vignan ...	[{'start': 2416, 'end': 2418, 'text': 'SSC'}, ...	[{'start': 1092, 'end': 1115, 'text': 'Special...	[{'start': 1801, 'end': 1841, 'text': 'indeed....	NaN	[{'start': 0, 'end': 10, 'text': 'Sharan Adla'}]	[{'start': 8('end': 90, 'text': '4yrs.'}]
64	Snehal Jadhav\nMumbai, Maharashtra - Email me ...	[{'label': ['Companies worked at'], 'points': ...	[{'start': 1794, 'end': 1799, 'text': 'Cisco '...	[{'start': 415, 'end': 1909, 'text': 'Architec...	NaN	[{'start': 387, 'end': 404, 'text': 'Shivaji U...	[{'start': 347, 'end': 384, 'text': 'B.E in El...	[{'start': 118, 'end': 137, 'text': 'L1 networ...	[{'start': 1077, 'end': 1119, 'text': 'indeed....	[{'start': 148, 'end': 153, 'text': 'Mumbai'},...	[{'start': 0, 'end': 13, 'text': 'Snehal Jadha...	NaN

Recommendations

Data Glacier
Your Deep Learning Partner

Training should be provided with larger data. Correct data must be obtained. In addition, models should continue to be produced with different approaches.

Thank You