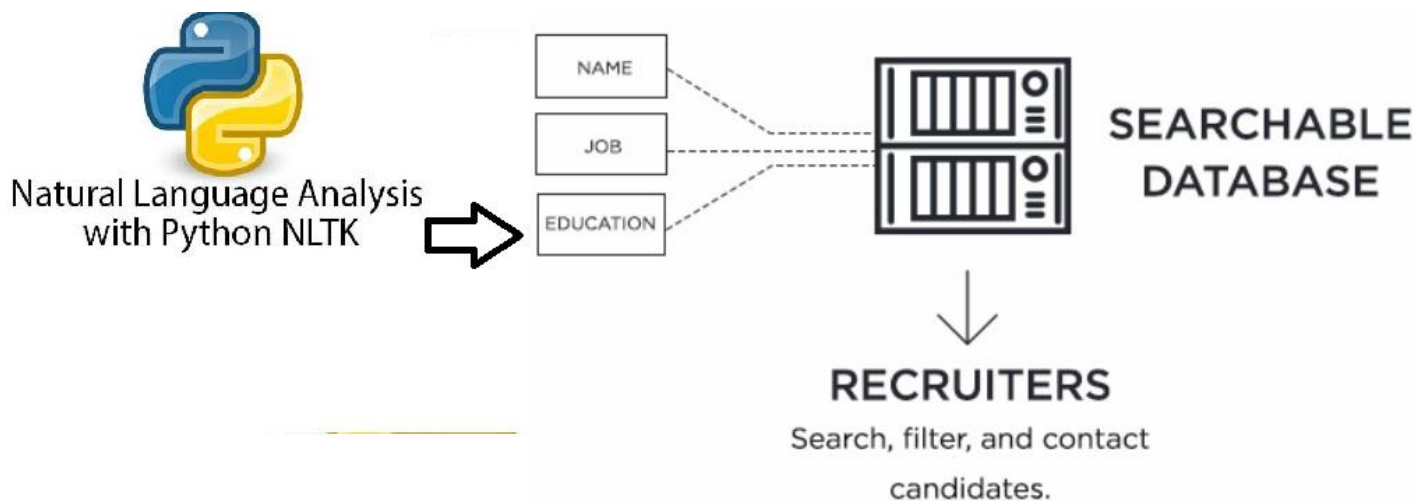


Automatic Resume Screening using Python Natural Language Processing Toolkit (NLTK) – Part I

Challenges for Recruiter in Candidate Resume Filtering: Going through candidates resume takes a lot of time, especially when you are dealing with hundreds of openings with multiple candidates for each opening. Various surveys reveal that job postings receive an average of 250 applications, so this kind of resume screening manually turns out to be a daunting task. Resume screening software lets recruiters quickly shortlist candidates that are the best fit for the job. However, there is a significant and recurring cost associated with resume screening software makes difficult to convince business community to buy such software. Other parameters like lack of flexibility with pre-packaged software makes it more complicated. Customization software in turn adds time and cost overhead.

Easing Recruiters job by using Python NLP toolkit: Python text mining using Natural Language Processing Toolkit seemed to be the solution to getting the important data without actually reading each and every resume. Using Python NLP toolkit can ease HR jobs in filtering resume with certain surrounding words, patterns, range and based sentence constructions.



Keyword Based Search: In the below example I used Python NLTK concordance function. NLTK provides the function `concordance()` to locate and print series of phrases that contain the keyword. However, the function only prints the output. The user is not able to save the results for further processing unless redirect the stdout. In the below I used `concordance` in candidate profile to locate profile with certain phrases and keywords. Below function will emulate the concordance function and return the list of phrases for further processing. It uses the NLTK concordance Index which keeps track of the keyword index in the passage/text and retrieves the surrounding words. This is a tiny example but demonstrate NLTK capabilities. Consider this example as starting point and we would build complete resume screen code by combining small pieces of codes in certain order. The code reads the resume and searches for multiple occurrence of words and the output can be redirected filtered folder. This ease manual resume processing job, the same code can be rewritten keeping multiple files as input with batch processing.

Keyword Based Search

Step#1 : Import NLTK

```
import nltk
```

Step#2 : Open File in Read Only Mode

```
file = open('Path\\ Resume.txt')
```

Step#3 : Read File

```
read_file = file.read()
```

Step#4 : Tokenizing Word

```
text = nltk.Text(nltk.word_tokenize(read_file))
```

Step#5 : Use concordance function

```
match = text.concordance('Java')
```

Python NLTK : Keywords Based Search

URL : <https://youtu.be/ickMjpJj3WY>



A concordance view shows us every occurrence of a given word, together with some context.