

Calculating the Match Rate of Twitter JobAds with Job Seekers' CVs

The project aims to inform job seekers with Tweets related to their profiles in real-time. In this way, we can let people have the news immediately without them having to follow each and every company one by one. In this project, we focused on software engineers.

To achieve this goal, we divided our work into three parts,

1) Get and process the dataset (Job_type and Job_seeker)

we first decided on the keywords that are essential for a software engineer by going through many CVs and job ads related to software engineers. Those collected skills (keywords) have been saved into an excel file manually and then saved into the Cassandra table(called job_type) by using Python. Then, we did web-scraping by using a Python library (BeautifulSoup) to get the CV-detail-page URL from the main search page [2] and then navigate to this URL. In each CV-URL, we searched on this page's source to match the essential skills and save a (0,1) list for each user in the job_seeker table in Cassandra. 200 records/CVs per job type were accessible (we created 3 job types: Software Engineer, Data Analyst, Systems Engineer and we only used Software Engineer since the process of the other two is the same)

```
# order by idx column
df_job_type.sort_values(by=['idx'])
```

	idx	data_analyst	software_engineer	systems_engineer
40	1	accounting	ad tech	2g
49	2	airflow	ajax	3com
134	3	aws	algorithm	3g
52	4	azure data lake	android	4g lte
12	5	big data	apache druid	5gnr
...
11	132		xml	
124	133		xml parser	
28	134		xpath	
68	135		xquery	
125	136		xslt	

136 rows x 4 columns

```
q_all_job_seekers = "SELECT * from job_seeker"
all_job_seekers = pd.DataFrame(session.execute(q_all_job_seekers)).sort_values(by=['user_id'])
all_job_seekers
```

	user_id	job_title	skillset
154	1	software_engineer	[1, 1, 1, 1, 0, 1, 0, 1, 1, 1, 1, 1, 1, 1, 0, 0, ...]
188	2	software_engineer	[0, 1, 0, 1, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, ...]
595	3	software_engineer	[0, 1, 0, 1, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, ...]
201	4	software_engineer	[0, 1, 0, 1, 0, 1, 0, 0, 1, 1, 0, 0, 0, 0, 0, ...]
46	5	software_engineer	[0, 1, 1, 1, 0, 1, 0, 1, 1, 1, 0, 1, 0, 1, 0, 0, ...]
...
277	597	systems_engineer	[1, 1, 1, 0, 0, 1, 1, 1, 1, 1, 0, 0, 1, 1, 1, ...]
434	598	systems_engineer	[1, 1, 1, 0, 0, 1, 1, 1, 1, 1, 0, 0, 1, 1, 1, ...]
164	599	systems_engineer	[1, 1, 1, 0, 0, 1, 1, 1, 1, 1, 0, 0, 1, 1, 1, ...]
483	600	systems_engineer	[1, 1, 1, 0, 0, 1, 1, 1, 1, 1, 0, 0, 1, 1, 1, ...]
510	601	systems_engineer	[1, 1, 1, 0, 0, 1, 1, 1, 1, 1, 0, 0, 1, 1, 1, ...]

601 rows x 3 columns

Connected to Test Cluster at 127.0.0.1:9042. [cqlsh 5.0.1 Cassandra 3.11.2 CQL spec 3.4.4 Native protocol v4] Use HELP for help. cqlsh> use jobsearch_keyspace; cqlsh:jobsearch_keyspace> select * from job_type;			
idx	data_analyst	software_engineer	systems_engineer
23		CircleCI	CICS
114	EDW	T-SQL	
53	PHP	JAVA	MBSE
110		Struts	
91		React	Y2K
128		Weblogic	
117		TSQL	
120		VBA	
55	Power BI	javascript	Network Protocols
129		Websphere	
33	J-Review	database	Firewall
132		XML	
5	Big Data	Apache Druid	SGNR
28	Gaming	Crystal Reports	Disaster Recovery
105		SQL	
123		Visual Studio	
42	MapReduce	GCP	IDMS

2) Kafka input data (Job_ads) by Twitter Streaming API

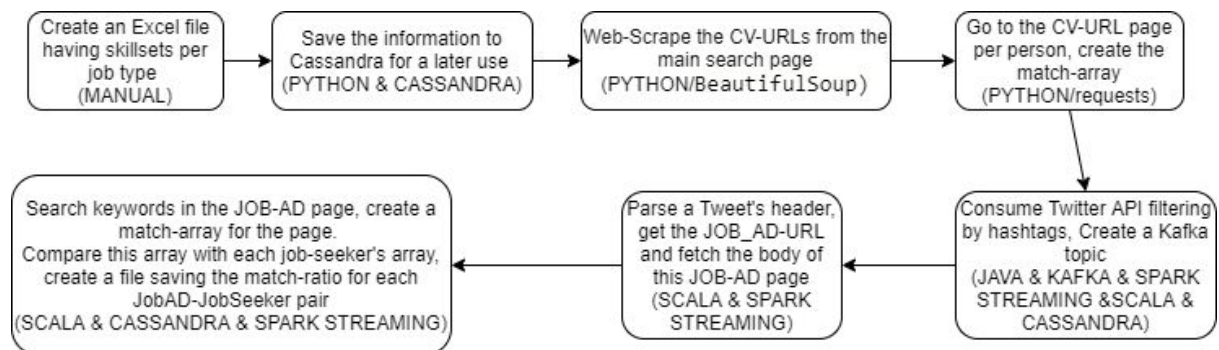
We built a Producer that will fetch the latest tweets in real-time on "software_engineer" and then send them to Kafka. To begin with, we generated Twitter API [1] Keys in order to get access to Twitter API. Then, we created a Maven Project and added Kafka, Twitter, and Gson dependencies in pom.xml. After basic configurations, we initialized our Twitter client and passkey, secrets, and token for authentication as well as the configuration of the producer. Then we passed the term (#SoftwareEngineer) which we want to track. Lastly, we made a connection to Twitter using the client and once we got a tweet, we sent it to Kafka using the producer.

3) Spark Streaming Consumer Process

At the final stage, in order to calculate the similarity rate between job_ads and job_seeker_cvs, we need two (0,1)lists. Firstly, we connect Spark with Cassandra to read the job_type skills and job_seeker skill list (list1).

Secondly, connect to Kafka to get the job_ad_URL, use http request to get the body of this URL, create another 0-1 list (list2) for the job-ad by checking if job-ad content contains the job_type skills. Then, get the common skills between the job-ad and the job seeker. Lastly, write the results (job-URL, userID, common match, user-match, ad match, ratio values) into a local file. (P.S. We needed to parse the header of the Twitter URL to get the final URLs complete string since Twitter provides a short version of the URL and with this URL, we can't get the body of the final page.)

WORK SUMMARY and USED TECHNOLOGIES



Result

Display the similarity rate between job_ads and job_seekers
(job-URL, userID, common match, user-match, ad match, ratio values)

https://www.monster.com/jobs/search?q=software-engineer&intcid=skr_navigation_nhps0_searchMain&jobid=218186617	24	4	122	25	0.2
https://www.monster.com/jobs/search?q=software-engineer&intcid=skr_navigation_nhps0_searchMain&jobid=218186617	81	4	122	25	0.16
https://www.monster.com/jobs/search?q=software-engineer&intcid=skr_navigation_nhps0_searchMain&jobid=218186617	153	5	126	25	0.2
https://www.monster.com/jobs/search?q=software-engineer&intcid=skr_navigation_nhps0_searchMain&jobid=218186617	126	5	125	25	0.2
https://www.monster.com/jobs/search?q=software-engineer&intcid=skr_navigation_nhps0_searchMain&jobid=218186617	162	5	126	25	0.2
https://www.monster.com/jobs/search?q=software-engineer&intcid=skr_navigation_nhps0_searchMain&jobid=218186617	27	4	128	25	0.16
https://www.monster.com/jobs/search?q=software-engineer&intcid=skr_navigation_nhps0_searchMain&jobid=218186617	36	4	121	25	0.16
https://www.monster.com/jobs/search?q=software-engineer&intcid=skr_navigation_nhps0_searchMain&jobid=218186617	108	5	125	25	0.2
https://www.monster.com/jobs/search?q=software-engineer&intcid=skr_navigation_nhps0_searchMain&jobid=218186617	135	5	125	25	0.2
https://www.monster.com/jobs/search?q=software-engineer&intcid=skr_navigation_nhps0_searchMain&jobid=218186617	117	5	125	25	0.2
https://www.monster.com/jobs/search?q=software-engineer&intcid=skr_navigation_nhps0_searchMain&jobid=218186617	9	4	108	25	0.16
https://www.monster.com/jobs/search?q=software-engineer&intcid=skr_navigation_nhps0_searchMain&jobid=218186617	18	4	117	25	0.16
https://www.monster.com/jobs/search?q=software-engineer&intcid=skr_navigation_nhps0_searchMain&jobid=218186617	48	4	122	25	0.16
https://www.monster.com/jobs/search?q=software-engineer&intcid=skr_navigation_nhps0_searchMain&jobid=218186617	57	4	122	25	0.16
https://www.monster.com/jobs/search?q=software-engineer&intcid=skr_navigation_nhps0_searchMain&jobid=218186617	21	4	118	25	0.16
https://www.monster.com/jobs/search?q=software-engineer&intcid=skr_navigation_nhps0_searchMain&jobid=218186617	3	0	39	25	0.0
https://www.monster.com/jobs/search?q=software-engineer&intcid=skr_navigation_nhps0_searchMain&jobid=218186617	12	4	112	25	0.16
https://www.monster.com/jobs/search?q=software-engineer&intcid=skr_navigation_nhps0_searchMain&jobid=218186617	84	4	123	25	0.16
https://www.monster.com/jobs/search?q=software-engineer&intcid=skr_navigation_nhps0_searchMain&jobid=218186617	93	5	125	25	0.2
https://www.monster.com/jobs/search?q=software-engineer&intcid=skr_navigation_nhps0_searchMain&jobid=218186617	102	5	125	25	0.2
https://www.monster.com/jobs/search?q=software-engineer&intcid=skr_navigation_nhps0_searchMain&jobid=218186617	75	4	123	25	0.16
https://www.monster.com/jobs/search?q=software-engineer&intcid=skr_navigation_nhps0_searchMain&jobid=218186617	30	4	120	25	0.16
https://www.monster.com/jobs/search?q=software-engineer&intcid=skr_navigation_nhps0_searchMain&jobid=218186617	39	4	121	25	0.16
https://www.monster.com/jobs/search?q=software-engineer&intcid=skr_navigation_nhps0_searchMain&jobid=218186617	111	5	125	25	0.2
https://www.monster.com/jobs/search?q=software-engineer&intcid=skr_navigation_nhps0_searchMain&jobid=218186617	66	4	123	25	0.16
https://www.monster.com/jobs/search?q=software-engineer&intcid=skr_navigation_nhps0_searchMain&jobid=218186617	15	4	114	25	0.16
https://www.monster.com/jobs/search?q=software-engineer&intcid=skr_navigation_nhps0_searchMain&jobid=218186617	42	4	121	25	0.16
https://www.monster.com/jobs/search?q=software-engineer&intcid=skr_navigation_nhps0_searchMain&jobid=218186617	51	4	122	25	0.16
https://www.monster.com/jobs/search?q=software-engineer&intcid=skr_navigation_nhps0_searchMain&jobid=218186617	24	4	120	25	0.16
https://www.monster.com/jobs/search?q=software-engineer&intcid=skr_navigation_nhps0_searchMain&jobid=218186617	33	4	121	25	0.16
https://www.monster.com/jobs/search?q=software-engineer&intcid=skr_navigation_nhps0_searchMain&jobid=218186617	6	4	85	25	0.16
https://www.monster.com/jobs/search?q=software-engineer&intcid=skr_navigation_nhps0_searchMain&jobid=218186617	179	5	126	25	0.2
https://www.monster.com/jobs/search?q=software-engineer&intcid=skr_navigation_nhps0_searchMain&jobid=218186617	152	5	126	25	0.2
https://www.monster.com/jobs/search?q=software-engineer&intcid=skr_navigation_nhps0_searchMain&jobid=218186617	188	5	126	25	0.2
https://www.monster.com/jobs/search?q=software-engineer&intcid=skr_navigation_nhps0_searchMain&jobid=218186617	170	5	126	25	0.2
https://www.monster.com/jobs/search?q=software-engineer&intcid=skr_navigation_nhps0_searchMain&jobid=218186617	143	5	125	25	0.2
https://www.monster.com/jobs/search?q=software-engineer&intcid=skr_navigation_nhps0_searchMain&jobid=218186617	161	5	126	25	0.2
https://www.monster.com/jobs/search?q=software-engineer&intcid=skr_navigation_nhps0_searchMain&jobid=218186617	197	5	127	25	0.2
https://www.monster.com/jobs/search?q=software-engineer&intcid=skr_navigation_nhps0_searchMain&jobid=03496e95-b3ed-45a7-a066-7e49f78b63cb	137	5	125	35	0.14285715
https://www.monster.com/jobs/search?q=software-engineer&intcid=skr_navigation_nhps0_searchMain&jobid=03496e95-b3ed-45a7-a066-7e49f78b63cb	146	5	125	35	0.14285715
https://www.monster.com/jobs/search?q=software-engineer&intcid=skr_navigation_nhps0_searchMain&jobid=03496e95-b3ed-45a7-a066-7e49f78b63cb	92	4	124	35	0.14285715
https://www.monster.com/jobs/search?q=software-engineer&intcid=skr_navigation_nhps0_searchMain&jobid=03496e95-b3ed-45a7-a066-7e49f78b63cb	101	5	125	35	0.14285715
https://www.monster.com/jobs/search?q=software-engineer&intcid=skr_navigation_nhps0_searchMain&jobid=03496e95-b3ed-45a7-a066-7e49f78b63cb	200	5	127	35	0.14285715
https://www.monster.com/jobs/search?q=software-engineer&intcid=skr_navigation_nhps0_searchMain&jobid=03496e95-b3ed-45a7-a066-7e49f78b63cb	173	5	126	35	0.14285715
https://www.monster.com/jobs/search?q=software-engineer&intcid=skr_navigation_nhps0_searchMain&jobid=03496e95-b3ed-45a7-a066-7e49f78b63cb	182	5	126	35	0.14285715
https://www.monster.com/jobs/search?q=software-engineer&intcid=skr_navigation_nhps0_searchMain&jobid=03496e95-b3ed-45a7-a066-7e49f78b63cb	191	5	126	35	0.14285715
https://www.monster.com/jobs/search?q=software-engineer&intcid=skr_navigation_nhps0_searchMain&jobid=03496e95-b3ed-45a7-a066-7e49f78b63cb	83	4	123	35	0.14285715
https://www.monster.com/jobs/search?q=software-engineer&intcid=skr_navigation_nhps0_searchMain&jobid=03496e95-b3ed-45a7-a066-7e49f78b63cb	155	5	126	35	0.14285715
https://www.monster.com/jobs/search?q=software-engineer&intcid=skr_navigation_nhps0_searchMain&jobid=03496e95-b3ed-45a7-a066-7e49f78b63cb	164	5	126	35	0.14285715
https://www.monster.com/jobs/search?q=software-engineer&intcid=skr_navigation_nhps0_searchMain&jobid=03496e95-b3ed-45a7-a066-7e49f78b63cb	110	5	125	35	0.14285715
https://www.monster.com/jobs/search?q=software-engineer&intcid=skr_navigation_nhps0_searchMain&jobid=03496e95-b3ed-45a7-a066-7e49f78b63cb	119	5	125	35	0.14285715
https://www.monster.com/jobs/search?q=software-engineer&intcid=skr_navigation_nhps0_searchMain&jobid=03496e95-b3ed-45a7-a066-7e49f78b63cb	128	5	125	35	0.14285715
https://www.monster.com/jobs/search?q=software-engineer&intcid=skr_navigation_nhps0_searchMain&jobid=03496e95-b3ed-45a7-a066-7e49f78b63cb	23	4	118	35	0.14285715
https://www.monster.com/jobs/search?q=software-engineer&intcid=skr_navigation_nhps0_searchMain&jobid=03496e95-b3ed-45a7-a066-7e49f78b63cb	95	5	125	35	0.14285715
https://www.monster.com/jobs/search?q=software-engineer&intcid=skr_navigation_nhps0_searchMain&jobid=03496e95-b3ed-45a7-a066-7e49f78b63cb	104	5	125	35	0.14285715
https://www.monster.com/jobs/search?q=software-engineer&intcid=skr_navigation_nhps0_searchMain&jobid=03496e95-b3ed-45a7-a066-7e49f78b63cb	131	5	125	35	0.14285715
https://www.monster.com/jobs/search?q=software-engineer&intcid=skr_navigation_nhps0_searchMain&jobid=03496e95-b3ed-45a7-a066-7e49f78b63cb	122	5	125	35	0.14285715
https://www.monster.com/jobs/search?q=software-engineer&intcid=skr_navigation_nhps0_searchMain&jobid=03496e95-b3ed-45a7-a066-7e49f78b63cb	194	5	127	35	0.14285715
https://www.monster.com/jobs/search?q=software-engineer&intcid=skr_navigation_nhps0_searchMain&jobid=03496e95-b3ed-45a7-a066-7e49f78b63cb	176	5	126	35	0.14285715
https://www.monster.com/jobs/search?q=software-engineer&intcid=skr_navigation_nhps0_searchMain&jobid=03496e95-b3ed-45a7-a066-7e49f78b63cb	77	4	123	35	0.14285715

How to Run the Code

Create a Keyspace:

```
create keyspace jobsearch keyspace with replication = {'class': 'SimpleStrategy', 'replication_factor': 1};
```

Create Related Tables:

```
create table job_type (Idx int, Software_Engineer text, Data_Analyst text, Systems_Engineer text, primary key (Idx));
```

```
create table job_seeker (User_ID int, Job_Title text, Skillset list<int>, primary key (User_ID));
```

Run Zookeeper:

```
$KAFKA_HOME/bin/zookeeper-server-start.sh  
$KAFKA_HOME/config/zookeeper.properties
```

Run Kafka:

```
$KAFKA_HOME/bin/kafka-server-start.sh  
$KAFKA_HOME/config/server.properties
```

Create a Topic

```
$KAFKA_HOME/bin/kafka-topics.sh --create --zookeeper localhost:2181  
--replication-factor 1 --partitions 1 --topic Itjob-tweets
```

Run the JAVA Code**Run the Scala Part**

```
sbt run
```

References

- [1] Twitter API <https://developer.twitter.com/en/docs/twitter-api/getting-started/guide>
- [2] <https://www.livecareer.com/resume-search/>