



SDAIA
الهيئة السعودية للبيانات
والذكاء الاصطناعي
Saudi Data & AI Authority

Stack Overflow Recommendation Question (NLP)

Ahmed Al-Muaybid
Abdulrahman Al-Rifae



stackoverflow

Agenda

Points to discuss:

- Introduction
- DATA
- EDA
- Topic Modeling
- Recommendation System
- Featured Work
- Conclusion

Introduction



Why Stack Overflow is important?

As any data since we all used the magical website called stack overflow those who don't know stack Overflow. Stack Overflow provides one of the largest question and answers platforms for programmers. Users post questions and their fellow peers try to provide solutions in the best way possible.



Introduction

In our project, we want to help the people ask the **right question in the correct way**. we will achieve that by using **natural language process (NLP)** to find the best format for the questions. In addition, a **recommendation system** can be built to help the user write the **optimal question** format that can help fellow peers understand the question

DATA

01 +60k Question

02 Kaggle

03 Multi-Language



A magnifying glass is positioned over a section of a mathematical table, focusing on the values for $\sinh x$ and $\cosh x$ when x is approximately 2.3. The table includes columns for e^x , e^{-x} , $\sinh x$, $\cosh x$, and $\tanh x$. The values are listed for x ranging from 1.0000 to 49.402.

e^x	e^{-x}	$\sinh x$	$\cosh x$	$\tanh x$
1.0000	1.0000	.00000	1.0000	.0000
1.1052	.90484	.10017	1.0050	.0998
1.2214	.81873	.20134	1.0201	.1959
1.3499	.74082	.30452	1.0453	.2913
1.4918	.67032	.41075	1.0811	.3799
1.6487	.60653	.52110	1.1276	.4621
1.8221	.54881	.63665	1.1855	.5374
2.0138	.49659	.75858	1.2552	.6062
2.2255	.44933	.88801	1.3374	.6692
2.4596	.40657	1.0247	1.4331	.7174
2.7183	.36926	1.1752	1.5431	.7616
3.0042	.33690	1.3356	1.6788	.8007
3.3201	.30851	1.5095	1.8414	.8353
3.6693	.28350	1.6984	1.9321	.8658
4.0552	.26137	1.9043	2.1517	.8828
4.4817	.24133	2.1293	2.3521	.8979
4.9530	.22290	2.3756	2.5777	.9113
5.4739	.20668	2.6456	2.8288	.9234
6.0496	.19230	2.9422	3.1077	.9344
6.6859	.17957	3.2682	3.4171	.9445
7.3891	.16834	3.6269	3.7611	.9538
8.1662	.15843	4.0219	4.1447	.9624
9.0250	.14972	4.4571	4.5729	.9703
9.9742	.14217	4.9370	5.0519	.9776
11.023	.13569	5.4669	5.5869	.9843
12.182	.13012	6.0472	6.1823	.9896
13.464	.12537	6.6807	6.7690	.9939
14.880	.12136	7.3703	7.4735	.9973
16.445	.11803	8.1219	8.2527	.9993
18.174	.11533	8.9416	9.1146	.9999
20.086	.11321	9.8367	10.068	.9999
22.198	.11165	10.815	11.122	.9999
24.533	.11054	11.876	12.287	.9999
27.113	.10977	13.028	13.575	.9999
29.964	.10932	14.281	14.999	.9999
33.115	.10917	15.643	16.573	.9999
36.598	.10918	17.115	18.313	.9999
40.447	.10932	18.707	20.236	.9999
44.701	.10957	20.429	22.362	.9999
49.402	.10991	24.691	24.711	.9999

EDA

Drop

Drop unneeded
columns

Function

Create helper
functions to help
us clean the
text(Lem)

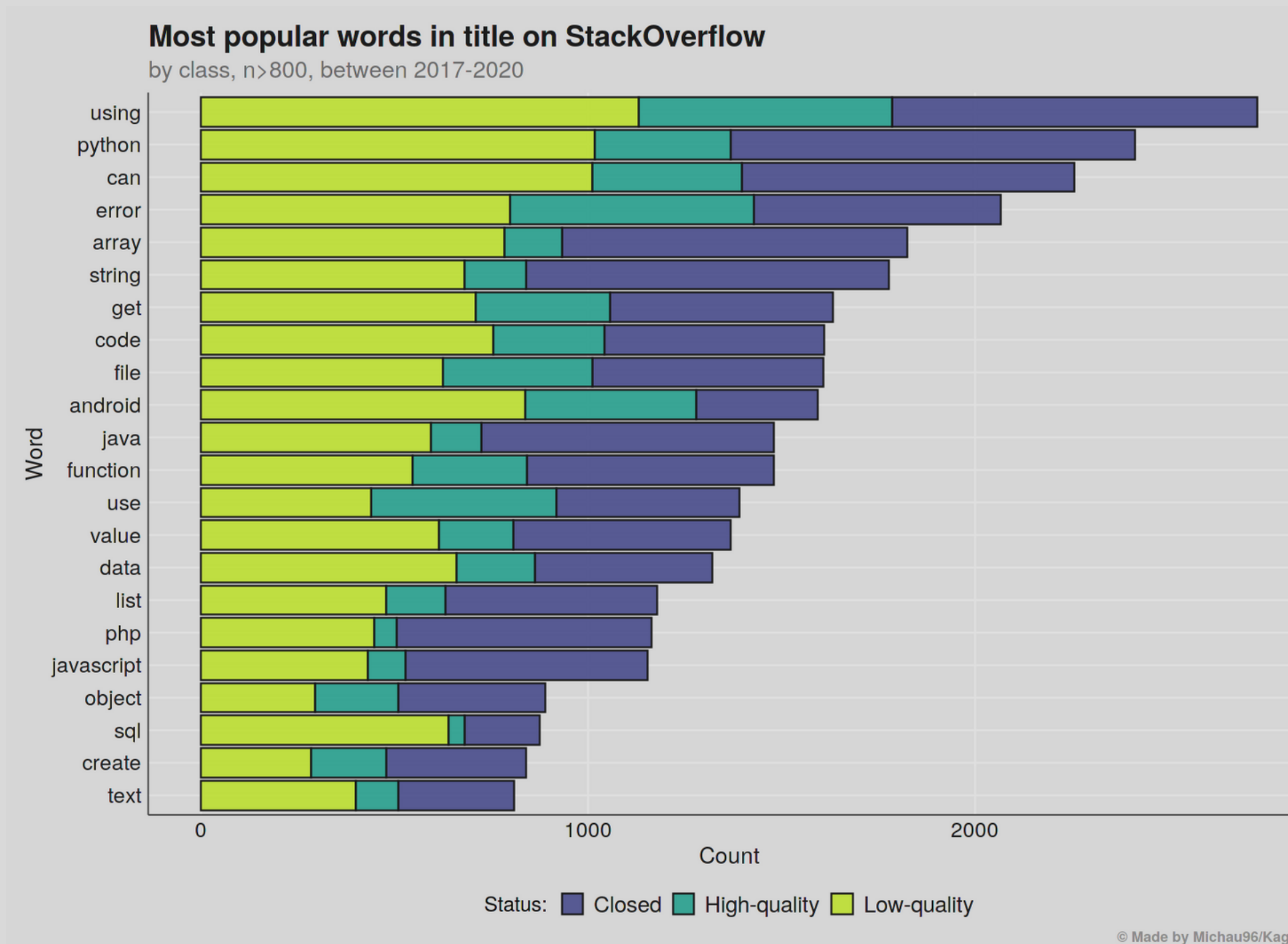
Words

removing the words
based on the frequency,
stop word,
LowerCase

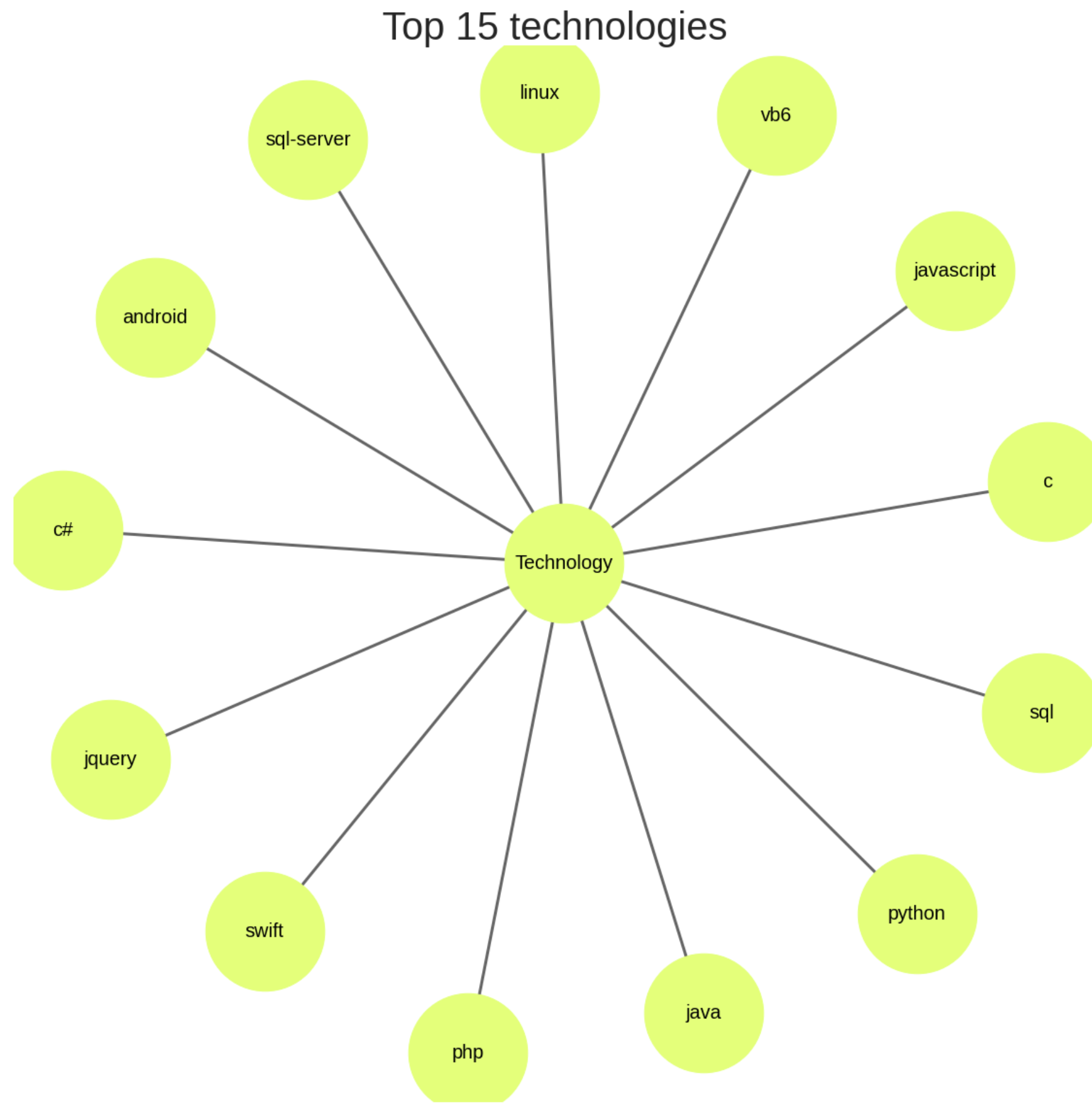
Language

Removing other
languages

EDA



EDA



Topic Modeling



Vectorization

TF-IDF with LSA,
CountVectorizer works
best with CorEX



CorEX

Select 3 topics



Topics

- Web Application
- CSS/Styling
- Related to data

Topics

Topics	Words
Topic 1 (Web Application)	(react,native), angular, page, application, (net, core)
Topic 2 (CSS/Styling)	(image,button, color, background), (studio, visual)
Topic 3 (Data base)	(class,mysql,query), (column, execl, table)



Topic Modeling for Recommendation System

Topic Modeling for Recommendation System

01 CorEx

The use of hyperparameter (anchor) to tune the model

02 Anchor Word

- GroupBy the rating quality
- Find most repeated words
- Create the anchor words list

03 Topics

- High Quality
- Low Quality Edit
- Low Quality Colsed

Recommendation System

1 Combinations

- Combine list of questions and the topics

2 Cosine Similarity

- Compare the question asked with the topic

3 Result

- Rating quality based on the question asked

Recommendation System

(Example)

Question	Result
why numpay dont work	LQ_EDIT
how to install react in windows	HQ
how to add row	LQ_CLOSED

Feature Work



Deploy the Model



**Recommend Alternative
Question**



Spell-Check

CONCLUSION

- 01 The most Searched Technology is JavaScript
- 02 Text fall in three topics normally





Thank you for listening
any questions?