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.

Inrtoduction

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

EDA

Drop

Drop unneeded columns

Function

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

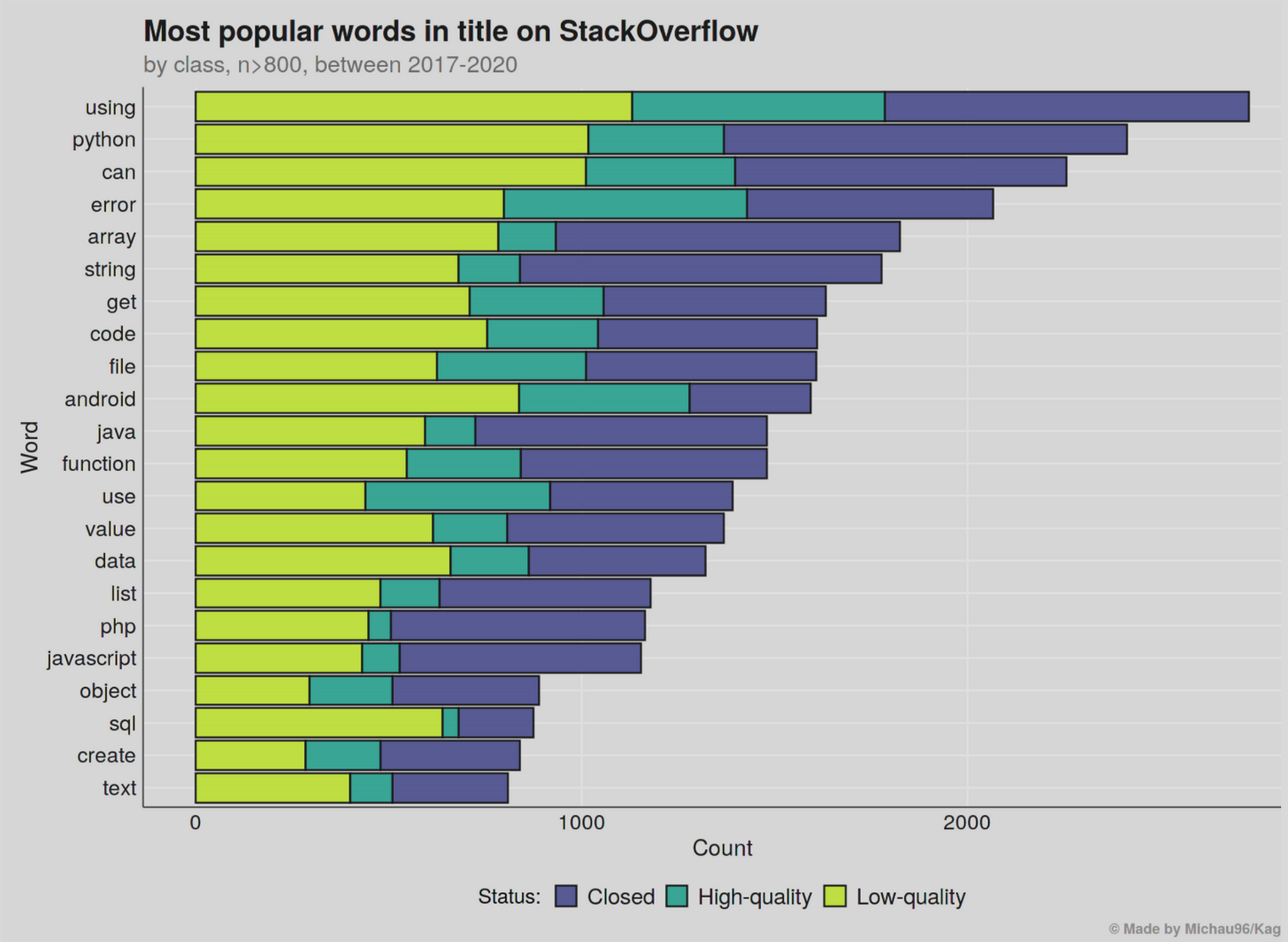
Words

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

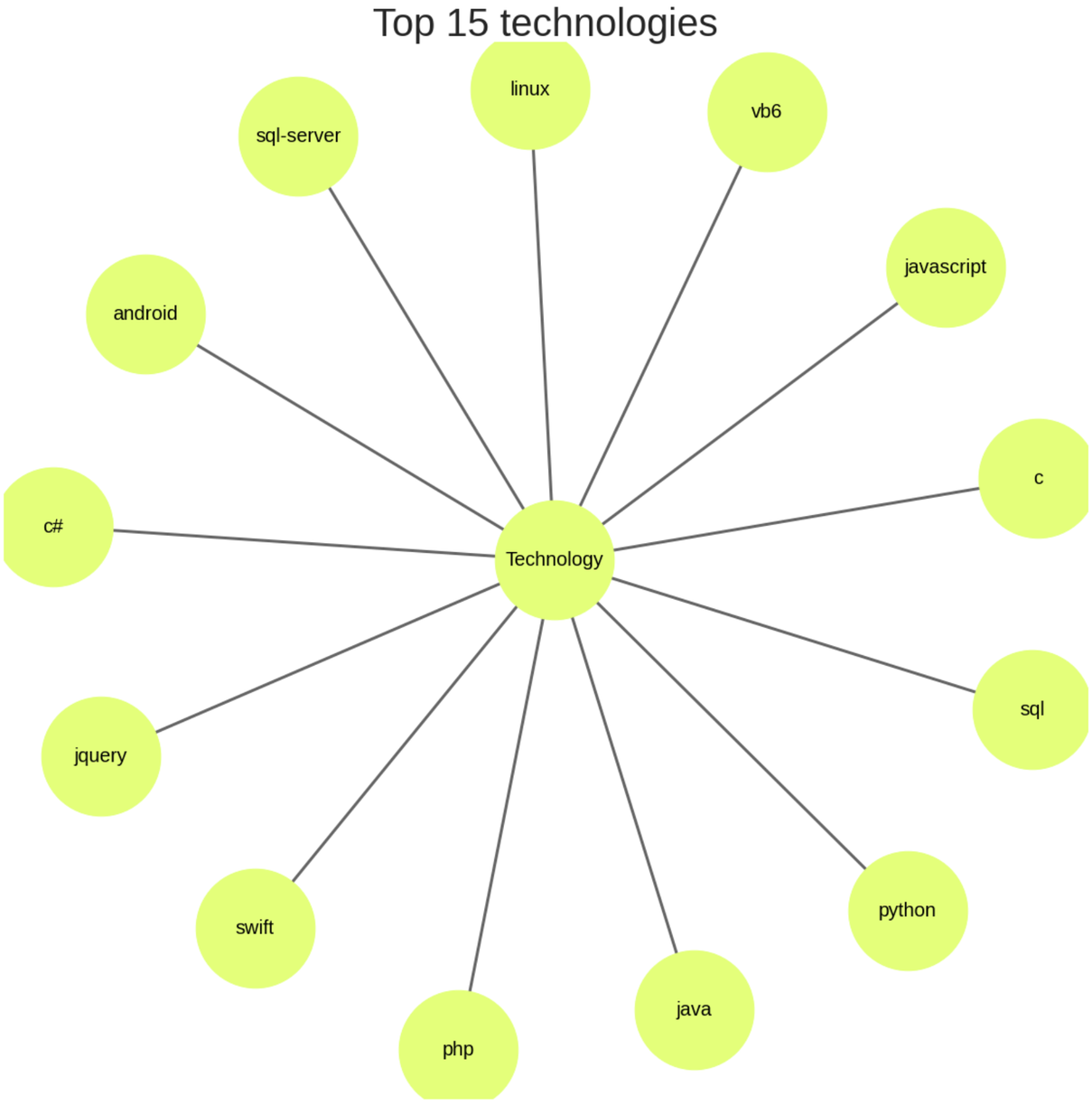
Language

Removing other languages

EDA


Most popular words in title on StackOverflow
by class, n>800, between 2017-2020
Word
using
python
can
error
array
string
get
code
file
android
java
function
use
value
data
list
php
javascript
object
sql
create
text
0
1000
2000
Count
Status:
Closed
High-quality
Low-quality
© Made by Michau96/Kag

EDA


Top 15 technologies
linux
vb6
sql-server
javascript
android
c
c#
Technology
sql
jquery
python
swift
java
php

Topic Modeling

Vectorizertion

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

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

03 Topics

- Hight 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?