Problem Statement 2

Identification of User Query on Stack Overflow Using Semantic Search

Stack Overflow is a question and answer site for professional and enthusiast programmers. It's built and run *by the community of developer,* as part of the Stack Exchange network of Q&A sites. A lot of content is present in form of stack overflow questions and answers, various studies point that developers face problems while development life cycles and they ask questions on stack overflow which gets answered by fellow developers across the globe.

In order for a new developer to understand a concept or solve an issue, it is very difficult to identify the problems. It involves domain experts in form of experienced software developers. The information present is overwhelming and at times can be too much to handle for a budding developer.

Problem Description

1. To identify most relevant questions to a query [text similarity]
2. Identify the matching tags and pick top relevant questions from stack overflow.
3. To identify top k solutions of the problem.
   (sentiment analysis of review content)
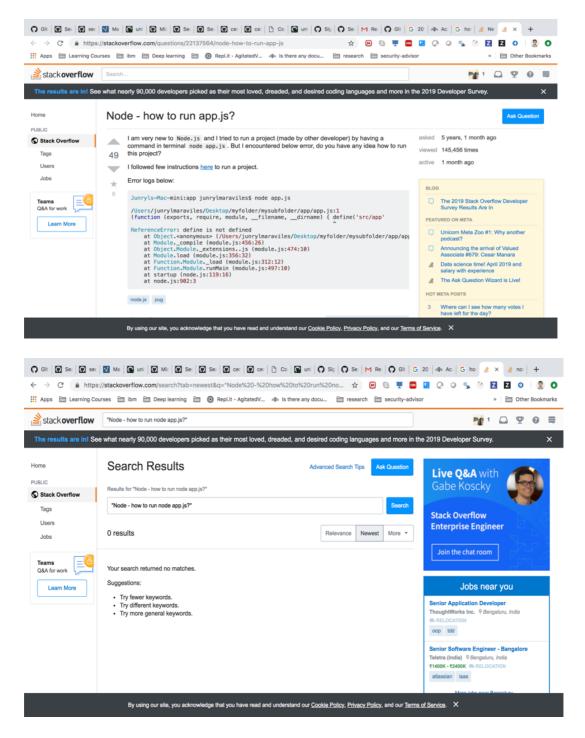
Data can be obtained using Stack overflow API.

Expectation
The solution should have a mechanism to identify relevant questions on stack overflow with tags and top k relevant answers that should be provided to user. So that he can understand the concept clearly and try to get his queries resolved more effectively and efficiently.

[scroll down to next page]

Listed below are the few challenges which can provide an insight about designing and finding the solution:

- **Stack overflow only gives exact match queries**
  (Google queries are not allowed)

Examples:
Query on stack overflow search:
1. "Node - how to run node app.js?" - No results.
2. "Node - how to run app.js?" – fetches a result.
   https://stackoverflow.com/search?q=%22Node++how+to+run+++app.js%3F%22

Evaluation Criteria

Accuracy of the answers obtained by the solution will be the mainstay of evaluation. The other evaluation parameters are listed on the hackathon landing page.

Tools & Technology

1. Python 2-3 or IBM cloud utilities python environment.
2. NLTK open source library.
3. Sentiment analysis open source library or Watson discovery.

Resources & References

- https://console.bluemix.net/catalog/
- https://stackoverflow.com/

Frequently asked questions

Q: What are the programming languages?
A: Python
Q: What are mobile platforms allowed?
A: NA
Q: Where to get free access to IBM Cloud?
A: Sign up on - https://www.ibm.com/cloud/
Q: Is there any documentation available to use IBM Cloud?
A: Yes, each service comes with elaborate documentation with step by step illustration to use the services available on IBM cloud, follow the VIEW DOCS, link available on each service.
Q: Is the knowledge of ML/DL is required?
A: No
Q: Is there any dataset provided?
A: No, there is no dataset made available with this problem statement.

Post your technical queries here.