TU Delft

SIMILAR REPOSITORY RECOMMENDER USING TAG HIERARCHIES

CSE 3000 - Author: Andrei-Cristian Ionescu (A.C.Ionescu-1@student.tudelft.nl) - Supervisor: Maliheh Izadi - Professor: Arie van Deursen

1. INTRODUCTION

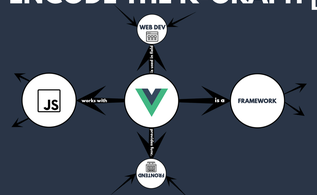
Developers do not want to reinvent the wheel when establishing a new software system. Open-source software repositories are packed with resources that may assist developers with their work. Our research intends to assist developers in locating such resources, allowing them to concentrate on the essential functions of an application or system rather than implementing fundamental functionalities from scratch.

3. RESULTS

	GitHub	Google	Baseline
Average Recommendation Time	0.1 sec.	1.6 sec.	0.1 sec.
Empty Entities	76%	19%	20%
Non-Empty Entities	24%	81%	80%
Average number of results	2.4	9.6	11.9
Average number of relevant items	0.69	2.02	3.03
Proportion of relevant items	0.28	0.21	0.25
MAP@3	0.13	**0.57**	0.48
Overall Recall@3	0.11	**0.49**	0.42

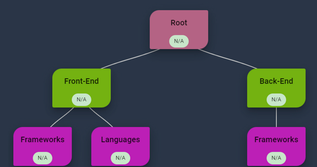
2. METHOD

1. ENCODE THE K-GRAPH [1]

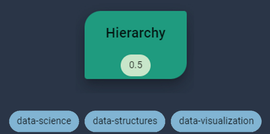


2. CLUSTER THE K-GRAPH

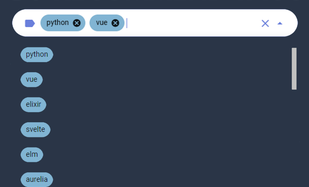
3. GENERATE HIERARCHY



4. HIERARCHY EVAL.



5. RECOMMEND TAGS

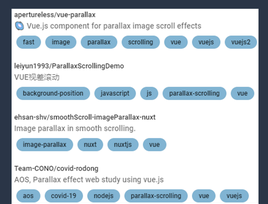


6. PREPARE SEARCH QUERY

7. FEED QUERY TO GOOGLE PROGRAMMABLE SEARCH API

8. SEARCH RESULTS THROUGH GITHUB SEARCH API

9. RETRIEVE TOP N SIMILAR REPOSITORIES



4. Conclusion

According to the findings, a similar repository recommender system based on tags and a hierarchy is valuable and accurate from the standpoint of an expert. The results also show that a tag hierarchy can be used to recommend similar tags to the users during the similar repository search procedure. During our study, we established some heuristics that allow us to assess the performance of any hierarchy in combination with the web-app [2]. Moreover, we developed and deployed a scenario-based evaluation framework [2] that allows the use of the hierarchy in searching for similar repositories using the Google and GitHub search approaches in comparison with our Baseline. The Google Search approach outperformed the Baseline by 18.75%, and the GitHub Search by more than 100%. In the future, we will focus on introducing the hierarchy as a new recommendation layer for filtering and reordering the recommendation results. Furthermore, we want to study the performance of a similar repository recommender system using more search engines such as Bing or DuckDuckGo and explore the implementation of the Google Knowledge Graph [3] in our Google approach.

References:

[1] - Izadi M., Nejati M., Heydarnoori A. "Semantically-enhanced Topic Recommendation Systems for Software Projects," Empirical Software Engineering, vol. 26, no. 5, 2021.
[2] - Ionescu, A. C. (2022, June). Hierarchy Visualiser. Hierarchy Visualiser. Retrieved June 19, 2022, from https://hierarchy-visualiser.herokuapp.com/#/evaluation
[3] - A. Singhal, "Introducing the Knowledge Graph: things, not strings," 05 2012. [Online]. Available: https://blog.google/products/search/introducing-knowl-