

Improving the Search for Architecture Knowledge in Online Developer Communities

Mohamed Soliman,¹ Amr Rekaby Salama,² Matthias Galster,³ Olaf Zimmermann,⁴
Matthias Riebisch⁵

Abstract: When architecting a software system, software engineers search for architectural solutions (e.g. technologies), which fulfill certain requirements. Current approaches for architecture knowledge repositories facilitate learning about different architectural solutions. Nevertheless, the rapid and continuous increase of solution alternatives makes it challenging to manually acquire architecture knowledge and to ensure that this knowledge is up to date. Our goal in this paper is to improve how software engineers search for architecturally relevant information in online developer communities. We developed a new search approach for architecturally relevant information using Stack Overflow as an example of an online developer community. Our search approach differs from a conventional keyword-based search in that it considers semantic information of architecturally relevant concepts in Stack Overflow. We implemented the search approach as a web-based search engine. To show the effectiveness of the search approach compared to a conventional keyword-based search, we conducted an experiment with 16 practitioners. To ensure realism of the experiment, tasks given to practitioners are based on real scenarios identified in a separate interview study. The experiment showed that the new search approach significantly outperforms a conventional keyword-based search.

The full paper of this extended Abstract has been published in [So18].

Keywords: software architecture; architecture knowledge; search approach; online developer communities; Stack Overflow

Software engineers face recently unprecedented challenges to learn and keep up to date with new architectural solutions (e.g. technologies). Architecture knowledge (AK) repositories facilitate learning about architectural solutions. However, repositories require manually curating and maintaining AK, which becomes out of date quickly with the fast evolution of AK. Online developer communities (e.g. Stack Overflow) succeeded to persuade software engineers to share their knowledge about software issues, which encompass not only coding problems but also architectural issues [So16]. Nevertheless, the unstructured nature of text makes it arduous to search for architecturally relevant information in developer communities. Moreover, traditional keyword-based search engines return many irrelevant results when searching for architectural information [Go17].

¹ Universität Hamburg, SWK, Hamburg, Deutschland soliman@informatik.uni-hamburg.de

² Universität Hamburg, NATS, Hamburg, Deutschland salama@informatik.uni-hamburg.de

³ University of Canterbury, Christchurch, New Zealand mgalster@ieec.org

⁴ HSR Hochschule für Technik Rapperswil, Schweiz olaf.zimmermann@hsr.ch

⁵ Universität Hamburg, SWK, Hamburg, Deutschland riebisich@informatik.uni-hamburg.de

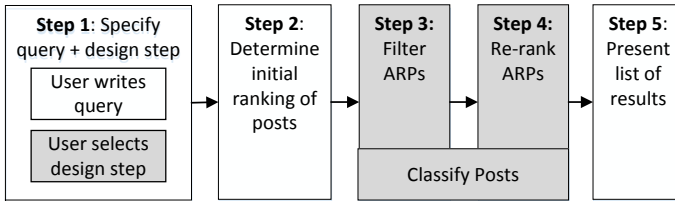


Abb. 1: Proposed search approach (gray boxes extend keyword-based search)

We propose a new search approach (see Abb. 1) to help finding architecturally relevant information on Stack Overflow. Steps 3 and 4 are our main contributions, which differentiate our enhanced search from a conventional keyword-based search. In step 3, we filter “architecture-relevant” posts from other types of posts. In step 4, we re-rank “architecture-relevant” posts based on their significance to support a certain architecture design step. To filter and re-rank posts in Stack Overflow, we developed a classification approach, which classify posts according to the type of discussed architectural solution and the purpose of a post [So16]. The classification approach combines two classifiers (Bag-of-Words and AK Ontology-based classification [SGR17]) using an ensemble learning. We conducted experiments using a corpus of Stack Overflow posts [So16] to evaluate the accuracy of classification. The average accuracy (F-Score) of the classification approach is 0.734. To compare the effectiveness of the proposed search approach to a conventional keyword-based search, we conducted an experiment with 16 practitioners. We asked participants to perform architecture design tasks and use both search engines to search for information that supports architectural design steps. Participants submitted a total of 422 queries. We calculated $Precision@k$ and $nDCG@k$ for both conventional and enhanced search approaches, with $@k_{1 \rightarrow 10}$. Our results show that our proposed enhanced search significantly improve the effectiveness of searching for architectural information in Stack Overflow.

Literatur

- [Go17] Gorton, I.; Xu, R.; Yang, Y.; Liu, H.; Zheng, G.: Experiments in Curation: Towards Machine-Assisted Construction of Software Architecture Knowledge Bases. In: IEEE/IFIP ICSA 2017. S. 79–88, Apr. 2017.
- [SGR17] Soliman, M.; Galster, M.; Riebisch, M.: Developing an Ontology for Architecture Knowledge from Developer Communities. In: IEEE/IFIP ICSA 2017. S. 89–92, Apr. 2017.
- [So16] Soliman, M.; Galster, M.; Salama, A. R.; Riebisch, M.: Architectural Knowledge for Technology Decisions in Developer Communities: An Exploratory Study with StackOverflow. In: IEEE/IFIP WICSA 2016. S. 128–133, Apr. 2016.
- [So18] Soliman, M.; Salama, A. R.; Galster, M.; Zimmermann, O.; Riebisch, M.: Improving the Search for Architecture Knowledge in Online Developer Communities. In: IEEE ICSA 2018. S. 186–18609, Apr. 2018.