# Exploring Web Search Engines to Find Architectural Knowledge

Mohamed Soliman,[1] Marion Wiese [2], Yikun Li,[3] Matthias Riebisch,[4] Paris Avgeriou[5]

**Abstract:**

Software engineers need relevant and up-to-date architectural knowledge (AK), in order to make well-founded design decisions. However, finding such AK is quite challenging. One pragmatic approach is to search for AK on the web using traditional search engines (e.g. Google); this is common practice among software engineers. Still, we know very little about what AK is retrieved, from where, and how useful it is. In this paper, we conduct an empirical study with 53 software engineers, who used Google to make design decisions using the Attribute-Driven-Design method. Based on how the subjects assessed the nature and relevance of the retrieved results, we determined how effective web search engines are to find relevant architectural information. Moreover, we identified the different sources of AK on the web and their associated AK concepts.

The full paper of this extended Abstract has been published in [So21].

**Keywords:** software architecture; architecture knowledge; search approach; Google

## 1 Summary

Architectural knowledge (AK) is crucial for software engineers to make architectural design decisions [BCK12]. However, finding architectural knowledge (AK) is a challenging task, because AK resides in multiple heterogeneous *AK sources*, such as developer communities. Moreover, each source of AK contains different *AK concepts* (e.g. design decisions). A pragmatic and common way to search for AK from different sources is to use web search engines (e.g. Google). However, there is little to no empirical evidence about which AK sources and AK concepts can actually be found by web search engines. Our main **goal** is to *explore which AK sources and AK concepts are retrieved by web search engines, and to gauge the effectiveness of web search engines to find relevant AK during the architectural design process*. To this end, we conducted an empirical study with 53 software engineers. The subjects used the most popular web search engine (i.e. Google) to find relevant AK concepts when conducting the steps of the Attribute-Driven Design (ADD) method [KC16].

[1] University of Groningen, The Netherlands m.a.m.soliman@rug.nl

[2] Universität Hamburg, SWK, Hamburg, Deutschland marion.wiese@uni-hamburg.de

[3] University of Groningen, The Netherlands yikun.li@rug.nl

[4] Universität Hamburg, SWK, Hamburg, Deutschland matthias.riebisch@uni-hamburg.de

[5] University of Groningen, The Netherlands p.avgeriou@rug.nl

We provide below an overview on the results of each research question:

*(RQ1) Which AK can web search engines support to find?* The results suggest that google retrieve the following AK sources: blogs (39%), technology vendor documentations (23%), scientific contents (13%), and forums like Stack Overflow (7%), source code repositories like Github (4.5%) and knowledge repositories (4%). Furthermore, issue tracking systems are not retrieved at all by Google. On the other hand, the results showed that blogs, forums, knowledge repositories, and scientific contents have the most AK concepts.

*(RQ2) Which AK concepts co-occur on the web?* Our results show that benefits and drawbacks have the highest correlation (0.651) to co-occur with each other. It is common to find web pages with benefits and no drawbacks, while it is rare to find web pages with drawbacks and no benefits. Moreover, lists of alternative solutions, are usually (0.355) accompanied with a comparison between them regarding their benefits and drawbacks.

*(RQ3) How well do web search engines support software engineers in following the ADD steps?* Our results show that Google supports software engineers to perform the ADD step "Select design concepts" better than the other steps like "Instantiate architecture elements". Moreover, we found that Google returns many distantly relevant web pages (e.g. general concepts on components design) to support the "Instantiate architecture elements" step.

*(RQ4) Which AK concepts make web pages more relevant for design decisions?* Our results show that web pages that contain solution benefits, solution drawbacks and made design decisions are most relevant for finding AK. Web pages with solution description' and alternatives have a lower probability to be highly relevant.

## 2  Data Availability

We provide our corpus online (`github.com/m-a-m-s/ICSA2021`) to facilitate replicating the study. Moreover, the corpus contains empirically classified and evaluated web pages and their respective AK sources and AK concepts.

## Literatur

[BCK12]  Bass, L.; Clements, P.; Kazman, R.: Software Architecture in Practice. Addison-Wesley Professional, 2012, ISBN: 0321815734, 9780321815736.

[KC16]  Kazman, R.; Cervantes, H.: Designing Software Architectures: A Practical Approach. Addison-Wesley Professional, 2016.

[So21]  Soliman, M.; Wiese, M.; Li, Y.; Riebisch, M.; Avgeriou, P.: Exploring Web Search Engines to Find Architectural Knowledge. In: 2021 IEEE 18th International Conference on Software Architecture (ICSA). IEEE, S. 162–172, März 2021, ISBN: 978-1-7281-6260-7.