

## University of Groningen

### Where and What do Software Architects blog?

Soliman, Mohamed; Gericke, Kirsten; Avgeriou, Paris

*Published in:*

IEEE 20th International Conference on Software Architecture, ICOSA

*DOI:*

[10.1109/ICSA56044.2023.00020](https://doi.org/10.1109/ICSA56044.2023.00020)

**IMPORTANT NOTE:** You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

*Document Version*

Publisher's PDF, also known as Version of record

*Publication date:*

2023

[Link to publication in University of Groningen/UMCG research database](#)

*Citation for published version (APA):*

Soliman, M., Gericke, K., & Avgeriou, P. (2023). Where and What do Software Architects blog? An Exploratory Study on Architectural Knowledge in Blogs, and their Relevance to Design Steps. In *IEEE 20th International Conference on Software Architecture, ICOSA* (pp. 129-140). IEEE.  
<https://doi.org/10.1109/ICSA56044.2023.00020>

#### Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

#### Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

*Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.*

# Where and What do Software Architects blog?

An Exploratory Study on Architectural Knowledge in Blogs, and their Relevance to Design Steps

Mohamed Soliman\*, Kirsten Gericke\* and Paris Avgeriou\*

\* Bernoulli Institute for Mathematics, Computer Science and Artificial Intelligence

University of Groningen, Groningen, The Netherlands

{m.a.m.soliman, p.avgeriou}@rug.nl, kirstykromhout7@gmail.com

**Abstract**—Software engineers share their architectural knowledge (AK) in different places on the Web. Recent studies show that architectural blogs contain the most relevant AK, which can help software engineers to make design steps. Nevertheless, we know little about blogs, and specifically architectural blogs, where software engineers share their AK. In this paper, we conduct an exploratory study on architectural blogs to explore their types, topics, and their AK. Moreover, we determine the relevance of architectural blogs to make design steps. Our results support researchers and practitioners to find and re-use AK from blogs.

**Index Terms**—Architecture knowledge, Architecture design decisions, blog articles

## I. INTRODUCTION

Software engineers need *architectural knowledge* (AK) [1] to make reasonable *architectural design decisions* [2]. For example, when software engineers work on component design (i.e. the structure and behavior of the system components), they often rely on their AK from previous projects, or from discussions with experienced architects. To understand the nature of this AK, researchers have explored various *AK concepts* [3], such as architectural solutions (e.g. patterns [4] and technologies [5]), as well as constraints, benefits and drawbacks of solutions [6].

Software engineers make design decisions in consecutive steps, and for each step, they need different AK concepts [7]. To exemplify this, let us consider an example of such a step-wise process, the Attribute Driven Design (ADD) [8], the most well-established architecture design process (see Section II). In the *identify design concept* step of ADD, AK concepts such as architectural solutions and quality attributes (e.g. performance and security) are considered. In the *select design concept* step, AK concepts such as benefits and drawbacks between alternative solutions are considered. In the *instantiate architecture elements* step, AK concepts such as components and connectors are considered. While AK is critical for those design steps, it is rather challenging to find AK. For instance, software engineers struggle to find design decisions from architectural documents, because design decisions are not systematically documented and formalized [9].

To address this problem, researchers have recently explored AK in software repositories [10] such as issue tracking systems [11], and mailing lists [12], as well as in certain web resources such as Stack Overflow [13] and technology documentations [14]. These contributions provide concrete

datasets, AK ontologies [15], and search approaches [7] that can help researchers and practitioners to find AK. Moreover, there is evidence that software engineers share their AK in technical articles within blogs (i.e. *architectural blogs*) [3]. Such blogs showed to be the most relevant source on the Web to perform architectural tasks, and the richest source of AK concepts, compared to forums (e.g. Stack Overflow), source code repositories and technology documentations [3].

Despite the value of architectural blogs in finding AK, software engineering researchers rarely explored blogs (e.g. [16]), and did not specifically explore architectural blogs and their contained AK. Moreover, architectural blogs are quite unstructured and disorganized [17]: they are scattered on the Web among thousands of websites and their *type* varies from personal blogs to blogs hosted by software companies and communities. This makes it hard to understand which blogs one should look for AK in. Furthermore, architectural blogs discuss multiple *topics* (e.g. comparing solutions), each involving different AK concepts. However, blog articles are not classified (e.g. using tags) [16], which makes it challenging to search for and subsequently reuse their contained AK concepts. Therefore, in this paper, we aim to *explore the types and topics of architectural blogs, as retrieved by Web search engines, and their relevance for software engineers during certain design steps*. This allows us to provide an overview of architectural blogs, which can be used by researchers and practitioners to find and re-use AK from architectural blogs.

To this end, we empirically analyzed the dataset of architectural blogs from Soliman et al. [3], and applied grounded theory to explore the types and topics of architectural blogs. Moreover, we applied the most popular topic modeling algorithm, namely Latent Dirichlet Allocation (LDA) [18] to explore architectural topics based on their contained AK concepts. Using two research methods (i.e. grounded theory and LDA) to identify topics of architectural blogs supports exploring topics from different perspectives, as well as data triangulation. Finally, we applied statistical analysis to identify significant co-occurrences between types and topics of architectural blogs, and the relevance of topics for certain ADD steps. In summary, we achieved the following contributions:

- The *types of blogs* that contain AK as retrieved by the most popular Web search engine Google.
- The *topics of blogs*, their contained AK concepts, and their significant co-occurrences with blog types.

- The *relevance* of architectural blog articles from certain topics in following ADD steps.
- A *dataset* of 718 classified architectural blog articles, based on their types and topics.
- A semi-automated *approach to identify architectural topics* using LDA and AK ontology.

Section II provides a background on ADD. Section III explains our study design, Sections IV, V, and VI present our results. Sections VII, VIII, IX discuss our results, threats to validity, and related work. Finally, Section X concludes the paper.

## II. BACKGROUND ON THE ATTRIBUTE DRIVEN DESIGN

In this paper, we evaluate the relevance of architectural blogs to follow the ADD steps [8] (see Section VI). We have selected the ADD process, as it is the most well-established architecture design process, it provides concrete steps, and it was previously applied by Soliman et al. [3]. In this section, we provide a brief summary of three out of the seven ADD steps, namely those that involve making design decisions, and thus require AK:

*Identify design concepts:* A list of alternative architectural solutions is identified. For example, a software engineer might identify a list of technology solutions (eg. Spark, Storm, Flink, etc.) when designing a stream processing engine.

*Select design concepts:* One design concept from the list of alternative solutions is selected. Each solution is compared and evaluated against requirements, constraints, quality attributes (e.g. security, performance, etc.), and considering the benefits and drawbacks of solutions. For example, a software engineer might require high performance of stream processing, she would therefore rather select Spark or Flink over Storm.

*Instantiate architecture elements:* Elements of the selected architectural solution are modified to achieve system requirements. For example, a software engineer might configure a reference architecture to meet a quality attribute such as add a custom layer, or specify relationships and responsibilities of elements of an architectural pattern [8].

## III. STUDY DESIGN

### A. Research questions

To achieve our aim (see Section I), we ask the following research questions (RQs):

**(RQ1)** *What types of architectural blogs do Web search engines retrieve?*

Software engineers write architectural articles in blogs (e.g. personal blogs) with different characteristics. For instance, some blogs are hosted by software companies, and focus on specific solutions (e.g. Big Data technologies), while other blogs are hosted by software development communities and discuss multiple topics on different solutions. We ask this RQ to determine types of blogs where AK is shared, and their characteristics as retrieved by Web search engines (e.g. Google). By answering this RQ, researchers can develop approaches that classify and index architectural blogs to find and re-use AK. Moreover, practitioners can make informed decisions on the types of blogs to share and search for AK.

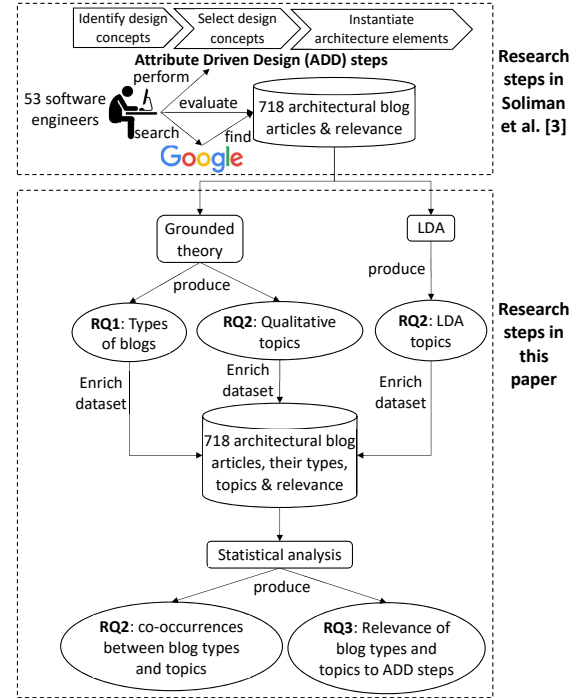


Fig. 1: High level view on the research process to answer RQs

**(RQ2)** *What kind of topics are discussed in the different types of architectural blogs?*

Blog articles discuss different topics (e.g. comparing solutions). Each topic involves different AK concepts. Thus, we ask this RQ to determine topics discussed in architectural blogs, their involved AK concepts, and the types of blogs (e.g. personal blog) that discuss each topic. These topics can guide practitioners to search for and re-use certain AK concepts. They can also help researchers to develop approaches that automatically classify architectural blogs based on their topics.

**(RQ3)** *How relevant are the topics of architectural blogs as retrieved by Web search engines to support practitioners follow the ADD steps?*

Some topics of architectural blogs could be more useful for practitioners to conduct certain design steps (such as the ADD steps in Section II). However, Web search engines (e.g. Google) retrieve blog articles based on keyword-searches, and thus might not retrieve relevant results for each design step [3]. Therefore, we ask this RQ to determine topics of architectural blogs that Web search engines do retrieve, as well as which topics are relevant for conducting the ADD steps, according to practitioners. By answering this RQ, practitioners could target their search towards more relevant architectural blogs to achieve their design steps effectively. Moreover, researchers can adapt their AK finding approaches by promoting architectural blogs with the highest relevance. Figure 1 shows an overview on the research process to answer the RQs.

### B. Dataset of architectural blog articles

We re-used the dataset of architectural blog articles from the study of Soliman et al. [3]. Following, we provide a

TABLE I: Architectural tasks

ADD step	Task description
Identify design concepts	<ul style="list-style-type: none"> <li>• For a realtime dashboard, identify middleware technologies which scale to &gt;100k users</li> <li>• Identify high performance Java JSON parsers for mobile app communication.</li> </ul>
Select design concepts	<ul style="list-style-type: none"> <li>• Compare interoperability and latency of RabbitMQ, Kafka, and ActiveMQ.</li> <li>• Compare data collectors, message brokers, and ETL engines for big data systems.</li> </ul>
Instantiate architecture elements	<ul style="list-style-type: none"> <li>• Search for technology features and component design to select deployment topology and routing.</li> <li>• Search for best practices regarding service decomposition to achieve high cohesion and low coupling.</li> </ul>

brief description of that study, while further details are in [3]. 53 software engineers used Google searches to perform six architectural tasks (see Table I) that correspond to the three ADD steps (see Section II). A complete description of each task is available online [19]. To perform the six tasks, software engineers evaluated each web-page according to its relevance to an architectural task in a five-level Likert scale:

- *Very High Relevance (5)*: fulfills more than one requirement.
- *High Relevance (4)*: fulfills one requirement of the task.
- *Medium Relevance (3)*: provides information to the task.
- *Low Relevance (2)*: is remotely relevant to the task.
- *No Relevance (1)*: has no relevance information.

The 53 software engineers in Soliman et al. [3] evaluated 2623 unique web-pages according to their relevance. Subsequently, 945 web-pages were classified as blogs and tutorials. These 945 web-pages provide the base dataset of this study. We first verified whether these 945 web-pages are indeed architectural blogs, and excluded the following web-pages:

- Web-pages with *no relevance (level 1 above)* to an architectural task as specified by practitioners. This is to ensure that all blog articles in the dataset are actually architectural.
- Blog articles in languages other than English.
- Web-pages that cannot be accessed anymore.
- Web-pages that include tutorials rather than blogs.
- Duplicated blog articles with different URLs.

As a result of this filtering step, a total of 718 unique architectural blog articles were used to answer the RQs.

### C. Apply grounded theory to identify types and topics of architectural blogs

To identify types and topics of architectural blogs, we manually analyzed all 718 of the architectural blog articles by following steps and best practices from grounded theory [20], [21]. In detail, we performed the following three steps:

1) *Open coding*: To identify initial types and topics of blogs, the first and the second authors randomly selected a significant sample of 300 architectural blog articles from the 718 articles, and independently inspected each blog article as well as the web-sites, in which the blog articles are hosted. Both researchers inspected the 300 architectural blog articles through 3 iterations. Within each iteration, both researchers independently assigned a type and a topic for each blog article, by focusing on the content of the blog articles and avoiding being biased from literature; subsequently they *wrote memos* to explain why a type or a topic of blog has been determined

for a certain article. For example, if all articles in a web-site are written by the same software engineer, then it is a personal blog, and if an article provides benefits and drawbacks of alternative solutions, then it is about comparing solutions. We share the memos for all blog articles online [19]. After each iteration, both researchers discussed the assigned types and topics of architectural blog articles, and performed *constant comparisons* to determine differences between the different types and topics of blogs. For instance, we determined the following attributes to differentiate the types of blogs:

- *Number of authors who write blog articles in a web-site*. For example, personal blogs tend to have a single author.
- *Host of the web-site*. For example, a blog article could be hosted by technology vendors or IT service companies.
- *Relationship between authors of articles and the hosting web-site*. For example, authors of articles could be independent or employees of the company that hosts a web-site.

For the topics, we determined the following attributes:

- *The purpose of an article*. Some articles discuss steps to design a system, while others provide steps to implement.
- *The shared AK concepts in an article*. Some articles provide benefits and drawbacks of solutions, while other articles provide a list of alternative solutions to solve a problem.
- *The number of discussed architectural solutions in an article*. Some articles discuss and compare multiple solutions, while other articles elaborate and evaluate a single solution.

As a result of this step, we defined initial types and topics of architectural blogs. During the third iteration, neither of the two researchers could determine new types or topics; consequently, this indicated *theoretical saturation*.

2) *Axial coding*: Based on the initial types and topics from the open coding step, the first two authors of this paper classified the rest of the 718 architectural blog articles: the first author identified the topics for the rest of the 718 articles, while the second author identified the types for the rest of the 718 articles. For some uncertain cases of articles, both researchers discussed the article to agree upon a single type or topic. During discussions, both researchers noticed that some types of blogs caused disagreements between them. Thus, we merged these types of blogs to ensure good agreement between researchers. For example, some community blogs have subtypes (e.g. general versus technology-specific) which could be confusing to differentiate; therefore, we merged community blogs in a single type. By the end of this step, we reached the final set of *types* and *qualitative topics* of architectural blogs (see Sections IV and V-A).

3) *Measure agreement*: To ensure the reliability of our qualitative analysis, we conducted a final agreement test: we randomly selected per author, a significant sample of 300 blog articles, which were not previously classified by that author. The first author independently classified the test sample regarding blog types, while the second author classified the test sample regarding topics. Based on this test sample, we calculated the kappa coefficient among the two authors, which are 0.81 regarding its type and 0.71 regarding the topics. These kappa values indicate good agreement beyond chance.

#### D. Apply LDA to identify topics of architectural blogs

In the previous section, we applied grounded theory to identify qualitative topics of architectural blogs. In this Section, we applied LDA to identify topics (i.e. *LDA topics*) in architectural blogs. Using LDA, we consider frequencies of terms and AK concepts within an article; these provide a different perspective on topics than qualitative analysis. We also note that we determined co-occurrences between the qualitative topics and LDA topics in Section III-E. For the LDA analysis, we followed steps and best practices for applying LDA in software engineering [18]. In summary, we applied 5 steps. We performed the first three steps iteratively; subsequently, we performed steps 4 and 5. These five steps are elaborated in the following.

1) *Pre-process blogs*: Blog articles contain textual contents, which can mislead the LDA algorithm, resulting in identifying irrelevant and inconsistent topics. Thus, this step aims to remove and process textual contents in blog articles to guide LDA towards architecturally-relevant topics. We removed the following terms from blog web-pages:

- HTML tags using the BeautifulSoup library<sup>1</sup>, stop words using the NLTK library<sup>2</sup>, and special characters (e.g. #,-).
- Too frequent words that appear in more than 95% of the blogs, and less frequent words that appear in less than 5% of the blogs, because these terms bias LDA to form separate topics. Other researchers have also followed this step [18].
- Generic keywords that do not refer to an AK concept but could mislead the LDA algorithm in creating separate topics for them, such as “microsoft”, “medium” and “software”. We expanded the list of generic keywords iteratively (see Step 3).
- Common terms related to source code implementation such as “method”, “procedure”, and “array”. The list of source code terms has been expanded iteratively (see Step 3). We removed source code terms, because we found that they bias LDA to create dedicated topics for blogs that contain source code elements; such topics are irrelevant to the actually discussed architectural topic.

After filtering out non-useful terms, we lematized terms to a single form. For example, “deciding” and “decided” are reduced to a single form “decide”. We used NLTK to lematize English terms. Moreover, we extended the NLTK dictionary to lematize architectural terms, which cannot be lematized by the default NLTK wordnet dictionary (e.g. “scalability” and “scalable” are reduced to a single form “scale”).

Subsequently, we found out that LDA was biased to identify topics based on their domain and technologies (e.g., Big Data and Middleware topics). Thus, to identify architecturally-relevant topics, independent of their domain, we replaced certain terms with tags that corresponds to their AK concepts. We considered specifically AK concepts from the ontology of AK in Stack Overflow [15]. We have decided on this ontology due to the possible similarity of AK between Stack Overflow and blogs. Table II presents the list of selected AK concepts

TABLE II: Examples of terms for each AK concept

AK concept	Examples of terms
<Technology>	ActiveMQ, SQS, JSON, Kafka, Maven, Flume
<Pattern>	layer, publish subscribe, Microservice, ESB, SOA
<Quality_attribute>	performance, accuracy, usability, security, throughput
<Requirement>	business, dashboard, chart, customer, market, financial
<Component>	application, server, client, back end, front end
<Connector>	retrieve, connect, consume, call, save, route, depend
<Connector_data>	socket, payload, message, dump, token, call, update

from the AK ontology [15], and examples of terms associated with each AK concept. In addition to replacing terms with AK concepts, we associated the frequency of unique terms in a blog article with their respective tag of AK concept. For example, if a blog post discusses three technologies: Kafka, RabbitMQ and ActiveMQ, these will be replaced with <Technology\_Solution\_1>, <Technology\_Solution\_2>, and <Technology\_Solution\_3>. In this way, LDA can differentiate between blog articles that contain different terms from specific AK concepts. For example, blog articles which compare different technology solutions would have <Technology\_Solution> tags with higher frequency, while blog articles that discuss different components will have <Component> tag with higher frequency. We note that we extended the list of terms associated with each AK concept iteratively (see Step 3). We provide online [19] lists of terms that correspond to each AK concept, as well as scripts to replicate this pre-processing step.

2) *Apply LDA*: After pre-processing blog articles, we experimented with a different number of topics ( $k$ ), between 5 and 20 topics, and applied the most common LDA parameters ( $\alpha = 50/k$  and  $\beta = 0.01$ ) [18]. We note that we further tuned LDA parameters in Step 4. After executing LDA on our dataset, each blog article is assigned to a single topic (with the highest coherence of terms and AK concepts). However, LDA does not describe or define each topic. Therefore, to understand each topic, we identified the top most frequently occurring terms and AK concepts in each topic. We used this list of terms and AK concepts in Step 3 to enrich the pre-processing step. Moreover, we used this list of terms and AK concepts in Step 5 to define *LDA topics* of architectural blogs.

3) *Update pre-processing*: In this step, we checked the list of terms from the LDA topics (see Step 2), and determined general terms and source code terms that could be removed. Moreover, we determined new architectural terms that refer to certain AK concepts, which could be replaced with their respective AK concepts. Accordingly, we updated the pre-process blogs step with new terms to refine the LDA topics, and repeated the first three steps (i.e. pre-process blogs, apply LDA, and update pre-processing) to ensure that we identified all frequent terms from blog articles. We stopped repeating the three steps, once no new terms appeared in the list of the most occurring terms of each topic, which needed to be either removed or replaced with an AK concept.

4) *Tune LDA parameters*: After establishing the pre-processing of blogs, we experimented with a different number of topics up to 20 topics, and calculated the coherence. We noticed that we got higher coherence for number of topics between three and six topics, while the coherence decreased

<sup>1</sup><https://www.crummy.com/software/BeautifulSoup>

<sup>2</sup><https://www.nltk.org>

for more than six topics. Therefore, we decided to explore six topics of architectural blogs. Moreover, we experimented with  $\alpha = 50/k$  to  $\alpha = 1/k$ , and calculated the density of topics among architectural blogs. Our experiments showed that the standard parameter of  $\alpha = 50/k$  provides a balanced density of topics among architectural blog articles.

5) *Define LDA topics*: LDA represents topics as a set of occurring terms and AK concepts. Using these terms and AK concepts, we provided a definition of each LDA topic (see Section V-B).

#### E. Apply statistical analysis to answer RQ2 and RQ3

As part of answering RQ2, we determine significant co-occurrences between qualitative topics, LDA topics and types of architectural blogs using the  $\chi^2$  significant test [22]. For each co-occurrence between a type or a topic, we calculated a 2x2 contingency table. For example between one type of blogs (e.g. personal blogs), and one topic of blogs (e.g. compare solutions), we calculate: number of co-occurring articles between the specified type and topic, number of articles with the specified type but with other topics, number of articles with the specified topic but with other types, number of articles with different types and topics. Among this 2x2 contingency table, we calculate the  $\chi^2$  value, and determine significant co-occurrences with  $\chi^2 > 10$  at p-value  $< 0.05$ . We present significant co-occurrences in Section V-C and Figure 2.

To answer RQ3, we used the relevance indicated by the 53 software engineers in Soliman et al. [3] to determine relevant articles for each ADD step. Specifically, we used descriptive statistics to calculate the number of relevant blog articles and their relevance for each topic when following the ADD steps. Moreover, we conducted a Kruskal-Wallis H test [23] to determine significant relevance between the different topics. We decided on the Kruskal-Wallis H test over the Anova test [24], because it can test abnormally distributed data as in our dataset. Further details about the significance test are shared online [19]. We present the results in Section VI.

#### IV. RQ1: TYPES OF ARCHITECTURAL BLOGS

We analyzed the architectural blogs in our dataset using grounded theory (see Section III-C), and identified the following types of architectural blogs (in parentheses their respective percentages in the dataset), as retrieved by Google:

- **Community blogs (43%)**: hosted by open publishing platforms (e.g. *dzone.com*), and authored by many different software engineers, who volunteer to share their experience in blog articles. The majority of community blogs host a wide range of architectural topics such as those in *dzone.com* and *medium.com*. But some community blogs focus on specific topics such as data integration as in *dataintegrationinfo.com*.
- **Technology vendor blogs (25%)**: hosted by specific technology vendors, such as commercial companies (e.g. SAP) or open source foundations (e.g. Apache). Blog articles focus on topics related to the products or technologies produced by the hosting technology vendor such as *blogs.sap.com* and *blog.rabbitmq.com*. The authors of articles are experts and developers of these technologies.

- **Personal blogs (15%)**: hosted and authored by individual software experts, and reflecting their personal experience such as articles in *martinfowler.com* and *alexanderdevelopment.net*.
- **IT service blogs (11%)**: hosted by IT service companies that provide services such as software development or consulting (e.g. *openlogic.com/blog*). Articles are posted by employees of these companies.
- **Magazines and newspapers blogs (3%)**: hosted by specialized magazines or newspapers (e.g. *opensourceforu.com*), where specific authors, possibly hired, write articles.
- **Educational blogs (3%)**: hosted by training organizations or universities such as *edureka.co/blog*. Articles are posted by students or educators.

#### RQ1 key takeaways:

- Software engineers share their AK in different types of blogs with different hosting and authorship policies.
- Google search results contain architectural articles mostly from community blogs, followed by technology vendor blogs, and then personal blogs.

#### V. RQ2: TOPICS OF ARCHITECTURAL BLOGS

We analyzed architectural blog articles using grounded theory (see Section III-C) and LDA (see Section III-D). Each approach produced separate but related topics. Subsequently, we determined significant co-occurrences between the topics from the two approaches, as well as the types of blogs (see Section III-E). Figure 2 shows an overview on the qualitative and LDA topics, and their significant co-occurrences with each other and with blog types according to the  $\chi^2$  significance test. In sub-sections V-A and V-B, we respectively explain the qualitative and LDA topics, and their contained AK concepts. Moreover, we support our explanations with examples of articles (see Section “References of blog articles”), referenced as [B#] (e.g. [B1]). Finally, we discuss the co-occurrences between the topics and types of blogs in sub-section V-C.

##### A. Qualitative topics of architectural blogs

We explain the qualitative topics (in parentheses their respective percentages in the dataset), as retrieved by Google (see Section III-B), as well as their contained *AK concepts* (underlined in the text).

- **Elaborate and evaluate a solution (28%)**: Articles in this topic elaborate and evaluate a single architectural solution, such as a technology and its features [B1], or a pattern and its elements [B2], or an architecture and its components [B3], or a design principle and its application [B4]. For example, in [B1], the author first elaborates the main features of Amazon SQS technology such as “*SQS is a distributed, queue-based, messaging service...It supports building (loosely coupled) integrations between two applications*”, then the author discussed abilities of Amazon SQS to achieve quality attributes such as performance and availability. Another example is [B2], where the author elaborates the Microservices architecture, and discusses its benefits and drawbacks.
- **List of related solutions (20%)**: Articles in this topic provide a list of related architectural solutions such as lists

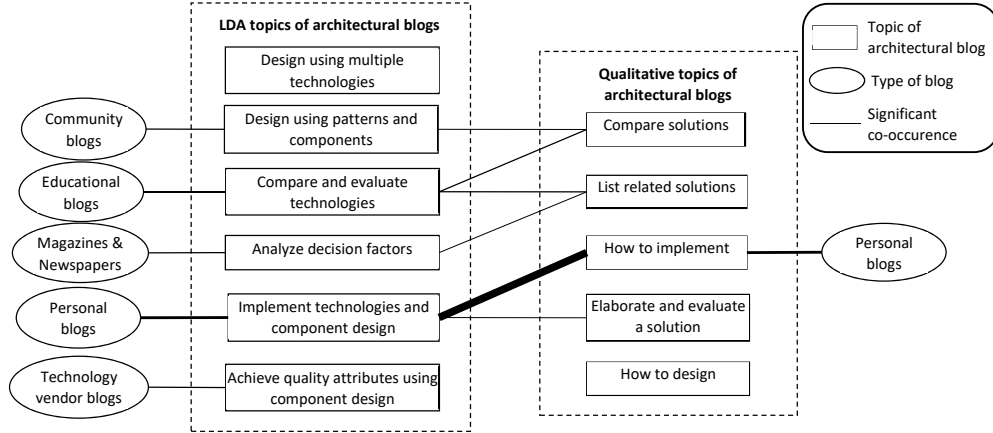


Fig. 2: Significant co-occurrences between LDA topics and qualitative topics of architectural blogs, and between topics and types of blogs. The line thickness corresponds to the significance of co-occurrence as measured by the  $\chi^2$  significance test.

of technologies [B5], patterns [B6], tactics [B7], design principles [B8], architectural components [B9], and best practices [B10]. Solutions could be either alternative solutions to solve a design issue, or solutions that complement each other to design an architecture of a system. Each solution could be associated with a description of its characteristics. In these articles, the solutions are not compared to each other. For example, [B10] provides a numbered list of 10 Microservices Best Practices, giving a short description of each item.

- **Compare solutions (20%):** Articles in this topic compare two or more architectural solutions such as technologies [B11] or patterns [B12], list their benefits and drawbacks, and explain in which cases each could be applied. For example, [B12] defines two solutions: Synchronous and Asynchronous request handling, states factors to decide on them (i.e. requirements and constraints), and states their benefits and drawbacks. Another example is [B11], which compares several JSON libraries (i.e. architectural solutions) based on their performance (i.e. quality attribute).

- **How to design (18%):** Articles in this topic discuss different steps or issues to design a system within a certain domain (e.g. Web applications [B13]), or to design a system based on specific architectural solutions such as technologies [B14] or patterns [B15]. The explanation of steps or issues are commonly supported with use-cases, and alternatives of architectural solutions. For example, [B16] presents a “Seven Step API Design Methodology”, within the domain of the Web, using the HTTP protocol. Another example is [B17], which explains how to design a routing topology using RabbitMQ, and discusses design issues, use cases and evaluations against quality attributes such as performance and scalability.

- **How to implement (14%):** Articles in this topic provide guidelines to implement a conceptual architectural solution (e.g. component design [B18] or pattern [B19]) through certain technologies, or provide guidelines to integrate different technologies together [B20]. For example, [B19] explains how to implement a RESTful web service, specifically in Java using JSON-B and JSON-P. Another example is [B20], which

explains how to integrate Apache Camel with the CData JDBC Driver to copy Dynamics CRM data to a JSON file on disk.

### B. LDA topics of architectural blogs

We define LDA topics based on the most commonly occurring terms and AK concepts in each topic, as identified by the LDA algorithm. Table III provides top terms, AK concepts and their frequencies (columns freq.) among articles for each LDA topic. Also, number of unique terms that belong to each AK concept per blog article (column terms/article).

- **Design using patterns and components (24%):** Articles in this topic contain terms that refer to designing an architecture, as well as names of architectural patterns, components and quality attributes. For example, [B21] discusses different patterns and configurations (e.g. number of layers) to design web applications. Another example is [B22], which discusses over 12 design patterns to design microservice systems.

- **Achieve quality attributes using component design (22%):** Articles in this topic contain terms that refer to quality attributes, as well as terms that refer to components, connectors and patterns. For example [B23] explains how to achieve performance requirements through different component design using Apache Kafka. Another example is [B24], which compares RabbitMQ and Apache Kafka regarding their abilities to realize a component design that achieves certain quality attributes (e.g. scalability and performance).

- **Implement technologies and component design (20%):** Articles in this topic contain names of technologies, and terms that refer to implementation details (e.g. “example”), as well as terms that refer to components and connectors. For example, [B25] explains implementation details for Microservice architecture using Spring technologies. Another example is [B26], which explains how to implement event-driven integration between components using Apache Camel and Kubernetes.

- **Analyze decision factors (16%):** Articles in this topic contain terms that refer to factors that influence design decisions. These involve business requirements, constraints (e.g. “time”), and benefits of solutions (e.g. using terms “free” and “easy”). For example [B27] discusses loading data to the cloud. Several



TABLE III: Top terms, AK concepts, their frequencies (freq.), and number of unique terms per article (terms/article) for LDA topics.

LDA Topic	Terms	freq.	AK concepts	freq.	terms/article
<i>Design using patterns and components</i>	architecture	2253	<Pattern>	4514	2 to 5
	design	1617	<Component>	4509	6 to 16
	need	1250	<Quality_attribute>	1226	2 to 4
	web	1103	<Connector_data>	1010	5 to 6
<i>Achieve quality attributes using component design</i>	approach	814	<Connector>	302	7
	time	1213	<Component>	7430	6 to 16
	exchange	1104	<Connector_data>	2031	4 to 10
	need	1021	<Pattern>	2041	2 to 4
	topic	1010	<Connector>	907	7 to 10
<i>Implement technologies and component design</i>	support	716	<Quality_attribute>	805	2 to 3
	create	1103	<Technology>	7411	6 to 31
	example	1121	<Connector_data>	2212	4 to 7
	version	801	<Connector>	1607	6 to 13
	type	721	<Component>	713	7 to 9
<i>Analyze decision factors</i>	test	451	<Requirement>	301	4
	time	1132	<Requirement>	8741	4 to 14
	real	901	<Component>	2028	6 to 18
	report	851	<Connector_data>	402	4
	analysis	801			
<i>Compare and evaluate technologies</i>	free	703			
	vs	1803	<Technology>	4721	6 to 23
	support	1103	<Requirement>	1328	5 to 11
	feature	946	<Component>	1316	6 to 13
	time	810	<Quality_attribute>	953	2 to 4
<i>Design using multiple technologies</i>	offer	510	<Pattern>	407	4
	question	601	<Technology>	1553	6 to 27
	developer	506	<Requirement>	608	4 to 12
	development	450	<Component>	206	10
	design	406			
	architecture	403			

decision factors are discussed such as productivity, infrastructure complexity and performance. Another example is [B28], which discusses the integration between two technologies: SAP and power Apps, and explain the business benefits from such integration in terms of real time visibility and efficiency.

- **Compare and evaluate technologies (13%):** Articles in this topic contain terms that refer to comparing technologies such as “vs”, “feature”, and names of different technologies, as well as terms that refer to quality attributes and requirements. For example, in [B29], the two technologies ActiveMQ and RabbitMQ are compared regarding their features and quality attributes, such as scalability and interoperability.

- **Design using multiple technologies (5%):** Articles in this topic contain terms that refer to designing an architecture, as well as terms that refer to requirements and technologies. For example, [B30] explains the architecture of web applications using Microsoft technologies.

### C. Co-occurrences of types and topics of architectural blogs

Figure 2 shows significant co-occurrences between the LDA topics (see Section V-B), the qualitative topics (see Section V-A), and the types of blogs (see Section IV). From the co-occurrences, we can observe the following:

- *Some LDA topics logically co-occur with their corresponding qualitative topics.* For instance, the LDA topic *Implement technologies and component design* significantly co-occurs with the *How to implement* qualitative topic. Both topics focus on the implementation of an architecture. Similarly, the LDA topic *Compare and evaluate technologies* significantly co-occurs with the *Compare solutions* qualitative topic. Both topics focus on comparing solutions. This confirms that our

proposed approach that uses LDA and AK ontology (see Section III-D) succeeded to identify some topics that correspond to their qualitative counterpart.

- *Some LDA and qualitative topics complement each other.* On the one hand, LDA topics focus on certain types of architectural solutions (e.g. either technologies or patterns). On the other hand, qualitative topics focus on the purpose of an architectural blog. For example, *Design using patterns and components* significantly co-occurs with *Compare solutions*. This indicates that software engineers discuss the comparison of patterns and components to design an architecture of a system in these articles. Another example is *Implement technology and architecture*, which significantly co-occurs with *Elaborate and evaluate solution*. This indicates that software engineers provide implementation details (e.g. source code) to elaborate a solution, and to evaluate a solution (e.g. provide source code to benchmark a technology).

- *Most blog types significantly co-occur with a single topic.* With the exception of IT service blogs, four blog types significantly co-occur with one topic, and one blog type (*Personal blogs*) co-occurs with two topics. For instance, articles in *Community blogs* significantly co-occur with the *Design using patterns and components* LDA topic. While articles in *Personal blogs* significantly co-occur with the *Implement technology and architecture* topic and *How to implement* topics. This finding can help software engineers to find AK. For example, to find AK about patterns and components, software engineers should better search in community blogs, while to find AK related to the implementation of an architecture, software engineers should better search in personal blogs.

- *Few LDA and qualitative topics do not significantly co-occur with other topics.* The LDA topics *Achieve quality attributes using component design* and *Design using multiple technologies*, as well as the qualitative topic *How to design* do not significantly co-occur with other topics. Thus, articles in these topics either rarely co-occur with each other or equally co-occur with several other topics, but not one to a significant degree. For example, the *How to design* qualitative topic involves long articles that discuss patterns, components, technologies, and decisions factors. These articles equally co-occur with the LDA topics, because they do not focus on a specific architectural solution.

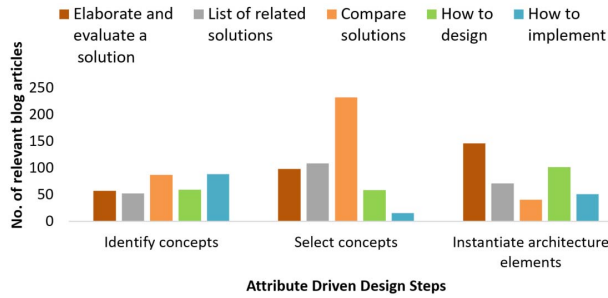
### RQ2 key takeaways:

- *Listing, elaborating, evaluating, and comparing architectural solutions* present the majority of AK (68%) in architectural blogs (see Section V-A).
- *Architectural patterns, component design and principles* are discussed in the majority (46%) of architectural blogs, followed by discussions on *technologies* and their implementations (38%) (see Section V-B).
- *Using LDA in combination with AK ontology* identifies reasonable architectural topics that correspond and complement the qualitative topics (see Section V-C).
- *Each type of blog* is specialized in one or two topics.

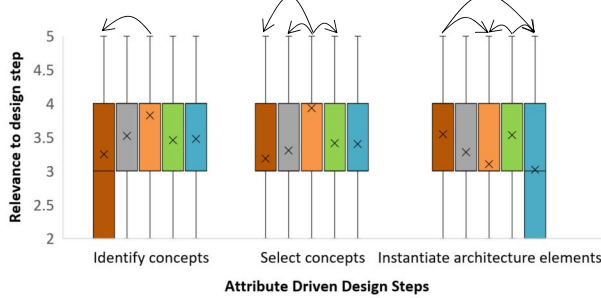


## VI. RQ3: RELEVANCE OF ARCHITECTURAL BLOG TOPICS TO ATTRIBUTE DRIVEN DESIGN STEPS

In the study of Soliman et al. [3], 53 software engineers evaluated the relevance of blog articles to perform tasks that apply the ADD steps (see Section III-B). We use this data to determine what topics of architectural blogs (see Section V) are retrieved by the Google search engine, and what topics are most relevant for software engineers to apply the ADD steps (see Section III-E). In the following sub-sections, we present the number of relevant blog articles (as retrieved by Google), and their relevance to each of the ADD steps, across the different topics (see Sections VI-A and VI-B).



(a) Number (No.) of relevant blog articles per design step as retrieved by Google



(b) Relevance of blog articles in each topic per ADD step, X presents the mean of relevance, and arrows refer to significant differences.

Fig. 3: Number and relevance of qualitative topics per ADD step

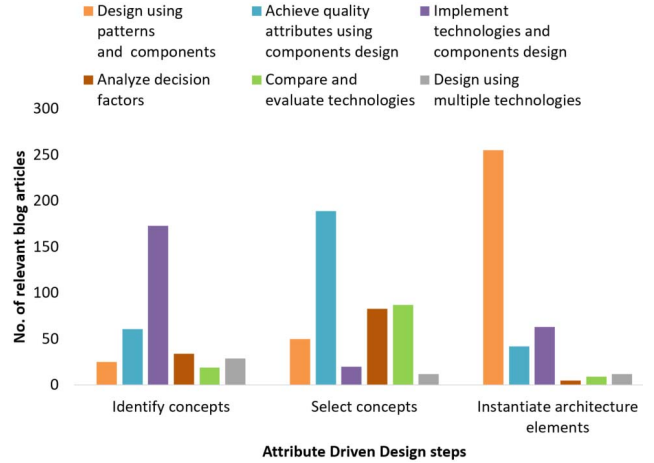
### A. Relevance of qualitative topics

Figure 3 shows the number of relevant blog articles for each qualitative topic as retrieved by Google, and their relevance as specified by practitioners. From Figure 3, we can observe that Google search results are different among the three ADD steps, and practitioners evaluate articles from different topics differently for each ADD step:

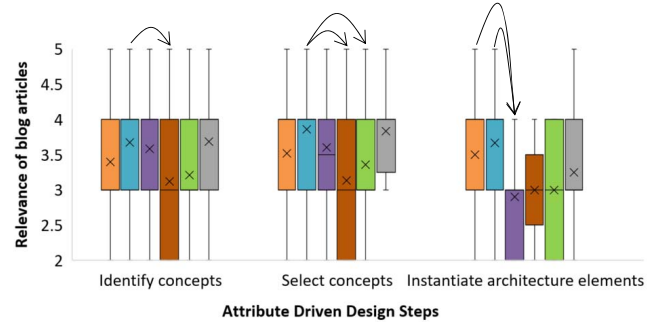
- For *Identify design concepts*, Google tends to retrieve the five topics of blogs similarly (see Figure 3a). However, our significance test shows that *Compare solutions* is significantly more relevant than *Elaborate and evaluate a solution* (see Figure 3b). Thus, Google tends to retrieve blog articles with low relevance from the *Elaborate and evaluate a solution* topic, which make it challenging for practitioners to find relevant AK for this step.
- For *Select concepts*, Google tends to retrieve articles from the *Compare solutions* topic significantly more than other topics (see Figure 3a). The Google search results are reasonable

for this step, and align with our significance test, which shows that articles from the *Compare solutions* topic are significantly more relevant to this step than other topics (see Figure 3b).

- For *Instantiate architecture element*, Google retrieves mostly articles from the *Elaborate and evaluate a solution* topic, followed by the *How to design* topic (see Figure 3a). Both topics have the highest relevance according to our significance test (see Figure 3b). However, Google retrieves some blog articles of low relevance from the *How to implement* and the *Compare solutions* topics, which negatively affect the effectiveness of Google to find AK for this step.



(a) Number (No.) of relevant blog articles per design step as retrieved by Google



(b) Relevance of blog articles in each topic per ADD step, X presents the mean of relevance, and arrows refer to significant differences.

Fig. 4: Number and relevance of architectural blog articles for each LDA topic, and for each Attribute Driven Design step

### B. Relevance of LDA topics

Figure 4 shows the number of relevant blog articles for each LDA topic as retrieved by Google, and their relevance as specified by practitioners. From Figure 4a, we can observe that Google tends to predominantly retrieve blog articles from a specific LDA topic to support an ADD step. Similarly to the qualitative topics, practitioners evaluate articles from different LDA topics differently for each ADD step (see Figure 4b):

- For *Identify design concepts*, Google retrieves most blog articles from the *Implement technologies and component design* topic. However, our significance test shows that the topic *Achieve quality attributes using component design* is significantly relevant to this step as well. Moreover, some

articles that belong to the *Analyze decision factors* topic have significantly low relevance for this ADD step. Thus, search approaches should consider to filter out blog articles from the *Analyze decision factors* topic to make this step effectively.

- For *Select design concepts* Google retrieves the majority of blog articles from the *Achieve quality attributes using component design* topic, followed by articles from the *Analyze decision factors* and *Compare and evaluate technologies*. However, our significance test shows that the topic *Achieve quality attributes using component design* is significantly more relevant than both topics *Analyze decision factors* and *Compare and evaluate technologies*. Thus, search approaches can filter out irrelevant articles from these two topics to ensure effective search for this ADD step.

- For *Instantiate architecture elements* Google tends to retrieve predominantly articles from the *Design using patterns and components* topic. However, our significance test shows that articles from the *Achieve quality attributes using component design* topic are also significant to this design step. In contrast, articles from *Implement technologies and component design* have significantly low relevance to this step, which could be filtered out by approaches to ensure effective search for AK.

#### RQ3 key takeaways:

- Practitioners evaluate the relevance of topics differently for each design step (see Figures 3b and 4b), and Google results are different among the ADD steps:
  - For *Identify design concepts* and *Select design concepts*, topics on comparing solutions, and achieving quality attributes are most relevant. But Google cannot distinguish highly relevant topics, and retrieves low relevant topics.
  - For *Instantiate architecture elements*, topics on elaborating and evaluating solutions such as patterns, and how to design are most relevant; Google retrieves articles on patterns, but misses other highly relevant topics.

## VII. DISCUSSION

### A. Implications for researchers

The results of RQ1 split architectural blogs into types as retrieved by Google. Researchers could use these types and dataset to develop specialized Web search approaches, which classify architectural blogs based on their types. For instance, an approach can allow practitioners to filter personal blogs from other types of architectural blogs, to facilitate searching for AK, because some types of blogs are specialized in certain architectural topics (see Figure 2). Furthermore, the types of blogs and our dataset facilitate future studies on AK. For instance, researchers recently analyze grey literature (including blog articles) to extract design decisions (e.g. on Microservices APIs [25]). Using the types of blogs and our dataset, researchers can collect a sample of architectural blog articles from specific types of blogs to explore certain topics.

The results of RQ2 help to understand the similarities and differences between the AK in architectural blogs and AK in Stack Overflow. In Table IV, we compare the AK in architectural blogs based on our results (see Section V), and

TABLE IV: Comparison of AK in architectural blogs (see Section V) and Stack Overflow architectural posts [26].

	Architectural blogs	Stack Overflow
<b>Types of architectural solutions</b>	46% components and patterns, 38% technologies, 16% decision factors	78% technologies and their features, 22% architectural components (incl. combinations with other solutions)
<b>Purpose of article or post</b>	48% elaborate, evaluate, compare solutions, 32% list solutions, 20% How to design or implement	50.7% evaluate solutions (incl. comparison), 40.7% synthesize solutions (incl. list solutions), 8.6% combine evaluation and synthesis

the AK in Stack Overflow based on Soliman et al. [26]. We could infer the following similarities and differences:

- The majority of articles in architectural blogs discuss patterns and components, while the majority of architectural posts in Stack Overflow discuss technologies and features.
- Most of the articles in architectural blogs and architectural posts in Stack Overflow evaluate and compare solutions.
- Architectural blogs involve articles that guide software engineers to design a system (i.e. How to design). These are complex articles with multiple steps and issues. On the other hand, posts in Stack Overflow focus on a single issue, and do not involve multiple steps or issues to design a system.

Based on the comparison between architectural blogs and Stack Overflow, researchers could benefit from both sources. For example, an approach can find AK on component design from blogs, and AK on technologies from Stack Overflow.

The results of co-occurrences between LDA topics and qualitative topics (see Section V-C) show that using LDA in combination with AK ontology (see Section III-D) identifies reasonable architectural topics that correspond and complement the qualitative topics. Thus, researchers could re-use this approach (LDA + AK ontology) to explore architectural topics in other sources (e.g. issue trackers [11] and mailing lists [27]).

The results of RQ3 can guide researchers to develop heuristics, which improve the effectiveness of AK search approaches. For instance, Soliman et al. [7] developed a heuristic-based specialized search approach to find relevant AK in Stack Overflow. Similarly, using the results of RQ3, researchers can develop heuristics to improve the search for AK in blogs by promoting highly relevant topics, and filtering low relevant topics. For example, when practitioners search for AK that pertains to the *identify design concepts* an approach can promote topics *Compare solutions* and *Achieve quality attributes using component design*, and filter out topics *Analyze decision factors* and *Elaborate and evaluate a solution*.

### B. Implications for practitioners

The results of RQ1 inform practitioners about the types of architectural blogs, which could help them to share and find AK on the Web. On the one hand, practitioners could decide on suitable types of blogs to share their AK. For instance, practitioners should better share AK in community blogs, because Web search engines find them better than other types. On the other hand, the types of blogs and our dataset can guide practitioners to search for AK on the Web. For instance, practitioners should go directly to specific websites for certain types of architectural blogs in our dataset, and search in these

websites directly, because some types of blogs are not well retrieved by Web search engines (e.g. personal blogs).

The results of RQ2 inform practitioners about the topics in architectural blogs. This can help practitioners to search for relevant articles in architectural blogs. For instance, practitioners could use terms and AK concepts of the LDA topics (see Table III) to search for specific architectural topics using keywords searches. Furthermore, the co-occurrences between blog types and topics (see Section V-C) guide practitioners to find certain architectural topics. For example, practitioners could search directly in community blogs (e.g. dzone.com) to find articles about architectural patterns and component design, because community blogs significantly co-occur with the LDA topic *Design using patterns and components* (see Figure 2).

The results of RQ3 guide practitioners on topics of architectural blogs, which are relevant to each of the ADD steps. Thus, if a practitioner requires AK in order to perform a specific ADD step, she can search specifically for the topics that are mostly relevant to this step. For example, regarding the *instantiate architecture elements* ADD step, practitioners can search for blog articles with topics *design using patterns and components* and *achieve quality attributes using component design*, because they have the highest relevance to this step.

### VIII. THREATS TO VALIDITY

1) *Construct validity*: One threat to construct validity is regarding the use of AK ontology [15], which map AK concepts into specific terms. The terms associated with each AK concept might not be complete, and thus might miss AK concepts in blog articles. Nevertheless, we followed an iterative process (see Section III-D) to enrich the AK ontology with new terms from architectural blogs to mitigate this threat.

2) *Reliability*: To identify types and topics of architectural blogs, we manually analyzed articles using grounded theory. This might involve a threat to the reliability of the study. Nevertheless, we carefully followed the steps of grounded theory, and discussed articles to ensure agreement among researchers. Finally, we measured the agreement among researchers (see Section III-C3), which indicated good agreement beyond chance. Regarding the LDA analysis, we provide the scripts and AK ontology online [19] to facilitate replicating the study.

3) *External validity*: One threat to the external validity is the limited number of analyzed architectural blog articles (i.e. 718 articles), which might not generalize to all architectural blog articles on the Web. Nevertheless, our sample is significant [28] with 95% confidence level and 3.58% error margin. Moreover, other studies that explored AK analyzed samples with comparable sizes: 858 Stack Overflow posts [26], 781 issues from issue trackers [11], and 980 decisions from mailing list [12]. Thus, our results provide a first hypothesis of AK in blogs, which can be well compared to related work.

### IX. RELATED WORK

We are not aware of any dedicated studies on architectural blogs and their contained AK. Thus, our study is the first study focusing on architectural blogs. In this section, we discuss

related work in the AK field, architectural grey-literature reviews, and studies on blogs in software engineering.

1) *Architectural knowledge*: Previous studies identified and modeled AK concepts on design decisions [2], their kinds [1], their reasoning [6], and solutions alternatives [29]. Although these studies identified AK concepts, they do not explore concrete sources of AK. Recent efforts explored AK in different software repositories and on the Web. For example, Gorton et al. [14] found AK regarding architectural tactics in technology documentation, and Bhat et al. [11] found AK regarding types of decisions in issue tracking systems. Furthermore, Bi et al. [13] explored quality attributes and architectural tactics in Stack Overflow. Recently, Fu et al. [27] developed an approach to identify decisions in mailing list. Moreover, Mahadi et al. [30] developed an approach to identify design discussions in pull requests. However, all previously mentioned approaches do not explore architectural blogs and their contained AK.

2) *Architectural grey-literature reviews*: Grey-literature may involve blog articles but also other sources (e.g. forums). Researchers recently utilized grey literature to answer research questions. For example, Soldani et al. [31] analyzed grey-literature to explore the benefits and drawbacks of microservices. While, Singjai et al. [25] analyzed grey-literature to determine relationships between Microservice APIs and Domain-Driven Design. However, these studies extract certain information (e.g. decision rules [25]) from blogs, but do not explore architectural blogs and its AK as one source of AK.

3) *Blogs in software engineering*: Blogs did not have much attention from software engineering researchers. Pagano et al. [16] was the first to analyze blog articles, and determined topics discussed by software developers such as features and domain concepts. Similarly, Parnin et al. [32] analyzed topics of software blogs (e.g. documentation, technology discussion), and surveyed practitioners to determine motivations and challenges to write blog articles. Recently, Williams et al. [33], [34] developed an approach to assess the credibility of blog articles. However, these studies analyzed blogs with different focus, and did not explore architectural blogs and AK.

### X. CONCLUSION AND FUTURE WORK

We aimed to explore architectural blogs, their types, topics and relevance to design steps. To this end, we analyzed architectural blog articles using qualitative and quantitative research methods. Our results show that architectural blog articles are shared in different types with different hosting and authorship policies, and discuss different topics that involve listing, elaborating, and evaluating solutions such as patterns, components and technologies. Furthermore, we found that some topics are more relevant for practitioners to make certain design steps than others. However, Web search engines (e.g. Google) do not always retrieve the most relevant topics for design steps. Our future work aims to develop approaches that automatically identify and classify blogs to improve the search for architectural knowledge in blogs.

## REFERENCES

- [1] P. Kruchten, P. Lago, and H. van Vliet, "Building Up and Reasoning About Architectural Knowledge," in *Quality of Software Architectures*, vol. 4214. Springer Berlin Heidelberg, 2006, pp. 43–58.
- [2] A. Jansen and J. Bosch, "Software Architecture as a Set of Architectural Design Decisions," in *WICSA*, 2005, pp. 109–120.
- [3] M. Soliman, M. Wiese, Y. Li, M. Riebisch, and P. Avgeriou, "Exploring Web Search Engines to Find Architectural Knowledge," in *2021 IEEE 18th International Conference on Software Architecture (ICSA)*. IEEE, 3 2021, pp. 162–172.
- [4] F. Buschmann, K. Henney, and D. C. Schmidt, *Pattern-Oriented Software Architecture, Volume 4: A Pattern Language for Distributed Computing*. Chichester, UK: Wiley, 2007.
- [5] M. Soliman, M. Riebisch, and U. Zdun, "Enriching Architecture Knowledge with Technology Design Decisions," in *Working International Conf. on Software Architecture*, 5 2015.
- [6] A. Tang, Y. Jin, and J. Han, "A rationale-based architecture model for design traceability and reasoning," *Journal of Systems and Software*, vol. 80, no. 6, pp. 918–934, 6 2007.
- [7] M. Soliman, A. R. Salama, M. Galster, O. Zimmermann, and M. Riebisch, "Improving the Search for Architecture Knowledge in Online Developer Communities," in *International Conference on Software Architecture 2018*. IEEE Inc., 7 2018.
- [8] R. Kazman and H. Cervantes, *Designing Software Architectures: A Practical Approach*. Addison-Wesley Professional, 2016.
- [9] C. Manteuffel, P. Avgeriou, and R. Hamberg, "An exploratory case study on reusing architecture decisions in software-intensive system projects," *Journal of Systems and Software*, vol. 144, pp. 60–83, 10 2018.
- [10] A. E. Hassan, "The road ahead for mining software repositories," *Proceedings of the 2008 Frontiers of Software Maintenance, FoSM 2008*, pp. 48–57, 2008.
- [11] M. Bhat, K. Shumaiyev, A. Biesdorf, U. Hohenstein, and F. Matthes, "Automatic extraction of design decisions from issue management systems: A machine learning based approach," in *Lecture Notes in Computer Science*, vol. 10475 LNCS. Springer Verlag, 2017, pp. 138–154.
- [12] X. Li, P. Liang, and Z. Li, "Automatic Identification of Decisions from the Hibernate Developer Mailing List," *ACM International Conference on the Evaluation and Assessment in Software Engineering (EASE '20)*, pp. 51–60, 4 2020.
- [13] T. Bi, P. Liang, A. Tang, and X. Xia, "Mining architecture tactics and quality attributes knowledge in Stack Overflow," *Journal of Systems and Software*, p. 111005, 5 2021.
- [14] I. Gorton, R. Xu, Y. Yang, H. Liu, and G. Zheng, "Experiments in Curation: Towards Machine-Assisted Construction of Software Architecture Knowledge Bases," in *IEEE/IFIP ICSA 2017*, 4 2017, pp. 79–88.
- [15] M. Soliman, M. Galster, and M. Riebisch, "Developing an Ontology for Architecture Knowledge from Developer Communities," *International Conference on Software Architecture*, 5 2017. [Online]. Available: <https://doi.org/10.1109/ICSAS.2017.31>
- [16] D. Pagano and W. Maalej, "How do open source communities blog?" *Empirical Software Engineering*, vol. 18, no. 6, pp. 1090–1124, 2013.
- [17] A. Williams and A. Rainer, "Toward the use of blog articles as a source of evidence for software engineering research," *Proceedings of the 21st International Conference on Evaluation and Assessment in Software Engineering (EASE'17)*, vol. Part F1286, pp. 280–285, 6 2017.
- [18] C. C. Silva, M. Galster, and F. Gilson, "Topic modeling in software engineering research," *Empirical Software Engineering* 2021 26:6, vol. 26, no. 6, 9 2021. [Online]. Available: <https://link.springer.com/article/10.1007/s10664-021-10026-0>
- [19] "Online material." [Online]. Available: <https://github.com/m-a-m-s/Explore-Architectural-Blogs>
- [20] A. Strauss and J. M. Corbin, *Basics of qualitative research: Grounded theory procedures and techniques*. Thousand Oaks, CA, US: Sage Publications, Inc, 1990.
- [21] K. J. Stol, P. Ralph, and B. Fitzgerald, "Grounded theory in software engineering research: A critical review and guidelines," in *Proceedings - International Conference on Software Engineering*, vol. 14-22-May-2016. IEEE Computer Society, 5 2016, pp. 120–131.
- [22] K. Pearson, "On a criterion that a given system of deviations from the probable in the case of correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling," pp. 157–175, 1900.
- [23] W. H. Kruskal and W. A. Wallis, "Use of Ranks in One-Criterion Variance Analysis," *Journal of the American Statistical Association*, vol. 47, no. 260, pp. 583–621, 1952.
- [24] R. A. Fisher, "Statistical Methods for Research Workers," pp. 66–70, 1992. [Online]. Available: [https://link.springer.com/chapter/10.1007/978-1-4612-4380-9\\_6](https://link.springer.com/chapter/10.1007/978-1-4612-4380-9_6)
- [25] A. Singjai, U. Zdun, and O. Zimmermann, "Practitioner Views on the Interrelation of Microservice APIs and Domain-Driven Design: A Grey Literature Study Based on Grounded Theory," *Proceedings - IEEE 18th International Conference on Software Architecture, ICSA 2021*, pp. 25–35, 3 2021.
- [26] M. Soliman, M. Galster, A. R. Salama, and M. Riebisch, "Architectural knowledge for technology decisions in developer communities: An exploratory study with StackOverflow," *Working International Conference on Software Architecture (WICSA)*, pp. 128–133, 7 2016.
- [27] L. Fu, P. Liang, X. Li, and C. Yang, "A machine learning based ensemble method for automatic multiclass classification of decisions," *ACM International Conference on Evaluation and Assessment in Software Engineering (EASE 2021)*, pp. 40–49, 6 2021.
- [28] K. A. Neuendorf, *The Content Analysis Guidebook*, 2nd ed. SAGE Publications.
- [29] O. Zimmermann, J. Koehler, F. Leymann, R. Polley, and N. Schuster, "Managing architectural decision models with dependency relations, integrity constraints, and production rules," *Journal of Systems and Software*, vol. 82, no. 8, pp. 1249–1267, 2009.
- [30] A. Mahadi, N. A. Ernst, and K. Tongay, "Conclusion stability for natural language based mining of design discussions," *Empirical Software Engineering*, 1 2022. [Online]. Available: <https://link.springer.com/article/10.1007/s10664-021-10009-1>
- [31] J. Soldani, D. A. Tamburri, and W. J. Van Den Heuvel, "The pains and gains of microservices: A Systematic grey literature review," *Journal of Systems and Software*, vol. 146, pp. 215–232, 12 2018.
- [32] C. Parnin, C. Treude, and M. A. Storey, "Blogging developer knowledge: Motivations, challenges, and future directions," *IEEE International Conference on Program Comprehension*, pp. 211–214, 2013.
- [33] A. Williams and A. Rainer, "How do empirical software engineering researchers assess the credibility of practitioner-generated blog posts?" *ACM International Conference on the Evaluation and Assessment in Software Engineering (EASE 2019)*, pp. 211–220, 4 2019.
- [34] A. Williams, M. Shardlow, and A. Rainer, "Towards a corpus for credibility assessment in software practitioner blog articles," *ACM International Conference on Evaluation and Assessment in Software Engineering (EASE 2021)*, pp. 100–108, 6 2021.

## REFERENCES OF BLOG ARTICLES

- [B1] Evaluating message brokers – amazon sqs. [Online]. Available: <https://neiljbrown.com/2017/08/26/evaluating-message-brokers-amazon-sqs/>
- [B2] Microservices architecture: What, when, and how. [Online]. Available: <https://dzone.com/articles/microservices-architecture-what-when-how>
- [B3] Components involved in creating a robust micro service architecture. [Online]. Available: <https://www.javacodegeeks.com/2016/05/components-involved-creating-robust-micro-service-architecture.html>
- [B4] Why is loose coupling between services so important? [Online]. Available: <https://www.ben-morris.com/why-is-loose-coupling-between-services-so-important/>
- [B5] Best 6 stock market apis for 2022. [Online]. Available: <http://www.columbia.edu/~tmd2142/best-6-stock-market-apis-for-2020.html>
- [B6] Design patterns for microservices. [Online]. Available: <https://ngkest.er123.medium.com/design-patterns-for-microservices-9340b72407bb>
- [B7] 11 ways to improve json performance & usage. [Online]. Available: <https://stackify.com/top-11-json-performance-usage-tips/>
- [B8] 5 fundamentals to a successful microservice design. [Online]. Available: <https://techbeacon.com/app-dev-testing/5-fundamentals-successful-microservice-design>
- [B9] Best json library for java. [Online]. Available: <https://www.javaguides.net/2019/07/best-json-library-for-java.html>
- [B10] What are 10 microservices best practices? [Online]. Available: <https://www.devteam.space/blog/10-best-practices-for-building-a-microservice-architecture/>
- [B11] Performance comparison of several common json libraries in java. [Online]. Available: <https://www.programmersought.com/article/8810377980/>

- [B12] Synchronous vs. asynchronous request handling. [Online]. Available: <https://blog.trigent.com/synchronous-vs-asynchronous-request-handling/>
- [B13] Rabbitchat - chat server in python using tornado, sockjs & rabbitmq - also talking about tcp sockets, websockets & python gil. [Online]. Available: <https://medium.com/@anirbanroydas/rabbitchat-chat-server-using-tornado-sockjs-rabbitmq-also-talking-about-tcp-sockets-914a79bbcd2f>
- [B14] Rest api best practices — design examples from java and spring web services. [Online]. Available: <https://dzone.com/articles/rest-api-best-practices-with-design-examples-from>
- [B15] Pattern: Decompose by business capability. [Online]. Available: <https://microservices.io/patterns/decomposition/decompose-by-business-capability.html>
- [B16] A web api design methodology. [Online]. Available: <https://www.infoq.com/articles/web-api-design-methodology/>
- [B17] Routing topologies for performance and scalability with rabbitmq. [Online]. Available: <https://spring.io/blog/2011/04/01/routing-topologies-for-performance-and-scalability-with-rabbitmq/>
- [B18] Microservices with spring. [Online]. Available: <https://spring.io/blog/2015/07/14/microservices-with-spring>
- [B19] Consuming a restful web service. [Online]. Available: <https://openliberty.io/guides/rest-client-java.html>
- [B20] Integrate with dynamics crm data using apache camel2. [Online]. Available: <https://www.cdata.com/kb/tech/dynamicscrm-jdbc-apache-camel.rst>
- [B21] How to design a web application: Software architecture 101. [Online]. Available: <https://www.educative.io/blog/how-to-design-a-web-application-software-architecture-101>
- [B22] Design patterns for microservices. [Online]. Available: <https://dzone.com/articles/design-patterns-for-microservices-1>
- [B23] Kafka performance tuning — ways for kafka optimization. [Online]. Available: <https://medium.com/@rinu.gour123/kafka-performance-tuning-ways-for-kafka-optimization-fdee5b19505b>
- [B24] Which service: Rabbitmq vs apache kafka. [Online]. Available: <https://www.cloudkarafka.com/blog/which-service-rabbitmq-vs-apache-kafka.html>
- [B25] Microservices with spring. [Online]. Available: <https://spring.io/blog/2015/07/14/microservices-with-spring>
- [B26] Event-driven integration on kubernetes with camel & keda. [Online]. Available: <https://tomd.xyz/kubernetes-event-driven-keda/>
- [B27] Etl vs. elt and the benefits of data transformation in the cloud. [Online]. Available: <https://solutionsreview.com/data-integration/etl-vs-elt-and-the-benefits-of-data-transformation-in-the-cloud/>
- [B28] Power apps and sap middleware integration to enhance the supply chain paradigm. [Online]. Available: <https://www.softwebsolutions.com/resources/powerapps-integration-with-SAP.html>
- [B29] Activemq vs rabbitmq: Key features and benefits. [Online]. Available: <https://www.openlogic.com/blog/activemq-vs-rabbitmq>
- [B30] Advanced architecture for asp.net core web api. [Online]. Available: <https://www.infoq.com/articles/advanced-architecture-aspnet-core/>