Ten Years of eHealth Discussions on Stack Overflow

Pedro Almir M. Oliveira[1][a], Evilasio Costa Junior[1][b], Rossana M. C. Andrade[1][c], Ismayle S. Santos[1][d] and Pedro A. Santos Neto[2][e]

[1]*Group of Computer Networks, Software Engineering and Systems (GREat), Federal University of Ceará, Ceará, Brazil*

[2]*Laboratory of Software Optimization and Testing (LOST), Federal University of Piauí, Piauí, Brazil*

Keywords: eHealth Trends, Stack Overflow, Mining Software Repositories, Topic Modeling.

Abstract: Over the past decade, we have seen growth in the usage of technologies in health. However, few papers are addressing the perspective reported by practitioners during the development of healthcare solutions. This perspective is relevant to identifying the most used strategies in this area and what challenges persist. Thus, this work analyzed eHealth discussions from Stack Overflow (SO) to understand the eHealth developers' behavior. Using a KDD-based process, we got and processed 6,082 eHealth questions. The most discussed topics include manipulating medical images, electronic health records with the HL7 standard, and frameworks to support mobile health (mHealth) development. Concerning the challenges faced by these developers, there is a lack of understanding about the DICOM and HL7 standards, the absence of data repositories for testing, and the monitoring of health data in the background using mobile and wearable devices. Our results also indicate that discussions have grown mainly on mHealth, primarily due to monitoring health data through wearables.

1 INTRODUCTION

The eHealth was defined in 2001 as a research field resulting from the intersection of medical informatics, public health, and business (Eysenbach, 2001). Since then, this research area has been strengthened as a more significant population-share turns their attention to well-being and health issues (Black et al., 2011). Furthermore, the increase of computational paradigms such as the Internet of Things (IoT) also contributed to the eHealth strengthening. In IoT, for example, things like smartwatches can monitor an elderly in his/her house, sending relevant information to physicians for improving the user care (Almeida et al., 2016) (Andrade et al., 2017). Also, there is a large number of initiatives in several subareas, *e.g.*, image processing, electronic health records, mobile health, and machine learning applied to health.

These initiatives usually result in new technologies with direct benefits for their users' quality of life (Gaddi et al., 2013), and their development creates many other valuable data such as the crowd knowledge built from technical discussions among practitioners working in this area (Ponzanelli et al., 2013). This knowledge is helpful in many ways, whether to boost the development overcoming issues already faced by other developers (Silva et al., 2019), or even to highlight demands to be addressed by researchers (Barua et al., 2014). Unfortunately, this crowd knowledge is often diffused in different software repositories (Kitchenham et al., 2015), and it needs to be systematically mined to become intelligible.

Among these repositories, the Question & Answer (Q&A) websites have a significant relevance due to their use by practitioners to discuss strategies to solve programming challenges (Beyer et al., 2019). A striking Q&A website is the Stack Overflow (SO), which had one of the largest technology enthusiasts communities. Also, SO data is public and can be accessed through the Stack Exchange Data Explorer tool[1]. In addition to being the widely used tool by technology professionals, SO is significantly used for scientific studies (Uddin et al., 2021).

Despite the relevance and applicability of the knowledge found on this kind of website by technology professionals who work directly with eHealth, we

[a] https://orcid.org/0000-0002-3067-3076
[b] https://orcid.org/0000-0002-0281-2964
[c] https://orcid.org/0000-0002-0186-2994
[d] https://orcid.org/0000-0001-5580-643X
[e] https://orcid.org/0000-0002-1554-8445

[1]Stack Exchange Explorer: data.stackexchange.com

Oliveira, P., Costa Junior, E., Andrade, R., Santos, I. and Neto, P.
Ten Years of eHealth Discussions on Stack Overflow.
DOI: 10.5220/0010801000003123
In *Proceedings of the 15th International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC 2022) - Volume 5: HEALTHINF*, pages 45-56
ISBN: 978-989-758-552-4; ISSN: 2184-4305

have not seen studies aiming to systematically mine the knowledge that emerges from the discussion of eHealth topics in Q&A websites. In the literature, most studies have used public databases to analyze and predict trends in this area from the perspective of eHealth end-users (Kwon et al., 2020), or from the researchers' point of view (Anonymous, 2020).

In this way, this paper presents an exploratory analysis on the eHealth Stack Overflow discussions aiming to understand the eHealth development community behavior considering the ICT professionals' perspective. This investigation can support researchers in understanding the difficulties faced by eHealth developers. It can also be helpful for practitioners who want to know which technologies are most used in eHealth. Thus, three Research Questions (RQ) were designed to guide our study:

- **RQ1:** What technologies have been discussed in eHealth?

  Rationale: This question is essential to analyze the software engineering artifacts used by eHealth developers. This information can assist practitioners in choosing the most suitable technologies for new eHealth projects.

- **RQ2:** What eHealth subjects have been discussed in the Stack Overflow?

  Rationale: Our interest with this question is to find hot topics related to eHealth discussions. Moreover, these topics can be used as a taxonomy to organize the knowledge present in the Stack Overflow about eHealth development.

- **RQ3:** What types of questions asked by developers related to eHealth are most recurrent on Stack Overflow?

  Rationale: This question aims to investigate what kind of questions are demanding more attention in this community. These demands can point out interesting challenges for future studies.

We highlight that this work contributes to researchers and practitioners, pointing out trends and open demands after ten years of eHealth development. For example, our data analysis shows a need for improvements in image processing and health records standards and mHealth frameworks optimized for background monitoring.

The paper outline is: in Section 2, we present a background on the strategies used in this research; in Sections 3 and 4, we detail our study design and discuss the results, respectively; in Section 5, we discuss some challenges and limitations; in Section 6, we bring the related works; and, finally, in Section 7, we present our final considerations and future works.

2 STACK OVERFLOW

EHealth has the potential to reduce costs and improve the quality of healthcare services. However, there are still open development challenges like high availability, scalability, fault tolerance, data management, interoperability, security and privacy, and user experience (Oliveira et al., 2021). This paper aims to analyze the eHealth community using Stack Overflow discussions and, consequently, observe how developers deal with these challenges.

We used the discussion definition proposed by (Bandeira et al., 2019) to filter our target questions in SO. This definition excludes questions that have not received answers and questions in which all answers were provided by the same user who asked the question. The rationale for using this definition is that we understand that questions classified as discussions represent relevant topics for the community.

Regarding Stack Overflow, we decided to focus on this database because of its strength and representativeness for the software development community[2]. Figure 1 presents the essential elements of the questions and answers on this platform.

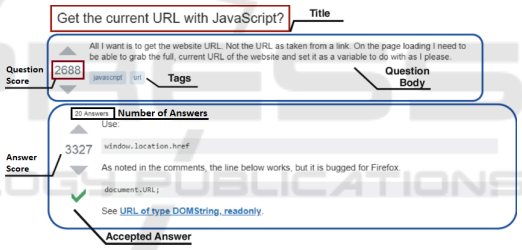

Figure 1: Q&A structure in SO (Mumtaz et al., 2019).

The title and the body contain information about the problem faced by the user. In addition, Q&A have tags used to categorize them and assist their retrieval. A question can have several answers provided by different users. The user who asked the question (the author) can accept one of the answers. Although it cannot be considered the best answer, the accepted answer usually represents the most suitable solution considering the author's opinion. Both questions and answers have a score computed by the difference between the up-votes and down-votes. In this work, we consider the question score to filter our dataset because it measures the relevance of a post to the community. Since this work represents an initial step in understanding this community, we have not considered the difference between up-votes and down-votes. Weighted correlations of these votes can be investigated in future work.

[2]SO Survey: insights.stackoverflow.com/survey/2019

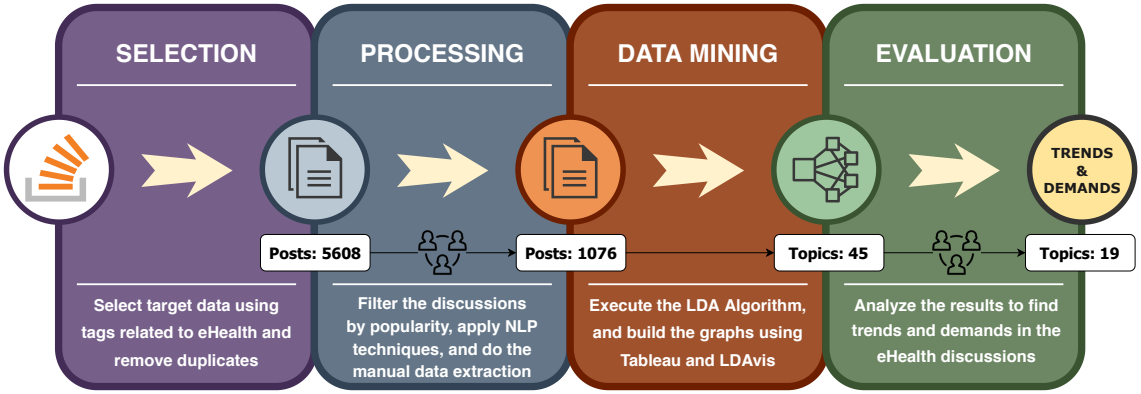

Figure 2: Our study design.

3 STUDY DESIGN AND OPERATION

This paper presents an empirical study focused on the discussions of the eHealth developers community. Thus, our study design was defined based on Knowledge Discovery in Databases (KDD) process (Fayyad et al., 1996). KDD was chosen given the need to extract knowledge from an unstructured database. Then, guidelines for data mining studies in software repositories were considered to adapt this process for evidence-based software engineering (Kitchenham et al., 2015). The following subsections detail the phases of our design study, which is illustrated in Figure 2.

3.1 Selection

We selected the English version of Stack Overflow as the data source due to its relevance to researchers and technology professionals (Beyer et al., 2019).

After the data source definition, we defined the search strategy. This activity required many trials to find a suitable strategy because i) it was not possible to use the main SO search field (that considers title and questions' body) since it can return inconsistent data; ii) the terms commonly used in the literature as synonyms for eHealth (according to MeSH controlled vocabulary[3]) did not return any data; iii) and there is no specific tag that characterizes eHealth questions.

After a comprehensive analysis, we realized a divergence between the MeSH terms and terms used by practitioners to categorize eHealth issues in SO. For example, the question "*How is it possible that Google Fit app measures number of steps all the time without draining the battery?*" – that should be classified as mHealth in MeSH – was tagged with the technology (Google Fit) and the platform (Android). We also tried many other tags defined by our expertise, but they did not return eHealth posts.

Thus, after this empirical analysis, we decided to use the term "health" as the starting point to execute a snowballing (Wohlin, 2014) search for tags. Initially, we searched for all tags related to the term "health" using Stack Overflow's tag search system[4]. For each tag found, we analyzed its definition to check if it is related to eHealth. The selected tags are shown in Table 1.

Table 1: Tags used to query the target data.

Subareas	Tags	Posts
DICOM	dicom, pydicom, fo-dicom, evil-dicom, dicomweb, niftynet, dcm4che	1,418
EHR	healthvault, intersystems-healthshare, hl7-fhir, ccd, cda, hl7, hapi, mirth, mirth-connect, hl7-v2, hl7-v3, hl7-cda, c-cda, hapi-fhir, fhir-server-for-azure, fhir-net-api, smart-on-fhir, openehr, nhapi, dstu2-fhir, hapi-fhir-android-library, snomed-ct, mdht, btahl7, medical, icd, hit	2,092
mHealth	health-kit, hkhealthstore, google-health, samsung-health, flutter-health, mapmyfitness, withings, heartrate, fitbit, hksamplequery, hkobserverquery, pedometer, google-fit, researchkit, samsung-gear-fit, strava, google-fit-sdk, wearables, hapi-fhir-android-library, carekit, medical, icd, hit	2,572
Total of Questions		6,082
Total of Questions without Duplicates		5,608

We found 6,082 questions using the selected tags (without date filters), and we got 5,608 after duplicate removal. The tags presented in Table 1 were classified into three subareas: DICOM, to represent questions about digital medical image processing; EHR, for issues focusing on Electronic Health Record; and mHealth, for questions about the development of Mobile Health applications. The authors proposed this classification following the tags' similarity and their

[3]MeSH: ncbi.nlm.nih.gov/mesh

[4]SO tag search system: https://stackoverflow.com/tags

community definitions. For sure, these are not the only subareas in eHealth. However, the tags found indicate that these three are the most discussed by developers on Stack Overflow.

3.2 Processing

After data collection, we performed three preprocessing activities. First, we carried out a filter considering the question popularity and the discussion definition (Bandeira et al., 2019). For the popularity filter, we used the third quartile of the question score. This filter's rationale is to reduce the noise that can be included by less popular questions or questions that only one user sent answers (Kavaler et al., 2013). We also adopted the third quartile to avoid bias regarding the definition of a hard threshold. Hence, our analyzes consider the questions that the community itself deemed most relevant. These filters reduced the number of questions from 5,608 to 1,112, implying a considerable data reduction. However, we decided to continue as our focus was on the most relevant questions using criteria defined by the eHealth developer community. Further analysis without this filter can be carried out in future works.

In the second activity, we performed a manual analysis to extract valuable data from each discussion. This process was conducted by two researchers with meetings to do the agreement check. This manual analysis was necessary because there is no automated method for classifying the question type and this information is essential to our research questions. For this classification, we decided to use the well known taxonomy proposed by (Treude et al., 2011) due to its characteristic of being able to be used in different study areas. This taxonomy has ten different types: how-to, discrepancy, environment, error, decision help, conceptual, review, non-functional, novice, and noise. For each discussion, we also done an open coding (Stol et al., 2016) to assist in identifying developers' concerns.

This manual analysis reduced the number from 1,112 to 1,076 due to the classification of 36 questions as noise (questions not related to eHealth but tagged with any of the tags defined in Table 1).

Since we are working with unstructured data, in the third activity, we used natural language processing techniques to improve LDA results (Thomas et al., 2014). These included removing code snippets, non-ASCII characters, punctuation, and words with less than three characters. Afterward, we chose the set of stop words from (Puurula, 2013) due to its size and availability on the Internet. Finally, we did not use word stemming algorithms because, in our empirical analysis, we realized that this process increases the difficulty of interpreting the topics.

3.3 Data Mining

For data mining, we performed a topic modeling with the LDA algorithm implemented by the Mallet tool (McCallum, 2002). As the number of topics depends on the problem investigated, and it is hard to define an idea number, we created models with different numbers of topics to empirically evaluate the most suitable one. We configured the tool to optimize hyperparameters every ten iterations and to train with 500 iterations. In the end, the chosen model had 15 topics for each eHealth subarea. The researchers hardly defined this number after a manual analysis of the models generated by the LDA, taking into account the trade-off between the model complexity and its representativeness.

3.4 Evaluation

In the evaluation phase, all results were analyzed by two researchers. The process of consolidating our open coding was carried out to define meaningful expressions that could characterize each discussion, highlighting recurring patterns.

Regarding LDA, initially, each researcher analyzed the topics independently using the LDAvis tool (Sievert and Shirley, 2014). Then, the divergences were discussed in meetings. Finally, after interpreting the initial 45 topics (15 topics for each eHealth subarea), we decided to group some of them, considering their semantic similarity. For example, Google Fit, HealthKit, Core Motion, and ResearchKit were grouped into mHealth Frameworks.

4 RESULTS AND DISCUSSION

We started our analysis considering 6,082 posts. After removing duplicates, we got 5,608. Then, applying filters related to popularity and the discussion concept, we obtained 1,076 relevant posts (our complete dataset can be accessed by bit.ly/3cHPEPL). From these posts, it was possible to create 45 topics later refined to meaningful 19 topics.

Before discussing the RQs, it is essential to highlight the increasing number of questions, especially in 2015 and 2016 (Figure 3.A). To better understand this behavior, Figure 3.B presents this data separated by the subareas: DICOM, EHR, and mHealth. Based on these graphics, we noted that the number of questions about DICOM and EHR presents a linear growth.

In contrast, between 2013 and 2015, the number of questions related to mHealth grew very sharply. This growth is probably associated with the launch of two frameworks to support mobile applications' development focused on healthcare: HealthKit (launched in September 2014) for the iOS platform and Google Fit (launched in October 2014) for the Android platform.

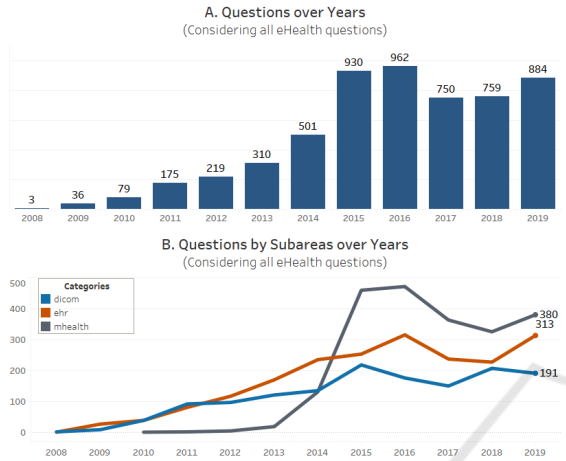

Figure 3: A. The number of eHealth questions over years. B. The distribution of questions considering the subareas. In this graphic, we considered the complete set with all questions to overview trends without filters.

4.1 RQ1: What Technologies Have Been Discussed in eHealth?

Regarding RQ1, we performed a manual data extraction to characterize the technologies discussed in this area. For this, we used an extraction form containing many fields, but not all fields presented significant data, probably because they are discussed in a transversal way. Thus, we discuss in this subsection the data for programming languages, operating systems, frameworks, API, libraries, and platforms.

As regard programming languages, the three most used languages for DICOM are Python (Van Rossum and Drake, 2011), C# (Hejlsberg et al., 2006), and Matlab (Herniter, 2000). But, many questions address Java (Arnold et al., 2000) and C++ (Stroustrup, 1984) too. For EHR, the most discussed languages are Java, C#, and Javascript (Flanagan and Matilainen, 2007). For mHealth, Swift, Java, and Objective-C (Knaster and Dalrymple, 2009) stand out. We observed that language usage is directly related to the development tools available in each subarea, as is the case of EHR with the HAPI Java API. Also, Python is widely used for image processing (Van der Walt et al., 2014). Swift and Objective-C are the basis for iOS development and Java for Android development.

Concerning operating systems and frameworks data, we found a few posts (27) in the DICOM and EHR subareas from which it was possible to extract these data. Hence, it was not possible to draw conclusions linking these technologies and subareas. However, for the mHealth subarea, we found a significant number of discussions bringing information about the used operating system and framework.

In the case of mHealth discussions considering operating systems (OS) and frameworks, we found 213 of 380 (56.05%) questions about Health Kit and 110 (28.95%) related to Google Fit. Together these frameworks represent 85% of mHealth questions. Both frameworks seek to facilitate the management of health and fitness data for users of smartphones and wearables. The ResearchKit framework, despite the low number of related questions, deserves attention because it is an open-source framework created to support medical research. However, we also realized that the framework is directly related to the operating system. Thus, we found a more significant number of questions about iOS, followed by Android OS.

Finally, Table 2 summarizes the eHealth APIs, platforms, and libraries more discussed in Stack Overflow. The sum of questions in this table is lower than the 1,076 used in the analysis because there are questions in which we could not extract all the data. Besides, to improve the table visualization, we suppressed technologies with just one question. However, our complete dataset is available online. This information can help practitioners choose the most suitable artifact, depending on the project's context.

4.2 RQ2: What eHealth Subjects Have Been Discussed in Stack Overflow?

For RQ2, we used the LDA to get the hot topics related to eHealth discussions from SO. As a result, we obtained a multidimensional model that correlates words and documents. Usually, the evaluation of this type of model is performed by specialists using the terms present in the topics. However, this activity is error-prone (Sievert and Shirley, 2014). Thus, many authors have proposed visualization methods to facilitate the evaluation of LDA models. Here, we used LDAvis (Sievert and Shirley, 2014).

We identified 45 topics, and they were refined to 19 more meaningful topics. In Figure 4, these critical topics are represented by the first tree level. The second level was used to give more details about the subjects. Together, these topics provide a high-level view of the subjects discussed by eHealth developers.

Regarding the DICOM subarea, most discussions deal with aspects such as Image Manipulation (34.81%) and Data Handling (22.24%). However, there is still a significant percentage of discussions about libraries (13.79%) and the DICOM Standard (Mildenberger et al., 2002) (13.77%), more specifically about Data Representation and the usage of UID tags. The python language is so prevalent in this subarea that we found a specific topic of questions about its application in medical image processing. We also identified a group of questions about extracting health information from images; questions focused on communication between DICOM servers, and discussions about the libraries used in this subarea.

Regarding the EHR, the most discussed subjects are HL7 (28.78%), Data formatting, storage and encoding (16.60%), and Communication issues (16.55%). We also have topics about Medical information (13.80%), FHIR (12.63%), Interoperability with Mirth Connect (6.70%), and OpenEHR (4.94%). As expected, we found many discussions about the HL7 and FHIR standards. These standards' relevance is due to their adoption by healthcare companies and by governments of many countries (Bender and Sartipi, 2013). FHIR emerged as an evolution for HL7v3, but due to difficulties in migrating to newer versions, many developers still use HL7v2 and HL7v3. Another important subject related to the standardization of electronic health records is OpenEHR. It provides open-source specifications for EHR systems (Kalra et al., 2005). This model has been strengthening as consumers, and healthcare professionals understand the benefits of a free and safe health data exchange.

For mHealth, the most discussed subjects are Framework (39.56%), Data Monitoring (26.81%), Development Issues (19.77%), Wearables (10.81%), and Security and Privacy Policies (3.05%). In this subarea, the step counting subject calls our attention due to its large number of questions. This behavior was unexpected once a large number of commercial applications already have this functionality. It was expected that this topic would be a consolidated knowledge in this community. Therefore, we can raise two hypotheses related to this concern: i) different mobile hardware generates different health measures, and commercial apps address these specificities internally; ii) there is a lack of approaches to ensure the correctness of measurements obtained by these apps.

Another subject that deserves attention is the difficulties faced when the developer needs to perform tasks in the background. Many developers still have problems using this feature, indicating a need for improvements in mHealth frameworks. We also observed a significant interest in monitoring health data such as heart rate and sleep using wearables. In fact, many studies have used wearables to detect atypical health conditions (Almeida et al., 2016). With the gradual transfer of this knowledge to the industry, dis-

Table 2: APIs, Platforms and Libraries discussed in eHealth subareas.

Subareas	Name	Description	# of questions
		APIs	
EHR	HAPI	HAPI is an open source HL7 parser for Java.	23
	nHAPI	It was developed from HAPI project to help .Net developers to manage HL7 messages.	5
	FHIR .Net API	It is used by developers to build FHIR client/server applications.	2
mHealth	Fitbit API	API that enables the communication with Fitbit devices like Fitbit Ionic.	7
	Withings API	This API allows the development of apps using the Withings devices.	4
	Strava API	It enables the development of apps that use the Strava athletes data.	2
	Google Fit REST API	This API helps developers to manage health data from Google Fitness Store.	2
		Platforms	
DICOM	.Net	.Net is a Windows platform used to build different types of applications.	11
	ClearCanvas	Platform for viewing, management, and distribution of DICOM images.	10
	NiftyNet	Open source CNN platform based on TensorFlow for medical image analysis.	6
EHR	Mirth Connect	It is used to connect healthcare systems using standard like HL7, FHIR, and others.	35
	.Net	.Net is a Windows platform used to build different types of applications.	10
	Node.js	Node.js is a platform to support the development of server-side JS applications.	6
mHealth	Node.js	Node.js is a platform to support the development of server-side JS applications.	5
	Xamarin	Xamarin is a .Net platform and enables the development of native mobile apps.	3
	Ionic	Platform for web developers who wants to build cross-platform mobile apps.	3
	.Net	.Net is a Windows platform used to build different types of applications.	3
		Libraries	
DICOM	pydicom	Python library to work with medical image datasets.	22
	dcm4che	Set of libraries used in DICOM healthcare applications.	12
	Fellow Oak DICOM	DICOM library in C# for .Net Platform.	12
	Simple ITK	Set of tools for image analysis and was originally developed in C++.	5
	GDCM	Grassroots DICOM (GDCM) is an open source library to work with DICOM standard	3
	JAI ImageIO	Java library that provide methods for image processing and image analysis.	3
	pynetdicom	Python implementation of the DICOM networking protocol.	2
	Papaya	Javascript medical image viewer.	2

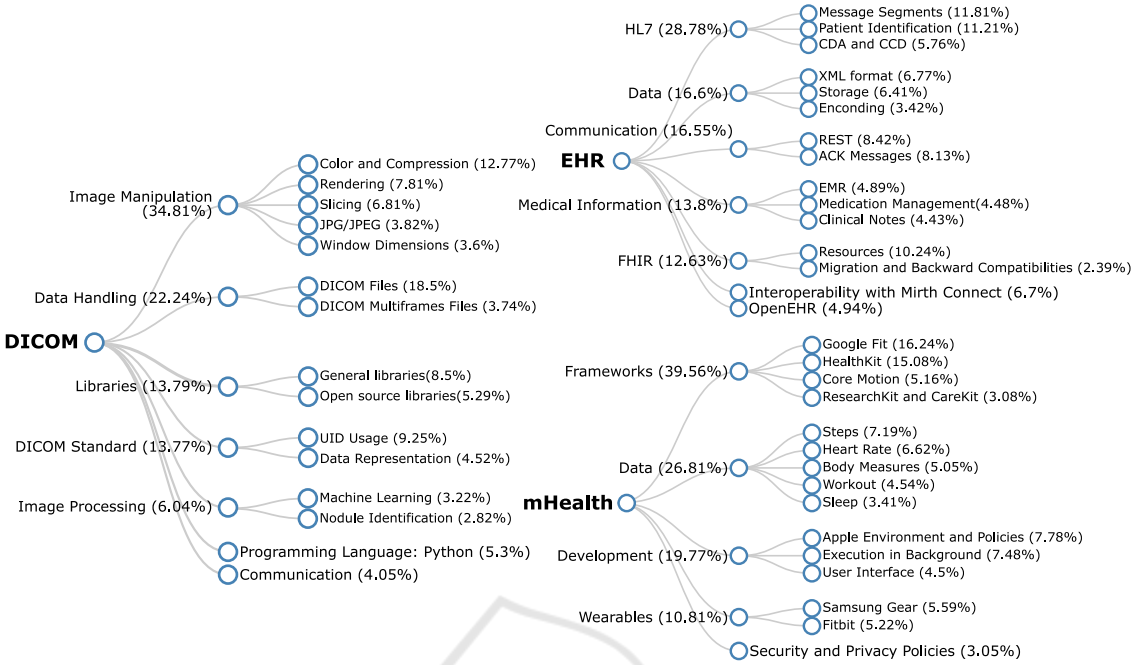

Figure 4: Most discussed topics in Stack Overflow by eHealth developers.

cussions about wearables should increase.

Finally, we expect that the number of user interface discussions (4.5%) should increase in the following years from the understanding that the user experience is a crucial factor in the acceptance of these new healthcare technologies (Zapata et al., 2015).

4.3 RQ3: What Types of Questions Asked by Developers Related to eHealth Are Most Recurrent on Stack Overflow?

Due to the Stack Overflow community's strengthening, the answers found in this website can be seen as an extension to the formal documentation for many technologies (Treude et al., 2011). However, an essential step in understanding the dynamics for a particular area is to observe what type of question developers ask (Beyer et al., 2019). In our case, we chose the taxonomy proposed by (Treude et al., 2011).

Thus, the first step to answer RQ3 was focused on the type of question that eHealth developers ask. Figure 5.A presents the types of questions considering each of the subareas. As expected, in all subareas, the how-to type presented the most number of questions. In fact, many developers use SO to find instructions about how to do a specific task, *e.g.*, *"How can I connect to the FitBit Zip over Bluetooth 4.0 LE?"*.

After the how-to type, the distribution of the other types for each subarea was different. For the DICOM subarea, the second question type with the highest number of questions was novice. We observed during the manual analysis that a large number of questions contain passages such as *"I'm new to processing DICOM images"*. This can indicate a substantial difficulty in starting working with medical image processing due to the complexity of the technologies and standards present in this area. The third biggest question type was decision help. Many DICOM questions seek opinions on the use of technologies or the best way to architect a solution (see Question ID:

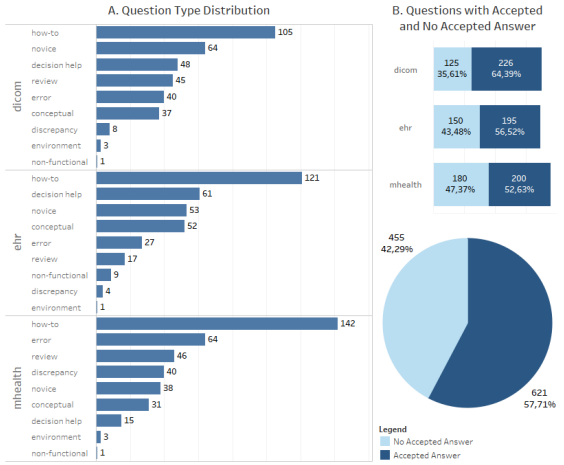

Figure 5: A. discussions by the question types; and B. questions with accepted and no accepted answer.

10390121). We found similar behavior in the EHR area, differing only in the order of the decision help and novice types.

For EHR, we also highlight a large number of questions of the conceptual type, *e.g.*, *"What Difference HL7 V3 and CDA?"*. For the mHealth subarea, we found 64 questions classified as error and 46 classified as a review. For this subarea, the questions are more practical, focusing on errors faced during development or requesting assistance with code review. For instance, we identified questions such as *"Why am I getting the error with the Ionic plugin healthkit?"*.

Figure 5.B shows the number of questions with accepted and no accepted answers considering each subarea and the global set. The hypothesis related to this information is that the higher the percentage of questions with an accepted answer, the more mature the community is about technologies that are being discussed. In this regard, DICOM has the most significant number of questions with accepted answers (64.39%), followed by EHR (56.52%) and mHealth (52.63%). Actually, mHealth is an area that has been gaining prominence in the last five years, and many aspects of the development of this type of solution still have not a community consensus.

Another finding of the RQ3 refers to the low number of questions focused on non-functional requirements (only 11 questions - 5 related to security and 6 related to performance). This fact is worrying due to the criticality of healthcare systems. These systems are expected to be developed with non-functional requirements such as security and performance.

After analyzing which type of discussions the eHealth developers are asking, we decided to deepen the investigation by conducting a pair-reviewed open coding process. This process aims to extract and systematize knowledge from textual data.

In DICOM, the main concerns are the use of unique identifiers (UID) of the DICOM standard, data for testing and validating applications, meaning and purpose of DICOM tags, manipulation of 3D images, and visualization of medical images in web browsers. As an example of questions related to these concerns, we have: *"Is it true that DICOM Media Storage SOP Instance UID = SOP Instance UID?"* (which has a score equals to 7 and 2,358 views); *"Where is possible download .dcm files for free?"* (with 30 as score and 58,540 views); *"How to decide if a DICOM series is a 3D volume or a series of images?"* (which has 9 as score and 3,767 views).

For EHR, the two main concerns are about the HL7 standard. First, similar to the DICOM, in EHR, there is a concern about data for testing and validating applications. Moreover, interoperability and the different versions of the HL7 standard still generate discussions. An interesting question to show these concerns is *"What Difference HL7 V3 and CDA?"*.

Regarding mHealth, the main concerns are associated with HealthKit usage, heart rate monitoring, security issues, and best practices to recover health data. In addition, we highlight the concern of dealing with data acquisition in the background due to its criticality for this type of application, its complexity related to battery consumption, and the recurrence of problems reported by developers. For example, the following question had more than 12,180 views and still has no accepted answer: *"Healthkit background delivery when app is not running"*. Another example of discussion close related to the main concerns in mHealth is *"How is it possible that Google Fit app measures number of steps all the time without draining battery?"*, which was visualized more than 74k times. By the way, this was the question with the most views in our dataset.

4.4 Findings Summary and Practical Implications

In this work, we analyzed ten years of the most popular eHealth discussion in the Stack Overflow. Our results can be used to understand the trends and difficulties faced by eHealth developers. It can also be helpful for practitioners who want to know which technologies are most used in the eHealth subareas. In this subsection, we summarize our findings and present their practical implications.

Concerning technologies (RQ1), each subarea has more adopted tools and, consequently, more discussions on Stack Overflow. For example, DICOM has as reference the library written in Python called pydicom. On the other hand, EHR has many questions about the open-source Java API called HAPI. Moreover, for mHealth, two frameworks stand out: HealthKit for the iOS operating system using Swift and Objective-C and Google Fit for the Android operating system using Java.

Regarding the subjects discussed in this area (RQ2), the results present many DICOM questions about image manipulation and data handling. In EHR, the most frequent subjects are the HL7 standard, data formatting, storage and encoding, and communication issues. For mHealth, we have discussions about frameworks (mainly Google Fit and HealthKit), data monitoring, development issues, wearables, and security and privacy policies.

In RQ3, we found that how-to questions are prevalent in all three subareas considered in this study. DICOM and EHR have many questions classified as

decision help and novice. This can indicate a greater barrier for developers who are starting in these subareas. mHealth, differently, has many questions that present specific errors or that request code snippets review. Another point that deserves attention is the low number of non-functional questions, which may indicate a low interest for non-functional requirements.

Deepening the RQ3 analysis, we found the following concerns: i) interpretation and usage of the DICOM and HL7 standards; ii) manipulation of 3D images; iii) DICOM web viewer; iv) data for testing and validating both DICOM and EHR applications; v) interoperability using Mirth Connect; vi) data types and background execution with HealthKit; vii) heart rate monitoring; viii) authentication and permission, and ix) best practices to query health data.

Finally, in addition to present a systematic analysis for the eHealth knowledge mined in SO and discuss a snapshot of ten years of discussions in this area, this study also has other practical implications:

Research Opportunities: this study can also represent a kick-off for further investigations regarding the gaps that were found, such as surveying with eHealth developers to understand better the difficulties faced when working with mHealth; or propose testing approaches focused on health tracking apps.

Towards an eHealth Development Taxonomy: our analysis pointed out that the eHealth discussions can be grouped into three major groups: DICOM, EHR, and mHealth. Specializing in this analysis, we found the topic tree (Figure 4). These topics can be used as a reference to classify new questions in order to improve the solution recommendations. For example, it is possible to use our LDA model to label GitHub eHealth open-source repositories and use that information to suggest possible solutions whenever a new question is registered on SO. This can increase the synergy among these software communities.

Beginners Guide: for practitioners who are starting to work with eHealth, this paper can support the decision about which tool to use. We have seen a strengthening of Python tools to deal with medical image processing, the consolidation of Java tools for EHR, and polarization between HealthKit and Google Fit for mobile health.

5 VALIDITY THREATS

The discussion of limitations and threats is essential when conducting evidence-based software engineering studies (Kitchenham et al., 2015). It helps to understand the outcome of confidence. In our study, we identified some threats to validity and sought to mitigate them through the conducting phase.

The first threat showed up during the selection of the data source and target tags. Despite the choice of just one database, we consider that the English version of Stack Overflow has high representativeness about developers' discussions. This website has also been used for the development of many scientific studies (Chen et al., 2019) (Beyer et al., 2019).

We sought to follow a systemic approach to reduce bias to choose the tags to compose our query. Initially, we considered eHealth synonyms defined by the MeSH vocabulary as tags. This controlled vocabulary has significant relevance for indexing papers in the life sciences, and it is used as a reference to support the building of search strings for systematic reviews (Lynch et al., 2019) (Salvador-Oliván et al., 2019). However, we faced a limitation concerning the different terminology used by researchers and practitioners. When we tried to use the terms suggested by MeSH, no results were found. With that, we decided to use a snowballing strategy to include all health-related tags. This strategy returned a significant number of questions. We understand that this decision can raise another threat once some questions could be left out of the study. However, after a detailed analysis of the questions, we considered them suitable to draw our conclusions. Also, the manual analysis removed noises, *i.e.*, questions whose tag does not match the content of the question. In this step, we found and removed 36 discussions classified as noise.

Regarding data volume, the filters applied in this study followed criteria already validated in the literature (Kavaler et al., 2013) (Bandeira et al., 2019). We chose to analyze only the most relevant questions for the eHealth community using the distribution of score metrics. Although small, we argue that our final dataset (1,076 questions) is highly representative. Further investigations can use the complete dataset.

We also sought to mitigate threats related to manual data extraction bias involving two researchers in this activity. We held meetings to discuss the results and the topic interpretation. In these meetings, we noticed some limitations. For example, the technologies identified in the Stack Overflow questions were classified in API, frameworks, libraries, or platforms, considering their description. However, we noted that these descriptions do not always follow definitions already established in software engineering. Thus, it would be interesting to adopt clear definitions for each type of software artifact in future work.

Another limitation of our analysis concerns the developers' profiles. In practice, there are several types of developers (such as mobile, web, and full-stack developers), and we did not make any distinc-

Table 3: Comparison among our proposal and the related works.

Work	Method/Technique	Data source	Perspective	Results
(Drosatos et al., 2017)	Topic Modeling with LDA	PubMed Papers	Researchers	Literature trends
(Drosatos and Kaldoudi, 2020)	Probabilistic techniques	PubMed Papers	Researchers	Literature trends
(Ahmed et al., 2019)	Systematic Review of Reviews	Review Papers	Researchers	Challenges
Our work	Topic Modeling with LDA	Stack Overflow Questions	ICT Professionals	Trends and Demands

tion between these profiles. Nevertheless, this point can represent an interesting opportunity from comparing our results, taking into account these profiles.

Regarding the topic interpretation, we sought to improve its reliability by conducting – with two researchers – an evaluation using different topic numbers. This evaluation considered the trade-off between model complexity and the topic's representativeness, and it was performed using the LDAvis.

6 RELATED WORK

In the literature, several papers use data mining techniques to analyze patterns in software repositories. There are also many studies seeking to map eHealth papers systematically. However, we did not find papers focused on observing this area's behavior from ICT professionals' perspectives and using the Stack Overflow as a data source. Thus, the papers listed in this section are related to our work by the research method or the study area.

The work developed by (Drosatos et al., 2017) has an objective similar to that proposed in our work. The authors modeled topics based on the PubMed Digital Library papers to extract trends in this literature. They used the MeSH controlled vocabulary to build the search string. They recovered 25,824 publications until December 2016. After the refinement stage, they got 19,825 papers. The LDA was applied on the titles, keywords, and abstracts, considering a number of 160 topics, which experts later reviewed. The most frequent topics found were wearables, randomized control trials, legal issues & ethics, eye disease, and teleconsultation. The difference to our approach lies in the purpose of characterizing the area from the professionals' perspective.

(Drosatos and Kaldoudi, 2020) used probabilistic techniques to analyze the literature related to the eHealth field. The authors considered the titles and abstracts of 23,988 articles (collected in PubMed Digital Library between December 31, 2017, and May 8, 2018) to compose the study corpus. The topic modeling identified 100 meaningful subjects into the service model, disease, behavior, and lifestyle categories. The results indicated a shift in focus from the DICOM to the mHealth subarea. We also observed an increase in Stack Overflow discussions focused on mobile health development, reinforcing this trend.

The work written by (Ahmed et al., 2019) presents a systematic review of reviews (*i.e.*, a tertiary study) carried out to identify research opportunities in this area. The authors analyzed 47 papers published in several digital libraries between January 2010 and June 2017. As a result, they highlight five challenging areas (stakeholders and system users, technology and interoperability, cost-effectiveness and start-up costs, legal clarity and legal framework, and local context and regional differences) and four areas of opportunity (participation and contribution, foundation and sustainability, improvements and productivity, and identification and application).

Regarding the stakeholder and system users' challenges, the authors mentioned the need to better integrate functional and non-functional requirements to the design and implementation of eHealth applications. Our results corroborate this point of view since there are still few discussions about non-functional requirements for eHealth within Stack Overflow. Concerning the technology and interoperability challenges, the authors noted that despite standards such as FHIR, its adoption is still slow and requires much effort. In fact, our data shows many questions in which the developers are looking for support to make decisions about EHR standards. There are also many discussions reporting difficulties with legacy systems that use older versions of HL7. Finally, unlike the other papers, our work performed a topic modeling with LDA to find trends and demands in Stack Overflow eHealth questions, taking into account the perspective of ICT professionals. Table 3 presents a comparison among our proposal and related works.

7 FINAL REMARKS

The eHealth term was defined more than 15 years ago. During this period, this area faced several changes driven by the emergence of new healthcare technologies. Recently, with the cost reduction of wearable devices and due to its ability to monitor many different aspects of its users' health, this area has gained new *momentum*. Thus, it is possible to find many papers proposing new technological artifacts and other stud-

ies that seek to map these advances systematically.

This work proposed an investigation into the discussions associated with eHealth development, considering the practitioners' perspective and using the Stack Overflow as the data source. Initially, we got 6,082 posts. Then, after removing duplicates and applying popularity filters, we found 1,076 discussions. So, we used manual extraction and topic modeling to understand the behavior of eHealth developers in SO. In this first study, we have done a descriptive analysis to understand the eHealth area from the developers' point of view. Using our results, it is possible to conduct further in-deep investigations on this area.

Moreover, we observed a growing trend regarding mHealth discussions. The data also revealed three clusters of questions in SO: DICOM, EHR, and mHealth. The most frequent discussions in the DICOM (that includes questions about digital medical image processing) and EHR (with issues focusing on Electronic Health Record) is related to the decision-making process during the development of solutions and the assistance to novices. In mHealth (that includes questions about mobile health applications), the discussions are more technical and specific, focusing on error resolution and code review. We also found just a few issues associated with non-functional requirements despite the relevance of safety, performance, and usability for health applications.

Regarding technologies, there is a direct correlation between programming languages and the most used artifact in the subareas. Python and the pydicom library for DICOM, Java and the HAPI API for EHR, Swift, and the HealthKit framework for mHealth stand out. After interpreting LDA's topics, we concluded that the most discussed subject in DICOM is image manipulation. In EHR, it is the HL7 standard, and for mHealth is the framework for development. Finally, the most significant concerns that arise from conceptual issues are understanding DICOM and HL7 standards, data for testing and validating, and the monitoring of health data in the background.

To conclude, we argue that this work can help practitioners know trends in this area, like the most used technologies. It can also help researchers identify opportunities such as improving DICOM and HL7 standards, the development of more suitable techniques for testing healthcare applications, the availability of datasets that assist these activities, and the investigation of why there are few questions about non-functional requirements in this area. Another research opportunity would be to investigate a tool to help developers deal with monitoring data in the background and optimize battery consumption.

CODE AND DATA AVAILABILITY

All codes and data are publicly available. The SO query can be accessed at the link: data.stackexchange .com/users/32389. Codes used to build the LDA models are available through the link: github.com/ pedroalmir/trends-ehealth. Also, it is possible to access the images attached to this paper (with a higher resolution) through the link bit.ly/3lhKhvP.

ACKNOWLEDGMENTS

The authors would like to thank CNPQ for the Productivity Scholarship of Rossana M. C. Andrade DT-2 (N^o 315543 / 2018-3), for the Productivity Scholarship of Pedro de A. dos Santos Neto DT-2 (N^o 315198 / 2018-4), and CAPES that provided to the Evilasio C. Junior a Ph.D. scholarship.

REFERENCES

Ahmed, B., Dannhauser, T., and Philip, N. (2019). A systematic review of reviews to identify key research opportunities within the field of ehealth implementation. *Journal of telemedicine and telecare*, 25(5):276–285.

Almeida, R. L., Macedo, A. A., de Araújo, Í. L., Aguilar, P. A., and Andrade, R. M. (2016). Watchalert: Uma evolução do aplicativo falert para detecção de quedas em smartwatches. In *Anais Estendidos do XXII Simpósio Brasileiro de Sistemas Multimídia e Web*, pages 124–127. SBC.

Andrade, R., Carvalho, R., de Araújo, I., Oliveira, K., and Maia, M. (2017). What changes from ubiquitous computing to internet of things in interaction evaluation? In *International Conference on Distributed, Ambient, and Pervasive Interactions*, pages 3–21. Springer.

Anonymous, B. (2020). Anonymous title (blind review). *PloS one*, 15(7):00.

Arnold, K., Gosling, J., Holmes, D., and Holmes, D. (2000). *The Java programming language*, volume 2. Addison-wesley Reading.

Bandeira, A., Medeiros, C. A., Paixao, M., and Maia, P. H. (2019). We need to talk about microservices: an analysis from the discussions on stackoverflow. In *16th International Conference on Mining Software Repositories*, pages 255–259. IEEE Press.

Barua, A., Thomas, S. W., and Hassan, A. E. (2014). What are developers talking about? an analysis of topics and trends in stack overflow. *Empirical Software Engineering*, 19(3):619–654.

Bender, D. and Sartipi, K. (2013). Hl7 fhir: An agile and restful approach to healthcare information exchange. In *Proceedings of the 26th IEEE int. symposium on computer-based medical systems*, pages 326–331. IEEE.

Beyer, S., Macho, C., Di Penta, M., and Pinzger, M. (2019). What kind of questions do developers ask on stack overflow? a comparison of automated approaches to classify posts into question categories. *Empirical Software Engineering*, pages 1–44.

Black, A. D., Car, J., Pagliari, C., Anandan, C., Cresswell, K., Bokun, T., McKinstry, B., Procter, R., Majeed, A., and Sheikh, A. (2011). The impact of ehealth on the quality and safety of health care: a systematic overview. *PLoS medicine*, 8(1).

Chen, H., Coogle, J., and Damevski, K. (2019). Modeling stack overflow tags and topics as a hierarchy of concepts. *Journal of Systems and Software*, 156:283–299.

Drosatos, G. and Kaldoudi, E. (2020). A probabilistic semantic analysis of ehealth scientific literature. *Journal of telemedicine and telecare*, 26:414–432.

Drosatos, G., Kavvadias, S. E., and Kaldoudi, E. (2017). Topics and trends analysis in ehealth literature. In *EMBEC & NBC 2017*, pages 563–566. Springer.

Eysenbach, G. (2001). What is e-health? *Journal of medical Internet research*, 3(2):e20.

Fayyad, U., Piatetsky-Shapiro, G., and Smyth, P. (1996). From data mining to knowledge discovery in databases. *AI magazine*, 17(3):37–37.

Flanagan, D. and Matilainen, P. (2007). *JavaScript*. Anaya Multimedia.

Gaddi, A., Capello, F., and Manca, M. (2013). *eHealth, care and quality of life*. Springer.

Hejlsberg, A., Wiltamuth, S., and Golde, P. (2006). *The C# programming language*. Adobe Press.

Herniter, M. E. (2000). *Programming in MATLAB*. Brooks/Cole Publishing Co.

Kalra, D., Beale, T., and Heard, S. (2005). The openehr foundation. *Studies in health technology and informatics*, 115:153–173.

Kavaler, D., Posnett, D., Gibler, C., Chen, H., Devanbu, P., and Filkov, V. (2013). Using and asking: Apis used in the android market and asked about in stackoverflow. In *International Conference on Social Informatics*, pages 405–418. Springer.

Kitchenham, B. A., Budgen, D., and Brereton, P. (2015). *Evidence-based software engineering and systematic reviews*, volume 4. CRC press.

Knaster, S. and Dalrymple, M. (2009). *Learn Objective-C on the Mac*. Springer.

Kwon, J., Grady, C., Feliciano, J. T., and Fodeh, S. J. (2020). Defining facets of social distancing during the covid-19 pandemic: Twitter analysis. *medRxiv*.

Lynch, B. M., Matthews, C. E., and Wijndaele, K. (2019). New mesh for sedentary behavior. *Journal of Physical Activity and Health*, 16(5):305–305.

McCallum, A. K. (2002). Mallet: A machine learning for language toolkit. *http://mallet.cs.umass.edu*.

Mildenberger, P., Eichelberg, M., and Martin, E. (2002). Introduction to the dicom standard. *European radiology*, 12(4):920–927.

Mumtaz, S., Rodriguez, C., and Benatallah, B. (2019). Expert2vec: Experts representation in community question answering for question routing. In *International Conference on Advanced Information Systems Engineering*, pages 213–229. Springer.

Oliveira, P. A. M., Andrade, R. M. C., and Neto, P. S. N. (2021). Iot-health platform to monitor and improve quality of life in smart environments. In *Conference on Computers, Software and Applications (COMPSAC) - 8th IEEE International Workshop on Medical Computing (MediComp 2021)*. IEEE.

Ponzanelli, L., Bacchelli, A., and Lanza, M. (2013). Seahawk: Stack overflow in the ide. In *35th Int. Conf. on Soft. Engineering (ICSE)*, pages 1295–1298. IEEE.

Puurula, A. (2013). Cumulative progress in language models for information retrieval. In *Proc. Australasian Language Technology Association Work. 2013 (ALTA 2013)*, pages 96–100, Brisbane, Australia.

Salvador-Oliván, J. A., Marco-Cuenca, G., and Arquero-Avilés, R. (2019). Errors in search strategies used in systematic reviews and their effects on information retrieval. *Journal of the Medical Library Association: JMLA*, 107(2):210.

Sievert, C. and Shirley, K. (2014). Ldavis: A method for visualizing and interpreting topics. In *Proceedings of the workshop on interactive language learning, visualization, and interfaces*, pages 63–70.

Silva, R., Roy, C., Rahman, M., Schneider, K., Paixao, K., and Maia, M. (2019). Recommending comprehensive solutions for programming tasks by mining crowd knowledge. In *2019 IEEE/ACM 27th Int. Conf. on Program Comprehension*, pages 358–368. IEEE.

Stol, K.-J., Ralph, P., and Fitzgerald, B. (2016). Grounded theory in software engineering research: a critical review and guidelines. In *Proceedings of the 38th Int. Conf. on Software Engineering*, pages 120–131.

Stroustrup, B. (1984). The c++ programming language: reference manual. Technical report, Bell Lab.

Thomas, S. W., Hassan, A. E., and Blostein, D. (2014). Mining unstructured software repositories. In *Evolving Software Systems*. Springer.

Treude, C., Barzilay, O., and Storey, M.-A. (2011). How do programmers ask and answer questions on the web?: Nier track. In *2011 33rd International Conference on Software Engineering (ICSE)*, pages 804–807. IEEE.

Uddin, G., Sabir, F., Guéhéneuc, Y.-G., Alam, O., and Khomh, F. (2021). An empirical study of iot topics in iot developer discussions on stack overflow. *Empirical Software Engineering*, 26(6):1–45.

Van der Walt, S., Schönberger, J. L., Nunez-Iglesias, J., Boulogne, F., Warner, J. D., Yager, N., Gouillart, E., and Yu, T. (2014). scikit-image: image processing in python. *PeerJ*, 2:e453.

Van Rossum, G. and Drake, F. L. (2011). *The python language reference manual*. Network Theory Ltd.

Wohlin, C. (2014). Guidelines for snowballing in systematic literature studies and a replication in software engineering. In *18th int. conf. on evaluation and assessment in software engineering*, pages 1–10.

Zapata, B. C., Fernández-Alemán, J. L., Idri, A., and Toval, A. (2015). Empirical studies on usability of mhealth apps: a systematic literature review. *Journal of medical systems*, 39(2):1.