The Making of Accessible Android Applications: An Empirical Study on the State of the Practice

Marianna Di Gregorio,[1] Dario Di Nucci,[2] Fabio Palomba,[1] Giuliana Vitiello[1]
[1]University of Salerno, Italy — [2]University of Tilburg/JADS, The Netherlands
madigregorio@unisa.it, d.dinucci@uvt.nl, fpalomba@unisa.it, gvitiello@unisa.it

***Abstract*—Context. Nowadays, mobile applications represent the principal means to enable human interaction. Being so pervasive, these applications should be made usable for all users: accessibility collects the guidelines that developers should follow to include features allowing users with disabilities (*e.g.*, visual impairments) to better interact with an application.**
Problem. While research in this field is gaining interest, there is still a notable lack of knowledge on how developers practically deal with the problem: (i) whether they are aware and take accessibility guidelines into account when developing apps, (ii) which guidelines are harder for them to implement, and (iii) which tools they use to be supported in this task.
Objective. To bridge the gap of knowledge on the state of the practice concerning the accessibility of mobile applications.
Method. Adopting a mixed-method research approach, we aim to (i) verify how accessibility guidelines are implemented in mobile applications through a coding strategy and (ii) survey mobile developers on the issues and challenges of dealing with accessibility in practice.
Limitations. Threats are represented by the size of the app sample and the number of answers to our survey study.

***Index Terms*—Empirical Software Engineering; Mobile Accessibility; Mobile App Evolution; Universal Design.**

I. Introduction

Mobile applications, a.k.a. apps, are nowadays used by billion users for any social and emergency connectivity [12]. For this reason, an ever-increasing population of users needs to interact with the functionalities they implement. This aspect represents a challenge for software maintenance and evolution research, which is called to devise novel instruments to support developers when evolving successful mobile apps. The pervasiveness of mobile applications has led researchers to reason more and more in terms of *accessibility*, giving rise to a research field that aims at making mobile apps usable for users affected by disabilities (*e.g.*, visual impairments) [5], [10] which affect over one billion people (around 15% of the world's population). The two main operating systems for tablets and smartphones, *i.e.*, iOS and Android, are equipped with pre-installed accessible functions, including screen reading functionalities as in the case of TalkBack for Android.

In the recent past, empirical investigations have been conducted to study how developers discuss the matter on StackOverflow [9] and how existing accessibility features support users with disabilities [4], [11]. Despite the recent advances in the field, there is still a notable lack of knowledge on the way developers approach the problem of accessibility and whether they implement the available guidelines to develop accessible applications. An improved understanding of these aspects is crucial to guide future software maintenance and evolution research efforts toward the definition of design, evolutionary, and testing techniques that can better support practitioners while developing mobile applications.

In this paper, we aim at bridging this gap of knowledge by presenting an empirical investigation into the making of mobile apps from the perspective of accessibility. We focus on Android not only because it has been the subject of previous accessibility studies, but also because still little is known of how to best engineer the problem in Android devices—as opposite to iOS and Apple in general, which provide an integrated set of devices and features to handle accessibility [2].

We discuss our plan toward this goal by defining two research questions to understand (i) whether and to what extent the available accessibility guidelines are implemented in Android applications and (ii) the developer's opinions about the matter. In so doing, we seek to elicit the state of the practice and the key issues and challenges faced by developers when dealing with accessibility.

II. Research Questions and Objectives

The *goal* of our empirical study is to understand the state of the practice of accessibility in mobile applications, with the *purpose* of providing an overview of how mobile developers currently deal with this problem as well as the issues and challenges they face when implementing accessibility guidelines.

We structure our investigation around two main research questions (**RQ**s). In the first place, we seek to understand how the existing accessibility guidelines are implemented in mobile applications, namely the extent to which developers adopt these guidelines when developing their apps. This goal leads to our first research question:

> **RQ$_1$.** *How are existing accessibility guidelines implemented in mobile applications?*

Once established how the accessibility guidelines are implemented, we then proceed with a finer-grained understanding of the developer's perspective as regards to the problem and, particularly, what are their opinions on (i) the issues and challenges of implementing accessible applications and (ii) the tools currently supporting them. An improved understanding

of these aspects aspect would lead to the definition of a set of developer's needs that the research community should further support. Hence, we pose our second research question:

RQ$_2$. *What are the developer's take on implementing accessibility guidelines in mobile applications?*

To address our **RQ**s, we plan to conduct mixed-method research [1], combining manual coding analyses with surveys and semi-structured interviews with developers [6]. The empirical study has an *exploratory* connotation and, as such, it must be seen as a *hypothesis-generating* investigation: in this setting, a set of hypotheses will be developed *after* the execution of the study based on the results achieved.

III. Research Protocol

This section presents our methodology to perform this study.

A. Material and Objects

The *objects* of the study are represented by (i) mobile applications and (ii) accessibility guidelines.

As for the former, we focus on the 50 top-rated Android apps coming from the AndroidTimeMachine dataset [3], which collects a reliable set of real open-source Android apps. We focus on these apps for two main reasons. On the one hand, we seek to analyze popular apps used by thousands, if not million users worldwide: this allows us to verify developers' behavior who should be more sensitive to accessibility issues given the number of users they can potentially attract. On the other hand, we have to limit the number of applications to consider because of the time- and effort-intensive manual activities that we need to perform to address our research questions (see Section III-C).

As for the latter, Table I reports the entire set of accessibility guideline categories currently available for the design of Android applications. Each category groups a set of guidelines to account for when considering a specific aspect of the mobile application (*e.g.*, *'Audio and Video'*). For the sake of understandability, we report in Table I a description of each category and a sample guideline it includes—we do not report all the guidelines because of space limitations.

The identified accessibility guidelines are a set of technological agnostic best practices for mobile web content, hybrid, and native apps. The guidelines are based on the content requirements of three de-facto standard providers of information on the matter, *i.e.*, the Android developer's documentation,[1] the World Wide Web Consortium (W3C) community,[2] and the BBC Standards and Guidelines academy.[3] We combine the three providers to create a comprehensive set of accessibility guidelines organized into 11 categories. Some guidelines are marked as *'MUST'* or *'MUST NOT'* depending on whether their implementation must be ensured or avoided. These guidelines are associated with specific, objective criteria that can assess their presence in a mobile app and can be implemented using the available mobile device technologies. Other guidelines are marked as *'SHOULD'* or *'SHOULD NOT'*: these represent less critical, yet important accessibility principles that should or should not be implemented. These guidelines are generally less testable and can be more subjectively interpreted by a user. In the final version of the paper, we will have a comprehensive background on these aspects.

[1]https://developer.android.com/guide/topics/ui/accessibility/index.html
[2]https://www.w3.org/TR/mobile-accessibility-mapping/
[3]https://www.bbc.co.uk/guidelines/futuremedia/accessibility/mobile

B. Subjects

The *subjects* of the study are developers of Android applications. We plan to involve both original and external developers of the applications object of the study. While the former can provide us with feedback on the implementation of the accessibility guidelines in their applications and their view of the problem, we believe that surveying a larger population of developers may provide additional insights into the issues and challenges of dealing with accessibility in practice. We plan to collect participants' background and demographic information to understand the representativeness of our results. We will follow the sampling strategies defined in literature [7] to define a sample that meets our goals.

C. Execution Plan

RQ$_1$. Accessibility guidelines in practice. To address **RQ$_1$**, we plan to manually test the considered applications to verify the implementation of accessibility guidelines—this strategy allows us to interact with an app and its accessibility services directly, much like a user would normally do. Overall, the guidelines to be verified are 54, divided into the 11 categories presented in Table I. To perform such a manual test, we plan to adopt a closed-coding strategy [8]: this is a systematic methodology that, in our case, involves the analysis of all the graphical user interfaces of an application and the subsequent labeling of the guidelines implemented as functionalities of the app, starting from a pre-established coding scheme represented by the set of guidelines available for Android applications.

More specifically, we create a data extraction form, implemented using an Excel sheet, to facilitate the verification of the guidelines. For each of them, the form contains four pieces of information: (i) the name of the guideline to verify, (ii) the procedure to follow to discover whether the guideline is implemented, *e.g.*, activate the notifications to verify that they are both visible and audible, (iii) the excepted visual/audio effect to observe in case the guideline is implemented, and (iv) the outcome to add once evaluating the guideline. The extraction process of an app will follow these steps and will be conducted by the first author of this paper:

Step 1 - Download: The author downloads the app from the Google Play Store on a Huawei Y5 smartphone.

Step 2 - Guideline identification: The author selects the next guideline to be tested and the corresponding instructions provided in the data extraction form.

Step 3 - Activation of accessibility features: Depending on the selected guideline, she activates the accessibility

TABLE I
ACCESSIBILITY GUIDELINES FOR MOBILE APPLICATIONS.

Guideline	Description	Example
Audio and Video	When creating interactive content, consider font size, style/position of controls, and how content is presented. If there is a strong need for the content to auto-play, the user should be aware of it and be able to set preferences to prevent it.	**Autoplay:** Audio should not be played automatically unless (1) the user is aware of it or (2) a Pause/Stop/Disable button is provided.
Design	For the best user experience, aspects such as clarity on color contrast, color and meaning, touch target sizes, content resizing, actionable elements, visible focus, content consistency, and adjustability should be properly designed.	**Colour contrast:** The color of the text and the content of the background must have sufficient contrast.
Editorial	Use of consistent labeling for buttons, links, and headings. Work closely with editorial colleagues to maintain consistency.	**Indicating language:** The language of a page or app must be specified, and changes in language must be indicated.
Focus	How content is visually presented can impact the order in which content is coded and, subsequently, the content order and focus order in which a user experiences the content, particularly users with alternative input methods such as keyboard or screen reader users.	**Focusable elements:** All interactive elements must be focusable and inactive elements must not be focusable.
Forms	Provide labels for all form inputs and ensure form layout and order is clear. Related form inputs should follow each other, and, if needed, the visual design should be applied to imply grouping.	**Form layout:** Labels must be placed close to the relevant form control and laid out appropriately.
Images	Avoid the use of images of text and those that do not covey key information solely through a background image.	**Background images:** Background images that convey information or meaning must have an additional accessible alternative.
Links	Design content layouts that facilitate grouping text and images as one link.	**Descriptive links:** Links and navigation text must uniquely describe the target or the function of the link.
Notifications	Design notifications to be inclusive and perceivable by all users. Where appropriate, include other feedback and assistance cues and prompts that might guide or encourage a user when needed.	**Inclusive notifications:** Notifications must be both visible and audible.
Scripts and dynamic content	Work from a basic core experience and progressively enhance this for more capable users.	**Controlling media:** Media that updates or animated content must have a pause, stop or hide control.
Structure	The design of the interface should convey the intended structure of the content. Identify headings, containers, and landmarks, working closely with UI/UX designers if needed.	**Unique page/screen titles:** All pages or screens must be unique and clearly identifiable.
Text Equivalents	The design of the non-textual content should describe their intent and not used to convey meanings.	**Alternatives for non-text content:** Alternatives must briefly describe the editorial intent or purpose of the image, object, or element.

function required to verify it if needed. Otherwise, she goes straight to the next validation step.

Step 4 - Element identification: The author exercises the app to identify the feature connected to the accessibility guideline, if available. For instance, this concerns the identification of the app's media in case the guideline refers to *'Audio and Video'* accessibility aspects. If identified, the author proceeds with the next step; otherwise, she goes back to Step 2 and continues with another guideline.

Step 5 - Verification of the guideline: Once the element is identified, the author determines if the guideline is implemented in the app. If so, she annotates the data extraction form putting, in the row corresponding to the considered guideline, a *'true'* in the fourth column. Otherwise, she annotates the column with *'false'*.

Using the above-described methodology, we will collect 50 EXCEL sheets, one for each application considered.

RQ_2 - Surveying mobile developers. To address RQ_2, we plan to conduct a survey study aiming at gathering insights regarding accessibility concerns from a broad audience of ANDROID developers. The survey is composed of three main sections—we report the full list of questions in Table II. The first one presents a total of nine questions about accessibility and how developers consider it in practice. We ask questions on the relevance of the problem, *i.e.*, how important is accessibility for the participants, what reasons would make them willing to implement accessibility features in their applications, and whether they are aware of the existence of guidelines to make an app accessible. Afterward, we continue with questions more related to the implementation of accessibility guidelines. In particular, how often developers implement them in their applications, how difficult they are to apply, and why. Finally, we ask participants to report up to five challenges they usually face when dealing with accessibility concerns and to report whether and which are the tools they use when performing the task.

In the second part of the survey, we allow participants to provide us with additional insights and feedback into the matter. They can leave their e-mail address if they are interested in receiving a summary of our findings and can

TABLE II
FULL LIST OF SURVEY QUESTIONS.

n.	Question	Evaluation Criterion
Section I. Accessibility of Android applications.		
1	In your opinion, how relevant is the problem of accessibility?	Likert scale from 1 (Not at all) to 5 (Very important)
2	Please, tell us more about your answer.	Open answer.
3	What makes you willing (or not) to implement accessibility guidelines?	Multiple Choice - it includes the *'Other'* option.
4	To what extent are you aware of the accessibility guidelines available for Android applications?	Likert scale from 1 (Not at all) to 5 (Very much)
5	To what extent do you follow accessibility guidelines when developing Android applications?	Likert scale from 1 (Not at all) to 5 (Very much)
6	Can you please rate how difficult it is for you to implement the following guidelines?	Likert scale from 1 (Not at all) to 5 (Very much) for each guideline.
7	For each guideline rated by the participant with 3/4/5 to question #6: 7.1. Can you please explain more about what makes it harder for you to implement the guideline?	Open answer.
8	What are the top 3 problems of dealing with accessibility in Android development?	Open answer.
9	What are the top 3 to 5 challenges you face when dealing with accessibility concerns?	Open answer
10	Do you use any tool to verify the implementation of accessibility guidelines?	Open answer
Section II. Further opinions.		
11	If you have further comments on the accessibility of Android applications and how you deal with the problem, feel free to comment more on it.	Open answer.
12	If you would like to receive a summary of our research results, please leave your e-mail.	Open answer.
13	Would you be willing to participate in a follow-up interview to better discuss the problem of accessibility in Android development?	Yes/No.
Section III. Background.		
14	What is your current job?	Multiple Choice - it includes the *'Other'* option.
15	What if your gender?	Multiple Choice - it includes the *'Other'* option.
16	How do you rate your expertise with programming?	Likert scale from 1 (Very poor) to 5 (Very high).
17	How do you rate your expertise with Android programming?	Likert scale from 1 (Very poor) to 5 (Very high).
18	What is your company size?	Multiple Choice - it includes the *'Other'* option.
19	What is your team size?	Multiple Choice - it includes the *'Other'* option.

express their consent to a follow-up interview aimed at further discussing the problem of accessibility in practice.

Finally, the third section of the survey concerns background information that we collect to understand our sample better and analyze the generalizability of our results.

The survey is designed to be done within 15/20 minutes and will be created using a GOOGLE survey module. Before releasing the survey on a large scale, we plan to perform a pilot with two developers of our contact network to evaluate if the survey is short and understandable enough to reduce the risk of having a low response rate and be appropriately filled out. According to the pilot results, we would need to adjust the text of some questions, add/remove some of them, or change the response type to make the questionnaire easier to understand or quicker to be compiled.

We plan to advertise the online survey using the personal social network accounts of the authors (*i.e.*, FACEBOOK, TWITTER, and LINKEDIN) to increase the number of participants and the overall response rate. At the same time, we are aware that the reliance on social media may negatively impact the selection of a valid sample. For this reason, we plan to complement social media with other sources to ensure the quality/completeness of the information gathered when addressing $\mathbf{RQ}_2$, still relying on a large sample of developers for our study. On the one hand, we will involve additional developers from our private contacts (*e.g.*, former University students or other practitioners that are currently mobile developers). On the other hand, we will advertise the survey on specialized practitioners' blogs—these represent powerful means to acquire information from developers who have a solid knowledge of programming. As an example, REDDIT contains more than 100 different subreddits dedicated to Android development, and that would potentially lead the survey to reach thousands of Android developers. We will keep track of the source used by participants to access the survey to better comment on the validity of the sample. To further stimulate the participation, we will allow the participants to indicate a non-profit organization of their choice to which we will donate 2 USD for the research against COVID19.

The answers received will be anonymized to preserve the privacy of participants. As a result of this study, we will have a clearer view of the relevance of accessibility in practice and the

major issues and challenges developers face when dealing with the problem. Depending on the answers received to question #13, we will plan follow-up semi-structured interviews with ANDROID developers. Their main goal will be to clarify ambiguous or contrasting answers received during the survey and to have a better picture of the current practices, issues, and challenges experimented by developers when dealing with accessibility in ANDROID environments. From a practical perspective, we plan to summarize the results of the survey to the interviewees and ask them to comment on the answers from which we could not derive a definitive outcome. The semi-structured interviews will be conducted through Skype, have a duration of 30/40 minutes, and be transcribed for further analysis.

D. Analysis Plan

Once gathered data from the closed-coding exercise and the survey study, we will proceed with their analysis.

As for $\mathbf{RQ}_1$, we will first provide descriptive statistics on the extent to which accessibility guidelines are implemented in the sample of ANDROID applications. As such, we will compute minimum, mean, median, standard deviation, and maximum number of accessibility guidelines implemented in the considered apps. Secondly, we will provide a finer-grained overview of each specific category of guidelines. We plan to discuss (i) to what extent each of them is present in the sample by reporting descriptive statistics, *i.e.*, minimum, mean, median, standard deviation, and maximum number accessibility guidelines for each category, and (ii) the relative and absolute frequency of implementation of the guidelines included in each category. Then, we will focus on the guideline requirements, *i.e.*, *'MUST'*, *'MUST NOT'*, *'SHOULD'*, and *'SHOULD NOT'*: in this case, we will aim at understanding whether developers take them into account, *e.g.*, if the *'MUST'* guidelines are implemented in the considered apps. Finally, we plan to verify the relation between the guidelines and the type of application considered. We will group the apps by category, as provided by the GOOGLE PLAY STORE, and compute descriptive statistics to grasp if some categories are more prone to accessibility concerns.

As for $\mathbf{RQ}_2$, we will first describe the background of survey participants by discussing the answers they provide in Section III of the survey. This detail will allow us to understand the sample of developers and reason about the generalizability of our findings. In the second place, we will distinguish the analysis procedures to use when considering closed and open questions. The former will be analyzed employing statistics: we will plot and discuss the distribution of answers provided by participants through the Likert scale evaluations. The latter will be subject of an *iterative content analysis*: in particular, we expect to conduct the following methodological steps:

Step 1 - Microanalysis: The first author of the paper will go through the content of the participant's answers and the possible semi-structured interviews. She will split sentences using standard text separators, *e.g.*, commas, and assign initial labels to each sentence: these labels represent the main concepts discussed by participants. Then, the three authors not involved so far will validate the initial labels assigned and provide feedback on how to improve them, for instance, by proposing to aggregate two semantically-similar labels. When performing this step, we will compute a measure of agreement between the labels assigned by the first author and those recommended by the other three.

Step 2 - Categorization: The first author will use the suggestions and feedback received in the first step to conduct a second iteration over the labels assigned. This step will result in a set of themes deemed important by participants when addressing each question of the survey.

Step 3 - Saturation: All the authors will be involved to reach a final agreement concerning the names and meanings of each label. This step will lead to a *theoretical saturation*, *i.e.*, the point in which no further labels are required because the existing ones already correctly represent the concepts expressed by the study participants.

The themes coming from this data analysis procedure will relate to each specific open question posed in the survey. We will discuss each theme and provide qualitative insights by presenting the most significant answers for a specific theme. In addition, when analyzing the answers to questions #8, #9, and #10 of the survey, we will also provide statistical data reflecting the number of times a specific issue/challenge/tool has been named by the participants, hence providing a kind of prioritization of the concerns and tools that developers have with respect to the problem of accessibility.

E. Limitations

Limitations are related to (i) the sample size of the considered apps and (ii) the number of answers received for our online questionnaire. In case our findings to $\mathbf{RQ}_1$ will not be definitive, we plan to extend the number of ANDROID applications. This extension will possibly result in a higher amount of manual work for the first author of the paper. In this case, such a work will be split between the first and second author to avoid overloading that would possibly lead to some imprecision in the closed-coding task. As for the number of survey answers ($\mathbf{RQ}_2$), we will involve practitioners from various sources to increase the number of answers and stimulate the participation by giving the possibility to donate for the research on COVID19: we expect that this societal commitment will lead practitioners to participate and further distribute the survey.

F. Publication of Generated Data

The data generated from our study will be publicly available through an online open-access archive (*e.g.*, Figshare). We plan to release raw data about the accessibility guidelines implemented in our dataset, the survey structure, the anonymized responses, and all scripts used for data analysis.

ACKNOWLEDGMENTS

Dario acknowledges the support of the European Commission grant no. 825040. Fabio is partially funded by the Swiss

National Science Foundation through the SNF Project No. PZ00P2_186090.

REFERENCES

[1] J. W. Creswel. Research design: Qualitative, quantitative, and mixed methods approaches. *University of Nebraska–Lincoln*, 2009.

[2] A. Darvishy. Accessibility of mobile platforms. In *International Conference of Design, User Experience, and Usability*, pages 133–140. Springer, 2014.

[3] F.-X. Geiger, I. Malavolta, L. Pascarella, F. Palomba, D. Di Nucci, and A. Bacchelli. A graph-based dataset of commit history of real-world android apps. In *Proceedings of the 15th International Conference on Mining Software Repositories*, pages 30–33, 2018.

[4] D. Kocieliński and J. Brzostek-Pawłowska. Improving the accessibility of touchscreen-based mobile devices: Integrating android-based devices and braille notetakers. In *2013 Federated Conference on Computer Science and Information Systems*, pages 655–658. IEEE, 2013.

[5] B. Leporini, M. C. Buzzi, and M. Buzzi. Interacting with mobile devices via voiceover: usability and accessibility issues. In *Proceedings of the 24th Australian Computer-Human Interaction Conference*, pages 339–348, 2012.

[6] P. H. Rossi, J. D. Wright, and A. B. Anderson. *Handbook of survey research*. Academic Press, 2013.

[7] L. Topp, B. Barker, and L. Degenhardt. The external validity of results derived from ecstasy users recruited using purposive sampling strategies. *Drug and alcohol dependence*, 73(1):33–40, 2004.

[8] P. Vaughn and C. Turner. Decoding via coding: Analyzing qualitative text data through thematic coding and survey methodologies. *Journal of Library Administration*, 56(1):41–51, 2016.

[9] C. Vendome, D. Solano, S. Liñán, and M. Linares-Vásquez. Can everyone use my app? an empirical study on accessibility in android apps. In *2019 IEEE International Conference on Software Maintenance and Evolution (ICSME)*, pages 41–52. IEEE.

[10] G. Vitiello, M. Sebillo, L. Fornaro, M. Di Gregorio, S. Cirillo, M. De Rosa, V. Fuccella, and G. Costagliola. Do you like my outfit? cromnia, a mobile assistant for blind users. In *Proceedings of the 4th EAI International Conference on Smart Objects and Technologies for Social Good*, Goodtechs '18, page 249–254, New York, NY, USA, 2018.

[11] B. N. Walker, B. J. Tomlinson, and J. H. Schuett. Universal design of mobile apps: Making weather information accessible. In *International Conference on Universal Access in Human-Computer Interaction*, pages 113–122. Springer, 2017.

[12] A. I. Wasserman. Software engineering issues for mobile application development. In *Proceedings of the FSE/SDP workshop on Future of software engineering research*, pages 397–400, 2010.