Nipping Inaccessibility in the Bud: Opportunities and Challenges of Accessible Media Content Authoring

Carlos Duarte
LASIGE, Faculdade de Ciências, Universidade de Lisboa
Portugal

Letícia Seixas Pereira
LASIGE, Faculdade de Ciências, Universidade de Lisboa
Portugal

André Santos
LASIGE, Faculdade de Ciências, Universidade de Lisboa
Portugal

João Vicente
LASIGE, Faculdade de Ciências, Universidade de Lisboa
Portugal

André Rodrigues
LASIGE, Faculdade de Ciências, Universidade de Lisboa
Portugal

João Guerreiro
LASIGE, Faculdade de Ciências, Universidade de Lisboa
Portugal

José Coelho
LASIGE, Faculdade de Ciências, Universidade de Lisboa
Portugal

Tiago Guerreiro
LASIGE, Faculdade de Ciências, Universidade de Lisboa
Portugal

ABSTRACT

Social media represents a large part of the content available on the Web. While the accessibility of the UIs of existing social media platforms has been improving, the same cannot be said about the accessibility of the content authored by their users. Specifically, the accessibility of multimedia content that is increasingly available given the ease of access to mobile devices with cameras. User research has revealed that accessible authoring practices are a foreign concept to most social media users, but also that they are motivated to adopt inclusive practices. Our work focuses on promoting awareness to accessible social media authoring practices and in assisting the authoring process. We have prototyped a Google Chrome extension and an Android application that can identify when a Twitter or a Facebook user is authoring content with images and suggests a text alternative for the image. By suggesting the alternative, we raise awareness to the accessible authoring process and make it easier for the user to include it in the tweet or post. Text alternatives may be suggested from different sources: descriptions entered by other users for the same image, analysis of the main concepts present in the image, or text present in the image, for instance. Our prototypes can also provide text alternatives on demand for images on any web page or Android application, not just social media. In this paper, we highlight some of the challenges faced to offer this support in different technological platforms (web and mobile), but also ones that are raised by the domain characteristics (e.g. detecting the same image, supporting different languages) and that can be addressed through AI based technologies.

WebSci '21 Companion, June 21–25, 2021, Virtual Event, United Kingdom

ACM ISBN 978-1-4503-8525-1/21/06.
https://doi.org/10.1145/3462741.3466644

CCS CONCEPTS

• **Human-centered computing** → **Accessibility**; **Social networking sites**; **Social media**.

KEYWORDS

accessibility, social media, visual content, user-generated content

ACM Reference Format:
Carlos Duarte, Letícia Seixas Pereira, André Santos, João Vicente, André Rodrigues, João Guerreiro, José Coelho, and Tiago Guerreiro. 2021. Nipping Inaccessibility in the Bud: Opportunities and Challenges of Accessible Media Content Authoring. In *13th ACM Web Science Conference 2021 (WebSci '21 Companion), June 21–25, 2021, Virtual Event, United Kingdom.* ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/3462741.3466644

1 INTRODUCTION

Social networks play a crucial role in connecting people, being more important than ever. Meanwhile, people with disabilities are once again being deprived of fully participating in a major aspect of society. Considering the collaborative aspect inherent in these platforms, to ensure that these environments are fully accessible, it is essential to guarantee that the user-generated content is accessible as well. One of the challenges lies in the amount of visual media content created by users that is currently not accessible to people with visual disabilities. Even though platforms are evolving strategies to address this problem, the features that are currently in place are still not sufficient to ensure that visually impaired users can properly understand an image [1, 7, 11, 14, 22, 23].

Major platforms have chosen to adopt different strategies to improve the accessibility of this content. For instance, Facebook provides automatically generated descriptions and allows its users to edit them [6]. Twitter, on the other hand, allows users to enter their own alternative descriptions [21]. Despite these differences, both implementations have one thing in common: most users are unaware of the existence of such tools, and when they do, they either have difficulties in finding them [7, 16, 18] or struggle with the creation of effective text descriptions. As for the automated

descriptions provided by Facebook, they are often very succinct (and incomplete), sometimes incorrect, do not provide enough contextual information and, consequently, blind users consider them insufficient [11, 14, 16, 20, 22, 23, 26]. As a result, most images in social networks either have no alternative description due to the lack of knowledge or effort employed by users, or have suboptimal automatic descriptions that are not able to fully describe an image to visually impaired users. The latter, however, is predicted to continue to be a major challenge due to the lack of context that automated solutions have on unique, personal images shared on social networks. For that reason, we argue that it is crucial to leverage a combination of automatic and manual descriptions – which may be iteratively improved – to increase the number and quality of alternative text descriptions.

Previous work has focused on bridging the gap through improving current AI-created descriptions to be more suitable to a social media context or by scanning for instances where an alternative text was once provided for the image [8, 13, 19, 20]. As presented in Pereira et al. [16], the process of creating accessible media content can be enhanced by a human-in-the-loop approach. Automated descriptions can be used as a resource to help users to provide an example of a suitable alternative description and the opportunity to improve the description provided by AI-based solutions. Including users in the process of authoring accessible content by embedding these features into the authoring process flow can raise awareness and thus inducing them to engage more frequently in such practices. For that, our approach aims to tackle the problem at its very root, providing enough resources to ensure that images are accessibly created from the start. For that, we developed the SONAAR prototypes exploring combinations between manual and automated approaches to improve the overall user experience for visually impaired users in social networks, both for authoring and consuming visual content. In particular, our main goal is to enrich the authoring process to support the creation of accessible media content.

The SONAAR prototypes are available as a Google Chrome extension and an Android mobile application and currently support authoring images on Facebook and Twitter. Using SONAAR prototypes, users are guided through a process of authoring accessible media content. We detect when users are uploading media content, indicate where they should include an alternative description, and offer suggestions of possible alternative descriptions. These suggestions come from a variety of sources such as the ones entered by SONAAR users, automatically generated from image analysis, or text recognized in the image, for instance. The deployed structure also allowed us to support a scenario where users consume visual content outside the scope of social networks. The SONAAR user, when in the presence of an image on a web page or on a mobile application screen, can request from our service a textual description for that image. The same suggestions offered on the authoring scenario are provided to this user.

In this paper, we first describe previous research on accessible media content and social networks to better contextualize our work. Next, we provide details about the deployment of SONAAR prototypes, including 1) the backend service responsible for supporting the operation of both the SONAAR browser extension and the SONAAR mobile application, 2) the features to support authoring of accessible content on social networks, and 3) the features to support accessible image browsing. Following that, we discuss the opportunities that our solution opens, and the challenges faced and in need of tackling.

2 RELATED WORK

This research work is related to prior work on 1) current strategies on alternative descriptions for images and 2) current state of the accessibility of social networks.

2.1 Alternative descriptions for images

The lack of alternative descriptions on the web is a longtime issue. The work conducted by Petrie et al. [17] back in 2005, identified that blind users felt frustrated by not only the lack of alternative descriptions but also by the poor quality of those descriptions provided. While different efforts were conducted meanwhile, visually impaired users are still facing the same barriers to interact with visual content on the web [1, 7, 11, 14, 16, 23]. This issue may be partly explained by the fact that most automatically generated alternative descriptions are based on image recognition approaches made by and for sighted users instead of on accessibility purposes [24]. This nuance may be exemplified by the ability to evaluate the accuracy of an alternative description. For instance, a sighted person can compare the image and the details provided by the automatically generated description and retain only what it is judged to be correct. On the other hand, blind users only have this piece of information to fully understand an image. One impact of this scenario is that blind users are very trusting of even incorrect automatically generated captions, often causing them to act based on inaccurate information [11, 19].

Another factor influencing the quality of alternative descriptions is the contextual information not provided by automatically generated descriptions. Firstly, visually impaired users have different needs on the details that should be provided depending on the source where the image is found [17, 20]. Secondly, most current AI solutions provide a list of concepts that may be contained in an image, but they do not convey the real intention and meaning of an image.

2.2 Accessibility and Social networks

Social networks are responsible for a great part of images being uploaded on the web nowadays. Concerning the accessibility of this content, some strategies have been employed by major platforms. Facebook, one of the most used social networks, currently provides two different approaches to improve the accessibility of images uploaded by their users [6]. The first one consists of embedding an automatic description for every image uploaded to the platform. Following that, users can also edit it to improve it or to provide a better description. However, most users are not aware of that possibility [7, 16, 18], leaving visually impaired users relying only on the automated descriptions generated by the platform – that, most of the time, do not provide enough contextual information to properly understand an image [11, 14, 16, 20, 22, 23, 26]. On the other hand, Twitter currently supports image descriptions provided by the users themselves [21]. However, it is also being reported as a hard-to-find feature, not drawing mainstream users' attention not

only to the existence of this feature but also to the importance of this practice [7, 16, 18].

The current situation of media accessibility in social networks is far from ideal. According to recent work, in a sample of more than 1 million images on Twitter, only 0.1% of them included an alternative description [7]. While the platforms themselves have been deploying new accessibility features at their own pace, a few other approaches to improve the accessibility of user-generated media content are also being explored. Gleason et al. [8] created TwitterA11y, a browser extension to add alternative descriptions to images on Twitter through different strategies such as text recognition, crowdsourcing, and reverse image search using Caption Crawler [9]. From another perspective, audio descriptions have also been presented as an alternative and promising way of providing image descriptions [4, 10].

Another scenario often ignored by major platforms is the one where visually impaired users author media content. On one hand, the accessibility features available are not fully accessible. Visually impaired users reported having difficulties in finding those features and, due to the constant interface updates, they have to be constantly coping with a new structure [16]. On the other hand, they do not find proper guidance for composing alternative descriptions for their images [16]. Even for platforms that provide automated descriptions such as Facebook, there is a reluctance to rely exclusively on them. In contrast to the trust placed in automatic descriptions when interpreting an image previously discussed, the same is not true in the process of authorship. According to Zhao et al. [26], the risks associated with mis-reacting to other people's content is much lower than that of mis-sharing their own information. Finally, users find themselves once again compromising their autonomy by asking the assistance of a trusted sighted contact [12, 26].

Despite all these efforts, blind users are still reporting image descriptions provided by their own author as having a better quality than automated ones [16]. Nevertheless, it is important to highlight the benefits of automatic approaches. Besides being fast and cheap, they allow deployment at a large scale compatible with the number of images being generated through current social networks [5, 8]. For that, we propose a mixed approach, benefiting from the advancements in automated image recognition and description to enhance human descriptions.

3 SONAAR

SONAAR combines image recognition techniques and human input to provide different sources of alternative descriptions for images in social networks. In our previous work, we identified the main barriers hindering mainstream users to provide accessible media content in their social networks and how they could be more engaged in accessible practices [16]. The structure deployed in SONAAR explores a human-in-the-loop and collaborative approach to 1) better assist mainstream users in the authoring process of accessible media content; 2) improve the general quality of image descriptions in social networks; 3) raise accessibility awareness of social networks users; and 4) discuss the benefits and feasibility of mixed approaches, by exploiting the strengths of each strategy: the agility and scalability of automatic descriptions, and the higher quality of the descriptions provided by the authors themselves.

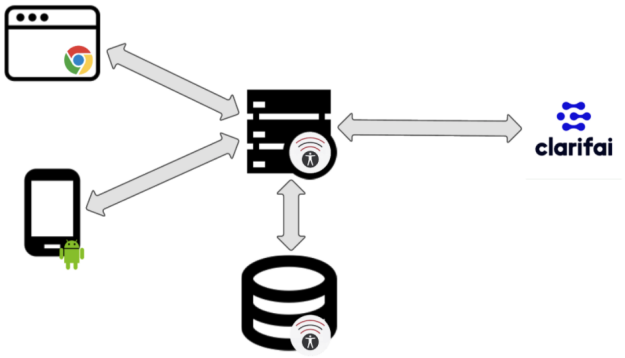

Figure 1: SONAAR system architecture

The SONAAR prototypes offer text descriptions for images. Different sources for the image descriptions are explored in the prototypes: the ones entered by users of social networks, automatically generated from image analysis, or text recognized in the image, for example. The instances where these suggestions are presented to the user are also varied: when a user is tweeting or posting an image on social networks the suggestion is automatic; on any other website or mobile application at the user's request. In the following paragraphs, we offer a short description of the backend service and the Android application and Chrome extension that implement these features.

3.1 Backend

The SONAAR backend supports the Android application and Chrome extension requests through a set of features that are independent of the request's origin. The backend is composed of a database that stores the known image descriptions. This is a simple database that connects an image identifier with the currently known descriptions for that image and the language the description is written in. In order to provide these descriptions upon a client's request, the backend needs to be able to search for an image in the database.

Image searching is achieved through the image recognition service provided by Clarifai[1]. We store images in Clarifai (and keep only their identifier in our descriptions' database) and use their image searching feature to search the image database for the image for which the description has been requested. Clarifai provides a similarity measure between the searched image and every image in the database. When an image has a similarity measure above the "same image threshold" we consider it to be the same image we have on the database. We have fine-tuned the "same image threshold" so that the same image is identified even if it has suffered small modifications, like a small crop and the addition of a watermark or signature. In the future, we want to explore this measure to assist us in identifying related images. For that, we can define a "related image threshold", which will allow us to identify images that are related but not the same. This knowledge will be useful for those instances where we haven't seen the image before, by allowing us to still offer a description of a related image.

[1] https://www.clarifai.com/

To assist in the preparation of image descriptions we use further features provided by Clarifai. The first one is the ability to provide a list of concepts related to the image that is searched. Clarifai provides us with a list of concepts, with, for each concept, a level of confidence in the accuracy of the result. We keep the concepts from this list that are above a "concept confidence threshold". These concepts offer us another way to create an image description. A second feature we use from Clarifai is the ability to identify any text present in the image. This is particularly relevant for the social network domain, where many posted images contain text (e.g. memes). The image's text is, very often, another good source for creating an image description.

A final source of image descriptions is the descriptions created by the users of the social networks themselves. When our front-end prototypes detect an image being posted with a description, that description is sent to the backend. The backend stores that description if it has not been stored before. If that description had already been stored, we increase a counter of the number of times it has been used. The language the description is written in is also stored. We use the Franc Natural Language Detection library[2] to identify the language of the description.

In summary, our sources for descriptions include: descriptions provided by users, image concepts identified by Clarifai, and any text in the image. Image descriptions are characterized by their language and by the number of times they have been used in tweets or posts.

In order to answer the client's request, the backend will have to decide which description or descriptions to send. To make that decision we explore three features. The client's request includes the language of the user's device or browser. With that information, we can limit our selection to descriptions in the same language. The other two features are combined to select which are the top descriptions to send back. The first feature is the number of times a description has been used. Heuristically, we can expect the most popular description to be the most adequate description for an image. However, if we apply this number without further consideration we might disregard newer descriptions that might be better, but since they are new, they will have a lower number of uses. The other feature is a quality measure of the description for the image. Our quality measure applies the algorithm described in Duarte et al. [5] which returns a metric of the similarity between the terms in the image description and several features in the image (including the concepts present in the image, concepts related to the image domain, and any metadata in the image). The algorithm classifies the semantic similarity between the image and a description on a scale from 0 to 1. The backend uses this additional information, combined with the usage count, to order the list of descriptions of the same language of the users' device or browser that are sent back to the client.

3.2 Supporting authoring accessible content on social networks

In this and the following section, we briefly describe the features of the Android application and Chrome extension, which operate in a similar fashion. Our main goal is to increase the availability of accessible media content on social networks. Two of the main reasons for social network users not authoring media containing posts in an accessible way are lack of knowledge about accessibility and the extra effort it requires, which might detract from the fast-paced nature of social browsing [16]. Our prototypes tackle these two reasons. By suggesting a text description for an image in a post, the prototypes raise awareness of the need for accessible authoring practices and make it easier for the content author to include a description.

The prototypes are capable of automatically detecting when the social network user is authoring content with images on Twitter and Facebook. The Chrome extension achieves this by inspecting the DOM of the web page and looking for the presence of elements with specific class attributes. The Android application inspects the elements on the screen and looks for values of specific attributes. After it has been detected that a user has uploaded an image in the authoring page or screen, a request is sent to the backend containing the image and the language of the user's device or browser. This automated authoring detection process has some limitations: it is dependent on Twitter and Facebook not changing their user interfaces; on Android, it is dependent on the language of the device, because some of the values that the prototype looks for are names of elements that change depending on the operating system's language; obtaining the image on Android requires capturing a screenshot of the device.

When the backend answers with the proposed descriptions for the image, the prototype makes the user aware of it. On Android, we use a notification to present the top-rated description to the user. In Chrome, the description is presented as an overlay window next to the field where the description is to be entered. On both platforms, the user can copy the description to the clipboard and paste it into the corresponding field in the authoring interface. If the backend sends more than one description, we offer the user a chance to see the extra descriptions. By selecting that option, the list of descriptions is presented to the user, and any can be selected.

Finally, the prototypes are able to automatically detect when the tweet or post is completed (i.e. the user activates the corresponding button on the interface). At this moment, the image's description is captured and sent to the backend, where it is stored as a new description (if the user created a new one or changes anything in one of the suggested descriptions) or, if it is not a new one, the number of uses of the description is incremented.

3.3 Supporting accessible image browsing

With the availability of image descriptions, as presented in the previous sections, we extended the functionalities of the prototypes to offer support for screen reader users that need to access image descriptions in any web page or mobile application.

The Android application registers itself as an application to which images can be shared. Whenever a blind user finds a shareable image on any Android application that does not have a description, or that has a description that the user does not consider adequate (e.g. the image filename), the user can request a description for that image. The prototype then lists the descriptions it has stored or created from the image's concept.

[2] https://github.com/wooorm/franc

The Chrome extension works in a similar way, but with a different scope. Given that images are not focusable by default on a web page, when a blind user finds any image for which a description is desired, activating the prototype on that web page results in all images on the web page being sent to the backend. On receiving an answer, the prototype modifies the page's DOM to make all images focusable and to insert the descriptions in the alt attribute of the corresponding image. The user can then browse the images on the page and listen to the descriptions for any image.

4 DISCUSSION

The SONAAR prototypes described above open up new opportunities for authoring media-based content on social networks in a more accessible way. In this section, we discuss how they can impact the accessibility of media but also some of the challenges that still need to be addressed.

4.1 Opportunities

Raise awareness and educate users. As presented in Pereira et al. [16], most social network users are not even aware of what digital accessibility is. These users do not know how people with visual impairments, for example, are able to access information provided through images. And, when they are aware of it, they do not know what to do to provide that information in the required alternative format. Often, people will become aware of this when they have a visually impaired friend or relative that can provide them with guidance. SONAAR can contribute to this awareness-raising effort. The current prototypes detect when an image is being posted on social media and notify the user of the possibility of inserting a text description of the image. For users that are not familiar with accessibility needs, this will still be a foreign concept. However, this represents an opportunity to educate users and we plan on expanding the current features of the prototypes, by providing materials that demonstrate the advantages of creating accessible content and instructions on how such content can be created. This, we expect, will promote accessible practices everywhere, not only on social networks, but will also lessen the burden on people with disabilities to promote these practices.

Reduce user effort. Most social network users that are aware of the possibility to provide an alternative description for their image consider that this activity takes too much time and effort, especially when considering the inherent agile nature of these platforms [16]. One source of this extra work is to compose an appropriate textual description for an image. We expect that the guidance and suggestions provided by the SONAAR prototypes will lessen this effort. However, users will still have to deviate from their usual authoring process flow. This is something that might be mitigated by the use of motivational content related to the benefits for all of creating accessible content.

Improve both automated and human-authored descriptions. The suggestions provided by our prototypes are the results of different approaches to obtain alternative descriptions. The SONAAR service also stores how many times a specific alternative description was chosen by the users. Therefore, we expect to have a large dataset of, not only possible suggestions, but also users' preferences on image descriptions. The information gathered may be used to complement current research on the quality, phrasing, and guidelines of image descriptions, as well as improving ML-based solutions for generating automated descriptions. With the SONAAR prototypes, we aim to investigate in the future how AI solutions may be useful to enhance human authored descriptions, maximizing the use of both approaches.

Leverage collaborative/community efforts. The SONAAR prototypes can also be extended to support a collaborative approach such as one based on crowdsourcing and voting capabilities. In this scenario, users with no visual impairments can contribute by providing or improving an alternative description for an image. These users would also be able to upvote or downvote alternative descriptions already provided. Such a solution can be especially useful for images that only have text descriptions that are judged to be of poor quality. In these instances, these images could be prioritized for a crowdsourced description generation.

Automatic descriptions to engage users. While SONAAR primarily aims to increase the accessibility of user-authored content, we also made it available for providing descriptions of images on request outside the social network context. We expect that many of these requests will not have a matching image on the SONAAR database, since images people share on their social network profiles won't likely be published elsewhere on the web. In this instance, SONAAR will only be able to provide the automatically generated image description, with less quality than human-authored descriptions. Nevertheless, this is an opportunity to engage users. For example, if a user judges an automatically generated description to be of poor quality, there can be an option for the user to improve it or to provide a new one, or, exploring the crowdsourcing scenario, to ask for a description from the crowd. It might not be possible to have the description in real-time to benefit the user, but overtime, the number of descriptions will increase and more users will benefit from the descriptions.

4.2 Challenges

Depend on an external source. Concerning the technical aspects of the solution proposed, the SONAAR backend relies on the service provided by Clarifai. Although our prototypes aimed to investigate the feasibility and benefits of such a structure, the scalability of our current service is dependent on an external service.

Frequent changes in the UI of social networks. Several challenges are raised due to the inherent volatility and agility of social networks. In order to detect when the user is authoring content with images, our prototypes inspect the interface of both Twitter and Facebook for specific attributes previously identified. However, the interfaces of these social platforms can be modified whenever Twitter or Facebook wishes, and it happens without any prior notice (so far we found out that this happens more frequently with Facebook than with Twitter). The UI changes can affect the elements and attributes that SONAAR scans for, causing our prototypes to no longer be able to recognize them. While these issues can be addressed by new SONAAR updates, it is still necessary to identify the frequency of these changes to assess the real impact on our prototypes.

Dealing with UI changes manually. Detecting and reacting to these changes in the interface are two issues that might be addressed

by human intervention, AI automated solutions, or a combination of both. For a human-supported solution, we plan on adding a feature to the SONAAR interface where users can report that SONAAR is no longer working, which is the immediate consequence when updates to the UI result in SONAAR no longer being able to identify the required elements. This feature can be extended to an interactive flow where SONAAR asks the user where a specific element in the interface is (e.g. the element that allows a user to select a photo to include in the post) and the user selects that element in the UI. By creating an overlay on the mobile screen, or intercepting events in the browser, we can identify the element and learn what changed in its properties. This would be repeated for all the required elements and would allow SONAAR to repair itself. The knowledge provided by one user could then be replicated to other users of the platform, ensuring that everyone gets an update quickly. While this solution is feasible to small scope changes to the UI, more profound changes that, for example, imply a different flow for creating tweets or posts with images, will probably not be fixable with this solution and would require another type of approach.

Dealing with UI changes automatically. A possible approach to the above problem is to rely on ML to identify the elements involved in the authoring of posts involving images. A similar problem was tackled by Zhang et al. [25] that trained a model to recognize elements on the iPhone screen and augment those with metadata to make them accessible to visually impaired users. For our goals, we need to identify elements and a sequence of interactions with those elements, which can be based on identifying elements in a sequence of screens. We can also envision expanding this approach to being able to identify the semantics of interaction so that it can be robust to more profound changes to the interface. Here we can expand on existing work that tries to identify the semantics of the UI [2, 3, 15].

Seamless expansion to other social platforms. This type of solution that would be able to automatically identify when a user is posting content with images, might also support an easier expansion of the image's description suggestion service to other social platforms. So far we have limited the suggestions service to only Twitter and Facebook due to the effort involved in the initial identification of the relevant elements in the mobile screens or web pages and the effort required to maintain the identification service up to date whenever there are updates to the UI of the social platforms. With a service that could respond dynamically to those changes, the effort needed to have it operate on more social networks would decrease considerably. Of course, the best way to eliminate the need to respond to UI changes would be to have the social networks integrate the service in their applications, therefore benefiting from descriptions made available elsewhere.

Dealing with different languages. The support for multiple languages is another challenge yet to tackle. The identification of key elements is also associated with, not only the language used on the social network, but also by the language of the user's device, in the case of Android mobile devices, and may vary according to it. This support needs also to take into account the different languages that alternative descriptions are being generated. Additionally, AI solutions could provide translations of image descriptions that are not available in the user's language. The same image will likely have descriptions stored in multiple languages. When a user needs a description that is not available in his or her language, but in another language, this solution would allow us to provide the user with a translated version of a quality description in a different language.

Ensure user privacy at all times. Finally, one important aspect has to be guaranteed during the whole process: user privacy. The current implementation does not take into account the origin of the image and descriptions, when deciding on what description to present when suggesting a description or answering a user's request for descriptions when browsing the web or using a mobile application. Users might share some images privately in a social network, and they expect these images and descriptions to stay private. A scenario where a user shares a photo and writes a description where persons in the photo are identified, and later that photo appears on a web page and through the description, it becomes possible to identify the persons in the photo, needs to be avoided. In our current implementation, users can control when the SONAAR service is activated and can enable it for a specific social network only.

5 CONCLUSION

Social networks are one of the biggest sources of user-generated content on the internet. However, most of this content is not accessible to users with visual impairments. This is due to multiple factors, like the lack of awareness of digital accessibility in general, to lack of knowledge about creating content in an accessible way, the lack of knowledge about how social network platforms allow users to author accessible content, or simply because it takes to much effort to write a text description for an image in a post. Existing solutions have tried to address this problem by automatically generating a description for images that have been posted on social networks. However, users with visual impairments have claimed these to be of lower quality than those created by humans. With SONAAR we want to explore a solution that promotes the authoring of accessible content by raising awareness of the need and benefits of creating accessible content, and that supports authoring by suggesting possible descriptions for an image.

In this paper, we described the features of our mobile and desktop prototypes that combine existing human authored descriptions, with automatically generated ones. The SONAAR prototypes explore AI-supported image recognition, text recognition in images, semantic similarity measures of text description and image concepts, and language identification. These provide results that allow us to offer suggestions of image descriptions during the authoring process in selected social networks, but also on request from users while browsing the web or using any mobile application. We also discussed some challenges that still need to be addressed in order to improve and expand the service, as well as opportunities that can be fulfilled in this domain with the adoption of additional AI-based features.

This work demonstrates how hybrid solutions, combining human authoring with AI-supported automatic generation and classification, can contribute to improving the overall accessibility of content published on the internet. Currently, we believe that these solutions are still able to offer a higher level of quality than fully automated ones. At the same time, the data that we collect will

allow improving the quality of automated solutions, therefore also contributing to their development.

ACKNOWLEDGMENTS

This paper was written with support from the SONAAR Project, co-funded by the European Commission (EC) through GA LC-01409741. This work was supported by FCT through funding of LASIGE Unit R&D, Ref. UIDB/00408/2020.

REFERENCES

[1] Julian Brinkley and Nasseh Tabrizi. 2017. A Desktop Usability Evaluation of the Facebook Mobile Interface using the JAWS Screen Reader with Blind Users. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 61, 1 (sep 2017), 828–832. https://doi.org/10.1177/1541931213601699

[2] Chunyang Chen, Sidong Feng, Zhengyang Liu, Zhenchang Xing, and Shengdong Zhao. 2020. From Lost to Found: Discover Missing UI Design Semantics through Recovering Missing Tags. *Proc. ACM Hum.-Comput. Interact.* 4, CSCW2, Article 123 (Oct. 2020), 22 pages. https://doi.org/10.1145/3415194

[3] Biplab Deka, Zifeng Huang, and Ranjitha Kumar. 2016. ERICA: Interaction mining mobile apps. *UIST 2016 - Proceedings of the 29th Annual Symposium on User Interface Software and Technology* (2016), 767–776. https://doi.org/10.1145/2984511.2984581

[4] João Marcelo dos Santos Marques, Luiz Fernando Gopi Valente, Simone Bacellar Leal Ferreira, Claudia Cappelli, and Luciana Salgado. 2017. Audio Description on Instagram: Evaluating and Comparing Two Ways of Describing Images for Visually Impaired. In *Proceedings of the 19th International Conference on Enterprise Information Systems*, Vol. 3. SCITEPRESS - Science and Technology Publications, 29–40. https://doi.org/10.5220/0006282500290040

[5] Carlos Duarte, Carlos M. Duarte, and Luís Carriço. 2019. Combining Semantic Tools for Automatic Evaluation of Alternative Texts. In *Proceedings of the 16th Web For All 2019 Personalization - Personalizing the Web*. ACM, New York, NY, USA, 1–4. https://doi.org/10.1145/3315002.3317558

[6] Facebook. 2020. How do I edit the alternative text for a photo on Facebook? https://www.facebook.com/help/214124458607871

[7] Cole Gleason, Patrick Carrington, Cameron Cassidy, Meredith Ringel Morris, Kris M Kitani, and Jeffrey P Bigham. 2019. "It's almost like they're trying to hide it": How User-Provided Image Descriptions Have Failed to Make Twitter Accessible. In *The World Wide Web Conference on - WWW '19*. ACM Press, New York, New York, USA, 549–559. https://doi.org/10.1145/3308558.3313605

[8] Cole Gleason, Amy Pavel, Emma McCamey, Christina Low, Patrick Carrington, Kris M Kitani, and Jeffrey P. Bigham. 2020. Twitter A11y: A Browser Extension to Make Twitter Images Accessible. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA, 1–12. https://doi.org/10.1145/3313831.3376728

[9] Darren Guinness, Edward Cutrell, and Meredith Ringel Morris. 2018. Caption Crawler: Enabling Reusable Alternative Text Descriptions using Reverse Image Search. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems - CHI '18*, Vol. 2018-April. ACM Press, New York, New York, USA, 1–11. https://doi.org/10.1145/3173574.3174092

[10] Lawrence H Kim, Abena Boadi-Agyemang, Alexa Fay Siu, and John Tang. 2020. When to Add Human Narration to Photo-Sharing Social Media. In *The 22nd International ACM SIGACCESS Conference on Computers and Accessibility*. ACM, New York, NY, USA, 1–3. https://doi.org/10.1145/3373625.3418013

[11] Haley MacLeod, Cynthia L. Bennett, Meredith Ringel Morris, and Edward Cutrell. 2017. Understanding blind people's experiences with computer-generated captions of social media images. *Conference on Human Factors in Computing Systems - Proceedings* 2017-May (2017), 5988–5999. https://doi.org/10.1145/3025453.3025814

[12] Reeti Mathur and Erin Brady. 2018. Mixed-Ability Collaboration for Accessible Photo Sharing. In *Proceedings of the 20th International ACM SIGACCESS Conference on Computers and Accessibility - ASSETS '18*. ACM Press, New York, New York, USA, 370–372. https://doi.org/10.1145/3234695.3240994

[13] Meredith Ringel Morris, Jazette Johnson, Cynthia L. Bennett, and Edward Cutrell. 2018. Rich representations of visual content for Screen reader users. *Conference on Human Factors in Computing Systems - Proceedings* 2018-April (2018), 1–11. https://doi.org/10.1145/3173574.3173633

[14] Meredith Ringel Morris, Annuska Zolyomi, Catherine Yao, Sina Bahram, Jeffrey P. Bigham, and Shaun K. Kane. 2016. "With most of it being pictures now, I rarely use it": Understanding Twitter's Evolving Accessibility to Blind Users. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA, 5506–5516. https://doi.org/10.1145/2858036.2858116

[15] Tam The Nguyen, Phong Minh Vu, Hung Viet Pham, and Tung Thanh Nguyen. 2018. Deep learning UI design patterns of mobile apps. *Proceedings - International Conference on Software Engineering* (2018), 65–68. https://doi.org/10.1145/3183399.3183422

[16] Letícia Seixas Pereira, José Coelho, André Rodrigues, João Guerreiro, Tiago Guerreiro, and Carlos Duarte. 2021. Barriers and Opportunities to Accessible Social Media Content Authoring. arXiv:2104.10968 [cs.HC]

[17] H Petrie, C Harrison, and S Dev. 2005. Describing images on the web: a survey of current practice and prospects for the future. *Proceedings of Human Computer Interaction International (HCII)* (2005), 1 – 10.

[18] Carolina Sacramento, Leonardo Nardi, Simone Bacellar Leal Ferreira, and João Marcelo dos Santos Marques. 2020. #PraCegoVer: Investigating the description of visual content in Brazilian online social media Carolina. In *Proceedings of the 19th Brazilian Symposium on Human Factors in Computing Systems*. ACM, New York, NY, USA, 1–10. https://doi.org/10.1145/3424953.3426489

[19] Elliot Salisbury, Ece Kamar, and Meredith Ringel Morris. 2017. Toward Scalable Social Alt Text: Conversational Crowdsourcing as a Tool for Refining Vision-to-Language Technology for the Blind. In *Proceedings of HCOMP 2017*. AAAI.

[20] Abigale Stangl, Meredith Ringel Morris, and Danna Gurari. 2020. "Person, Shoes, Tree. Is the Person Naked?" What People with Vision Impairments Want in Image Descriptions. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA, 1–13. https://doi.org/10.1145/3313831.3376404

[21] Twitter. 2020. Twitter Accessibility. https://twitter.com/TwitterA11y/status/1265689579371323392

[22] Violeta Voykinska, Shiri Azenkot, Shaomei Wu, and Gilly Leshed. 2016. How Blind People Interact with Visual Content on Social Networking Services. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing - CSCW '16*, Vol. 1. ACM Press, New York, New York, USA, 1582–1593. https://doi.org/10.1145/2818048.2820013

[23] Gill Whitney and Irena Kolar. 2020. Am I missing something? *Universal Access in the Information Society* 19, 2 (jun 2020), 461–469. https://doi.org/10.1007/s10209-019-00648-z

[24] Shaomei Wu, Jeffrey Wieland, Omid Farivar, and Julie Schiller. 2017. Automatic Alt-text. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*. ACM, New York, NY, USA, 1180–1192. https://doi.org/10.1145/2998181.2998364

[25] Xiaoyi Zhang, Lilian de Greef, Amanda Swearngin, Samuel White, Kyle Murray, Lisa Yu, Qi Shan, Jeffrey Nichols, Jason Wu, Chris Fleizach, Aaron Everitt, and Jeffrey P. Bigham. 2021. Screen Recognition: Creating Accessibility Metadata for Mobile Applications from Pixels. arXiv:2101.04893 http://arxiv.org/abs/2101.04893

[26] Yuhang Zhao, Shaomei Wu, Lindsay Reynolds, and Shiri Azenkot. 2017. The Effect of Computer-Generated Descriptions on Photo-Sharing Experiences of People with Visual Impairments. *Proceedings of the ACM on Human-Computer Interaction* 1, CSCW (dec 2017), 1–22. https://doi.org/10.1145/3134756