Verbalizing Images

Lisa Tang and Jim Carter

Dept. of Computer Science, University of Saskatchewan,
Saskatoon, Saskatchewan, Canada
lisa.tang@usask.ca, carter@cs.usask.ca

Abstract. Although a picture is worth a thousand words, how can you communicate its meaning and content in less than 250 words when that is all you have? Images are often used to convey information, supplement textual content, and/or add visual appeal to documents. The Usability Engineering Lab (USERLab) at the University of Saskatchewan developed an approach for generating informative alternative text for all types of images. This paper describes the approach and reports on the results of applying the approach by developers, content providers, usability and accessibility specialists, and the general public users.

Keywords: alternative text, images, description, captions.

1 Introduction

Textual content within a document often makes reference to images within the same document without explicitly explaining what is being communicated by the image. Alternative text helps people (both sighted and non-sighted) understand the image content by providing equivalent information to an image in a textual form. Alternative text can be beneficial in the following scenarios:

- when multi-tasking and interacting with the computer aurally (e.g. while driving)
- when using a device where the picture cannot easily be seen (e.g. on a smart phone in environments producing glare on the screen)
- when the person has a visual disability and relies on screen reader technologies
- when the person cannot comprehend the image.

While alternative text is beneficial to have, how should it be composed? There is a need for guidance concerning what to describe about an image. The purpose that an image serves and its context in the document can help determine what to describe. An image could encapsulate a vast amount of information. While it is desirable to keep as much information as possible, there may be too much for a person to process. Some of the information may also be irrelevant to the situation or context.

People need guidance on how to identify important information, how to rate the information, and how to transform the information into comprehensible prose. This paper proposes an approach for guiding people to create informative and descriptive alternative text for images and other graphical content. This work is being developed into ISO/IEC technical report 20071-11[b].

C. Stephanidis (Ed.): Posters, Part I, HCII 2011, CCIS 173, pp. 394–398, 2011.

2 Background

While there is much guidance on containers for providing alternative text [a], there is currently a lack of literature and research regarding the *composition* of alternative text, that is, what information to include and how to put the information together. We investigated the research areas of library science, image indexing, captioning, audio description, art description, and tactile representation for guidance.

Types of Information. It is important to know what information to convey to the user before beginning to write alternative text. The types of information that can be communicated through an image can be categorized into "What", "Who", "Where", "When", "Why", "How", and "How much".

Intended Versus Actual Information. An image chosen to convey certain information may be interpreted differently by a user, leading to miscommunication. It is important to ensure that the intended information is clearly communicated.

General Versus Specific. Information about an image can be generic or specific, yet accurately describes the image. For example, a general description could be "dog" and a specific description could be "poodle". While both uses the same number of words, the specific description is more descriptive. Vivid, succinct, and imaginative words provide the user with a better mental image within a limited time.

Importance Levels. Different pieces of information can have different levels of importance. It is important that the user receives essential information (that they want or need) as quickly as possible. Then if they so choose, they can request for additional, less significant, information. Considering time or space limitations, it is important to identify the essential information and present it first.

Purpose and Context. Depending on the purpose and context of an image, different pieces of information become important. The alternative text used in one instance might not be applicable in another. Therefore, a standard alternative text might not be appropriate to be used for any given image and a new version of alternative text may need to be generated with each use of the image.

3 Method for Identifying Information for Images

The creation of suitable alternative text should be based on a thorough understanding of the image, its components, and the document that contains it. This can be done by applying the following approach:

1. **Justification of the Image.** The purpose of each image is identified and described. This step will influence which image components and information are important for the user to know. This answers the question "Why?"
2. **Identification of the Image Components.** Image components are identified depending on the purpose of the image. Image components may be a person, place, object, or area of the image that the user should know about. This step is necessary to properly identify information about the image that may be of importance to the

user. Identifying image components is an iterative process. While identifying the image (or component) content, it may become necessary to separate that image (or component) into several image components, and so on.

3. **Identification of the Image (or Image Component) Content.** The content of an image and its components are described based on the purpose and context of the image, along with its importance. This involves answering the question "What?" Identifying content often involves expert knowledge in the field of the image. For example, an art historian would have expert knowledge on paintings from the 1800s and can interpret the painting. However, those without expert knowledge can also identify content based on what they see in the image without interpretation.
4. **Elaboration of the Content.** Elaboration focuses on identifying specific details that might be significant in understanding an image or component. This involves answering applicable questions relating to "What?", "Who?", "When?", "Where?", "How much?", and "How?" from a comprehensive set of questions that can elaborate on a wide range of images. Not all of the questions will apply to every image, so this step should focus on significant questions and information.
5. **Organization of Alternative Text Information.** Information obtained from steps 1 to 4 is organized to improve its readability and allocated to the short or long descriptions of an image or to the document textual content. This involves removing redundancies; considering the importance of each piece of information; ensuring flow with the rest of the document; and reorganizing the information for readability.
6. **Evaluation of Alt-Text.** The resulting alternative text should be evaluated by someone other than the person who created it to verify that it suitably describes the image within the context of the document within which it is contained. Ideally, this would involve actual user testing. However, it is important that this step not be omitted due to lack of available users or resources. Evaluation by a colleague or friend is better than no external evaluation at all.

4 Prototype Tool

To make the approach easier (and thus more likely) to use, we created an online prototype tool that guides individuals through the first four steps of the approach. The prototype tool lets the user add and delete image components, save responses to questions about each component, and review the information. The prototype tool provides an opportunity to further improve the guidance in ISO/IEC 20071-11.

The prototype tool presents the set of questions that the individual should consider about the image. While it is not required to provide a response to every question in the set, it is highly recommended that all the questions be considered. To ensure that a question is at least considered, the prototype tool forces the individual to provide a response, even if only "No / Not Applicable".

The set of questions can be answered from a depth-first or a breadth-first approach. To facilitate both approaches, there are options to expand or collapse each category of questions individually, as well as to expand or collapse all categories at once.

The individual can complete the approach and answer questions over several sessions. The prototype tool saves the information provided by the individual into a database for review between sessions and ensures that the information is accurate to the best of their knowledge.

5 Evaluations

We conducted two research studies: to evaluate the approach in a document format and to evaluate the prototype tool. Both studies were evaluated by four user groups: developers, accessibility specialists, content specialists, and general public Internet users. Each participant was given the approach (in the form of a document or prototype tool), five images to describe, and a feedback form.

5.1 Procedure Document Evaluation

After analyzing the descriptions based on quality and quantity, we found that most developers performed as well or better than the other user groups and generally identified a higher quality of information. The participants felt that the document was logical, but they were unclear as to how to apply the approach. This suggests that the focus needs to change from developing guidance to supporting its application.

Half of the time (51.2%), participants did not break the image down into components. This may be because the participants did not understanding the importance of components or they did not know how to identify components. There was an expressed need for examples in order to better understand how the approach should be applied. For the evaluation, the examples were deliberate removed from the document in order to see how the approach would be interpreted.

Half of the participants focused on the general questions (e.g., “What is in the image?”) when describing the image and did not consider the detailed questions. Very few participants considered questions regarding relationships, location, or “When”. Identifying image components would also help to identify more detailed information regarding the image.

The three-level scale for rating the importance of a piece of information raised many questions and concern. The participants were not clear as to when each rating should be used and felt that they did not have the necessary knowledge to make that judgment.

Despite the difficulties the participants experienced in applying the approach, many were capable of writing informative alternative text. To improve their experience, many participants recommended that a tool be created.

5.2 Prototype Tool Evaluation

The prototype tool was tested by the same individuals from the first study. Participants used the prototype tool to apply the approach on a different set of five images. The participants felt that the tool-based approach was easier to understand than the document. The comments and feedback were focused on the phrasing and presentation of the questions in order to improve the user experience.

Based on the web page format, the participants did not feel the same freedom to write long descriptions. The responses tended to be more concise and less descriptive. Also, because the prototype tool forced the participants to consider all general and specific questions about each image component, the approach took on average over an hour to complete for each image.

In the existing structure of questions, some questions (such as those related to actions) exist across multiple categories (such as "Who" and "Logical Relationships"). Reorganization of the questions may help identify applicable questions faster and decrease the amount of time spent on the approach. The questions can also be rephrased to extract detailed information quickly.

Several participants (23.5%) commented that there was too much content within a single web page. They were overwhelmed by the amount of information to process and the number of tasks to complete. They suggested that the image components be identified on a separate page and the set of questions be presented over several pages.

Ideally, the approach would be completed within five to fifteen minutes. With the additional improvements described above, a tool can help extract essential information in a timely manner while generating higher quality descriptions.

6 Future Work

Our procedure for creating alternative text for images has proven to be a feasible approach that can help in creating good alternative text. Some of the recommendations for the prototype tool were implemented and the prototype tool is now publically available [c]. Further research and development is currently underway in the following aspects:

- Evaluate the quality of the resulting alternative texts generated by the procedure.
- Identify information currently missing from the set of information types.
- Identify which information is of greater importance for certain types of images.
- Clarify the guidance on levels of importance.
- Enhance the prototype tool to filter non-applicable questions.
- Provide additional guidance on transforming the information into alternative text.

Acknowledgements. We would like to thank Microsoft Canada for their support.

References

[1] ISO/IEC WD 20071-11 Information technology — User interface component accessibility — Guidance on creating alternative text for images (2010)
[2] Usability Engineering Lab (USERLab). Alternative Text Prototype Tool (2010), http://userlab.usask.ca/AltTextTool
[3] W3C. Web Content Accessibility Guidelines (WCAG) 2.0 (2010), http://www.w3.org/TR/WCAG20/Overview.html