

Verbalizing Images

Lisa Tang and Dr. Jim Carter, USERLab, University of Saskatchewan, lisa.tang@usask.ca, carter@cs.usask.ca

If a picture is worth 1000 words, what might those words be?

Introduction

While there is guidance on the technical methods of presenting alternative text (such as using the alt and longdesc attributes of an HTML image tag for web pages) [1], there is no guidance on what information to include in the actual alternative text. Alternative text is a textual representation of an image and can be helpful in the following situations:

- Multi-tasking and listening to the computer at the same time
- Using a screen reader due to a visual disability
- The screen is too small to see the entire image
- The user cannot comprehend the image

Some guidance exists for specific types of images or situations. However, it is sometimes hard to determine how and if those examples apply to each unique case. Therefore, guidance for images in general is needed to ensure a wider coverage of images.

ISO/IEC 20071-11 [2] describes a procedure that focuses on identifying the important information based on the purpose and context of the image and composing alternative text based on the information. The image information is collected through a series of questions categorized by "What", "Who", "When", "Where", "Why", "How Much", and "How". It is recognized that not all of the questions in each category would be applicable or important to every image. However, the questions help identify a wide range of information about an image that may be important to a given image.

The procedure is not specific to a particular image type or purpose, therefore, it can be applied to a large range of images and situations. By answering the applicable and important questions and identifying the important information, it is possible to create alternative text that is informative and useful to users.

Procedure

Step 1. Justification of the Image



Step 2. Identification of the Image Components



Step 3. Identification of the Image Content



Step 4. Elaboration of the Image Content



Step 5. Organization of Alternative Text Information



Step 6. Evaluation of Alternative Text

Step 1: Explain the purpose that the image serves by being in the document.

Step 2: Break the image down into parts in order to focus on different parts of the image and identify more specific details.

Step 3: High level information relating to "What" is identified according to the purpose and context of the image.

Step 4: Elaborate on the content identified in Step 3, relating to "Who", "When", "Where", "How much", and "How".

Step 5: Arrange the information identified in a way that flows with the rest of the document text.

Step 6: The resulting alternative text is evaluated to verify that it suitably describes the image within the document context.

Example

Image:

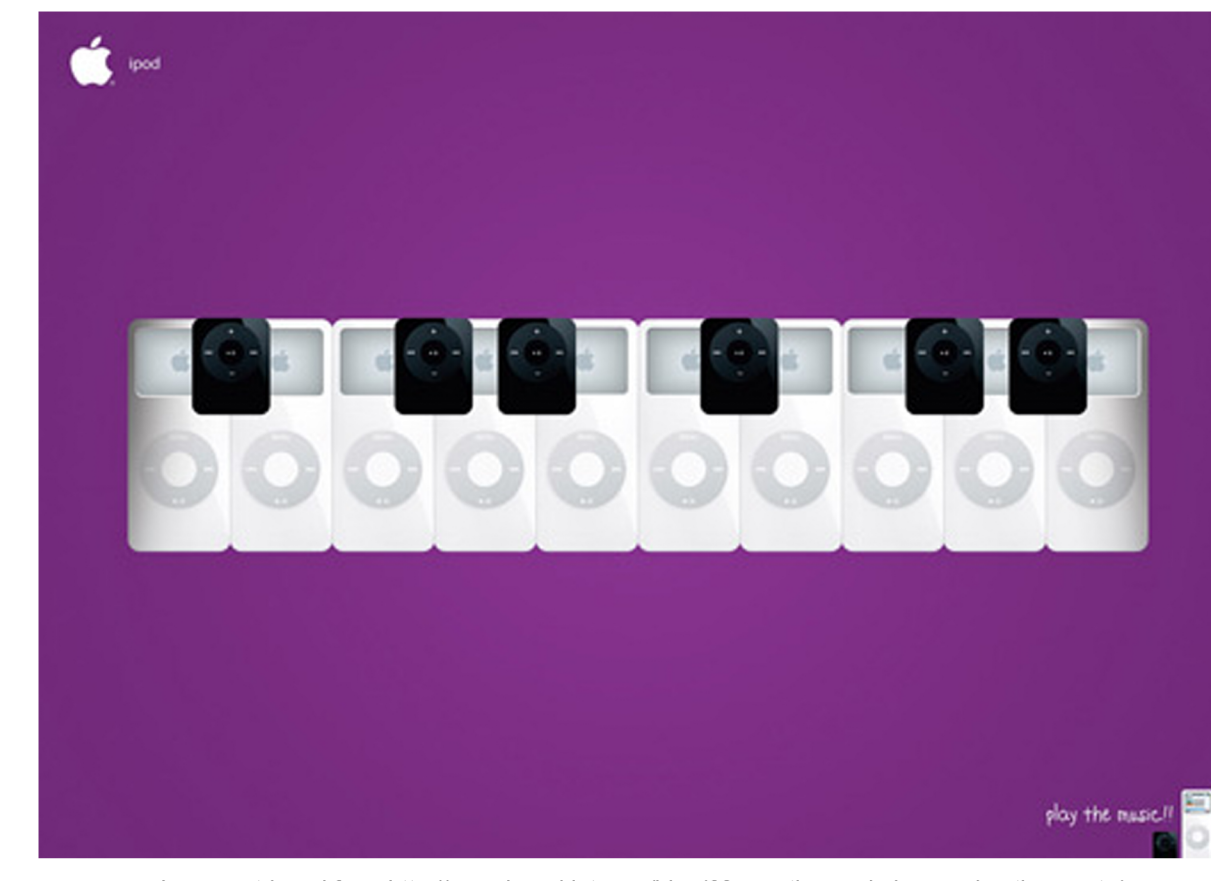
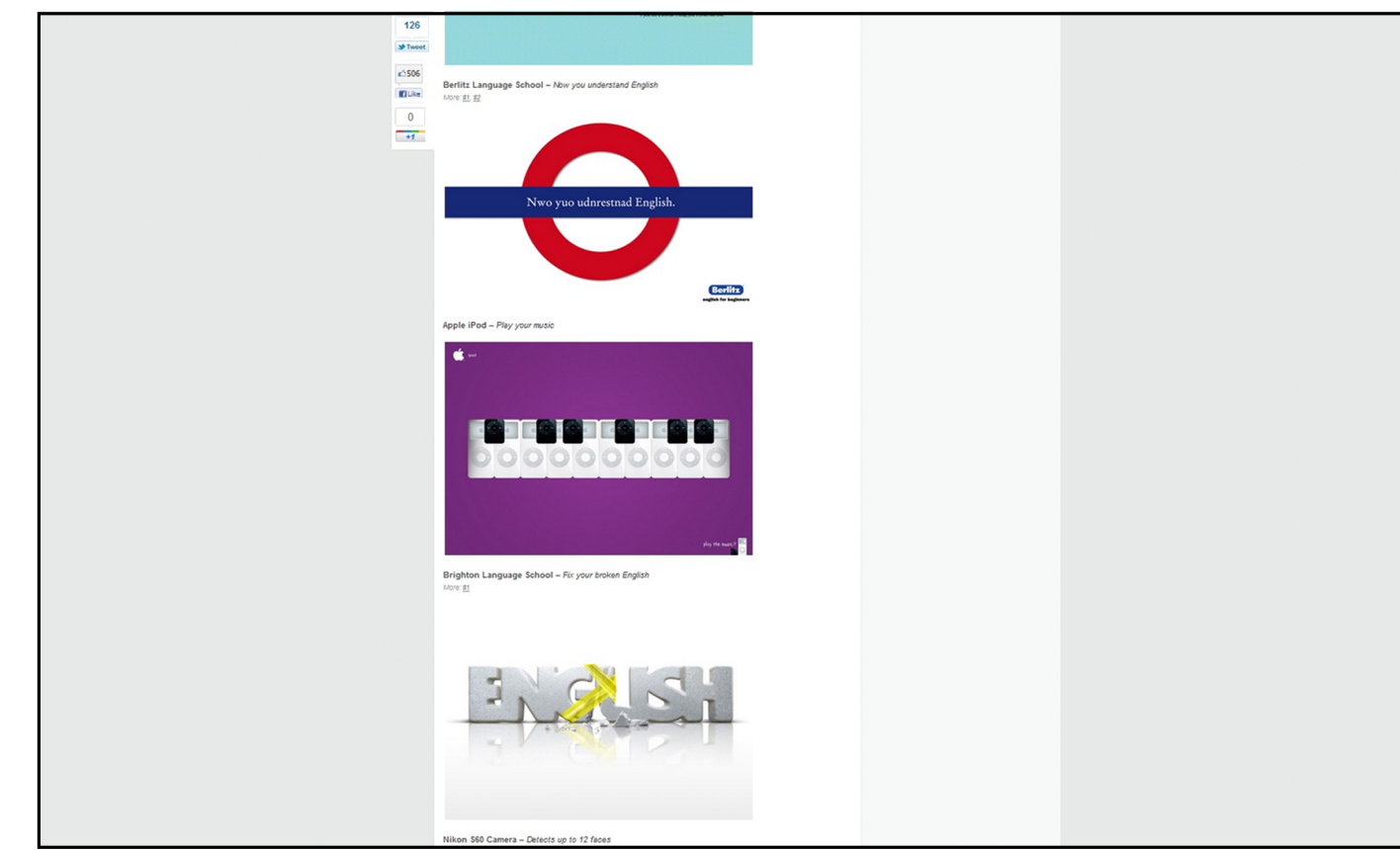


Image in Context:

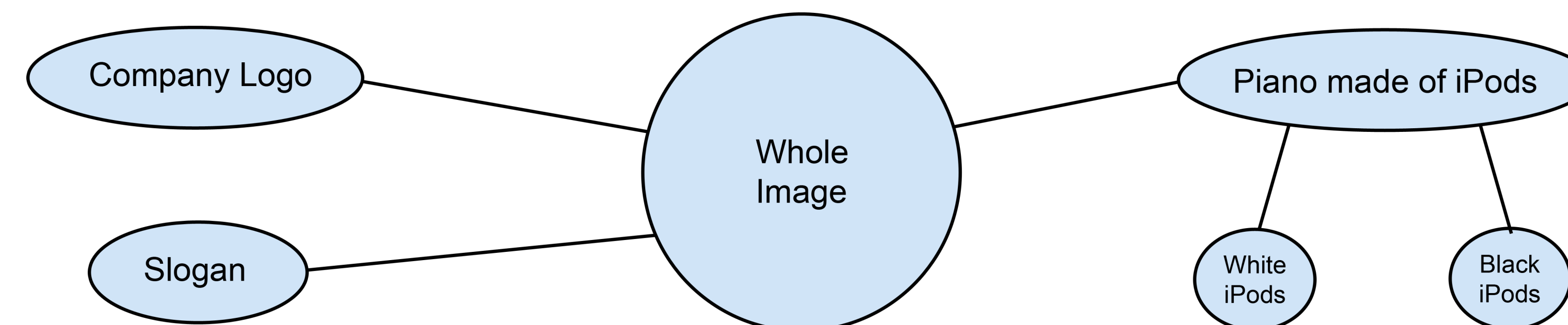


Step 1: Justification of the Image

"This image is used to inform the user by promoting iPods in a creative manner."

Step 2: Identification of Image Components

There are 4 major components and 2 minor components:



Step 3 & 4: Identification and Elaboration of Image Content

<p>Whole Image:</p> <p><i>What is the image or image component representing or symbolizing?</i> It's a metaphor that you can use iPods to play music, like a musical instrument. It's a double entendre on the word "play".</p> <p><i>Where is the image located with regards to the object that contains it?</i> Below its caption, the tenth image in the list.</p> <p><i>How is the user supposed to use the image?</i> Mouse over and drag the image.</p> <p><i>What is intended to result from interacting with the image?</i> Share image via social networks.</p>	<p>Piano Made of iPods:</p> <p><i>What is the shape of the component?</i> A piano</p> <p><i>What is the brand/model of the object?</i> iPod nano and iPod shuffle</p> <p><i>What are the colors of the component?</i> Nanos are white, shuffles are black</p> <p><i>Where is the component relative to other entities?</i> The shuffles are placed on top of the nanos, spaced out to look like a piano.</p> <p><i>What is the quantity?</i> 10 white nanos, 6 black shuffles</p>
<p>Company Logo:</p> <p><i>What is the color of the component?</i> White</p> <p><i>Where is the component located?</i> Top left corner</p> <p><i>What is the shape of the component?</i> An apple with a bite taken out of the right side.</p> <p><i>What does the text say?</i> "iPod"</p> <p><i>Where are the components relative to each other?</i> The apple is to the left of the text.</p>	<p>Slogan:</p> <p><i>Where is the component located?</i> Bottom right corner</p> <p><i>What does the text say?</i> "Play the music!"</p> <p><i>What does the object represent?</i> iPod nano, iPod shuffle</p> <p><i>Where are the components relative to each other?</i> The text is on the left, the iPod shuffle in the middle, iPod nano on the right</p>

Research Results

Three research studies were conducted:

- (1) Without the procedure
- (2) With the procedure in a document format
- (3) With the procedure as a prototype tool

Study 1 consisted of only Developers, while Studies 2 and 3 consisted a mixture of Developers, Content Providers, Usability Specialists, and Internet Users. In all three research studies, the participants were separated into two groups to describe and identify information about two different sets of five images (A and B). The participants in Study 2 also participated in Study 3. Those who applied the procedure to images in Set A for Study 2 then applied the procedure to images in Set B for Study 3, similarly for those who had Set B images for Study 2.

We found that most developers performed as well or better than the other user groups in identifying information about an image. However, the results also showed that the number of questions answered and the number of words written was highly dependent on the individual rather than user group. Table 1 shows the average number of questions answered and number of words written by each user group.

Table 1. Comparison Between User Groups in Identified Information

	Developer	Content Provider	Usability Specialist	Internet User
Questions Answered (Document)	8.3	6.7	7.2	6.0
Words Written (Document)	93.8	81.7	70.0	45.3
Questions Answered (Tool)	25.2	24.6	45.9	23.2
Words Written (Tool)	107.3	244.1	242.4	77.0

Comparing the results between research studies, we found that the procedure was helpful in identifying information about an image. Participants provided the least amount of information and limited range of information when they were not given the procedure to help identify information. Participants an increased amount of information and range of information with the help of the procedure. Even more information and range of information was identified when participants were forced to answer every question when using the prototype tool. Figures 1 and 2 show the number of questions answered (range of information) and the number of words written between research studies.

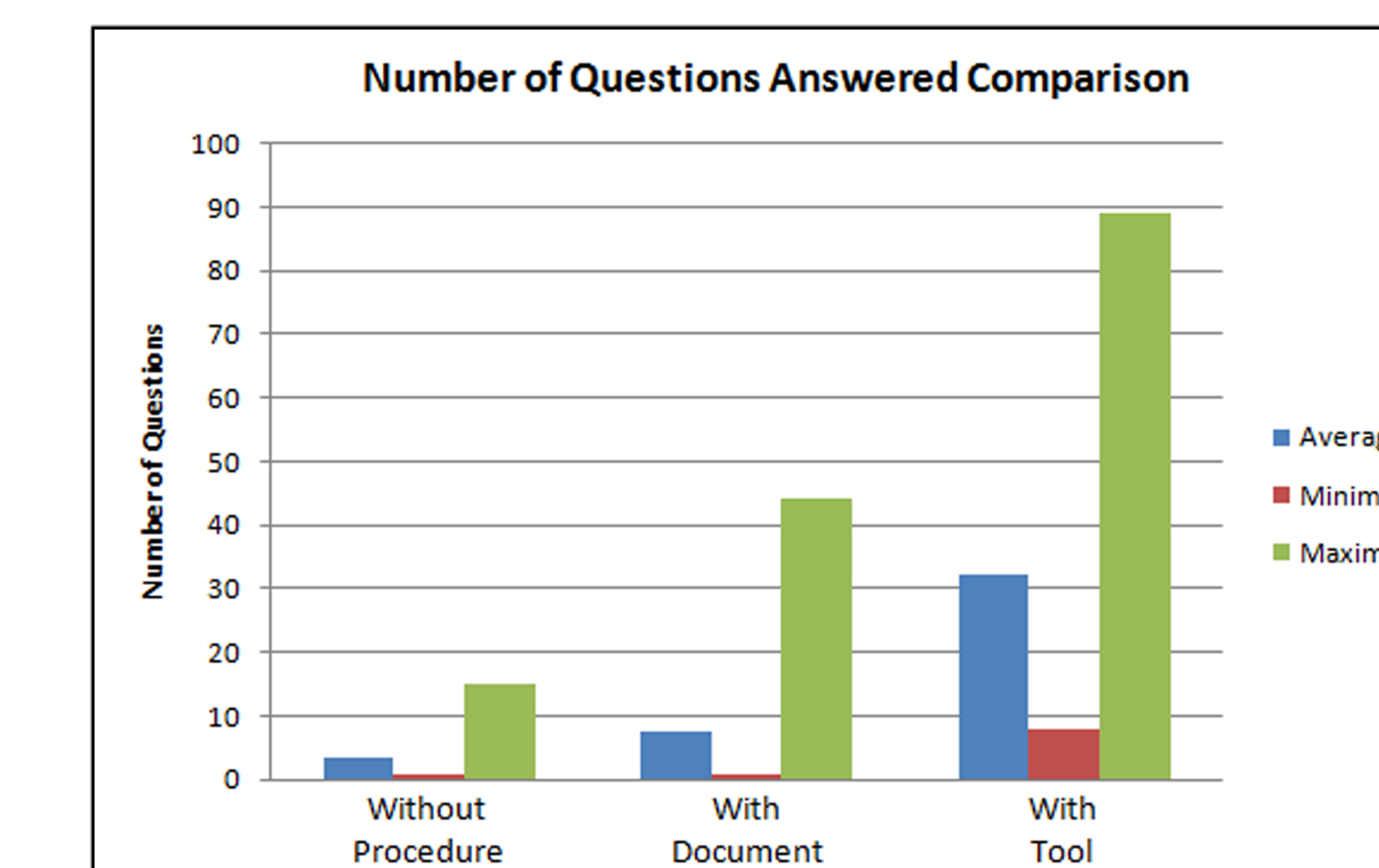


Figure 1. Number of Questions Answered Comparison

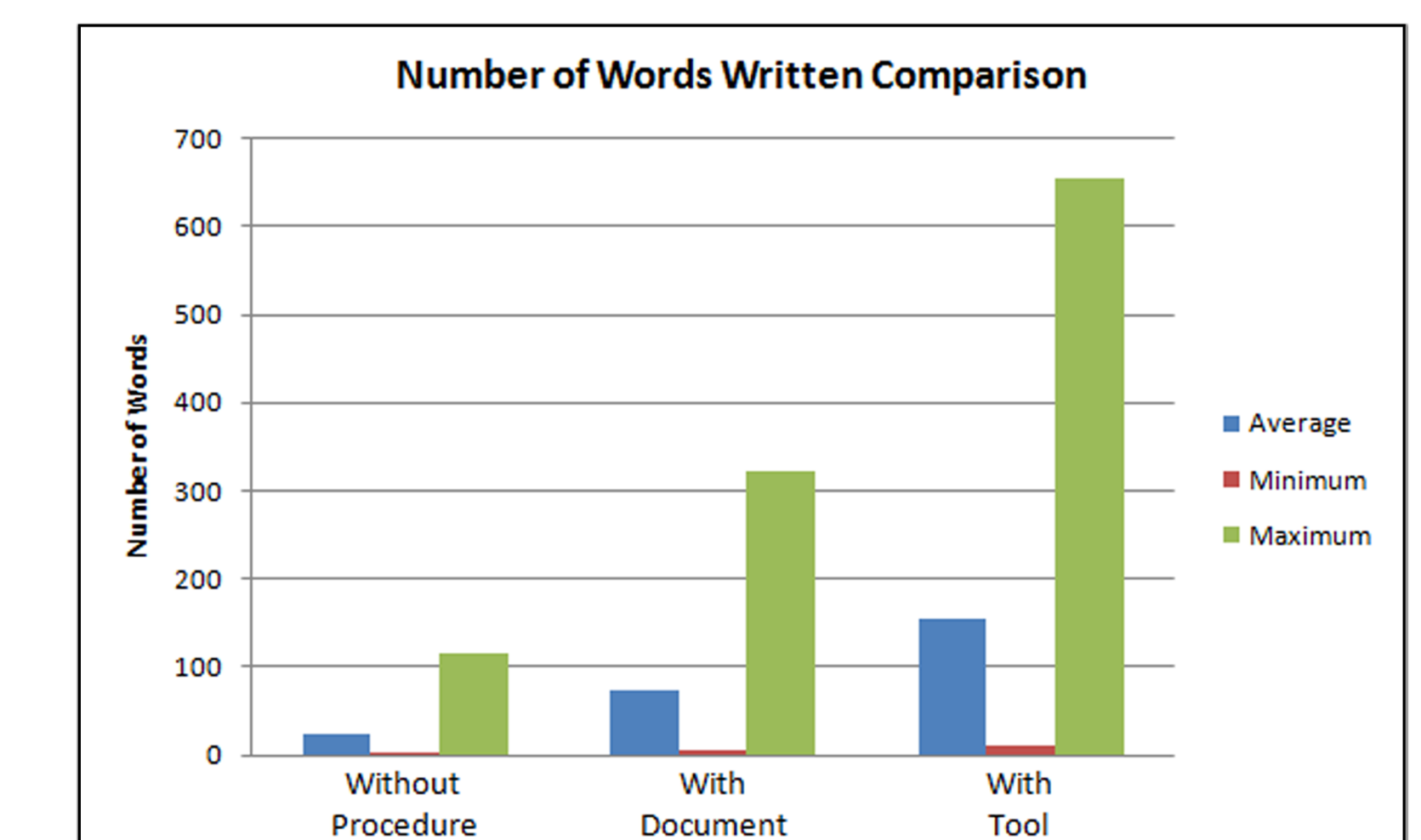


Figure 2. Number of Words Written Comparison

Considering the amount of time it took to identify information about an image, using a prototype tool proved to be helpful. While the number of words written per minute remained the same when applying the procedure with a document and a prototype tool, the average number of questions answered per minute improved (from 0.38 to 1.02). This shows that a greater number and range of questions can be answered (and information identified) in the same amount of time using the prototype tool. Improvements to the prototype tool could greatly improve the rate of output and decrease the amount of time necessary to create informative and useful alternative text.

We would like to thank our sponsors Microsoft Canada for their support.

References:
 [1] W3C. Web Content Accessibility Guidelines (WCAG) 2.0. (2010) <http://www.w3.org/TR/WCAG20/Overview.html>
 [2] ISO/IEC WD 20071-11 Information technology — User interface component accessibility — Guidance on creating alternative text for images. (2010)