Describing and Assessing Image Descriptions for Visually Impaired Web Users with IDAT

Julius T. Nganji, Mike Brayshaw and Brian Tompsett

Abstract People with visual impairments, particularly blind people face alot of difficulties browsing the web with assistive technologies such as screen readers, when websites do not conform to accessibility standards and are thus inaccessible. HTML is the basic language for website design but its ALT attribute on the IMG element does not adequately capture comprehensive image semantics and description in a way that can be accurately interpreted by screen readers, hence blind people do not usually get the complete description of the image. Most of the problems however arise from web designers and developers not including a description of an image or not comprehensively describing these images to people with visual impairments. In this paper, we propose the use of the Image Description Assessment Tool (IDAT), a Java-based tool containing some proposed heuristics for assessing how well an image description matches the real content of the image on the web. The tool also contains a speech interface which can enable a visually impaired individual to listen to the description of an image that has been uploaded unto the system.

1 Introduction

The primary language for authoring web pages is the HyperText Markup Language (HTML). HTML is used to position resources, including images on web pages

Julius T. Nganji
Distributed Reliable Intelligent Systems (DRIS) Lab, Department of Computer Science, University of Hull, Cottingham Road, Hull, HU6 7RX, United Kingdom. email J.Nganji@hull.ac.uk

Mike Brayshaw
Department of Computer Science, University of Hull, Cottingham Road, Hull, HU6 7RX, United Kingdom e-mail: M.Brayshaw@hull.ac.uk

Brian Tompsett
Department of Computer Science, University of Hull, Cottingham Road, Hull, HU6 7RX, United Kingdom e-mail: B.C.Tompsett@hull.ac.uk

as well as provide a description of some media. Image description in HTML is achieved through the ALT and LONGDESC attributes of the IMG element. Whilst ALT only enables a short description of the image within the <IMG >tag as seen in Listing 1,

Listing 1 A sample <img>element with an ALT attribute

```
<p>
<img src= "images/man.jpg" alt="a man in a corn field"/>
</p>
```

LONGDESC could be used for longer descriptions of the image within the <IMG>tag as represented in Listing 2 or could be used to specify the location of a description of the image, in the form of a URI as shown in Listing 3. However, it is not supported by some web browsers and some assistive technologies [2].

Listing 2 A sample <img>element with a longdesc attribute

```
<p>
<img src="images/man.jpg" alt="a man in a corn field"
longdesc="imagedescription.html"/>
</p>
```

Assistive technologies such as screen readers interact with web pages to read the content to people with visual impairments. Nevertheless, most websites do not provide adequate text alternatives to images [2] hence images are not well interpreted to blind people.

Listing 3 A sample <img>element with a d link

```
<p>
<img src= "images/man.jpg" alt= "a man in a corn field"
longdesc="imagedescription.html"/>[a href="imagedescrip
tion.html" title="a detailed description of the image"/>
D</a>
</p>
```

The Web Content Accessibility Guidelines (WCAG) 2.0[1], a W3C recommendation of 11 December 2008 advises web content creators to provide text alternatives for any non-text content and this is also reiterated by [16]. According to WCAG 2.0 guidelines,

> the objective of this technique is to create a text alternative that serves the same purpose and presents the same information as the original non-text content. As a result, it is possible to remove the non-text content and replace it with the text alternative and no functionality or information would be lost. This text alternative should not necessarily describe the non-text content. It should serve the same purpose and convey the same information. This may sometimes result in a text alternative that looks like a description of the non-text content. But this would only be true if that was the best way to serve the same purpose.

[1] http://www.w3.org/TR/WCAG20/

For screen readers to convey the correct information about an image to a visually impaired person, image description needs to be detail enough to provide a complete and true description. For instance, the image may have been taken in a specific location, during a specific season and may contain a number of objects (people, animals, etc.) of different sex (male, female) who may be displaying different emotions (smiling, frowning, etc.). All these in addition to other features of the image are things that are seen and appreciated by people with normal eye sight but which are hidden from visually impaired users.

The semantic web [11] is more meaningful and can enable machines to understand information and to communicate with each other in ways that have not been possible before. An ontology is "a specification of a representational vocabulary for a shared domain of discourse"[20]. Ontologies can facilitate semantic image description which can ease understanding of images on the web. Even with that, the ontology needs to contain comprehensive information about the image.

This paper therefore seeks to encourage web developers to present accurate and comprehensive image descriptions for visually impaired people by providing ten heuristics that will help in accomplishing this. The heuristics could be used as classes in an ontology to describe the image. The remainder of this paper is as follows: section 2 looks at the problems associated with an inaccessible web for people with visual impairments, section 3 reviews image description while section 4 explains the considerations for a good image description. Section 5 on the other hand provides a tool for assessing image descriptions. Section 6 presents some limitations and future work. The paper concludes in section 7 with a summary.

2 The difficulties of describing images on the web

For most people with normal eyesight, browsing the web is a painless task because most websites have been designed for people with normal vision. They can readily click on objects and follow links to other resources, even for websites that do not conform to accessibility and usability norms. This is, however, not the case for people with visual impairments, particularly those with congenital blindness who may have to rely on assistive technologies such as screen readers to browse through websites. Whilst the use of screen readers is helpful in browsing web content, they read out web content sequentially, and so could take a long time to reach the information required by the user thus [18] developed a multimodal interface to improve web accessibility for visually impaired people. When web designers do not adhere to web accessibility guidelines, visually impaired people browsing the web with screen readers will be excluded from accessing information since the assistive technology will be unable to interpret the content to the user [19]. [17] also noted some problems people with visual impairments face when browsing the web, such as problems associated with linking from one document to another given that they cannot follow visual cues, and the increasing use of multimedia and interactive documents which are difficult for such people to access.

Providing an inaccessible website means discriminating against disabled people by denying them access to online services. Most governments today have legislation to prevent discrimination of disabled people, with prosecutions for those who do not adhere to such legislations. The Americans with Disabilities Act[2] (ADA) of 1990 requires employers to provide "reasonable accommodation"and mechanisms for "effective communication"to workers with disabilities [17] while in the UK, the Disability Discrimination Act [3] (DDA) of 1995 and 2005 requires service providers to make "reasonable adjustments"to their services to meet the needs of disabled people.

We noted earlier that screen readers are unable to describe an image to a user if no description is provided. However, in the event where a description is provided but is insufficient or wrong, the visually impaired user will still be unable to understand the image. This calls for good practice in image description and employment of alternative technology to improve understanding of the image. Consequently, the problems associated with image description by assistive technologies need to be tackled with cutting edge technology, which we will explore in section 5.

3 Review of image description

Much work has been done on image analysis with focus on retrieval. In the mid-1980s, Sara Shatford [8] extended Erwin Panofsky's [9] *pre-iconographical, iconographical* and *iconological* model of describing an art work. Shatford further subdivided the three levels into *who, what, when* and *where*. Recent research has shown the lack of adequate image description [2] hence, [3] created a web mediator that automatically adds ALT tags based on an analysis of the image contents. [12] developed a classification framework for classification of image descriptions by users and found that users preferred general descriptions rather than specific or abstract descriptions. Most research on media semantics has focused on semantically annotating media to facilitate search and retrieval [e.g. 5, 10]. Nevertheless, there is a lack of research towards adequately describing images for assistive technologies to interpret to people with visual impairments.

4 Comprehensive image description

With the advent of the semantic web, ontologies have been employed in image annotation. The Resource Description Framework (RDF) [6] or the Web Ontology Language (OWL) [7] can be used to convey a meaningful description of an image because of their semantics. However, to comprehensively describe images for

[2] http://www.ada.gov/

[3] http://www.dwp.gov.uk/employer/disability-discrimination-act/

assistive technology interpretation, the image description needs to take into consideration some important heuristics that capture the full meaning and description of the image as represented in Fig. 1. This follows a *Who, What, When, Where* and *How* approach. For instance, *who* asks the questions relating to the people in the image, while *what* relates to other non-human objects including buildings, trees, automobile, etc. including their descriptions such as colour. *When* on the other hand asks questions related to time such as when the picture was taken (time, season, etc.) while *where* seeks to find out the location such as where the image was taken, the positions of various objects in the image, etc. *How* relates to actions, emotions, etc.

These heuristics could be used as classes in an ontology to describe images. An image's description can be represented on three levels: text, structured, and ontology-based structured similar to [13]. Considering Fig. 2, the text description could be as in Listing 4. The textual description serves to provide a description of the image in natural language, which could be understood by humans.

Listing 4 Textual description of Fig.2

```
A male of black African origin wearing some lenses and a
colourful African shirt is standing amongst some maize
plants in front of a non plastered house with a
protective window in Buea, Cameroon, holding and staring
at a green maize fruit still on the plant with both
hands
```

Although the above description could be well understood by humans, machines might not readily understand this, causing assistive technology interpretation to be a difficult task. Semantic web technologies can overcome the limitations of information retrieval using keyword-based methods [14] as it makes information more meaningful to people by making it more meaningful to machines [15]. Similarly, web ontologies could be employed to overcome the challenges of describing images with HTML. Nevertheless, that is not the focus of this paper which seeks more to use a tool to encourage comprehensive image descriptions and to assess the accuracy of the descriptions.

5 Image description assessment

Tools have been developed to evaluate various contents of websites. The Healthcare Website Assessment Tool (HWAT) is one of such tools used to evaluate the quality of osteoporosis websites [4]. Similarly, to evaluate image descriptions, we propose the Image Description Assessment Tool (IDAT) which adopts ten heuristics earlier described with various quality indicators and weighting scores as represented in Table 1.

Each heuristic has a weighting score of 10 and the extent to which the image description conforms to that heuristic is thus scored on a maximum of 10. The total weighting score will thus be 100 for an image that contains a scenario with ten

heuristics. People describe images differently and in their interpretation of images have various qualities or heuristics. It is therefore imperative to make provisions for this in calculating the accuracy of the description as in the following formula.

Let xi be the weighting score on 10 for an individual heuristic
and n be the total maximum weighting score.

Then the percentage accuracy (P) of the image description to the original image could be derived as follows:

$$P = \sum xi/n \times 100 \quad (1)$$

Let us consider the following description (see Fig. 2 caption) of an image which was taken during the rainy season in Buea, a small town in Cameroon.

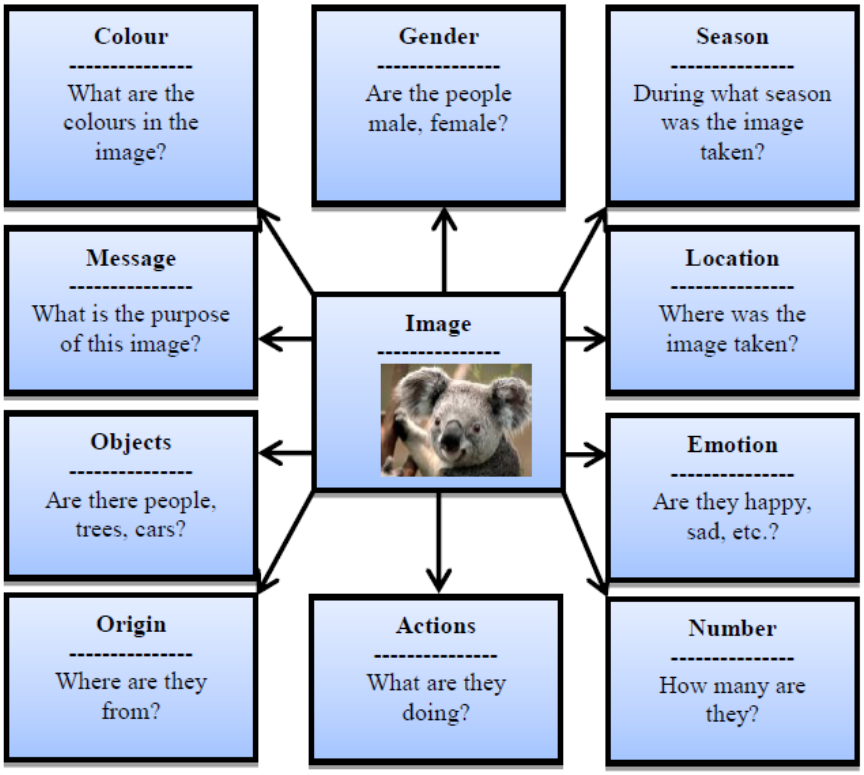

Fig. 1 Comprehensive image description heuristics

Table 1 Image Description Assessment Tool

Heuristic	Quality Indicator	Maximum Weighting Score
Location	Where was the image taken? Does the description of the image adequately convey this information?	10
Objects	What are the objects in the image (people, trees, cars, buildings, etc.)? How well does the image description convey this information?	10
Gender	What is the gender of the people and the animals in the image? Are they male or female and does the description convey this message?	10
Colours	What are the various colours in the image? Do tree leaves colours show it is autumn, spring, summer, winter, or is it rainy or dry season and does the image description adequately convey this?	10
Emotions	What emotions (smiling, frowning, etc.) are expressed by people, animals, etc. in the image and does the image description adequately convey this information?	10
Numbers	How many of those objects are in the image and does the image description adequately convey this information?	10
Message	What is the image all about? How well does the image description adequately convey this message?	10
Origin	What is the origin of the people, animals, etc. in the image and does the image description convey this information?	10
Action	What is the state of motion of the objects in the image? Are they moving, still, etc. and are these well represented in the description?	10
Season	During what season was this image taken and if applicable, is this represented in the description?	10

To assess how close the description is to the original image, IDAT can be applied as in Table 1. When a heuristic is present in the description, 1 is recorded and a weighting score given on a scale of 10. Where a heuristic is absent in the description but the image clearly shows the presence of that heuristic, 1 is recorded and a score of 0 is given. On the other hand, where a heuristic is absent in the description and the image clearly shows that the heuristic cannot be applied to the description, 0 is recorded and no score is attributed. Based on these conditions, we can now assess the description of Fig. 2 as shown in Table 2.

Where the image was taken cannot be determined by looking at the image, hence no weighting score is given. The season has not been mentioned. The maize plant

Table 2 Image description with IDAT

Heuristic	Present or absent (1 or 0)	Weighting score/10)	Maximum weighting score
Location	0	-	-
Objects	1	10	10
Gender	1	10	10
Colours	1	0	10
Emotions	0	-	-
Numbers	1	10	10
Message	1	5	10
Origin	0	-	-
Action	1	8	10
Season	0	-	-
Total		53	60

requires rainfall to grow during the rainy season but it is also possible to be grown during the dry season if it is well watered. That description scores zero for colours as the green colour of the plant and other dominant colours in the image have not been mentioned. The message is not 100% accurate because the man in the image is not holding a camera. It follows from the equation that the percentage accuracy of the description is

$$P = 53/60 \times 100 = 88.33 \tag{2}$$

Fig. 2 A male of black African origin wearing some lenses is standing amongst some maize plants in front of a house and holding a camera with both hands.

The accuracy of the description is thus 88.33%.

A screenshot of IDAT is seen in Fig.3.

6 Limitations and future work

In assessing image description with IDAT, the tool relies on the assessor's ability to interpret the image, which in itself may be faulty. The need for a tool to automate the process has been raised and further research will investigate this.

IDAT implementation is nearly completed and once the process is over, user testing will be done and the results used to improve the tool. This system allows users to upload any image and to provide a description of the image with a text to speech output of the description as depicted in Fig.3. It also allows the person describing the image to be able to listen to the description of the image while describing it. Finally, IDAT enables users to compute the percentage accuracy of an image's description and hence can be used to encourage web designers/developers to provide comprehensive image descriptions for visually impaired web users.

7 Conclusion

This paper has discussed the limitations of providing image descriptions with HTML for assistive technology (screen reader) interpretation and introduced the need to comprehensively describe images for visually imapired web users by proposing ten heuristics incorporated in an Image Description Assessment Tool (IDAT) which calculates the degree of accuracy of an image description compared to the original image.

This will hopefully provide a means of increasing web accessibility for people with visual impairments as it will encourage adherence to web accessibility standards.

Acknowledgements The authors would like to thank the Department of Computer Science, University of Hull, UK for funding that enabled this research to be carried out and presented. Many thanks to the anonymous reviewers of this paper for their comments and to Shawulu H. Nggada for his insightful comments on refining and improving IDAT.

References

1. Espinosa Peraldi, S.,Kaya, A., Melzer, S., and Wessel, M. Towards a Media Interpretation Framework for the Semantic Web. *2007 IEEE/WIC/ACM International Conference on Web Intelligence*, **1**5(5), 795–825 (2007)

2. Petrie, H., Harrison, C. and Dev, S. Describing images on the web: a survey of current practice and prospects for the future. In: Universal Access in HCI: Exploring New Dimensions of Diversity, Volume 8, *Proceedings of the 3rd International Conference on Universal Access in HumanComputer Interaction, 22–27 July 2005, Las Vegas, Nevada*). New Jersey: Lawrence Erlbaum Associates, (2005)
3. Keysers, D., Renn, M. and Breuel, T.M. Improving Accessibility of HTML Documents by Generating Image-Tags in a Proxy. In *Proceedings of the Ninth International ACM SIGACCESS Conference on Computers and Accessibility*, pp. 249–250, (2007).
4. Lewiecki, E. M., Rudolph, L. AKiebzak, G. M., Chavez, J. R. and Thorpe, B. M. Assessment of osteoporosis-website quality, Osteoporos Int, **17**, 741–752, (2006)
5. Yang, H.C. and Lee, C.H. Image semantics discovery from web pages for semantic-based image retrieval using self-organizing maps. Expert Systems with Applications, **34** (1), 266–279, (2008)
6. Lassila, O. and Swick, R.: Resource Description Framework Model and Syntax Specification. World Wide Web Consortium (1999) Available: http://www.w3.org/TR/1999/REC-rdf-syntax-19990222. Cited 10 Mar 2011
7. McGuinness, D.L. and Van Harmelen, F.: Web Ontology Language Overview. World Wide Web Consortium (2004) Available: http://www.w3.org/TR/owl-features/ Cited 10 Mar 2011
8. Shatford, S.: Analyzing the subject of a picture: a theoretical approach. Cataloging Classification Quarterly, **6**(3),39-62 (1986)

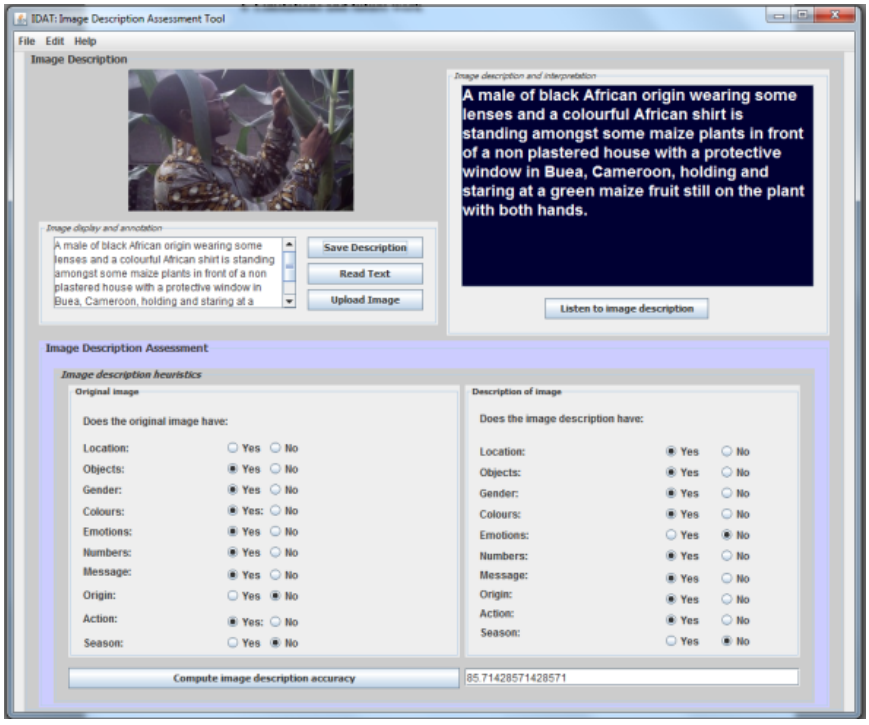

Fig. 3 IDAT showing some features including entering image description, listening to the description and calculating the image description accuracy

9. Panofsky, E.: Studies in Iconology: Humanistic Themes in the Art of the Renaissance. New York: Harper Row, pp. 5–9, (1972)
10. Elahi, N., Karlsen, R. and Akselsen, S.: A Context Centric Approach for Semantic Image Annotation and Retrieval, computationworld, pp.66–668, 2009 Computation World: Future Computing, Service Computation, Cognitive, Adaptive, Content, Patterns, (2009)
11. Berners-Lee, T., Hendler, J. and Lassila, O.: The Semantic Web, Scientific American, pp. 35–43, (2001)
12. Hollink, L., Schreiber, G., Wielinga, B., and Worring, M.: Classification of User Image Descriptions. Intl J. Human Computer Studies, **61**(5), 501–626, (2004)
13. Ruotsalo, T.: Methods and Applications for Ontology-Based Recommender Systems. Doctoral Diss., Aalto University, (2010)
14. Castells, P., Fernndez, M and Vallet, D.: An Adaptation of the Vector-Space Model for Ontology-Based Information Retrieval. IEEE Transactions on Knowledge and Data Engineering, **19**(2), 261–272, (2007)
15. Ohler, J.: The Semantic Web in Education. Educause Quarterly, **31**(4), 7–9, (2008)
16. Baguma, R., and Lubega, J.T.: Web Design Requirements for Improved Web Accessibility for the Blind. In: Fong, J., Kwan, R and Wang, F L. (eds.) Hybrid Learning and Education. Lecture Notes in Computer Science, pp. 392-403. Springer, Heidelberg (2008)