Improving Accessibility of HTML Documents by Generating Image-Tags in a Proxy*

Daniel Keysers
German Research Center
for Artificial Intelligence (DFKI)
Kaiserslautern, Germany
daniel.keysers@dfki.de

Marius Renn
German Research Center
for Artificial Intelligence (DFKI)
Kaiserslautern, Germany
marius.renn@dfki.de

Thomas M. Breuel
Computer Science Dept.
Technical University
Kaiserslautern, Germany
tmb@informatik.uni-kl.de

ABSTRACT

The widespread use of images without ALT tags on web pages reduces accessibility for the visually impaired. We present a system that automatically adds ALT tags based on an analysis of image contents.

Categories and Subject Descriptors

K.4.2 [**Computers and Society**]: Social Issues—*Assistive technologies for persons with disabilities*

General Terms

Human Factors, Algorithms

1. INTRODUCTION AND RELATED WORK

Missing alternative textual descriptions for images on web pages constitute an important aspect of reduced accessibility [6]. For example, when using a text-to-speech device for web-pages, "if no ALT tag is provided for an image, the user simply hears the word 'image' when they mouse over it" [7]. The W3C Web Content Accessibility Guidelines include as "Guideline 1" the recommendation to "Provide equivalent alternatives to auditory and visual content" with the first checkpoint being "Provide a text equivalent for every non-text element (e.g., via `"alt"`, `"longdesc"`, or in element content)" [8]. Not following this guideline reduces the usability of web pages not only for the visually impaired, but also for users of cell phones, low-bandwidth connections, or text browsers. However, only 39.6% of significant images of high-traffic websites were found to be labeled with alternative text in a recent study [2].

Our approach to making web pages more accessible is the creation of a web mediator that automatically adds missing `alt`-tags to `img`-tags [3]. Mediation for improving accessibility has been used by a number of existing services, e.g., by the BBC [1]. Mediated annotation of images in web pages has been independently developed as part of the WebInsight project [2]. WebInsight attempts to infer good `alt`-tags based on image context and link structure of the web page, optical character recognition (OCR), and manual human labeling.

*Work partially funded by the BMBF (German Federal Ministry of Education and Research), project IPeT (01 IW D03).

ASSETS'07, October 15–17, 2007, Tempe, Arizona, USA.
ACM 978-1-59593-573-1/07/0010.

Human labeling has a number of limitations, however, including cost and availability of labelers, as well as privacy and security issues. Our project has therefore aimed from the start at completely automated mediation for vision impaired users.

We treat the creation of ALT tags as the problem of automatic metadata creation or automatic tagging of images based on their content. The assumption is that the tags that users naturally supply for images in existing photo collections (e.g., Flickr, personal photo databases) contain a lot of information about the image itself. Alternatively, since the system is data driven, tagged image collections created manually for systems like WebInsight can also be used as the tagged image collection used to train our system.

Our fully automatic tagger[1] is based on image similarity measures as they are used in image retrieval and a collection of already tagged images. Similar system have recently also been developed by other groups (e.g. [5]). Using content-based similarity, for a given image without textual description, we can find a set of visually similar images in our labeled data base. The descriptions of the similar images can then be used to predict a suitable description for the new image. By adding category labels, we can also supply category names for image types that occur often in web pages like 'logo', 'icon', or 'photo'.

In addition to the automated tagging results, we also perform a simple rule-based image analysis in order to generate additional textual descriptions of the image, giving the user information about image size, dominant colors, and image type.

2. GENERATING IMAGE-TAGS

Figure 2 shows an overview of the architecture of our system. An HTTP proxy analyzes the requested and fetched HTML documents. It then fetches the embedded images and asks the image tagger for appropriate descriptions. There are several ways to implement the proxy, of which we chose to use the Apache web server and PHP scripting. Generated tags are cached to speed up the response if the same image is to be processed more than once.

Figure 1 shows example results of automatically generated `alt`-tags viewed as tool-tips on mouse-hovering (the mouse pointer is not shown). You can observe one interesting result in the last image, which is labeled with the term "horses". This is due to the fact that there are no images of soccer players in our labeled image database, but several of horses

[1]A short video demo of the tagger is available at `http://demo.iupr.org/tagger`

Figure 1: Examples of automatically generated ALT-tags.

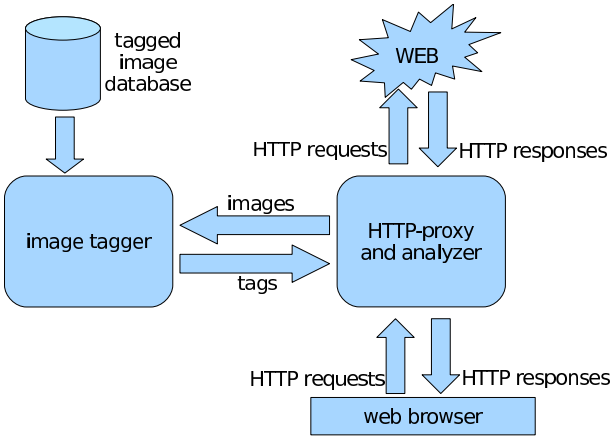

Figure 2: Architecture of our system.

on green grass. Thus the tagger predicts the label "horses" incorrectly here. This illustrates the dependency on a well-labeled image data base for this approach to work well.

To automatically add descriptive tags to an image, a system was designed that assigns a category and a set of tags to an arbitrary image. Together, these describe the query image, where the category focuses more on the type of image (e.g. photo, icon, graphic), while the tags give more information on the actual image contents. The system design follows a client-server architecture: the server, which is a slightly modified version of the FIRE image search engine [4], houses a large database of images and image features, for which the tags are known in advance. To evaluate image similarity, we use a weighted combination of Tamura texture features and RGB-color histograms, in both cases using the Jensen-Shannon-divergence as the distance measure. This combination has been shown to yield good results in comparison to various other approaches [4]. For a new image, a k-nearest-neighbor search through the image database is performed. The nearest neighbors' tags and categories are used along with their similarity scores to determine the output tags and categories using a voting scheme.

3. CONCLUSION

We have shown a system that performs first steps towards automated addition of missing `alt`-tags to images in HTML documents. The system already appears to be a useful alternative to manual tagging regarding both cost and privacy.

In fact, even though our system does not attempt to perform a deep semantic analysis of the input images–and such an analysis is still far beyond the state of the art in image understanding–our shallow, retrieval-based approach already appears to capture some of the information that users require in these kinds of applications [6], including color, location, and the presence of people. Even seemingly high-level properties, like emotion and atmosphere, can be captured through low-level analysis since they are often expressed through choice of color, focus, and brightness.

Currently, we are aiming at improving the system by evaluating further features and adapting the tagging method to the image type, by constructing better training data sets, and by incorporating OCR output using the open source OCR system OCRopus (`ocropus.org`). Furthermore, WebInsight [2] and the data it produces could be combined very well with our approach.

Perhaps the most important next step will be the creation of standard data sets and the performance of user studies to determine how large the gap between manual and automated annotation is, and what changes to the system lead to demonstrable improvements in user satisfaction and productivity.

4. REFERENCES

[1] BBC. Betsie (BBC education text to speech internet enhancer). Dec. 1999. Retrieved June 20, 2007 from http://www.bbc.co.uk/education/betsie/about.html.

[2] J.P. Bigham, R.S. Kaminsky, R.E. Ladner, O.M. Danielsson, and G.L. Hempton. Webinsight: making web images accessible. In *Proc. 8th int. ACM SIGACCESS conf. on Computers and Accessibility*, pages 181–188,Portland, OR, 2006.

[3] S.S. Brown and P. Robinson. A world wide web mediator for users with low vision. In *CHI Human Factors in Computing Systems Workshop No. 14*, 2001.

[4] T. Deselaers, D. Keysers, and H. Ney. Features for image retrieval: A quantitative comparison. In *DAGM 2004, 26th Pattern Recognition Symposium*, LNCS 3175, pages 228–236, Tübingen, Germany, Aug. 2004.

[5] J. Li and J.Z. Wang. Real-time computerized annotation of pictures. In *MULTIMEDIA '06: Proc. 14th annual ACM int. conf. Multimedia*, pages 911–920, Santa Barabra, CA, 2006.

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

[7] J.T. Richards and V.L. Hanson. Web accessibility: a broader view. In *WWW '04: Proc. 13th int. conf. World Wide Web*, pages 72–79, New York, NY, 2004.

[8] W3C. Web content accessibility guidelines. May 1999. Retrieved June 20, 2007 from http://www.w3.org/TR/WAI-WEBCONTENT/.