WebInSight: Making Web Images Accessible

Jeffrey P. Bigham, Ryan S. Kaminsky, Richard E. Ladner,
Oscar M. Danielsson and Gordon L. Hempton
Department of Computer Science and Engineering
University of Washington
Seattle, WA 98195 USA
{jbigham|rkamin|ladner|osda02|ghempton}@cs.washington.edu

ABSTRACT

Images without alternative text are a barrier to equal web access for blind users. To illustrate the problem, we conducted a series of studies that conclusively show that a large fraction of significant images have no alternative text. To ameliorate this problem, we introduce WebInSight, a system that automatically creates and inserts alternative text into web pages on-the-fly. To formulate alternative text for images, we present three labeling modules based on web context analysis, enhanced optical character recognition (OCR) and human labeling. The system caches alternative text in a local database and can add new labels seamlessly after a web page is downloaded, resulting in minimal impact to the browsing experience.

Categories and Subject Descriptors

K.4.2 [**Social Issues**]: Assistive technologies for persons with disabilities; H.5.2 [**Information Interfaces and Presentation**]: User Interfaces

General Terms

Design, Human Factors

Keywords

Web accessibility, web studies, transformation proxy, optical character recognition

1. INTRODUCTION

Blind users do not currently have equal access to the web. Images are used in navigation bars, as form buttons and to display textual and visual content, but, unless web authors provide alternative text for these images, blind users employing screen readers and refreshable Braille displays are left to guess the images' contents. In our studies we found that a large fraction of images lack alternative text. For example, of the significant images found on the homepages of the 500 most high-traffic websites[1], only 39.6% were assigned alternative text.

Illustrating the problem, the homepage of the UCLA Computer Science Department (Figure 1) contains 30 images that should have alternative text, but only two (6.67%) were assigned any text at all. As a result, a blind user may have difficulty navigating this page. As another example, the University of Michigan Computer Science and Engineering Department provides a listing of all faculty on its website along with their contact information (Figure 2). Unfortunately, each e-mail address is presented as an image without equivalent alternative text, presumably in a misguided attempt to thwart e-mail harvesters.[1] The unintended consequence, however, is that blind users who want to e-mail these professors either cannot do so or must find their e-mail addresses through other means. The problems highlighted here are characteristic of many sites throughout the web.

The W3C accessibility guidelines recommend that each image be assigned a textual equivalent[23], and the HTML standard includes easy ways to provide such *alternative text* with the `alt` and `longdesc` attributes of the `img` tag. The `title` attribute is another non-standard method of providing alternative text that is recognized by many accessibility tools, such as screen readers and refreshable Braille displays. The negligence of web authors in providing alternative text is one cause of web inaccessibility. Complicating matters, the proper selection of alternative text is considered by many to be more of an art than a science, which increases the difficulty of construction and verification of alternative text. Images that are crucial for understanding or navigating a page should have appropriate alternative text. Images that only serve to enhance the visual appeal of a page should have a zero-length `alt` attribute defined to make this intention clear[20]. Deviation from these accessibility guidelines generally leads to web pages that are less accessible.

Images that either have an action associated with them (i.e., links or buttons) or are multicolored and larger than a certain size are of particular concern. When such images do not have alternative text, accessibility can be severely reduced. The WebInSight system presented in this paper targets such significant images and provides a mechanism for automatically inserting appropriate alternative text. It does this by processing web requests and transforming the returned pages on-the-fly. As part of the transformation process, it coordinates three novel, underlying image-labeling

[1]All of the e-mail addresses listed on the page were also found in the Google index at least once, meaning e-mail harvesters could look elsewhere to find them.

ASSETS'06, October 22–25, 2006, Portland, Oregon, USA.

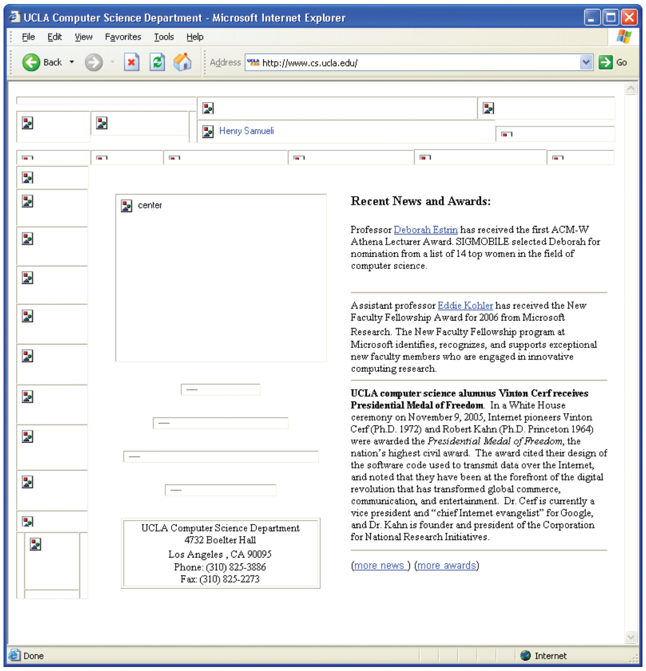

Figure 1: A screenshot of the UCLA Computer Science Department's homepage with images removed. Only 2 of the 30 significant images on the page have alternative text, including only one of the linked images.

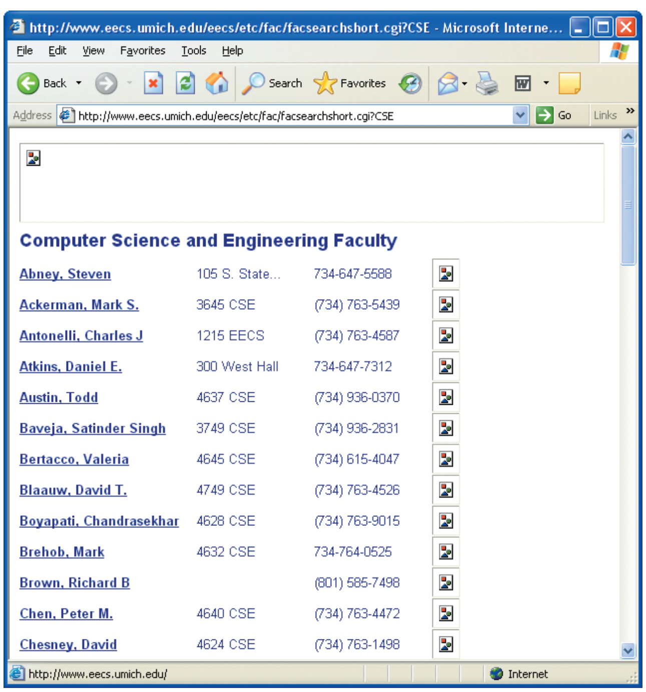

Figure 2: A screenshot of the directory listing for the Computer Science and Engineering faculty at the University of Michigan. With images turned off, the right column of e-mail addresses is inaccessible.

modules targeted to this domain, utilizing methods based on enhanced web context analysis, optical character recognition (OCR) and human labeling.

In this paper, we present a series of web studies that demonstrate the observed problem, discuss the architecture and implementation of WebInSight and, finally, describe the labeling modules used to assign alternative text to arbitrary images on the web.

2. RELATED WORK

To the best of our knowledge, our system is the first to automatically generate alternative text for web images and dynamically add this alternative text to the parent web page. However, it is not the first system to perform web personalization for particular users or user groups nor is it the first to do so via a transformation proxy[3]. Petrie *et al.* observed that the lack of proper alternative text needs to be addressed for web accessibility, but hoped only to encourage web designers to provide it[17]. Ahn *et al.* suggested that alternative text could be stored in a centralized database and be added to web pages by a browser plugin or extension[22].

Three previous web studies found that less than half of `img` tags found on the web contain an `alt` attribute. The first found that 24.9% [22] of images in a random selection of web pages were labeled. In the second 47.7% and 49.4% [7] of images found in two random sets of web pages derived from Yahoo and Google were labeled, respectively. These studies considered random web pages uniformly, but the web is known to contain a few very popular websites and many unpopular websites[5]. Considering each web page equally during analysis may not as accurately capture the experience of a user compared with a study that weights popular pages more highly. These studies also failed to distinguish between images of differing importance. The third study found that 45.8%[2] of nearly 6300 images contained in 100 homepages considered were assigned alternative text[17]. This study manually differentiated images based on their function and, as a result, only considered a small sample of web pages. The web studies presented in the next section seek to improve upon the shortcomings in these previous studies.

Much previous work exists on the difficult task of deriving textual labels for arbitrary images for keyword-based image retrieval tasks. These textual labels are also an attractive source of alternative text for web images. One focus has been on content-based image retrieval, but determining an accurate textual label for an arbitrary image using computer vision techniques is generally unsolved[9]. Most image retrieval systems instead rely largely on image context rather than image content. In such work, a valuable source for labels is derived from pre-existing alternative text, meaning that these systems are best at indexing the images that are already accessible[18]. Work in this area constructs textual labels by statistically associating images with nearby text. While useful for building an index for image retrieval, users of WebInSight are unlikely to benefit from automated natural language analysis of contextual text to which they already have immediate access. In contrast, WebInSight uses text found on the pages pointed to by linked images, which is not directly available to users without leaving the page, to formulate alternative text from context for images.

Another approach is to have humans label images. The

[2]This number has since been changed by the authors to 32%.

ESP Game[21] and Phetch[22] are computer games designed to generate keywords and explanatory sentences for arbitrary images. Photo-sharing sites, such as Yahoo's flickr[8], are becoming quite popular and provide a mechanism for users to caption their photos. These captions are already being used as alternative text. Related work has also suggested semi-automated semantic labeling of images in order to narrow the space of possible labels for use in the semantic web[15]. Instead of always requiring human labeling, WebInSight allows users to chose when to incur this added cost. Furthermore, for an important subset of images, including images not already in its database, WebInSight is capable of generating alternative text automatically.

3. WEB ACCESSIBILITY STUDIES

To gauge the current accessibility of web images, we designed and executed a series of web studies that measure the prevalence of alternative text currently on the web. Ideally, our studies would determine if appropriate alternative text has been provided for each image, but, unfortunately, automated studies cannot perfectly judge this. Instead, we placed images into two categories that are generally easily distinguishable: images that are used to convey content and images that are used for decorative or structural purposes. We call these images *significant images* and *insignificant images*, respectively.

Significant images provide information to a user and add to the content of a web page. As examples, these images can be part of a website's navigation menu, a map describing the layout of a building, a picture accompanying a news article, or an e-mail address. Significant images should have alternative text that is a textual equivalent to the content or function of the image. Insignificant images are used to add visual appeal or to assist in web page layout and should have a zero-length alternative text because their function is inherently visual. Longer alternative text for insignificant images adds little value to the browsing experience and can detract from the important content of the page when the page is viewed serially with a screen reader[20]. The specific methods used by each study to determine significance are presented later along with the description of each study.

While we do not attempt to measure the specific appropriateness of the alternative text provided for images, our studies improve upon previous work[22, 7, 17] in three important ways.

First, our studies distinguish between significant and insignificant images and do so automatically. While significant images should be assigned descriptive alternative text, insignificant images should be assigned a zero-length `alt` attribute to indicate their insignificance, according to accessibility guidelines[20]. Ahn *et al.* and Craven ignored significance and treated spacer GIFs and images forming a navigation menu equally. Petrie *et al.* manually differentiated images into five categories based on their function, but did so for images found on only 100 homepages. Automatically determining the significance of images allows us to properly measure the presence of insignificant images with alternative text and significant images without alternative text on a large number of webpages.

Second, our studies consider multiple methods of providing alternative text for images. Previous studies counted only the presence of the `alt` attribute of the `img` tag as properly assigned alternative text. To address this, we record the presence of the `title` and `longdesc` attributes, which are often used to convey alternative text and are properly utilized by most screen readers. We also counted linked images that occurred within the same anchor tag with text as being properly labeled even if no alternative text is directly specified. Screen readers can correctly associate alternative text specified in this manner.

Finally, our studies explicitly consider the popularity and importance of websites to determine the effect the accessibility of each is likely to have on the average user's experience. In our first series of studies we explicitly choose websites based on their observed traffic and importance. In the second, we consider all traffic generated by our department and count all observed `img` tags. In effect, this weights each image by its popularity, better approximating the average user's web experience.

For our studies, we used the following labeling criteria. A significant image was properly labeled if its associated `img` tag contains a non-zero length `alt`, `title` or `longdesc` attribute, or if the image is contained in the same anchor (link) tag as additional text. An insignificant images was properly labeled if it contained a zero-length `alt` attribute. In our studies, we first focused on university, government and high-traffic websites. Next we looked at the traffic generated by members of the University of Washington Computer Science Department.

3.1 Important Websites

We first examined five high-traffic and important groups of websites. These include the homepages of *(i)* the 500 most high-traffic international sites[1], *(ii)* the top 100 international universities[2], *(iii)* the 158 Computer Science departments listed in the Taulbee Report[24], *(iv)* the 137 U.S. Federal agencies listed on whitehouse.gov, and *(v)* the 50 U.S. states plus the District of Columbia.

To gather the image data for this portion of the study, we created a web crawler that loads each web page with Internet Explorer. Using the Document Object Model (DOM), we extract all relevant information from each image on the web page, including those images that have been dynamically loaded. This method provides more accurate data than simply parsing the HTML text of a web page as we also have access to images loaded using dynamic scripting.

We then analyzed the image data collected for each group of web pages to determine which images should be classified as significant. Images containing more than one color and having both width and height greater than 10 pixels are classified as significant. Images that are clickable, either because they are contained within anchor tags or because of defined script events, are also considered significant. Our criteria for classifying an image as significant for this portion of the study was determined from manual observation and we have found it to be quite accurate.

Once the significant images are identified, we examine each for alternative text according to the labeling criteria described in the introduction to this section. We then compute the aggregate image statistics for each homepage to determine the percentage of significant and insignificant images that are properly labeled. Table 1 summarizes this information for each group.

The results show that the web pages of the U.S. states and U.S. federal agencies have the highest percentage of labeled significant images. Federal agencies in the United States

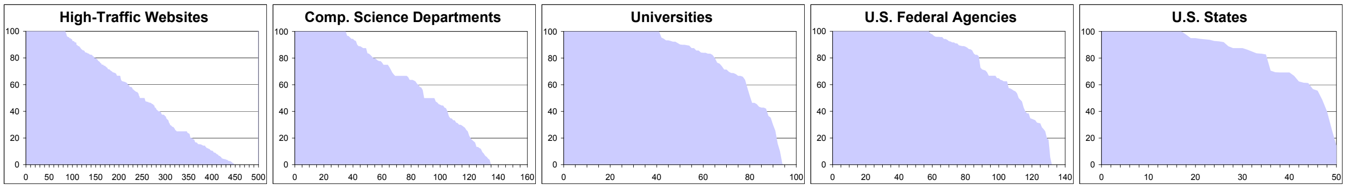

Figure 3: Study results for five groups of websites. Each bar graph shows the percentage of significant images labeled for each web page in the group and is arranged in descending order.

Group	Sig.	Insig.	> 90%	N
High-traffic	39.6	27.4	21.8	32913
Computer Science	52.5	41.6	27.0	4233
Universities	61.5	70.2	51.5	3910
U.S. Federal Agencies	74.8	66.6	55.9	5902
U.S. States	82.5	77.1	52.9	2707

Table 1: The percentage of significant images with alternative text, percentage of insignificant images with zero-length alternative text, percentage of web pages with greater than 90% of significant images labeled for each of five important website groups and the total number of images observed.

are required to make their web content accessible pursuant to Section 508 and many U.S. states have adopted similar policies[20]. The percentage of insignificant images correctly labeled with a zero-length **alt** attribute follows a similar pattern. Only 7.3% of insignificant images observed were assigned non-zero-length alternative text, which matches with our expectations and suggests that a common mistake is to assign insignificant images no alternative text instead of zero-length alternative text. These results are in contrast to the results of Petrie *et al.* who used a different methodology to choose web pages and found that 76.9% of significant images and 12.8% of insignificant images were assigned alternative text[17].

Figure 3 contains a bar graph for each group that presents the percentage of significant images that were labeled on each individual web page. This figure shows, for instance, that 87 of the 500 high-traffic web pages had 100% of significant images labeled, and that all of the state web pages had at least some significant images labeled.

3.2 Department Traffic

We next examined all web traffic generated by members of the University of Washington Department of Computer Science and Engineering during a period of approximately one week. This was made possible using a machine sitting on the network between the computer science department and the Internet. Because of privacy concerns we only kept counts and no identifying information. For this reason, we define a significant image as one that is clickable either because it is contained within an anchor tag or because of a defined script event.

This study should most accurately represent statistics on the accessibility of pages as experienced by the average user because each image is weighted by its observed popularity. The study ran for approximately one week and observed all unencrypted web content entering the department during this period. This captured 11,989,898 images, and, of those, 4,889,948 (40.8%) were significant. Of these images, 63.2% were assigned alternative text.

4. WEBINSIGHT SYSTEM

The WebInSight system addresses the accessibility shortcomings that we observed in the studies we conducted. When a user loads a web page, the system retrieves alternative text from its database for each image when it is present and requests that alternative text be calculated for any image not already in its database. It then dynamically inserts alternative text into the web page, instead of relying upon web authors to have added accessibility information. To calculate alternative text automatically the system utilizes OCR and web context labeling. While many images lend themselves to these automatic labeling, images that cannot be labeled automatically can be sent to human labeling services if the user desires. The system, including the labeling processes, operates with only a small delay, allowing WebInSight to provide alternative text for most images, including those that have not been previously viewed.

4.1 Architecture

The WebInSight system consists of two main components: a transformation proxy that sits between the user and the Internet, and a labeling framework that is queried to supply alternative text for arbitrary images (See Figure 4). A blind user accesses the web through the proxy, which arranges for images to be labeled on his or her behalf. When alternative text is present in the database for an image viewed by a user of the system, the proxy automatically appends the alternative text and the name of the module that generated it to the value of the `alt` attribute of the associated `img` tag. When users reach the image on the page, the alternative text is then available.

If alternative text is not immediately available in the database, then it is calculated by the labeling framework. Because calculating alternative text for all images on a web page may take some time (generally less than a few seconds per image), the proxy inserts Javascript code into web pages that allows them to dynamically query the labeling framework after the main content has loaded. This allows users to begin reading a web page without requiring them to wait for the labeling framework to attempt to assign alternative text to all of the images. Because the modifications made to the web page are not immediately apparent, the system appends an information bar to the page informing the user that the page has been altered by WebInSight. The bar is appended to the end of the page to be easily accessible but unobtrusive to users employing serial interfaces, but is visually stylized to appear at the top of the page to be readily apparent to users employing graphical displays. The

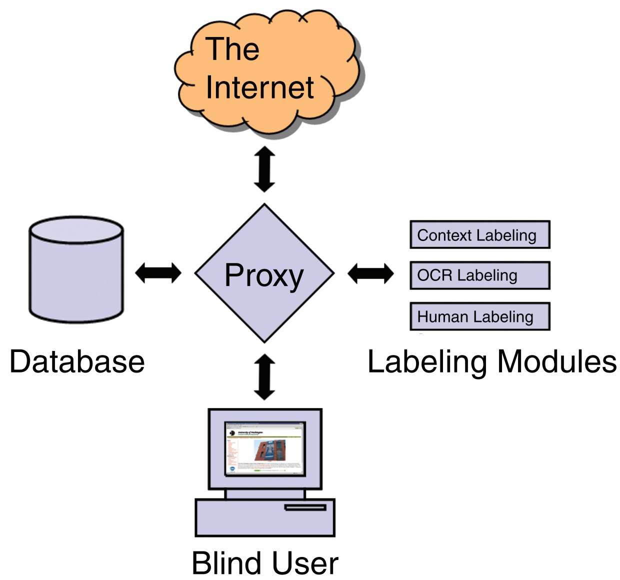

Figure 4: The architecture of WebInSight.

bar is dynamically updated as additional alternative text is requested and added.

While current screen reader products have limited support for dynamic HTML content, such support is expected to be a critical feature in future releases as this technology is becoming prevalent on the web. Current versions of popular screen reading software can be configured to reevaluate the page DOM upon changes. Other groups have recognized the importance of making dynamic web content accessible and have proposed extensions that would facilitate its use[10].

To setup and use the WebInSight system, a user simply configures his or her web browser to connect via the WebInSight proxy (a simple change in both the Internet Explorer and Firefox web browsers). WebInSight requires the user to specify login information to prevent misuse of the system. A proxy-based implementation has a number of advantages that make it easy to both use and administer. Because the proxy is centralized, updates and changes can be managed from one location and will immediately be propagated to all users. In addition, each user can create a customized experience through the web interface provided by the proxy. WebInSight could also be implemented as an Internet Explorer plugin or Firefox extension without substantial changes to how the system transforms web pages, calculates alternative text, or interacts with the database. This decentralized solution does not have the proxy as a bottleneck and has the additional benefit of working with secure sites. The database currently contains a record of which images were viewed, but does not contain a record of who viewed which image. No personal information or user-specific browsing data is stored, preserving privacy expectations.

4.2 Implementation

The transformation proxy is an enhanced version of the open source webserver, Apache HTTP Server. The main additions are two modules which target requests for HTML and image content. The first captures web pages and alters them to include both cached alternative text and the code allowing pages to be dynamically updated as described previously. The second module calculates the MD5 hash of each observed image and records the mapping between this hash and the image's URL in a local database. An MD5 hash serves as a unique identifier for each image, irrespective of its location on the web. This allows us to detect when the image at a given URL changes and allows us to recognize images that appear in more than one web location.

5. IMAGE LABELING

At the core of the WebInSight system are the image-labeling modules that provide a mechanism for labeling arbitrary web images. Content-based image labeling of arbitrary images is beyond the state-of-the-art in computer vision [9]. Consequently, our system takes the approach of combining many modules, each with different strengths, costs and capabilities to tackle the problem. Below we describe those modules that are currently implemented.

5.1 Web Context Labeling

The context in which an image appears has previously been leveraged to reveal its contents[11, 14]. Just as anchor text is often an accurate summary of the linked page[6], the converse is also often valid: summaries of pages linked to by images appearing within anchor tags often accurately describe the images. As an example, web authors often use an image of the logo of an organization to link to its homepage and the name of the organization usually appears in the title of its homepage. In such cases, the text in the title of the linked page accurately describes the contents of the image.

WebInSight retrieves the contents of pages linked to by images within anchor tags and uses this text to formulate alternative text. Currently, it does so by returning text found in `title` or `h1` tags on the linked page. These text strings are often a succinct summary of the page on which they are found. We have discovered empirically that taking the longer of the strings, up to 50 characters, is a good heuristic for choosing the text to use.

More advanced methods of summarizing web pages are available[4, 13], but such systems target the production of longer descriptions than what is desired for alternative text. Many operate by selecting existing sentences or phrases from the document, and our method simply chooses between text found in the `title` and `h1` tags, which we have found to generally be short, accurate summaries. Often, even when the method does not produce alternative text that matches the image, it produces alternative text that matches the function of the image. The user still benefits because of the value in knowing what is behind the unlabeled link.

5.2 OCR Image Labeling

Many of the web's significant images contain some form of text[19]. In an analysis performed by Kanungo *et al.*, it was determined that 42% of images on the web contain text. Petrie *et al.* found that 79.4% of graphical text images on the homepages of 100 major organizations were assigned alternative text[17]. Because graphic text images convey information, even 20% of such images lacking alternative text is a problem for web accessibility. Examples of images that typically contain text include image buttons, navigation bars, informational banners and e-mail addresses. Considering the high accuracy of OCR techniques[16], it would follow that much of this text should be extractable. However, the performance of OCR is closely tied to its application on

Highlight	Resulting Image	Text Retrieved
None	Upgrade Now	qmmpmm.
	Upgrade Now	d i
	Upgrade Now	W .tgwkim
	Upgrade Now	Upgrade Now
	Upgrade Now	-

Table 2: The OCR originally produces incorrect output on this image. Color segmentation yields an image that the OCR can correctly process.

black and white documents scanned at known resolutions. This poses a problem when performing OCR on web images as they are found in a variety of resolutions, colors, image formats and compression levels. We address this below.

To implement this labeling module we used Nuance OmniPage Capture SDK, which provides an extensive API for OCR processing. The system properly formats the input image for the OmniPage OCR engine, which in turn is highly configurable. With the correct image format and OCR engine configuration, we can accurately capture text from many web images. However, different images require different formatting options and OCR engine settings to achieve the most accurate OCR output. For a given image we can automatically conduct a structured search for optimal settings and use a confidence measure produced by the OCR engine to differentiate between good and bad results. Searching for the optimal settings does not produce acceptable output for all images, as shown by the results at the end of this section.

The output of the OCR is verified by a custom spell checker. First, a dictionary is searched to verify the OCR output. If there are no matches, the output is passed to Google's spell-checking API, which can verify and correct many words that do not appear in traditional dictionaries. If this still fails to properly verify the output and if the string is sufficiently long as to make a chance occurrence on the web unlikely, the system issues a web query for the output. If any results appear, we assume that the output is valid. This multi-tiered verification allows the system to correctly verify a variety of strings including misspellings and e-mail addresses.

To enable the OCR to properly extract text from more images, we implemented a color segmentation process similar to that explored by Jain *et al.* that we apply to images as a preprocessing step[12]. For this process, we use a color histogram-based algorithm to identify the major colors in an image. Once all these colors are discovered, we create a set of black and white *highlight images* for each identified color. To create a highlight image we color all shades of the major color black and all other pixels white. The OCR engine can then be fed each highlight image separately and in many cases one of these highlights will create the ideal OCR situation of black text on a white background. The OCR labeling module chooses among the outputs for the highlight images using a score output by the spell checker.

Correct Output Without Segmentation	Correct Output Only With Segmentation	Correct Output Unattained
Web Search	About Us	APPLY ONLINE TO GRADUATE PROGRAM
Administration	Research Center	MOST WATCHED
Key Features	Edit	THREAT LEVEL ELEVATED

Table 3: Examples of input images for which correct alternative text was obtained with the OCR, was obtained with the OCR employing segmentation, and was not obtained.

Color segmentation can improve the accuracy of OCR output versus simply performing the OCR without color segmentation preprocessing (see Table 2). Segmentation seems to be of most benefit on images where the text is shadowed or when the contrast between the text and the background is poor. Images with text in different colors are even more difficult to process correctly.

The performance of the OCR and the usefulness of segmentation were evaluated and the results are below. The test set consisted of 100 multicolored images that contained text collected from the web (mainly images for navigation, buttons and information banners). The output was considered correct if all text in the image was contained in the output string. Overall, the OCR correctly processed 52% of the images when segmentation was not used. When segmentation was enabled, 65% of images were correctly processed. Thus, segmentation helped on 13% of all images or on 27% of the images where normal OCR failed. Some examples of images from the test set are shown in Table 3.

5.3 Human Labeling

The automatic labeling methods mentioned earlier work well for the large subset of significant images, but they do not apply to all web images. To generate labels for these images, the WebInSight system provides a mechanism for users to request that an image be sent to a labeling service for labeling by humans.

The ESP Game[21] and Phetch[22] are two recent computer games designed to effectively encourage humans to label images with keywords and descriptions. Both are modeled such that a service could be built to allow users to request alternative text and receive a reply with an acceptable latency. Because of the benefit to blind users, such a service may also be eligible for charitable or governmental support. WebInSight currently implements a simple web-based application for humans to label images.

Because the cost of labeling is greater for services that involve humans, WebInSight provides a framework that allows users to decide when to incur this additional cost. To facilitate user requests, the WebInSight system inserts a link immediately following each image that users can select to request additional alternative text. The link clearly states its function as a means to request additional labeling for an image and it clearly identifies the associated image. The presence of these request links can be toggled depending on a user's preferences. The current state of the system is reflected in the information bar inserted by WebInSight.

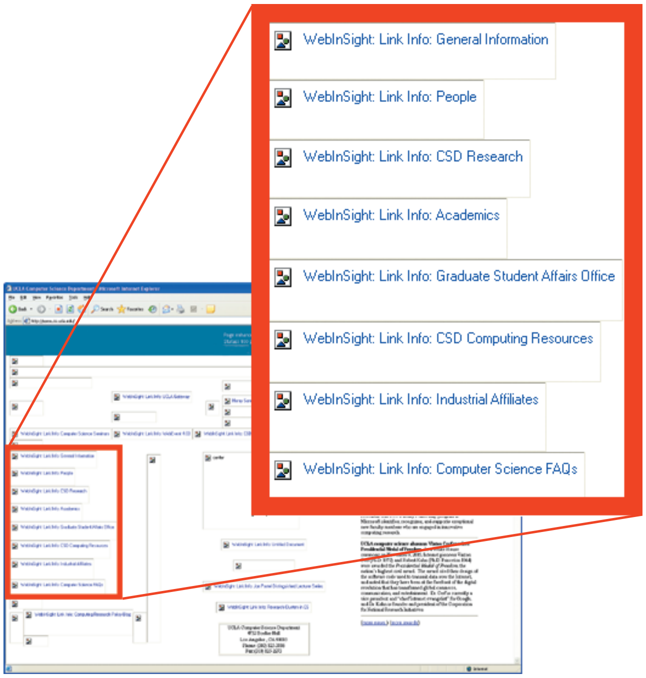

Figure 5: The UCLA Computer Science Department homepage as presented by WebInSight.

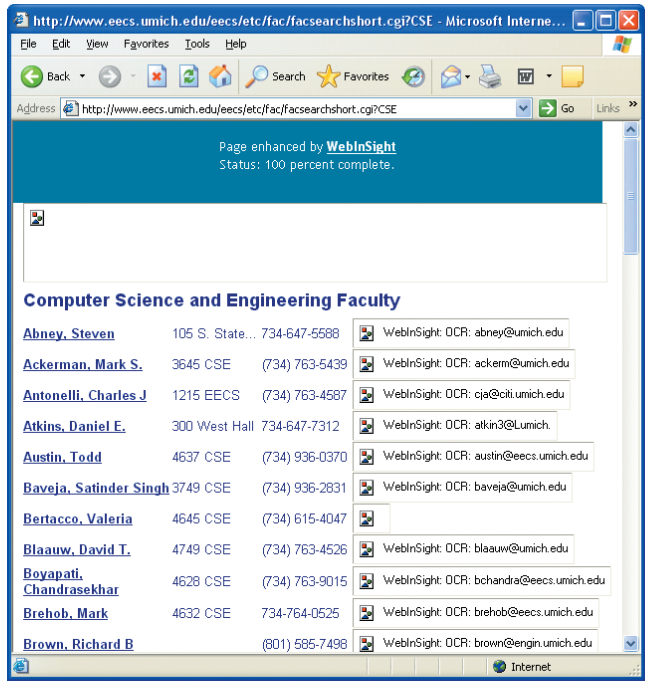

Figure 6: A screenshot of the directory listing for the Computer Science and Engineering faculty at the University of Michigan. WebInSight recovers many of the hidden email addresses using OCR.

6. RESULTS

WebInSight is capable of automatically providing alternative text for a significant portion of the most important web images. To evaluate the performance of our system we reran the series of studies presented earlier over important websites through WebInSight. We then randomly selected 2500 significant images that were not originally provided alternative text, and, of those, WebInSight automatically provided alternative text for 1079 (43.2%). We manually verified the alternative text provided by the system and found that it was correct on 1015 (94.1%) of the images. We found that labels provided for linked images are particularly reliable because they reflect the contents of the `title` and `h1` tags of the linked page. The results are promising because they indicate that almost half of unlabeled, significant images can be labeled automatically. The rest can be labeled by humans using the human labeling module.

Next, we revisited the example web pages offered in our introduction. Figure 5 shows the UCLA Computer Science Department's homepage as viewed using WebInSight. The system provided valid labels for 18 of the 21 images that served as links on this page by using context-based image labeling. The errors occurred because one document had the title "Untitled Document", another pointed to a PDF file which our system cannot currently parse, and the last pointed to a document without a title.

Figure 6 shows the faculty directory on the University of Michigan Computer Science and Engineering web page as viewed through WebInSight. Fifty-two of the 65 email images were correctly recognized by the OCR labeling module and inserted by the system. For the majority of the others, the OCR software gave no output. When the output was incorrect, the spell checker suppressed insertion because the text did not appear on the web. As a result of WebInSight, both of these web pages are more accessible to blind users.

7. FUTURE CHALLENGES

We hope that as we continue to improve and refine our system, it will make the web more accessible to a large number of users, but challenges remain.

Given the recent legal troubles faced by image search engines, copyright issues and responsibility for the labels produced must be considered. Another important consideration is misuse of the system. The human labeling methods are most successful when many labelers are available and private companies are likely in the best position to offer such a service. However, such companies may want to retain control of their labels, potentially limiting access. Users also could potentially exploit the system to build their own image index or to facilitate denial-of-service attacks.

Many of these problems can be mitigated by requiring users to login to the system and/or by limiting usage. A question that remains, however, is who should be allowed to access such a system. Potential ideas include making the system a government service, a subscription-based service, or relying on a trusted third party for verification.

We also hope to improve our system to automatically generate alternative text for even more images and to combine the results from the various labeling modules in a way that is most helpful to users. To this end we plan to first conduct a focus group in order to determine what users want out of WebInSight and use their guidance to shape its future. A vital concern is the impact to users when the system provides incorrect alternative text. We plan to conduct user studies to quantify the value of the alternative text provided by the system and use this information to decide which alternative text is likely to help and when to present it.

An extension of our system currently under development is a tool for web authors that will provide suggestions for alternative text and coordinate the labeling of images across an entire site. We see this as a promising future direction because automatically assigned alternative text will not be as accurate as those provided by well-informed humans. By using our system to provide suggestions, we hope to lower the cost to web authors, enticing them to provide alternative text for their images by making it more convenient and resulting in more accessible web pages for blind users.

8. SUMMARY

The lack of alternative text for many web images is a web accessibility challenge for blind users. We first conducted a series of web studies that demonstrated the current problem. Next, we introduced WebInSight, a system capable of automatically creating and inserting alternative text without negatively impacting the user's browsing experience. We have described three labeling modules that combine novel components to automatically and accurately label arbitrary images on the web. Finally, we presented an evaluation of the system that shows that WebInSight is capable of producing correct alternative text for nearly half of the unlabeled images in a large collection of web pages.

9. ACKNOWLEDGMENTS

This research was funded by National Science Foundation grant IIS-0415273. We thank Luis von Ahn and Scott Rose for their important early guidance and feedback. We also thank Steve Gribble for his invaluable assistance in collecting accessibility statistics from our department's web traffic.

10. REFERENCES

[1] Alexa web search – data services, 2006. `http://pages.alexa.com/prod_serv/top_sites.html`.

[2] ARWU2005-Top 500 World Universities, 2005. `http://ed.sjtu.edu.cn/rank/2005/ARWU2005TOP500list.htm`.

[3] R. Barrett, P. P. Maglio, and D. C. Kellem. How to personalize the web. In *Proceedings of the Conference on Human Factors in Computing Systems CHI'97*, 1997.

[4] A. L. Berger and V. O. Mittal. OCELOT: a system for summarizing web pages. In *Research and Development in Information Retrieval*, pages 144–151, 2000.

[5] L. Breslau, P. Cao, L. Fan, G. Phillips, and S. Shenker. Web caching and zipf-like distributions: Evidence and implications. In *INFOCOM (1)*, pages 126–134, 1999.

[6] S. Brin and L. Page. The anatomy of a large-scale hypertextual Web search engine. *Computer Networks and ISDN Systems*, 30(1–7):107–117, 1998.

[7] T. C. Craven. Some features of alt text associated with images in web pages. *Information Research*, 11, 2006.

[8] Flickr. Yahoo Inc., 2006. `http://www.flickr.com/`.

[9] D. A. Forsyth and J. Ponce. *Computer Vision: A Modern Approach*. Prentice Hall, Englewood Cliffs, NJ, 2002.

[10] B. Gibson and R. Schwerdtfeger. Dhtml accessibility: solving the javascript accessibility problem. In *Assets '05: Proceedings of the 7th international ACM SIGACCESS conference on Computers and accessibility*, pages 202–203, New York, NY, USA, 2005. ACM Press.

[11] J. Hu and A. Bagga. Identifying story and preview images in news web pages. In *Proceedings of the Internation Conference on Document Analysis and Recognition (ICDAR)*, 2003.

[12] A. K. Jain and B. Yu. Automatic text location in images and video frames. In *Proceedings of the 14th International Conference on Pattern Recognition (ICPR)*, volume 2, pages 2055–2076, Washington, DC, USA, 1998. IEEE Computer Society.

[13] J. Kupiec, J. O. Pedersen, and F. Chen. A trainable document summarizer. In *Research and Development in Information Retrieval*, pages 68–73, 1995.

[14] Y. Lu, C. Hu, X. Zhu, H. Zhang, and Q. Yang. A unified framework for semantics and feature based relevance feedback in image retrieval systems. In *Proceedings of the eighth ACM international conference on Multimedia*, pages 31–37, 2000.

[15] O. Marques and N. Barman. Semi-automatic semantic annotation of images using machine learning techniques. In *International Semantic Web Conference (ISWC)*, 2003.

[16] G. Nagy, T. A. Nartker, and S. V. Rice. Optical character recognition: An illustrated guide to the frontier. In *Proceedings of SPIE Vol. 3967, Document Recognition and Retrieval VII*, pages 58–69, January 2000.

[17] 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.

[18] S. Sclaroff, M. L. Cascia, and S. Sethi. Unifying textual and visual cues for content-based image retrieval on the world wide web. *Comput. Vis. Image Underst.*, 75(1-2):86–98, 1999.

[19] C. H. L. T. Kanungo and R. Bradford. What fraction of images on the web contain text? In *Proceedings of the International Workshop on Web Document Analysis*, September 2001.

[20] J. Thatcher, P. Bohman, M. Burks, S. L. Henry, B. Regan, S. Swierenga, M. D. Urban, and C. D. Waddell. *Constructing Accessible Web Sites*. glasshaus Ltd., Birmingham, UK, 2002.

[21] L. von Ahn and L. Dabbish. Labeling images with a computer game. In *Proceedings of Computer Human Interaction (CHI) 2004*, April 2004.

[22] L. von Ahn, S. Ginosar, M. Kedia, R. Liu, and M. Blum. Improving accessibility of the web with a computer game. In *CHI '06: Proceedings of the SIGCHI conference on Human Factors in computing systems*, pages 79–82, New York, NY, USA, 2006. ACM Press.

[23] W3C: Web Accessibility Initiative, 2006. `http://www.w3c.org/wai/`.

[24] S. Zweben. 2004-2005 taulbee survey. *Computing Research News*, 18(3):7–17, May 2006.