CAPTCHA Accessibility Study of Online Forums

ABSTRACT

The rise of online forums has benefited disabled users, who take advantage of better communications and more inclusion into society. However, even with accessibility laws that are supposed to provide disabled people the same equal access as non-disabled users, sites have erected technical barriers, such as CAPTCHAs, that prevent users from taking full advantage of site capability. This study analyzes 150 online forums to determine if sites use CAPTCHAs, and what types are used. Each variety presents accessibility problems to disabled users and the results of the research show that most sites use text-based CAPTCHAs, but rarely provide alternatives that would help users with visual disabilities. The research presents alternatives that site designers may wish to consider in order to allow more disabled users to access their sites.

KEYWORDS: CAPTCHA, Accessibility, Web design

1. Introduction

The growth of Web-based social sites and online forums has led to a greater inclusion of disabled persons into the online community, as they have access to more communications and information than would occur in traditional networking. While the growth of these services has been beneficial to the disabled community, it has not come without problems, such as accessibility issues with CAPTCHAs.

The idiom CAPTCHA stands for ‘Completely Automated Public Turing test to tell Computers and Humans Apart.” These are computer-generated checks that attempt to tell humans and computers apart by presenting specific tests that humans should be able to pass, but computers could not. CAPTCHAs came into being because problems Web owners encountered with automated spam-bots gaining access to Web services that they should not have. Although they can help site owners prevent spam, they pose problems for the visually impaired Web community. According to Schlaikjer (2007), screen readers do not support text-based CAPTCHAs, and deaf users will have issues with audio types. Therefore, it is important that site operators weigh the relative security benefits compared to accessibility problems. Some online communities, such as banking, may warrant more stringent safeguards while other public forums may decide that greater accessibility is more important than including CAPTCHAs.

This research analyzes 150 online forums to determine if they use CAPTCHAs, and to what extent various types are utilized. The study attempts to answer the following research questions:

1. What percentage of online community forums use CAPTCHAs?
2. Which types of CAPTCHA systems are most prevalent?
3. Based on the results, are there categories of disabled users that would have trouble accessing these sites?

The study starts with a literature review of the disabled online community, followed by legal aspects and a review of CAPTCHAs. Next, the research functionality is covered followed by an explanation of survey results. Finally, implications of this study are highlighted along with suggestions for site owners on how to effectively manage CAPTCHAs.

2. Framework for Web Accessibility
2.1 Growth of Web Disability Market

The United Nations (UN) estimated in 2006 that approximately 10 percent (650 million) of the world’s population has some form of disability (UN 2006). The estimated number of disabled Web users is difficult to quantify, but a study by the U.S. National Institute on Disability and Rehabilitation estimates 10 percent of Web users are disabled (Salamone, 2001) while more recent study in 2004 by the Danish Center for Accessibility puts the estimate at 25 percent (Rønn-Jensen, 2004). Irrespective of the specific study, the number of disabled people using Web sites is a sizable number.

The number of disabled users using social sites and online forums has experienced a tremendous growth as there are many benefits. Disabled people find that social networks and forums are a better way to communicate with other disabled viewers as well as the greater non-disabled community in a more equal setting. Disability Now Magazine indicates that “social-networking site can be a boon for disabled people, because they remove some of the barriers faced by disabled people who want to socialize and network…and for people with learning difficulties who may have difficulty with face-to-face communication” (Whittle, 2007). These sites and online forums can not only provide the ability to better communicate and form online communities, but can help disabled users better access to information from others. However, although this online community has grown and many make use of social sites, disabled users do find accessibility a problem.

A 2007 study was done by the U.K. national computing and disability charity, AbilityNet, using 100 disabled end user testers. They provided feedback on accessibility issues on five of the most popular social sites including Facebook, MySpace, YouTube, Yahoo and Bebo. Part of the study included testing of CAPTCHAs, and they found that most of these sites did not allow a login or signup without using a CAPTCHA. Yahoo and Facebook offered audio CAPTCHAs, but were found to be of such poor quality that they were considered unusable (AbilityNet, 2008).

2.2 Legal Aspects

Web owners need to review the disability and accessibility laws for their specific country regarding how to properly design their systems. There are 19 countries or federations that have enacted their own laws on Web accessibility, including the European Union (EU), France, Germany, India, Switzerland and the U.K. (W3C, 2006). Of these countries, the EU and U.K. have enacted legislation that requires government Web sites to comply certain industry accessibility checkpoints (Shi, 2006), which would include design of CAPTCHAs.

There have been lawsuits against firms regarding accessibility of Web sites, and several recent instances regarding litigation against the use of CAPTCHAs. In 2004, Ramada.com and Priceline.com agreed to implement a variety of accessibility standards to permit users of assistive technology to more easily navigate their sites (WebAIM, 2009). In 2009, U.S. retailer Staples agreed to update its Website to conform with Web Content Accessibility Guidelines (WCAG) promulgated by the Web Accessibility Initiative (WAI) of the World Wide Web Consortium.

Although litigation against inaccessible use of CAPTCHAs has occurred against some companies, other firms have bypassed legal problems by actively responding to pressures from the disabled community. In 2009, Twitter responded to pressure from the blind user online community about problems with their audio CAPTCHAs locking blind users out of the account creation process. After complaints from the disabled community, Twitter made technical upgrades to their system (Blind Access Journal, 2009). This does indicate that not all issues need to progress to the litigation level, as disabled people and firms can effectively work together to solve accessibility issues.
2.4 CAPTCHA types

There are several different types of CAPTCHA systems that the disabled community can use. The most common is the text-based (character recognition) CAPTCHA where the viewer needs to read a series of letters or numbers randomly generated by the system. The viewer then types those text digits into a text box. The results the user types into the text box must match the predefined generated text digits, or the person will be prevented from advancing further. An example of a text-based CAPTCHA system is shown in Figure 1, where the system generates letters on the left side (ekHggg), and the user would have to enter them in the blank text box.

![Text-based CAPTCHA](image)

Figure 1: Text-based CAPTCHA

The second most common type of CAPTCHA is a sound based system. According to Sauer et al (2008), there are two formats for sound-based CAPTCHAs. The first would be the ‘spoken words or numbers’ and the second would be sounds related to an image.

![Combination of text and audio CAPTCHA](image)

Figure 2: Combination of text and audio CAPTCHA

A third type of system is the implicit CAPTCHA. This method and the audio type are options used as alternatives for people with disabilities or to provide another type for users having difficulty reading text. With implicit CAPTCHAs, users only have to make one click instead of typing in a series of text or listening to an audio. In this method, for example, the user is show an image, such as a picture of a mountain, and the user is asked to click on the pinnacle of the mountain. Or a number of words are shown in a pictures and the user is asked to click on a specific word (Shirali-Shahreza, et al., 2007).

2.5 CAPTCHA and Accessibility
Although the use of CAPTCHAs may be beneficial towards lowering online security risks, they pose a variety of challenges to users with both cognitive and physical disabilities. Hochheiser (2008) explains that individuals with disabilities such as Downs Syndrome may have difficulty with various security mechanisms including passwords, CAPTCHAs and security questions. Thus, people who have mental challenges may not be able to use CAPTCHA systems without assistance from others.

The most common type of CAPTCHA is character-based where the system generates images. However, these types of CAPTCHAs are not accessible to screen readers, putting visually-impaired readers at a disadvantage. In order to overcome this, site developers attempt to provide a viable alternative – the audio CAPTCHA.

According to Schlaikjer (2007) the use of audio CAPTCHAs poses several problems. First, the failure rate of humans on audio systems is much higher than text-based systems. A case can be made for the amount of background noise when trying to listen to the challenge. Schlaikjer also explains that another problem with the audio is that the sequence of digits are randomly generated, thus causing people to have little contextual information to complete the task. This is different from text-based CAPTCHAs where humans can take advantage of the context to help classify any particularly obscured character within a word. Also, issues with language may pose challenges, such as non-native speakers trying to decipher specific accents. It may also be problematic for those without sound cards or don’t have required sound plugins (May 20005).

A test done by CNET News reporters on Microsoft’s Hotmail audio CAPTCHA found that the audio file was unintelligible by all testers (Festa, 2003). A more comprehensive test completed in 2008 by a group of researchers on audio CAPTCHAs found that for disabled users, only 46 percent of them could successfully pass an audio CAPTCHA test, far less than the 90 percent success rate that humans are supposed to have for solving CAPTCHAs (Sauer, et al., 2008).

According to May (2005), people with hearing impairments may have trouble with this method. Users could also find audio CAPTCHAs inconvenient because of the auditory nature of the method where users often have to write down the code after they hear the prompt, and then transfer the code into the computer, a two-step process. May also explains that some audio implementations are JavaScript-based, or designed in a way that some blind users may not be able to access them. Another problem is when disabled users attempt to use public computers, which may not be equipped with soundcards.

According to Yan & El Ahmad (2008) there is no standard graphical symbol for representing an audio CAPTCHA. Often, sites use a speaker symbol (as shown in Figure 2), although this is not universal. The authors explain that Google uses a disability symbol while Microsoft’s site uses a speaker symbol. Another issue that people with disabilities face with CAPTCHAs is their use on mobile devices. Because of the smaller size of mobile devices, attempting to read and then type in the password challenge is more difficult than using a normal keyboard. Lin (2009) states that good accessibility design of CAPTCHAs on mobile devices is a concern and must be effectively reviewed if the CAPTCHAs are to be effective on small devices. Accessibility issues on mobile devices are expected to become even more prevalent as the number of people using these devices rises. Already, 40 percent of Japanese adults use mobile devices to access the Internet (Ipsos News Center, 2006), which will require CAPTCHA designers to come up with better ways for disabled consumers to use the systems.

3. Methodology
The research in this paper was accomplished through analyzing 150 online forums to determine if the sites used CAPTCHAs and determining if they incorporated accessibility into the CAPTCHA functionality. The project consisted of three phases:

- Choosing a list of online forum sites to test
- Choosing what information to collect
- Collecting and analyzing the data

3.1 Choosing Online Forums

The first portion of this study was to choose five categories of forums to test. It was decided to concentrate on service and financial types of forums, as these would be types that disabled as well as non-disabled would both use. The following industry forums were picked:

a) Webmaster, b) money making, c) stocks d) gambling and e) government. With the number of Web sites estimated to be 186.7 million at the end of 2008 (Gruener, 2009), the number of Web designers and developers needing help to create sites has increased. This has lead to an increase in online forums dedicated to 'Webmaster' issues. Web designers are in the forefront of designing accessible Web sites, so it would follow that forums dedicated to Web design issues should follow W3C accessibility standards. In order to choose a list of the top 30 forums, a Google search of the keywords 'Webmaster forums' and 'top Webmaster forums' was run. A large list was compiled, and the researcher had to view over 100 sites in order to come up with 30 forum-based sites. It should be noted that the search did result in some sites that were not forums, so it was essential that sites were viewed before being compiled into an Excel spreadsheet list.

The next two types of forums chosen were financial-based sites. Forums dedicated to money making advice were chosen with a Google search of ‘money making forums’ and ‘finance advice forums.’ These forums deal with generic money making advice such as affiliate programs, buying and selling products and services, business legal advice, and personal finance. Next, stock advice forums were chosen from a Google search of ‘stock market forum.’ The fifth forum group chosen was for government sites. To populate the list of 30 sites, two Google searches were completed, one for the term ‘government forums’ and another for ‘government message boards.’

The top 30 sites for the Google searches were compiled. This would include forums related to that topic, and could include sites from a variety of countries. For the government sites, these could include a variety of government types, such as federal, local and state sites for a variety of countries.

3.2 Choosing What Information to Collect

The second phase of this project was to determine what type of information to collect for this study. First, each site was to be reviewed to determine if a CAPTCHA was found. The next step was to review what type of CAPTCHAs were in the site. If a CAPTCHA was found, the default type was a character-based CAPTCHA where the viewer had to type in either text or a number into the blank text field. Some CAPTCHA designs do allow for audio-based functions, so the number of audio types would also be compiled.

If there was a sentence in the page indicating that disabled users could contact the firm if they were unable to use the CAPTCHA, that option was also recorded in the results. The ability to press a ‘refresh’ button to reset the CAPTCHA text or numbers was compiled. Finally, the number of implicit CAPTCHAs was compiled. For example, there could a ‘no-spam’ verification question such as ‘what is the result of 1+4,’ A human could type in the correct answer of ‘5,’ but a spam bot would have trouble with this type of verification question.
3.3 Collecting and Analyzing the Data

For this study, each site was reviewed to determine if a CAPTCHA was present. In most cases, if a CAPTCHA was found, it was usually found on the registration page. However, if a CAPTCHA was not found, other pages of the site had to be reviewed as well. However, in almost all cases where a CAPTCHA was found, the CAPTCHA was found on the forum registration page.

If a CAPTCHA was found, then the researcher determined if both a character-based and an audio CAPTCHA were found. All CAPTCHAs that had audio functionality used the same design shown in Figure 2, which was the reCAPTCHA program service (reCAPTCHA, 2009). Contact information, refresh capability and the number of implicit CAPTCHA uses were collected. It should be noted that several issues were found in the collection process that were not expected in the original methodology, and should be noted.

First, three sites had registration GIF images that did not appear properly. These image links were broken, thus causing a problem for continuing the registration process since the users would not be able to see the text within the image. Second, one site had a problem with the refresh image. When the GIF image with clicked to refresh the image, the GIF link was broken and the image did not appear.

The third abnormal situation that occurred during data collection was that one site had a two-part CAPTCHA process. This site first had an image-based CAPTCHA where the viewer was offered a choice of four images and asked to pick the object that best described a ‘flower.’ The next CAPTCHA was further along in the registration process and involved using a character-based CAPTCHA. This was the only site that did use the type of CAPTCHA requiring the viewer to choose an image.

4. Analysis of Results

Table 1 shows testing results of 150 online forums for CAPTCHA usage. The first two columns indicate the specific function that was compiled as well as the number found and the percentage result. The next five columns contained the number and percentage results for each of the five Web forum industries used in this study. Finally, the last column of Table 1 shows the total numeric results for all 150 forums. The five rows contain the CAPTCHA functional information compiled including: a) whether a text CAPTCHA was used, b) whether an audio CAPTCHA was used in the site, c) whether contact information was provided for disabled users, d) whether the CAPTCHA image could be refreshed for that viewing session and e) whether an implicit CAPTCHA was included in the site.

Results show that most of the online forum sites (61 percent) did use a CAPTCHA function, with the top range of 80 percent for gambling sites to a low of 20 percent for government site, and any site that did have a CAPTCHA did use text/character based CAPTCHAs. The number of sites that offered disabled users the audio CAPTCHA option was lower, with an average of 14 percent of the site offering this service. Money-making sites had the most at 23 percent, while no government sites had audio CAPTCHAs. An average of 19 percent of forums had verbage that indicated that disabled users should contact the firm if they had trouble using the CAPTCHA or needed assistance with registering. One-third of gambling sites had this information while none of the government forums did.

Almost half (43 percent) of sites did have the ability for viewers to refresh a CAPTCHA. This is important if the user could not easily read the initial image and needs to refresh the image, thus getting a new text option. Most of the gambling sites (53 percent) had this option while no government sites did. Finally, few forums offered the option of secondary anti-spam
functionality, such as systems that asked viewers to answer a question such as “2 + 2 = ?” Twenty six percent of Webmaster sites had this, while no government sites did.

Table 1. CAPTCHA Testing Results

<table>
<thead>
<tr>
<th></th>
<th>Web</th>
<th>Money</th>
<th>Stock</th>
<th>Gamble</th>
<th>Govt</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Text CAPTCHA</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number</td>
<td>25</td>
<td>18</td>
<td>17</td>
<td>25</td>
<td>6</td>
<td>91</td>
</tr>
<tr>
<td>Percent</td>
<td>.83</td>
<td>.60</td>
<td>.56</td>
<td>.83</td>
<td>.20</td>
<td>.61</td>
</tr>
<tr>
<td><strong>Audio</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number</td>
<td>6</td>
<td>7</td>
<td>4</td>
<td>4</td>
<td>0</td>
<td>21</td>
</tr>
<tr>
<td>Percent</td>
<td>.20</td>
<td>.23</td>
<td>.13</td>
<td>.13</td>
<td>0</td>
<td>.14</td>
</tr>
<tr>
<td><strong>Contact</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>10</td>
<td>0</td>
<td>28</td>
</tr>
<tr>
<td>Percent</td>
<td>.20</td>
<td>.20</td>
<td>.13</td>
<td>.33</td>
<td>0</td>
<td>.19</td>
</tr>
<tr>
<td><strong>Refresh</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number</td>
<td>21</td>
<td>15</td>
<td>12</td>
<td>16</td>
<td>0</td>
<td>64</td>
</tr>
<tr>
<td>Percent</td>
<td>.84</td>
<td>.5</td>
<td>.4</td>
<td>.53</td>
<td>0</td>
<td>.43</td>
</tr>
<tr>
<td><strong>Implicit</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number</td>
<td>8</td>
<td>4</td>
<td>4</td>
<td>1</td>
<td>0</td>
<td>17</td>
</tr>
<tr>
<td>Percent</td>
<td>.26</td>
<td>.13</td>
<td>.13</td>
<td>.03</td>
<td>0</td>
<td>.11</td>
</tr>
</tbody>
</table>

5. Implications

Evaluation results show that the majority of online forums sites (61 percent) do use CAPTCHA systems when requiring potential users to sign up to be able to write forum messages. All sites used a text-based testing system, which could prove detrimental to users with some disabilities. Some sites do provide an alternative for users with visual impairments by giving them the option to use audio CAPTCHAs (14 percent) or implicit CAPTCHAs (11 percent). However, overall use of alternate forms of CAPTCHAs to meet the needs of disabled users is quite low. Thus, people with disabilities are put at a disadvantage when attempting to use many online forum sites.

In order to provide better accessibility for CAPTCHA systems, some researchers have come up with alternate technical solutions that designers should consider deploying. Some of these methods are in the early stage of development, and may not yet be a viable method, but others could be successfully implemented.

Chow (et al, 2008) proposed and tested a clickable CAPTCHA which uses a 3-by-4 grid or matrix. The users would be asked to identify within the grid the challenge. For example, Figure 3 shows a clickable CAPTCHA where the user’s task is to identify the three valid English words (in this case monster, grass and nation) by using a touchpad, stylus or phone keypad to choose the three correct cells. This specific system was designed for mobile systems using a touch-based system, but the methodology could be adapted to Web-based systems using a computer screen and mouse. The researchers in this study found that clickable CAPTCHA were easier for humans to solve compared to textual CAPTCHAs. The
study also found it took 30 percent less time to solve clickable CAPTCHAs than regular text-based English language CAPTCHAs.

Figure 3: Example of clickable CAPTCHA (Chow, et al, 2008)

Another new CAPTCHA technology was recently tested on 184 participants by researchers at Rochester Institute of Technology in 2008. This is video-based technology where CAPTCHAs are generated from YouTube videos. Users must submit three words (tags) that match tags taken from the video taken from the YouTube database. The challenge is passed if any of the three submitted tags match the author-supplied video tags. The researchers found that video CAPTCHAs have comparable usability and security to existing CAPTCHAs and 60 percent of respondents found video CAPTCHAs more enjoyable than traditional methods (Kluever & Zanibbi, 2009).

Almost 20 percent of sites in this study did provide contact information in case disabled users have trouble with using CAPTCHAs. In most cases, the contact will be a direct link to an email address. This method does increase the amount of time and frustration of setting up an account as a series of emails would have to be exchanged. However, Google does take this one step further by providing a more robust option for disabled viewers who have issues with the audio CAPTCHA. Google will automatically create a user account after the user clicks on a hidden registration link and fills in account information (Google, 2009).

Markkola & Lindqvist (2008) proposed a voice CAPTCHA for Internet Telephony. Once a user encounters a CAPTCHA, they would make a free call through Skype and then be redirected to a CAPTCHA service where they would be walked through the verification process. This system is a work-in-process and has not been implemented, but does present another possible option.

It should be remembered that although this research suggest methods to enhance the online experience for disabled users, especially those with sight problems, effective design of CAPTCHAs could also benefit other users. Some CAPTCHA systems, such as the Gimpy text displays, are difficult for many people to read. Audio CAPTCHAs also pose problems for non-impaired people because of noise distortion and language problems. Thus, it behooves firms to consider other forms of CAPTCHA-based systems that will serve both as a method of securing the system but also enable humans to easily pass the CAPTCHA test without undue effort. It is suggested in this study that if Web owners do decide to use CAPTCHAs in order to preserve security, they should implement several different CAPTCHA options so
that disabled users can choose and be able to use any of them. Most sites implement one type of text-based systems, and some include audio as well. However, this study has shown that both of these have issues with some disabled users as well as presenting some issues with non-disabled viewers. Therefore, it may be advisable to consider offering three options to help all Web surfers.

The entire issue of Web-site protection against spam and hacking attacks has led security developers to use CAPTCHAs as one approach to securing the systems. However, studies have shown that CAPTCHA systems can easily be broken by a variety of attacks. Therefore, it should be considered whether the use of CAPTCHAs is even a viable method of protection considering the ease of penetration and the problems with accessibility. It may be advisable to review other methods of security that do not require CAPTCHAs. For example, it may be advisable to consider account locking mechanisms for password logins. If the number of login attempts is greater than three attempts, then the system would automatically prevent another login attempt. This would bypass the need for CAPTCHAs and prevent spam bots from accessing blog and forum areas.

6. Conclusion

The results of this study show a dearth of online forums that provide for full levels of accessibility for their disabled viewers who wish to sign up for the services. Although it is admirable that site owners are worried about security of their systems and implement CAPTCHA technology to mitigate security vulnerabilities, these same technologies contribute to limiting accessibility of a sizable number of their users. A majority of sites in this study have only implemented character-based CAPTCHAs, which cannot be used by visually impaired people. A minority of sites do give surfers the ability to choose character or audio CAPTCHAs, but even audio CAPTCHAs are not always easy for disabled people to use. It is suggested that site owners first determine if the security of CAPTCHAs is really secure enough to justify their use, or if other technologies can instead be implemented. If the decision is made to utilize CAPTCHAs, then it may be beneficial to implement a system that will give users a choice of a variety of character, audio and logic-based tests that they can decide to use based upon their abilities. Firms need to realize that it is legally and ethically important to provide fully accessibility to their systems. Finally, with the increasing number of disabled people using these sites, firms can benefit economically by catering to their disabled constituents.

8. References


