CAPTCHA: The story behind those squiggly computer letters

June 15 2012, By Stacey Burling

If you use the Web, you have probably encountered an annoying invention called a CAPTCHA.

They're the squished-up, stretched and squiggled, color-blotched collections of letters that often must be deciphered before sending an email, posting a comment, or buying a ticket.

Is that an "i" or an "l"? you wonder. A zero or the letter "O"? Maybe you see three letters where it seems there should only be two. You tilt your head. You scoot your chair back and squint. You wonder if you need new glasses.

You might also wonder if these things are getting harder - maybe too hard for people with aging eyes and brains.

The CAPTCHA was created at Carnegie Mellon University in 2000. The name is short for Completely Automated Public Turing test to tell Computers and Humans Apart. Websites need CAPTCHAs to guard against the "bots" of spammers and other computer underworld types.

"Anybody can write a program to sign up for millions of accounts, and the idea was to prevent that," said Luis von Ahn, a Carnegie Mellon professor who was part of the CAPTCHA team. The little puzzles work because computers are not as good as humans at reading distorted text. Google says that people are solving 200 million CAPTCHAs a day.

Over time, though, the bad guys' computers have been getting smarter and, well, people have not. The CAPTCHAs have to get harder for users, because they're easier for the computers.

"It's an arms race between site owners and spammers; users lose," said Jeremy Elson, a researcher at Microsoft Research who has developed a CAPTCHA called Asirra. It uses pictures of dogs and cats.

Von Ahn said there were now "probably hundreds" of different kinds of CAPTCHAs. He worked on one of the biggies, reCAPTCHA. Google bought that one and now offers it for free. Users have to decipher two words for reCAPTCHA. One of them, usually the easier one, is lifted from an old book. A computerized scanner has failed to read it properly, and reCAPTCHA users get a chance to do the job right, thereby helping Google digitize books.

Von Ahn said he thinks some kinds of CAPTCHA have been getting harder. ReCAPTCHA is harder than it was in 2000, but it has been at about the same difficulty level for the past two years. On average, he said, people spend nine seconds solving a reCAPTCHA, and 92 percent of them get it right. In 2000, the success rate was 97 percent. The letters will be made more distorted when too many spammers start getting in.

Von Ahn said he did not know how many people give up when they see a hard CAPTCHA or ask for new words. He also did not know whether older people had more trouble than young, but there's reason to wonder.

Robert Sergott, a neuro-ophthalmologist at Wills Eye Hospital in Philadelphia, said seniors were more likely to have cataracts, glaucoma and macular degeneration - eye diseases that can make vision blurry, especially when there is low contrast between letters and their background. Older people read best when there's high contrast and more space between letters, pretty much the opposite of what some

CAPTCHAs offer.

"A lot of younger people have visual problems too," Sergott said. "I've had errors doing it. I think everybody has. How are you going to balance security without making this an impossible task for certain individuals?"

Rachel Greenstadt, a computer-science professor at Drexel University who specializes in the intersection between artificial intelligence and security, said there were audio alternatives to the written CAPTCHAs. ReCAPTCHA's uses spoken words and a lot of background noise. They're "even harder to solve, and they're easier to break," she said.

In 2009, Harry Hochheiser, an assistant professor of biomedical information at the University of Pittsburgh, did a small study of audio reCAPTCHAs. It involved five blind people, including one with some residual vision. They got the audio CAPTCHAs right 45 percent of the time, and it took them 65 seconds to complete the task.

He says he's not sure what the solution is, but he wonders whether some websites need so much security. "It's quite possible that there are people out there who are getting discouraged by the difficulty," he said.

He pointed out that some politicians require people to solve CAPTCHAs before sending them email. What about The Philadelphia Inquirer? he asked. The paper lets readers send the editor an email without solving a CAPTCHA, but they are used for some tasks on The Inquirer's website, philly.com.

L. Jean Camp, who teaches informatics at Indiana University in Bloomington, focuses on how difficult computer security is for most people to understand.

"Security technologies tend to be designed by people who are young,

male and extremely experienced with computers," she said.

Companies are not taking older computer users seriously, she said. "I know of no technology company, none, that has employed a gerontologist. None. Which to me is amazing," she said.

The solution to the CAPTCHA problem is for companies to invest more in detecting spam, she said. "It's just cheaper and easier to say to the human, 'No, you solve this.' " She said some spammers now employ people in foreign countries to solve the CAPTCHAs.

Drexel's Greenstadt sees a silver lining in the growing difficulty of CAPTCHAs. It's a "triumph for artificial intelligence and optical character recognition," she said.

Creating a better CAPTCHA is tough. "The computer has to be able to generate the problem and check if it's right, but not solve it, and the human has to be able to solve it," she said.

Von Ahn says things are far from the crisis point. Most people can solve the CAPTCHAs, even if they have grown up with a different alphabet.

He lets us in on a little secret: Users don't have to be perfect. The computers know that some letters look the same, and they give users the benefit of a doubt. Even dyslexics do OK.

"We allow you to be a little bit wrong, and spammers know this too," von Ahn said.

He says some of us are overthinking, then typing while nervous. That only ups the odds of mistakes. "I'll tell you the trick," he said. "Type what you see. Whatever. Don't think about it too much."

Von Ahn's current project is duoLingo, a way to crowdsource document translation and learn a new language at the same time. He's out of the CAPTCHA business now, but he says humans can probably beat the machines for another 10 years. "I'm certain it will happen at some point that computers are as good at this as humans," he said. "At that point, we'll have to figure something else out."

(c)2012 The Philadelphia Inquirer
Distributed by MCT Information Services

Citation: CAPTCHA: The story behind those squiggly computer letters (2012, June 15) retrieved 24 April 2024 from https://phys.org/news/2012-06-captcha-story-squiggly-letters.html