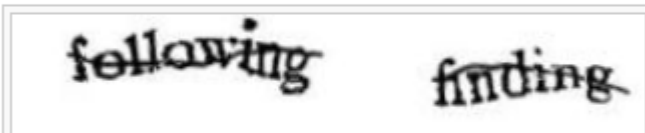


CAPTCHA

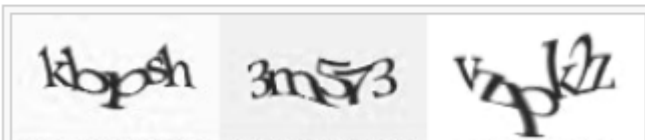
Every web surfer, in the course of his or her browsing, has been forced to stop and perform this weird little task: look at a picture of some wavy, ghostly, distorted letters and type them into a box. Sometimes you flub it and have to retype the letters, but otherwise you don't think about it much. That string of letters has a name; it's called a CAPTCHA. And it's a test. By correctly transcribing it, you have proved to the computer that you are a human being.



Early CAPTCHAs such as these, generated by the [EZ-Gimpy](#) program, were used on [Yahoo!](#). However technology was developed to read this type of CAPTCHA.^[1]



A modern CAPTCHA, rather than attempting to create a distorted background and high levels of warping on the text, might focus on making [segmentation](#) difficult by adding an angled line.



Another way to make segmentation difficult is to crowd symbols together, as in Yahoo's current CAPTCHA format.

This electronic hoop you have to jump through was invented in 2000 by a team of programmers at Carnegie Mellon University. Somebody at Yahoo! had gone to them, complaining that criminals were taking advantage of Yahoo! Mail--they were using software to automatically create thousands of e-mail accounts very quickly, then using those accounts to send out spam. The Carnegie Mellon team came back with the CAPTCHA. (**It stands for "completely automated public Turing test to tell computers and humans apart"**; no, the acronym doesn't really fit.) The point of the CAPTCHA is that reading those swirly letters is something that computers aren't very good at. If you can read them, you're probably not a piece of software run by a spammer. Congratulations--you can have an e-mail account.

The CAPTCHA caught on, and now it's all over the Web. Luis von Ahn, an assistant professor at Carnegie Mellon who was part of the original CAPTCHA team, estimates that people fill out close to 200 million CAPTCHAS a day. But you should pause when you see one--it's one of the rare moments when the invisible war being waged between spammers and programmers becomes visible to you, the prey. "Of course," says Von Ahn, "this has been a little bit of an arms race with spammers, because now there's a huge incentive for spammers to try to get around CAPTCHAS." You can bypass them, using brute force, for example, though it'll cost you. Go to a website like GetAFreelancer.com and you'll see dozens of ads placed by spammers and other bad actors, who hire whole teams of people to read and type out CAPTCHAS, all day, by hand, by the thousands. ("How the hell can they still maintain a profit margin?" Von Ahn wonders. "This is amazing to me!")

You can also get around CAPTCHAS by being clever. They work only because there are things computers can't do, and there are fewer and fewer of those things all the time. Headlines on tech blogs regularly announce the cracking of CAPTCHAS--Gmail's, Hotmail's, Yahoo!'s. Von Ahn doubts the headlines are true--and companies aren't eager to confirm this kind of rumor--but it's possible for an amateur, poorly conceived CAPTCHA to be hacked. (He gives an example: a CAPTCHA in which each letter was always formed out of the same number of pixels. All the malware had to do was count the pixels in a letter to figure out which letter it was looking at.)

The faster that software evolves, and the harder it gets to distinguish between people and computers, the faster CAPTCHAS have to change. They might soon involve identifying animals or listening to a sound file--anything computers aren't good at. (What's next? Tasting wine? Composing a sonnet?) Von Ahn is confident that the good guys are still ahead for now, but the point at which software can reliably read CAPTCHAS is probably as few as three to five years away.

In the meantime, Von Ahn has figured out a way to take advantage of all the spare brainpower hundreds of millions of people expend deciphering wiggly letters. He has teamed up with the Internet Archive, a San Francisco nonprofit that uses computers to digitally scan books and put the text online, where it can be accessed for free. When its scanners find a word they can't read, they automatically turn it into a CAPTCHA that gets exported to a website in need of one. A human reads it and transcribes it, and the results get sent back to the scanner and added to the archive. It's nice to know we humans are still good for something.