Scaling Spam Eradication Using Purposeful Games: Die Spammer Die!

Friday, October 17, 2008 at 1:01AM

Todd Hoff in spam

Update: As expected I'm undergoing a massive spam attack for speaking truth to dark powers. This is the time to be strong. Together we can make a change. What change you may ask? I can't say, just change and lots more change. Let's link arms together and bravely stand against the forces of chaos for a better yesterday and a better tomorrow.

CAPTCHA doesn't work. Even Google can't make CAPTCHA work (Spammers Choose GMail). And even if CAPTCHA worked it wouldn't really work because CAPTCHA solving markets (Inside India's CAPTCHA solving economy) have evolved where for a mere $2 you can buy 1000 human broken CAPTCHA's. And we know once the free market tackles a problem that's it. Game over :-)

Making ever more clever CAPTCHA programs won't outwit and outlast the CAPTCHA solving markets. Until Skynet evolves the only way to defeat humans is with humans.

Using Games to Get Humans to Do Work (like CAPTCHA) for Free

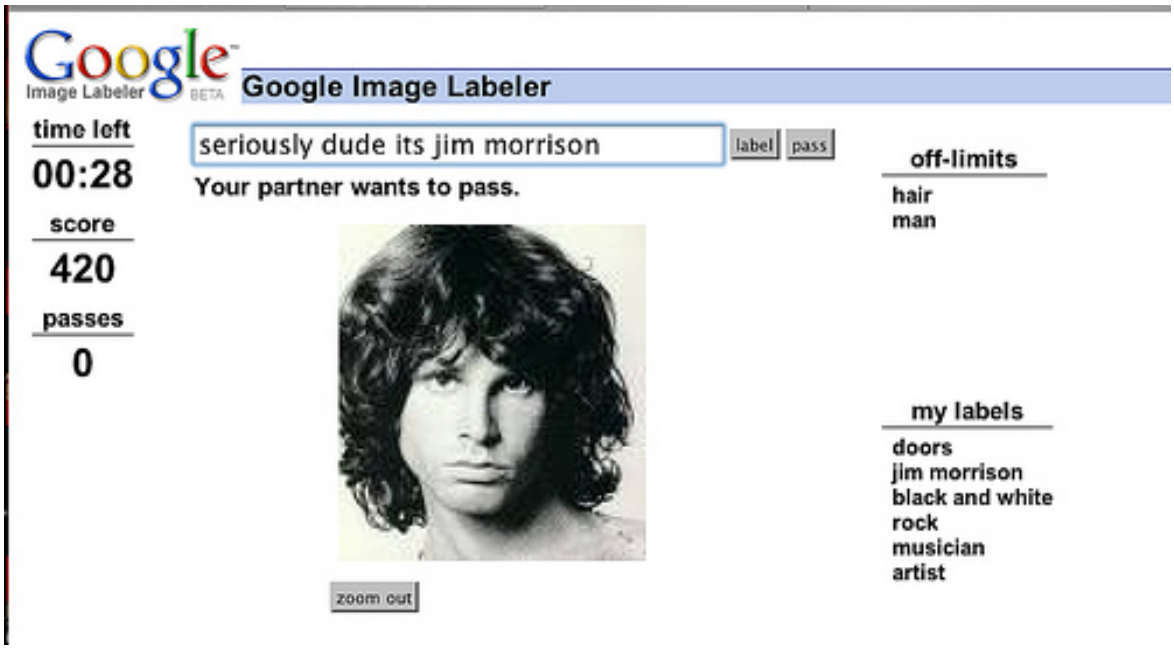
How do we harness the power of humans to do battle with the CAPTCHA solving networks, without, of course, paying them anything? We make it a game! In particular we make a *Game With a Purpose* (GWAP). Read all about GWAPs in Designing games with a purpose. A GWAP is a game *in which people, as a side effect of playing, perform tasks computers are unable to perform.*

Google's Image Labeler

A good example GWAP is Google's Image Labeler, a game in *which people provide*

meaningful, accurate labels for images on the Web as a side effect of playing the game; for example, an image of a man and a dog is labeled "dog," "man," and "pet.". Now this sounds like work. And it is. But because it's made into a game people will do it for free!

An example Labeler session looks like:



In the game two people are matched at random to label the same set of images. Points are awarded when you and your partner match labels. Top scores are kept so you can earn your label street cred. But can't people cheat? GWAP games include cheating detection mechanisms, but we won't go into detail here, see *Designing games with a purpose* for cheater foiling strategies.

ESP Game, Tag a Tune, and Squigl

More games can be found at the GWAP Home Page. They have the *ESP Game* which is like Labeler. *Tag a Tune* is a game where players hear tunes, describe them, and through the description guess if they are listening to the same tune.

In *Squigl* partners see an image and a word. Using the mouse each player traces the object described by the word in the image. Winning is when both players trace the same image. Here's what a Squigl session looks like:

So you see the pattern. Players are picked from a pool. They are asked to do some task that's hard for computers to do. The task must be structured so that winning enables the system to learn something valid while providing a feeling of game play for the humans. Points are awarded and scores are kept to keep the poor human slaves playing.

Creating a Spam Catcher Game

With the basic ideas in place let's create a game for identifying and filtering out comment spam. According to *Designing games with a purpose* this appears to a be an **output-agreement** type game, which has the following structure:

- **Initial setup**. Two strangers are randomly chosen by the game itself from among all potential players;
- **Rules**. In each round, both are given the same input and must produce outputs based on the input. Game instructions indicate that players should try to produce the same output as their partners. Players cannot see one another's outputs or communicate with one another; and
- **Winning condition**. Both players must produce the same output; they do not have to produce it at the same time but must produce it at some point while the input is displayed onscreen.

Simple enough. But comments exist as a part of blogs, websites, microblogging engines, and other programs. Any game has to interface with live systems.

Integrating the game with a comment system might work something like:

- User comments are sent from an originating system to a decentralized game comment queue.
- Comments are pulled from the queue as new games start. Posts are stripped of identifying information and presented to the players.
- Points are allocated if both players agree that a comment is spam or not spam within a very short period of time. With comments latency is the name of the game so they need to be processed as fast as possible.
- Comments and the spam judgments are sent back to the originating system for handling.

It's not too hard too imagine such a system being used for content other than comments and for making judgments like age appropriateness and other subtle criteria that could be communicated using site meta data.

One UI idea it to make the game like a first-person-shooter. Spam is blasted into a 1000 pieces. Oh that would be rewarding, but you can also imagine all the usual game type mechanisms to keep people interested. An accuracy feedback loop would be useful to rate players so less accurate players could be dropped from the game.

Players would be recruited from the general population. Another good source of players is the site owners and the site participants who's sites are the source of comments. This would be sort of **Internet Comment Tax** for keeping the Internet safe and sane.

I, for example, would sign up to process 500 comments a week in order to have HighScalability.com comments processed by the game. Everyone else taking advantage of the system could pledge a number that made sense for their site. This would provide a ready pool of motivated players and docents to keep the game running efficiently.

A nice widget system would make it possible to play the game from any site.

The Final Move

Spam crushes many sites. Many site owners don't even allow comments anymore because of the time it takes to deal with spam, which is a shame, because without interactivity the internet might as well be a newspaper. We can't let those spammers win! A system like the Spam Catcher Game might be able provide the human oversight, quick latency, and high throughput needed to out compete the CAPTCHA solving networks. The game is finally afoot!

Related Articles

- GWAP Home
- Designing games with a purpose
- Inside India’s CAPTCHA solving economy
- Spammers Choose GMail
- Google's Image Labeler

- Google Crashing
