Automatic CAPTCHA detection

Intermediate version of the project

Martin Sopko

2022

Can be CAPTCHA solved by computer?

CAPTCHA Completely Automated Public Turing test to tell Computers and Humans Apart is well know system to protect websites from malicious bot programs. But is still effective?

According to the research provided by Apostolis Zarras, Ilias Gerostathopoulos, and Daniel Méndez Fernández [1]are ML services capable to break CAPTCHA. They used publicly available ML services and that means that anyone who has at least some knowledge how to use these services can create a automated CAPTCH solver.

Abhijeet Chougule, Harshal Tupsamudre and Sachin Lodha[2] proved another more disturbing information about how easily text-based CAPTCHA can be broken and how many state institutions in India still use this system. They attacked 14 major Indian websites and got 90% accuracy on 10/14 websites. Between them was website of State Bank of India and Indian Railways.

What kind of automatic CAPTCHA detection system exist?

Since we know that nowadays CAPTCHA systems can be broken, we need to develop better systems. One of options is make these tests harder or longer but that is affecting users what we don't want to. This is reason why we should look at some systems that don't use standard CAPTCHA but are harder for computer to break it.

One of existing system with this definition is GESTCHA based on determining if is user human or not by gestures. Ahmed Iqbal Pritom, Md. Abdullah Al Mashuk, Somi Ahmed, Nazifa Monira and Md. Zahidul[3] Islam developed system by collecting data from angular velocity and gyroscope. With collected data they trained ML algorithms to recognize basic gestures of users. This test is much easier for users and doesn't take so much time.

Unfortunately, this test work just for smartphones. How can we make easier CAPTCHA test for users using normal computer?

[1] Can Today's Machine Learning Pass Image-Based Turing Tests?; authors: Apostolis Zarras, Ilias Gerostathopoulos & Daniel Méndez Fernández; year: 2019; International Conference on Information Security
[2] Revelio: A Lightweight Captcha Solver Using a Dictionary Based Approach; authors: Abhijeet Chougule, Harshal Tupsamudre & Sachin Lodha; year: 2020; International Conference on Information Systems Security
[3] GESTCHA: a gesture-based CAPTCHA design for smart devices using angular velocity; authors: Ahmed Iqbal Pritom, Md. Abdullah Al Mashuk, Somi Ahmed, Nazifa Monira & Md. Zahidul Islam; year: 2022; Multimedia Tools and Applications

Automated CAPTCHA system for web applications

How will our system work?

Our automated CAPTCHA system should perform detection if is user human or computer without human knowledge. Main thing that is different between bots trying to make 100 fake users and real user is mouse. Bots don't need to move with mouse because they found input fields by their location in code. Most of inputs field have their id attribute named by which information should be inserted into this field. Humans usually use their mouse to find field and by click on this field before entering any information. Our system counts with this behavior.

We created simple registration formular in angular application. Then using the mouse events we handle where user click on page. Every click is compared with previous and if is different from previous we increased number of clicks. For our registration formular which has 6 fields and one checkbox we decided that when user clicked more than six times it is human.

```
this.mousedown$ = fromEvent(this._el.nativeElement, 'mousedown');
        this.mousedown$.subscribe((e) => {
          this.x = e.x;
          this.y = e.y;
          console.log('clicked');
          console.log(this.x,this.y);

          if(this.lastx!=this.x||this.lasty!=this.y){
            this.numOfClicks++;
          }
          if(this.numOfClicks>6){
            this.isDisabled=false;
          }
          this.lastx=e.x;
          this.lasty=e.y;
          console.log(this.numOfClicks);
        })
```

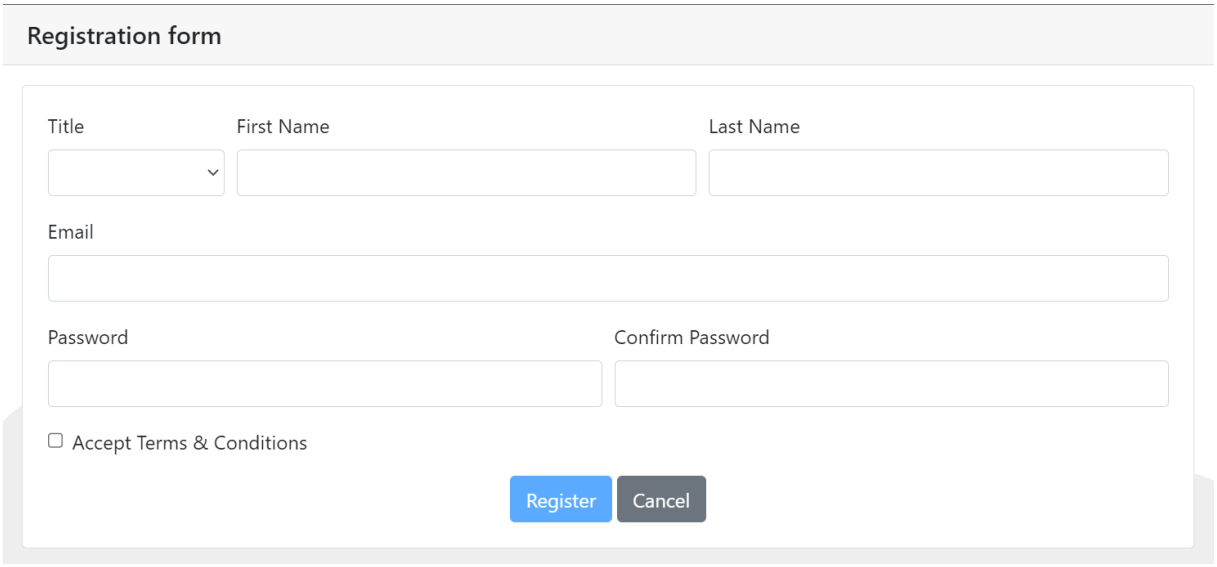


Figure 1Registration formular

We set register button to disabled until system recognize user as human. After successfully recognition system allow user to register.

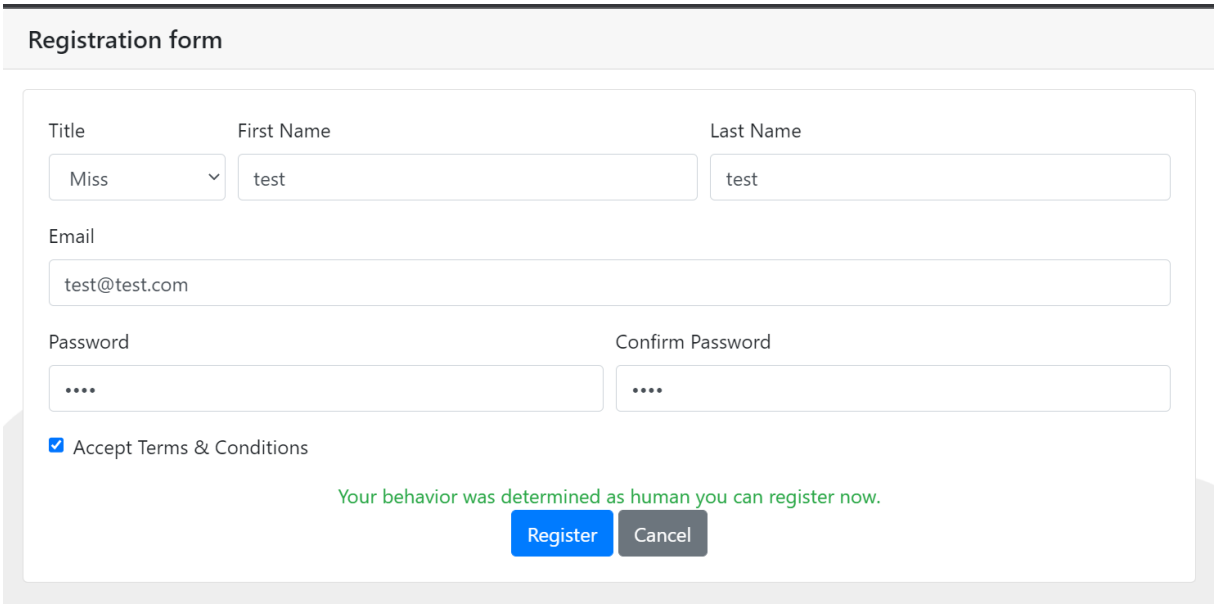


Figure 2succesful recognition

As you can see in Figure 2 when user enter all information registration button is enabled and can be registrated.

For testing purpose, we are logging in to console every click and number of recognized clicks from user.

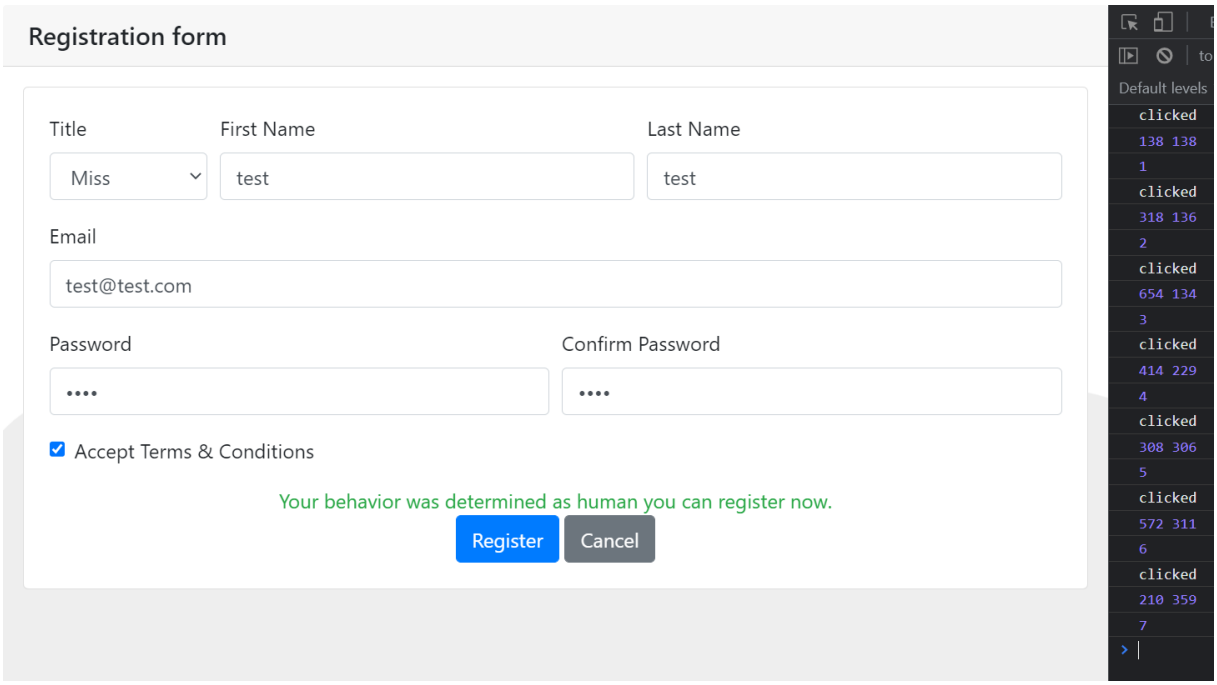
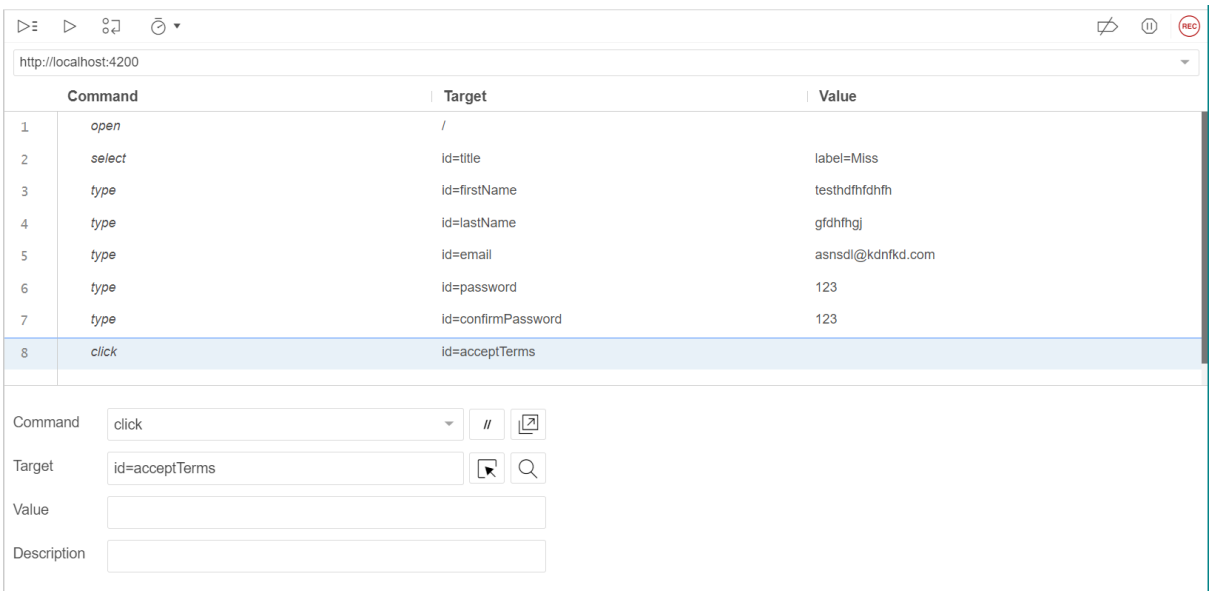


Figure 3Console

Now came question if computer can solve this problem. For our small testing we created automated test in Selenium IDE with following commands:



Then we run this automated test which simulates our bot. From test we got this result:

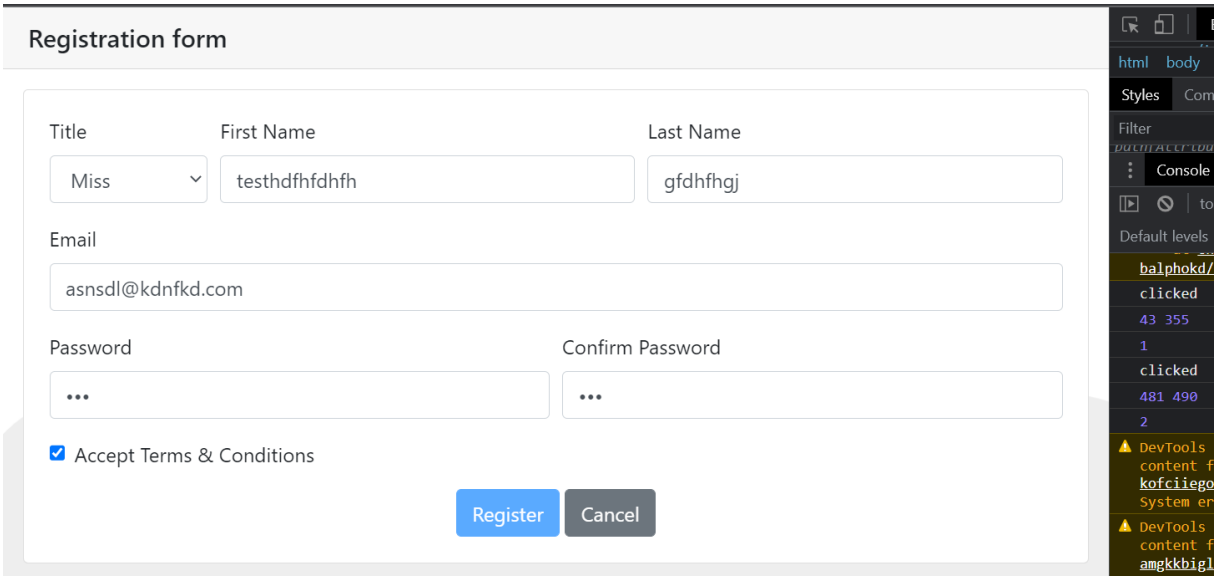


Figure 4Automated test

From result we can see our system recognized only two clicks. As we were testing, we found another problem and it is that some users are not clicking on field but using tab key to navigate to next field. In that case our system says they are not humas. Our next focus is include another detection for this cases.