**END TERM REPORT**

***by***

**Naman Jain, Akshay Pratap Singh, Archit Nagpal**

Section: K19QK

Roll Numbers: 20, 67, 11

Github link:- <https://github.com/naman2001jain/text-based-captcha-generator.git>



**Department of Intelligent Systems,**

**School of Computer Science Engineering,**

**Lovely Professional University, Jalandhar**

November, 2020

**Student Declaration**

This is to declare that this report has been written by our team members. No part of the report is copied from other sources. All information included from other sources have been duly acknowledged. I/We aver that if any part of the report is found to be copied, I/we are shall take full responsibility for it.



Signature:

Name: Akshay pratap singh

Roll Number: 67

Place: LPU

Date: 24/10/2020

**TABLE OF CONTENTS**

**Text based captcha Generator PAGENO. 1**

|  |  |
| --- | --- |
| **1. Background and objectives of project assigned ................** | **1** |
| **1.1 Abstract**.............................................................................. | 1 |
| **1.1.1 Introduction.....................................................................** | **3** |
| **1.1.2 Background.....................................................................** | **5** |
| **1.2 Captcha and turing...........................................................** | 7 |
| **1.2.1 Why it was needed.............................................................** | **7** |
| **1.2.2 Captcha Logic..........................................................** | **9** |
| **1.2.3 How to generate.............................................................** | **10** |
| **2 Description of Project ...........................................................** | **14** |
| **2.1 Implementation ....................................................................** | **17** |
| **2.2 Advantages and disadvantages..........................................** | **18** |
|  | **20** |

|  |  |  |  |
| --- | --- | --- | --- |
| |  |  | | --- | --- | | **2.4 Conclusion.........................................................................** | **17** | |  |

**Background and objective of the project**

Abstract

CAPTCHA (completely automated public test to tell computers and

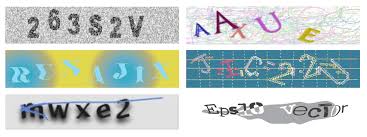
lumens apart), a technique used by a computer to tell if it is interacting with a human or another computer. Because computing is becoming pervasive, and computerized tasks and services are commonplace, the need for increased levels of security has led to the development of this way for computers to ensure that they are dealing with humans in situations where human interaction is essential to security.

Activities such as online commerce transactions, search engine submissions. Web polls, Web registrations, free e-mail service registration and other automated services are subject to software programs, or bots, that mimic the behaviour of humans in order to skew the results of the automated task or perform malicious activities, such as gathering e-mail addresses for spamming or ordering hundreds of tickets a concert

In order to validate the digital transaction, using the CAPTCHA system the user is presented with a distorted word typically placed on top of a distorted background. The user must type the word into a field in order to complete the process Computers have a difficult time decoding the distorted words while humans can easily decipher the text.

Introduction

A CAPTCHA is a program that protect websites and application against bots by generating and grading tests that humans can pass but current computer programs cannot. For example, humans can read distorted text as the one shown below, but current computer programs can't:



A CAPTCHA or Captcha is a type of challenge-response test used in computing to ensure that the response is not generated by a computer. The process usually involves one computer (a server) asking a user to complete a simple test which the computer is able to generate and grade. Because other computers are unable to solve the CAPTCHA, any user entering a correct solution is presumed to be human. Thus, it is sometimes described as a reverse Turing test, because it is administered by a machine and targeted to a human, in contrast to the standard Turing test that is typically administered by a human and targeted to a machine. A common type of CAPTCHA requires that the user type letters or digits from a distorted image that appears on the screen. The term "CAPTCHA" (based upon the word capture) was coined in 2000 by Luis von Ahn, Manuel Blum, Nicholas J. Hopper (all of Carnegie Mellon University).

Background

The need for CAPTCHAs rose to keep out the website/search engine abuse by bots. In 1997, AltaVista thought ways to block and discourage the automatic submissions of URLs into their search engines. Andrei Broder, Chief Scientist of AltaVista, and his colleagues developed a filter. Their method was to generate a printed text means only that only humans could read and not machine readers. Their approach was so effective that in an year, Spam-add-ons were reduced by 95% and a patent was issued in 2001.

In 2000, Yahoo's popular Messenger chat service was hit by bots which pointed advertising links to annoying human users of chat rooms. Yahoo, along with Carnegie Mellon University, developed a CAPTCHA called EZ-GIMPY, which chose a dictionary word randomly and distorted it with a wide variety of image occlusions and asked the user to input the distorted word.

In November 1999, slashdot.com released an online poll asking which was the best graduate school in computer science (a dangerous question to ask over the web ). As is the case with most online polls, IP addresses of voters were recorded in order to prevent single users from voting more than once. However, students at Carnegie Mellon found a way to stuff the ballots by using programs that voted for CMU thousands of times CMU's score started growing rapidly. The next day, students at MIT wrote their own voting program and the poll became a contest between voting bots". MIT finished with 21, 156 votes, Camogie Mellon with 21,032 and every other school with less than 1,000, Can the result of any online poll be trusted? Not unless the poll requires that only humans can vote.

Captcha and Turing test

CAPTCHA technology has its foundation in an experiment called the Turing Test. Alan Turing, sometimes called the father of modern computing, proposed the test as a way to examine whether or not machines can think - or appear to think - like humans. The classic test is a game of imitation. In this game, an interrogator asks two participants a series of questions. One of the participants is a machine and the other is a human. The interrogator can see or hear the participants and has no way of knowing which is which in the interrogator is unable to figure out which participant is a machine based on the responses, the machine passes the Turing Test.

Of course, with a CAPTCHA. the goal is to create a test that humans can pass easily but machines can't. Is also important that the CAPTCHA application is able to present different CAPTCHAS to different users. If a visual CAPTCHA procced a static image that was the same for every user, it wouldn't take long before a spammer spotted the form, deciphered the letters, and programmed an application to type in the correct answer automatically.

Most, but not all, CAPTCHAS rely on a visual test. Computers lack the sophistication the human beings have when it comes to processing visual data, We can look at an image and pick out patterns more easily than a computer. The human mind sometimes perceives patterns even when none exist, a quirk we call pareidolia Ever see a shape in the clouds or a face on the moon? That's your brain trying to associate random information into patterns and shapes.

But not all CAPTCHAS rely on visual patterns. In fact, it’s important to have an

alternative to a visual CAPTCHA. Otherwise, the Website administrator runs the risk of franchising any Web user who has a visual impairment. One alternative to visual test Indian audible one. An audio CAPTCHA usually presents the user with a series of spoken letters or numbers. It's not unusual for the program to distort the speaker's voice, and it's also common for the program to include background noise in the recording. This helps that voice recognition

Why Captcha was needed?

**It was originally created to prevent the following issue:-**

* **Multiple number of votes in the online polling.**
* **To prevent the spam emails.**
* **Preventing the dictionary attack(Brute force password cracking).**
* **Search engine Bots**
* **Tampering with ranking systems.**
* **To make sure the visitors count on the website consist of real people**

Captcha Logic!

Step -1: The CAPTCHA image (or question is generated. There are different ways to do this The classic approach is to generate some random text, apply some random effects to it and convert it into an image.

Step -2: It is not really sequential During step-1, The original text (pre-altered) is persisted somewhere, as this is the correct answer to the question. There are different ways to persist the answer as a server-side session variable, cookie, file, or database entry.

Step -3: The generated CAPTCHA is presented to the user, who is prompted to answer it.

Step -4: The back-end script checks the answer supplied by the user by comparing it with the persisted (correct) answer. If the value is empty or incorrect we go back step-1: a new CAPTCHA is generated. Users should never get a second shot at answering the same CAPTCHA.

Step -5: If the answer supplied by the user is correct, the form post is successful and processing can continue. If applicable. the generated CAPTCHA image is deleted.

How to generate the Captcha!



The code is written that way it selects the 6 digits, in those position number 3rd and 5th are character and remaining are the integers and they’re all small letters none is capitalised.

Then a distorted image is generated in the front page:



Graphical user interface, application

Description automatically generatedNow only human eyes could understand the text distorted text which is to be entered in the box marked below to qualify and authenticate for the next step: the use case can be unlimited which depends upon the user’s website or the particular program which enable to download something.

Refresh Button to Refresh the captcha

Insert the text here



Description and Implementation of the project

Some python scripts:-

* main.py
* interface.py
* func.py

main.py script is used here to run the project.

Interface.py script is having a bunch of code which builds the interface of a great design. There is something like login interface which is also having a captcha image and we can refresh it through refresh button just right to the “GO” button. Also when the user answers correctly then the color and text of the status bar is changed(In case of correct answer the text is changed into “Successful” and background is changed to green color, In case of incorrect answer the text is changed into “Invalid” and background is changed to red color).

Some of the functions are build in interface.py script.

* Refresh() to refresh the captcha if captcha is not recognized by the user.
* Chkauth() to check if entered captcha is matching with generated captcha or not.
* Login\_screen() which contains the GUI functionality.

Widgets are packed in the frames by two ways:-pack and grid

Func.py script is containing the script of generating the captcha text and same captcha image is also created named as cap\_img.png

Func.py can be said as the engine of our project.

There is a folder named as captcha containing the python scripts how captcha image is created of any text.

So there is a function named as generate(text) text is passed as an argument and through write function a .png file is built.

Advantages and disadvantages

**Advantages:**

* To Distinguish between the user and find out if it is human or a computer bot.
* Make the online marking and voting and counting more genuine.
* The captcha makes the online stores and web shop more secure for the user when making payment and will help the websites too.
* Reduces the number of span and fraud messages and illegal activity by identifying the user.
* Stops the abuse of free services provided by the Companies so that needy will always get the services.

**Disadvantages:**

* Makes the Process slow wherever it was used: login, signup, download etc
* Person with the disability will struggle if the special support is not provided by the developers.
* Not every time we will be able to understand the distorted words and this may consume time.
* Difficulties with certain browsers regarding the support system.

Conclusion

In conclusion, The internet is does not only have User base of good people. The security is required more than ever on the internet and hence the measures are also taken. The web sites are provided to help users, give them more information or make their work easier, by CAPTCHA makes reaching the web sites services more complicated. At last, we need some way to tell CAPTCHA that special needs users are human not bots. Sites with attractive resources and millions of users will always need access control systems that limit the bad using of them. At that level, it is reasonable to employ many concurrent approaches, including audio and visual CAPTCHA, to do so. However, it must be noted that users with disabilities can interact with a given resource in a reasonable amount of time.

Other Conclusions:

1. CAPTCHAS are any software that distinguishes human and machine.

2. Research in CAPTCHAS implies advancement in Al making computers understand how human thinks.

3. Internet companies are making billions of dollars every year, their security and services quality matters and so does the advancement in CAPTCHA technology.

4. Different methods of CAPTCHAS are being studied but new ideas like

Re CAPTCHA using human time on internet is amazing.

BONAFIDE CERTIFICATE

Certified that this project report “Text Based Captcha Genrator ” is the bonafide work of “Naman jain, Akshay Pratap singh, Archit Nagpal” who carried out the project work under my supervision.

### <<Signature of the Supervisor>>(Due to Covid 19, signature is exempted )

Name: Dr. Dhanpratap Singh

<<Academic Designation>>

Id: 25706

Department: school of computer science and engineering