Università degli Studi di Padova
Corso di Ingegneria del Software 2022/2023

CAPTCHA: Umano o Sovrumano?

Distinguere se un utente di una procedura è un umano o un robot.

ZUCCHETTI
IL SOFTWARE CHE CREA SUCCESSO

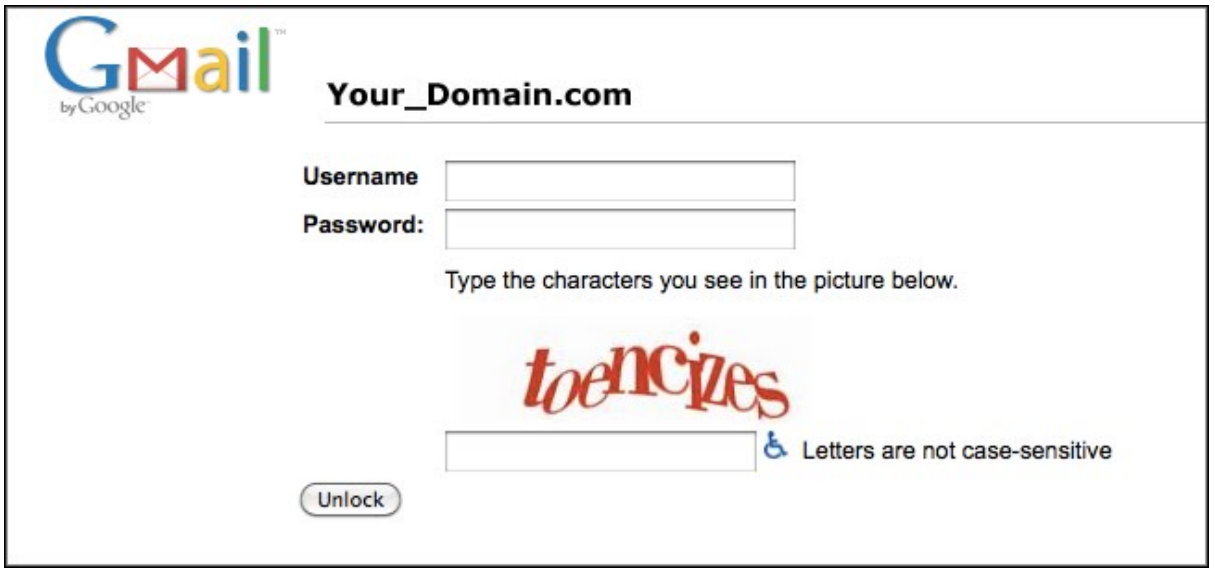
Oggetto dell’appalto

Le procedure moderne sono ormai ospitate in sistemi cloud. Questo le rende raggiungibili sia dai nostri utenti, ovunque essi siano e con qualunque device si connettano, ma anche da strumenti informatici robotizzati che le attaccano con fini poco amichevoli.

E’ quindi importante sviluppare dei metodi che permettano di distinguere se la persona che sta interagendo con il nostro sistema è effettivamente una persona fisica o dimostra i comportamenti di uno strumento automatico.

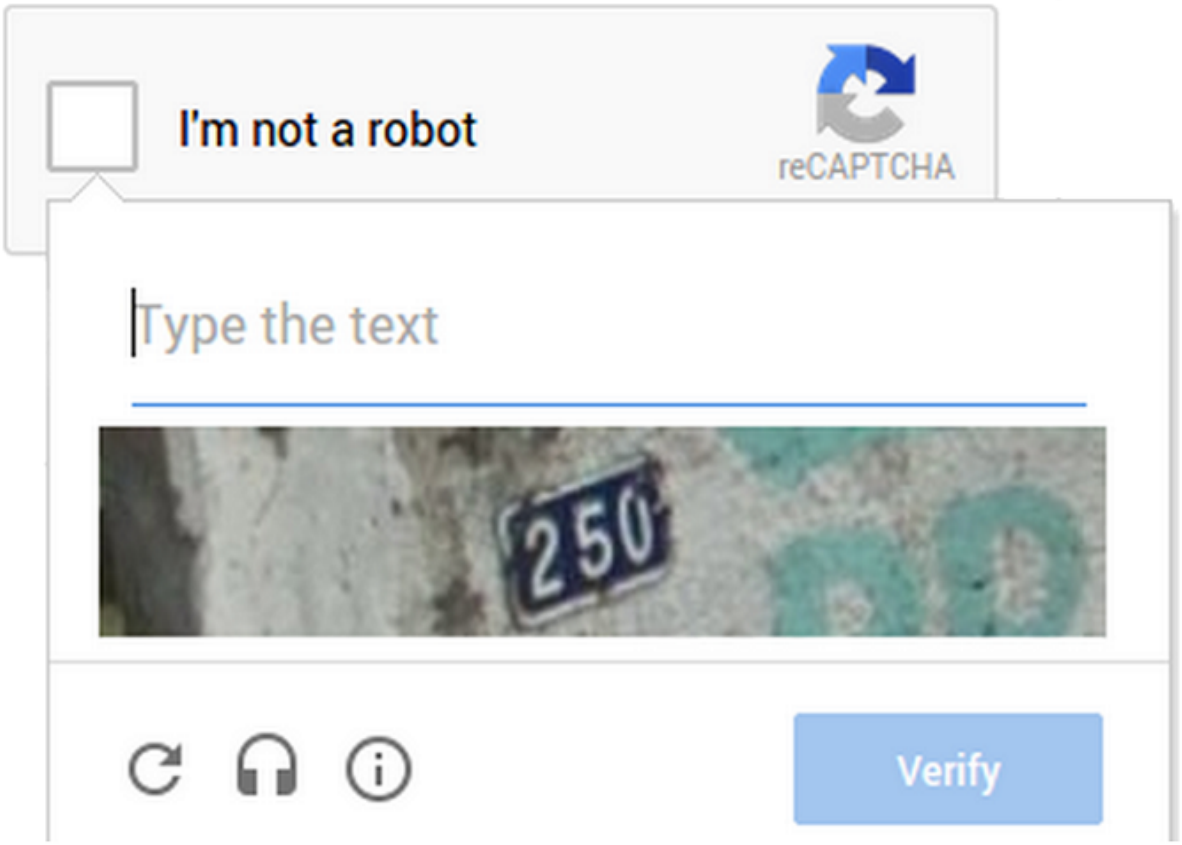
Nel tempo sono stati sviluppati vari metodi per questo scopo, battezzato con l’acronimo CAPTCHA che sta per “Completely Automated Public Turing test to tell Computers and Humans Apart”.

I primi CAPTCHA avevano la forma riportata nell’immagine seguente.



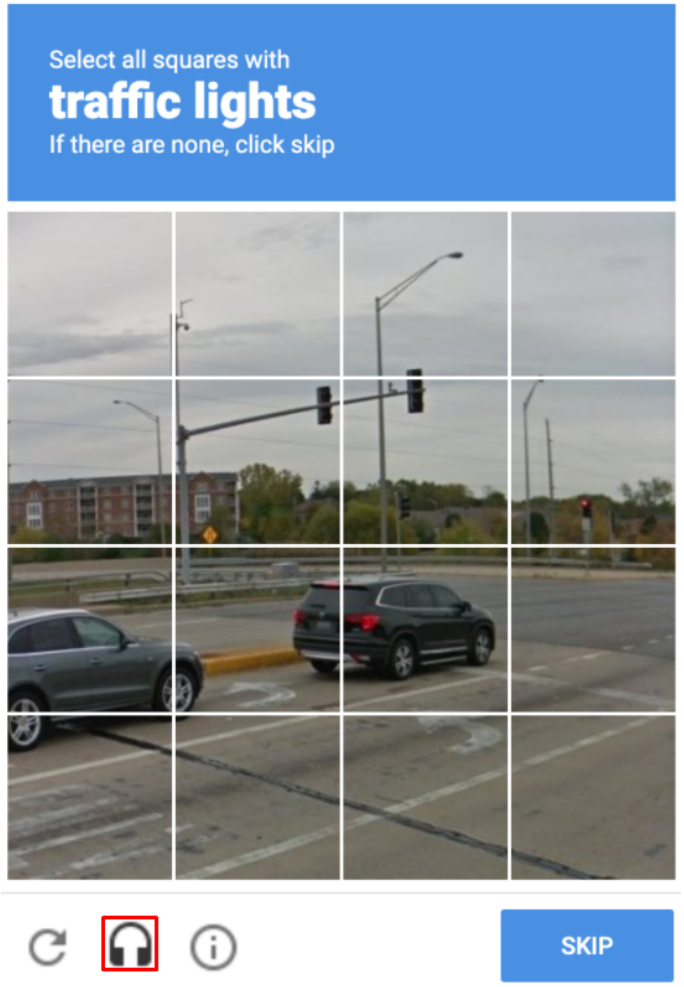
Si può apprezzare una scritta in rosso leggermente alterata, in modo che un umano riesca ancora a leggerne le lettere, mentre un bot si presume non abbia questa capacità. Si può già notare che ci sono degli umani che non sono in grado di fare questa attività, per cui si vede il simbolo dei portatori di handicap che avranno accesso ad un’altra forma di verifica.
Poi i sistemi di riconoscimento ottico delle lettere sono molto migliorati e questo metodo non è più considerato robusto.

Sono allora apparsi sistemi con la forma che vediamo nell’immagine che segue.



Questo sistema richiede di riconoscere un numero presente in una foto. E' apparso dopo che Google ha creato Street View, l'applicazione che ha fotografato gran parte delle strade di tutto il mondo. Viene il leggero sospetto che gli utenti ignari abbiano aiutato "Big G" a riconoscere i numeri di casa che non erano stati distinti dai sistemi automatici di lettura delle foto.

Poi lo scenario è ancora cambiato ed oggi vediamo richieste di questo tipo:



Chissà cosa staranno cercando di farci distinguere … a chi può servire avere immagini di semafori, strisce stradali, passaggi pedonali, autobus, moto, camion? Mah, forse qualche sospetto ci viene in mente, ci sono applicazioni che in effetti devono saper distinguere queste cose e c'è un gran bisogno di immagini ben catalogate …

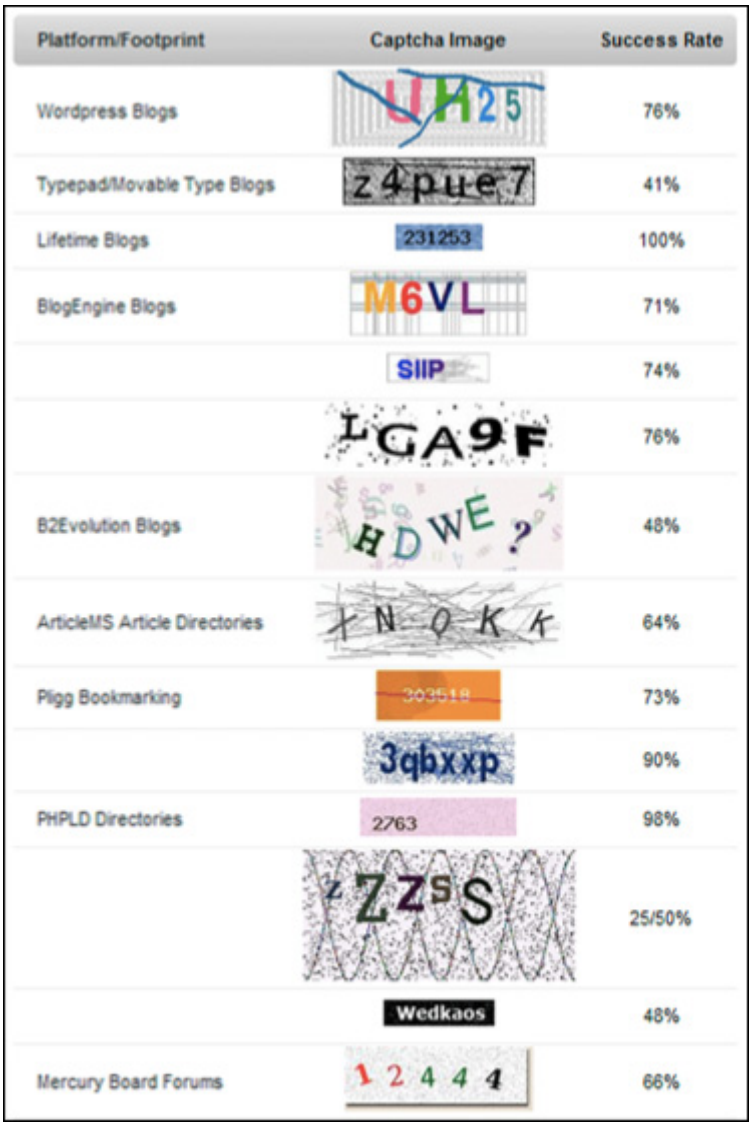
Ci sono molte operazioni che richiedono questa distinzione e portano a requisiti diversi. Vediamo con degli esempi:

1. Registrazione in un sito: chi gestisce un'applicazione di solito desidera che a registrarsi sul sito siano delle persone e non dei bot. Il CAPTCHA serve ad evitare che si registrino a migliaia utenti fasulli che in realtà non sono poi interessati ai servizi offerti.
2. Pubblicazione di contenuti: è comune l'esperienza di forum impestati da pubblicità non pertinenti con la discussione. Un bot che riesce ad accedere risponde in una conversazione con il suo paylod pubblicitario, disturbando gli utenti reali.
3. Login: un bot può tentare attacchi "brute force" provando migliaia di password per un determinato profilo utente di cui ha determinato l'esistenza.

Già da questi esempi si vede come scopi diversi possano richiedere funzionalità di CAPTCHA diverse: nei primi due è necessario stoppare i bot in senso assoluto, solo una persona reale può registrarsi o postare messaggi, nel terzo caso lo scopo del CAPTCHA è mitigare attacchi “brute force” che si immagina possano essere portati solo da un bot.

Ma come dicevamo i sistemi di CAPTCHA vengono superati dalle tecnologie, quindi i primi due sono abbastanza accettabili se falliscono (imponendo un costo e una complessità all'attaccante), mentre se il terzo viene eluso diventa molto pericoloso.

Il presente capitolato propone di sviluppare una applicazione web minimale che implementi un sistema di verifica per distinguere umani da robot.



Platform/Footprint	Captcha Image	Success Rate
Wordpress Blogs		76%
Typepad/Movable Type Blogs		41%
Lifetime Blogs		100%
BlogEngine Blogs		71%
		74%
		76%
B2Evolution Blogs		48%
ArticleMS Article Directories		64%
Pligg Bookmarking		73%
		90%
PHPLD Directories		98%
		25/50%
		48%
Mercury Board Forums		66%

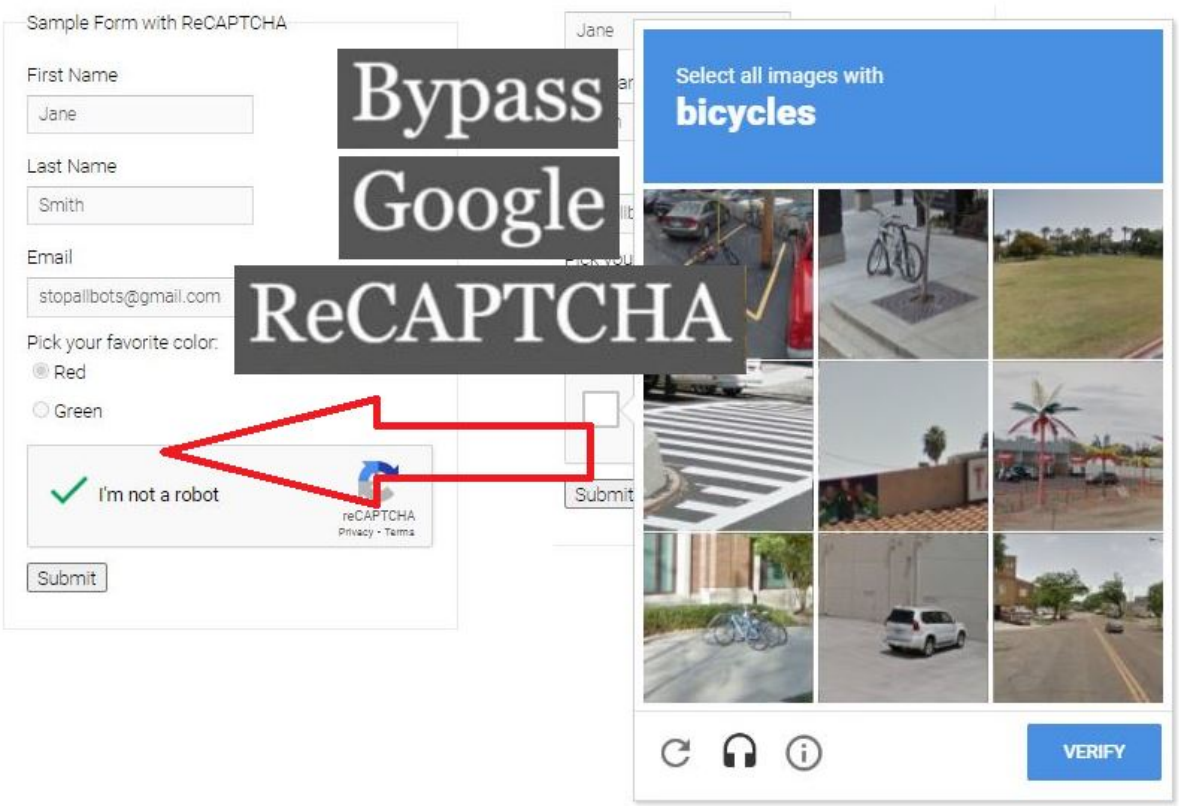
Caratteristiche e Requisiti Obbligatori

Sviluppare una applicazione web costituita da una pagina di login che presenti un sistema in grado di distinguere un utente umano da un robot.
L' applicazione dovrà utilizzare HTML/CSS/JavaScript lato client e la parte server deve essere sviluppata utilizzando Java o PHP.

Il sistema di CAPTCHA potrà essere una libreria Open Source, un servizio (obbligatoriamente gratuito) fruibile via web, un programma originale sviluppato dal gruppo di lavoro.

Realizzata l'applicazione dovrà essere prodotta una verifica che dimostri che il sistema di CAPTCHA non è eludibile chiamando in modo diretto la componente server senza aver utilizzato la parte client. Cioè il server deve essere costruito in modo che un attaccante che legga il codice sorgente della pagina di login non possa presentarsi con dati falsificati che siano accettati supinamente dal server.

Deve quindi essere condotta una analisi sulle tecnologie utilizzate, al fine di indicare quali sviluppi futuri di Machine Learning, Intelligenza Artificiale o qualunque altro tipo possano con il tempo rendere inefficace il sistema di verifica.

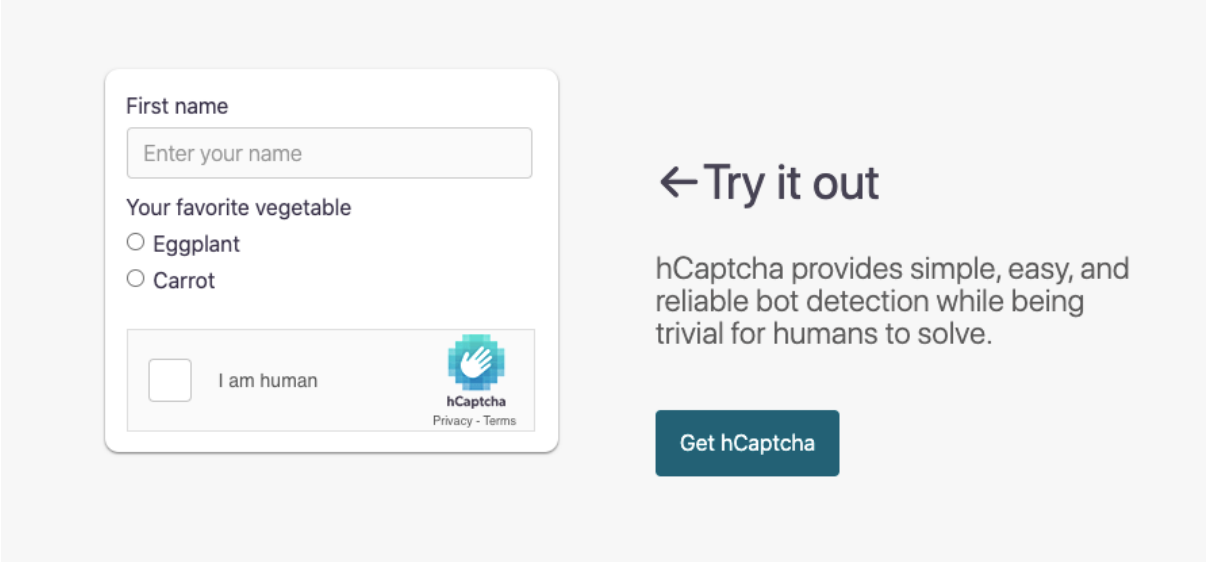


Requisiti Opzionali

Il tema proposto si presta a molte estensioni. L' applicazione minima richiesta è costituita dalla sola pagina di login, si può estendere questa applicazione includendo:

1. Form di registrazione di un nuovo utente.
2. Mini-forum che accetta contenuti prodotti dagli utenti dell'applicazione.
3. Pagina di ricerca sul forum con verifica CAPTCHA.

Per ogni scopo potrebbe essere studiata la reale necessità ed utilizzato un sistema di CAPTCHA diverso caso per caso.



ZUCCHETTI
IL SOFTWARE CHE CREA SUCCESSO

Suggerimenti

I metodi per realizzare un sistema di CAPTCHA sono molteplici:

1. Immagini distorte
2. Risposte a domande
3. Selezione di oggetti
4. Rotazioni di oggetti tridimensionali
5. Piccoli puzzle

ma un sistema di CAPTCHA potrebbe fallire per problemi dell'utente umano:

1. Disabilità
2. Cultura
3. Lingua
4. Evoluzione tecnologica

solo per nominarne alcuni.

Potrebbero esserci degli utenti non in grado di superare il test perché vedono male i colori o non capiscono il compito che viene richiesto.

I sistemi che non prevedono un servizio esterno devono inoltre tener conto che è necessario disporre di tanti elementi diversi quando viene proposto il problema: se infatti si dispone di poche immagini o di un set limitato di quiz l'attaccante potrebbe enumerarli e quindi istruire opportunamente un sistema di attacco automatizzato.

La pagina di Wikipedia riporta molte informazioni sull'argomento.

https://it.wikipedia.org/wiki/CAPTCHA

Una raccolta di librerie è a questo link:

https://github.com/ZYSzys/awesome-captcha

qui si trovano sia librerie per realizzare il sistema dei CAPTCHA che programmi di attacco ai CAPTCHA esistenti.

Il sistema di gran lunga più utilizzato è reCaptcha di Google, reperibile a questo indirizzo:

https://www.google.com/recaptcha/about

https://developers.google.com/recaptcha/intro

Una libreria interessante con un approccio originale, molto diverso dagli altri sistemi, è mCaptcha.

https://mcaptcha.org/

In questo sito viene offerto un servizio per “rompere” i CAPTCHA:

https://2captcha.com/

Variazioni ai requisiti

In corso d'opera non sarà possibile variare/modificare i requisiti minimi (obbligatori per accettare il prodotto). Sarà invece possibile variare i requisiti opzionali, in quanto saranno i gruppi vincitori dell'appalto a modificarli / eliminarli / aggiungerli.

Documentazione

Il progetto dovrà essere supportato dalla documentazione minima richiesta per il corso di Ingegneria del software e dovrà essere fornito un manuale per l'utilizzo ed un manuale per chiunque voglia estendere l'applicazione.

Garanzia e Manutenzione

L'azienda Zucchetti SPA è interessata a questo progetto come dimostrazione della fattibilità dell'obiettivo utilizzando le tecnologie web. Costituirà titolo preferenziale nella valutazione delle proposte la pubblicazione del progetto sul sito “github.com” o altri repository pubblici, in conformità con i relativi requisiti di natura open-source, per favorire la continuità del prodotto risultante.

Rinvio

Per tutto quanto non previsto nel presente capitolato, sono applicabili le disposizioni contenute nelle leggi e nei collegati per la gestione degli appalti pubblici.

Contatti

Per qualsiasi informazione la persona di riferimento in Zucchetti è Gregorio Piccoli all’indirizzo email gregorio.piccoli@zucchetti.it