

YAPPTCHA: Yet Another Picture Promoted CAPTCHA with Spam Stopping, Image Labeling and SIFT Accomplishment

M. Babaei*, M.B. Ghaznavi Ghouschi**, A. Noori***

*Tehran University, m.babae@ut.ac.ir

**Shahed University, ghaznavi@shahed.ac.ir

***Qazvin Islamic Azad University, alirezanoori@hotmail.com

***Abstract** We present YAPPTCHA, a new CAPTCHA (Completely Automated Public Turing test to tell Computers and Humans Apart) as a security measure to recognize human users. The proposed system uses images instead of distorted text to label images as a valuable output. YAPPTCHA is generated from images on the Web. In order to solve this type of CAPTCHA, users must type two words for description of two images. When users pass the challenge, the provided meaningful labels are used to determine the content of images. In addition, semantic graphs for labels and images are created and it will make it possible to develop an image semantic search engine accordingly. Due to usage of images in this system, and its architecture, it is highly secure compared to its counterparts. In a user study involving 60 participants, YAPPTCHA's word accuracy is measured to be 98.18% while 61.26% of attempts of the users were successful.*

Keywords: CAPTCHA, Ontology, Image labeling, Security

1. Introduction

In recent years, security has been an important aspect of the Web. Many operations, such as banking transactions, registration, etc. are being performed on the Web. As a result, the Web administrators are often struggling to protect their websites against malicious attacks. One of the most popular and widespread security measures is CAPTCHA (Completely Automated Public Turing test to tell Computers and Humans Apart). It is a service that prevents bots and automated scripts from abusing websites. Email providers take advantage of CAPTCHAs in their registering forms; weblogs use it to prevent automated programs from spamming their website, and so on.

There are a number of CAPTCHA systems; however, all of them face security problems. Currently reCAPTCHA [1] is the most resistant system against attacks; nevertheless, it has been cracked by Stanford University researchers [2].

The proposed system, called YAPPTCHA, aims to solve this problem by using images instead of distorted text. In addition to the security aspect, the labels provided by the users to pass the challenge can be used to determine the content of images grabbed from the web. This is similar to reCAPTCHA's approach which uses the provided words to digitize books.

The rest of this paper is organized as follows. Section 2 reviews related works. YAPPTCHA is described in detail in section 3. Experiments and their results are given in section 4, while section 5 concludes.

2. Related Work

A CAPTCHA is a program which is designed based on Automated Turing Test [3]. Alan Turing in 1950 [4] introduced it to test a machine's ability to exhibit intelligent behavior [4]. A machine will pass the test successfully if a person outside of a room cannot distinguish it from a human in the room only based on some textual interactions and their replies.

In the rest of this section some of the most effective and popular CAPTCHA systems and their features are discussed. At the end of this section, a comparison of YAPPTCHA and those systems is presented.

2.1 Overview of CAPTCHA Projects

Based on the Turing test, CAPTCHAs create challenges which are not easily performed by computers. This usually involves an artificial intelligence problem, such as natural language processing, character recognition, speech recognition and image understanding. Due to this fact, there are a variety of different CAPTCHA systems.

In 1997, Moni Naor presented the first idea based on Turing test to distinguish a computer from human, which was not published [3]. This manuscript contains several crucial notions and intuitions which led to creation of