

Design of Accented Character-based CAPTCHA with Usability Test for Online Transactions

Olanrewaju, O. T.

Dept. of Computer Science, Federal College of Animal
Health and Production Technology, Moor Plantation
Ibadan, Nigeria
ayotundetaiwo@gmail.com

Osunade, O.

Dept. of Computer Science
University of Ibadan
Ibadan, Nigeria
seyiosunade@gmail.com

Abstract—Websites serve as the primary interface on the Internet for transactions such as subscription, downloads, database access and storage. Websites are however, sources of security breaches to information systems that are attached to them. Several techniques have been developed to provide security on websites such as secured socket layer (ssl) and CAPTCHA systems. CAPTCHA is an authentication system for verifying human identity during online transactions. Text, mathematical operations, images and audio have been used to develop CAPTCHA systems. The basis of each system has been limited thus leading to successful attacks and compromised systems. In this work, the aim is to integrate accented characters into the CAPTCHA code generation mechanism and test the usability of the developed system on a website. The results indicate successful generation and user acceptability.

Keywords—CAPTCHA; Human Computer Interaction; Accented characters; website security

I. INTRODUCTION

Security is a main issue for users of technology products and a topic of interests for researchers in mathematical sciences and computer science. The ubiquitous nature of information technology has increased the need for security considerations in many digital products, services and transactions. CAPTCHA (Completely Automated Public Turing test to tell Computers and Humans Apart) is an authentication scheme that allows a digital device e.g. computer or electronic service e.g. e-commerce website, to judge whether a user is a Human or an automated Computer program. The development of automated programs called bots that could access and interact with web services in the same way as humans led to the adoption of CAPTCHA as a means of authentication. Most commercial websites use a type of CAPTCHA for authentication to reduce automated attacks from bots. The different types of CAPTCHAs include Gimpy, Question, Aural, and reCAPTCHA which have evolved over the years.

The CAPTCHA schemes available have used text, numbers, images, audio, pictures and several combinations and formats e.g. such as distorted / slanted text, blurred images, to ensure that the CAPTCHA code generated is difficult for the bot to crack. The text-based CAPTCHA has been the most popular because of the universality of the English language.

The English characters do not exceed 26 hence complex automated programs can attack them. However, the inclusion of accented characters in the repertoire of the CAPTCHA system increases the complexity of the code and introduces usability issues.

Thus, this work aims to design and implement a CAPTCHA authentication system using both accented characters with English characters with background image manipulation. The resulting CAPTCHA code generated is expected to be more secure, suitable for touch screen devices and acceptable to users.

II. RELATED WORKS

Novel CAPTCHA designs from a usability perspective were proposed by [1]. The proposed designs factored in cognitive issues and enhanced the difficulty level for the software robot. For example, the hybrid question CAPTCHA uses image(s) of a product with a question asking about the product while the user is expected to respond with text. Other proposed CAPTCHAs include Awareness CAPTCHA, Domain CAPTCHA, Context CAPTCHA and Matching/Puzzle CAPTCHA. The cognitive factors introduced in the design made it usable for the end user.

Colour is an attention getting mechanism, initiates recognition and comprehension, arouses user interest, and serves as a defense against automated attacks, [2] examined the implications of colour in CAPTCHAs for usability and security. He further explained that automated program was designed to extract text from text-based CAPTCHA colour schemes such as Gimpy-r, EZ-gimpy, Megaupload, Securimage, FreeCap, Cryptographp, BotBlock and BotDetect. The attacks on these schemes were successful since colour was not implemented as a security feature. The authors concluded that colour schemes for text-based CAPTCHAs should be simple and focused on usability.

There are different design factors of image CAPTCHAs. Reference [3] evaluated the effect of image layouts, quantities, sizes, tilting angles and colours when image CAPTCHAs are deployed. They proposed a four-dimensional method, i.e. eye-tracking, efficiency, effectiveness and satisfaction, for the usability evaluation of image CAPTCHAs. The research work

used the Asirra image CAPTCHA scheme. The results indicate that only the trapezoid image layout was not preferred; four to nine images was the ideal quantity for display in the image CAPTCHA; bigger sizes for the images were satisfactory; the tilting angles of 0 and 45 were the most preferred for the images; and coloured images gave the most satisfaction to users. The research outcomes are based on controlled experiments and may be different in real life.

There are different techniques through which various CAPTCHA scheme can be attacked, [4] examine current segmentation techniques used to attack various CAPTCHA schemes. He further explained that new CAPTCHA scheme was designed to be used on touch-screen devices without the need for keyboard input. The new CAPTCHA scheme requires the user to find the locations of the characters provided in the challenge character set of the generated CAPTCHA code, then drag-and-drop each challenge character onto the correct character in the image. Tests carried out to determine the usability and robustness of the new CAPTCHA scheme indicated positive feedback from the participants.

Three different categories of CAPTCHAs (text-based, game and face biometrics) are evaluated for usability by [5]. The results showed user preference for text-based CAPTCHA followed by the gamified CAPTCHA, while the biometric scheme was the least preferred although the speed of completion was low. The work also highlighted that participants preferred certain CAPTCHA schemes under different context or scenario.

A CAPTCHA that employed Persian/Arabic characters was developed by [7] which are widely used by all Muslim communities and usually complex due to the difference of letter size, lack of space between words, variation of formats, etc. Their study indicated that, integrating those inherent complexities with random noises and backgrounds, it can hardly be recognized by commercial Arabic OCR (Optical Characters Recognition) software and therefore secure. Meanwhile, user study demonstrated a high recognition rate of 90% for human beings.

The application of Korean syllables in text-based CAPTCHAs were also explored in which each character was rotated, split, distorted and combined with a background image that contained random lines and special effects. They concluded that such a design is secure and effective for Korean users [9].

The comparison of the solving time and correction rate of text CAPTCHAs using English and non-English Characters were also discovered by [6]. The study implied that for native speakers, the correction rate of native-language CAPTCHAs is better than that of English ones, while the solving time are similar for CAPTCHAs based on both languages.

CAPTCHA authentication systems are only as effective as the corpus i.e. characters, numbers, images. Most text-based CAPTCHA authentication systems that use the English character set as the corpus for their authentication have been compromised by attackers. Conventional keyboards do not provide users with accented characters in their layout thus making it difficult to respond to a CAPTCHA system using

accented characters. In this work the integration of accented characters into the corpus is expected to increase the non-vulnerability of text-based CAPTCHA authentication systems.

III. METHODOLOGY

This research work is quantitative and consists of two phases. The first phase is the design of a CAPTCHA system using accented characters, while the second phase is the usability testing of the accented CAPTCHA system.

The design of the accented CAPTCHA generation system has three modules that are interlinked. Figure 1 shows the three modules and the relationship between the modules.

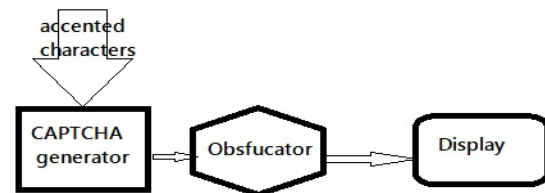


Figure 1: Accented CAPTCHA authentication system

1. CAPTCHA Generator: This is a JavaScript code that randomly generates the five characters of the CAPTCHA. It uses characters from the string "ááèèééíííóóòóúúúşşøø" in addition to non-accented characters. Each accented CAPTCHA code must be five characters long. Four character codes are generated to verify human interaction with the authentication system. The generation algorithm is being patented in Nigeria.

2. Obsfucator: This module that takes the randomly generated code i.e. CAPTCHA and adds a background image. The combination of the background image and generated accented code uses a matrix for splitting and random rotation for merging.

3. CAPTCHA Display: This module displays the generated accented CAPTCHA code the user and accepts the user input for authentication. There are three sub-modules operating within this module. They are:

- Virtual Keyboard: This electronic keyboard provides the accented characters using the format of a standard 101- keyboard. This is used to aid text entry of accented characters.
- Response checker: It determines whether the user has submitted a response for the challenge posed; and
- Match Response: The user input is checked for correctness. The input should match the pre-obfuscated string generated by the CAPTCHA code generator.

The usability tests were carried out in a controlled laboratory environment. The participants got familiar with the experiment apparatuses by means of solving five CAPTCHAs developed for testing purposes. In the actual test, the participant was left alone in the laboratory and different types of CAPTCHA design presented sequentially through a web interface. The test included 12 randomly generated CAPTCHAs and participants were asked to recognize, input and submit the characters shown. After the submission of each CAPTCHA result, a record will be generated on the remote server, indexing the solving time, the user input values and whether the CAPTCHA was solved correctly. The webpage also refreshed automatically after submission and the participant was directed to solve the next CAPTCHA till the end of the task cycle.

For each type of CAPTCHA design, 200 records were collected to obtain the average solving time and correction rate for that type of design. Participants were also interviewed to learn their comments and emotional feelings about the different CAPTCHA designs.

IV. RESULTS

The results of the implementation of the accented CAPTCHA system and usability tests are presented using screenshots, tables and charts. Discussion of the results follows thereafter.



Figure 2: CAPTCHA code generated

The output from the CAPTCHA generator and the Obfuscator is shown in Figure 2. The displayed code contains accented and non-accented characters.

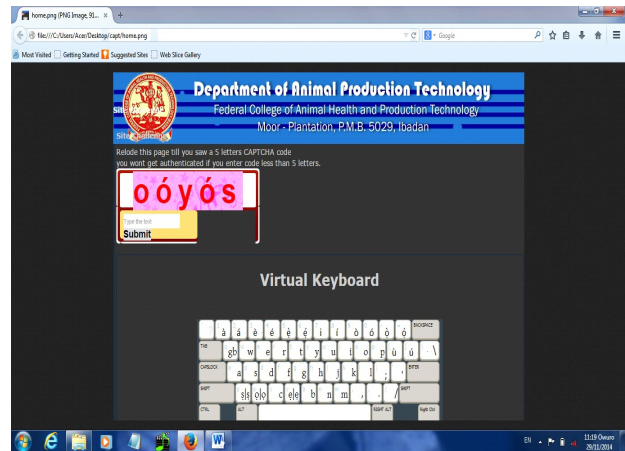


Figure 3: CAPTCHA code entry dialog box

The accented CAPTCHA system is used for authentication on an institutional website. The virtual keyboard allows users with conventional English keyboards to input accented characters and submit either with a mouse or touch-screen.

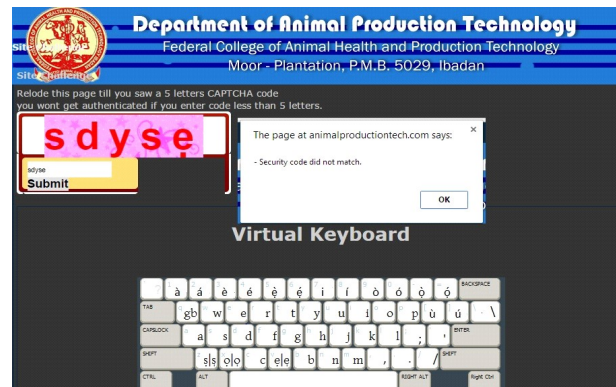


Figure 4: Wrong accented CAPTCHA code entry

The Match Response sub-module of the CAPTCHA Display checks the correctness of each code entry. In Figure 4, the user response does not match the code generated by the system. The system gives a warning to the user and regenerates another accented CAPTCHA code.

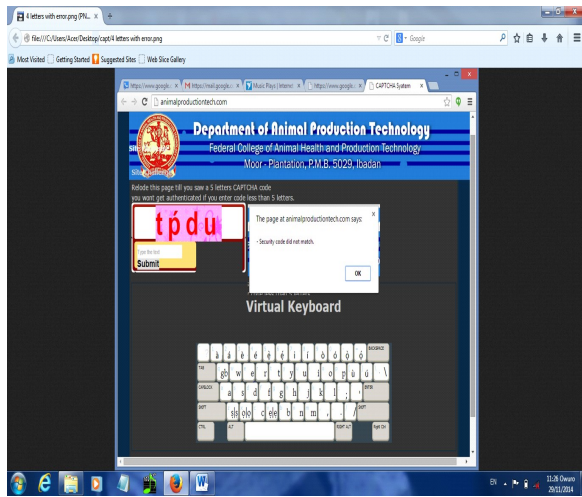


Figure 5: Four letter CAPTCHA code generated

The accented CAPTCHA code must have five characters for authenticity. In some cases, accented CAPTCHA with less than five characters are generated to verify user intelligence and interaction as in Figure 5. The user has to refresh the webpage until a five character CAPTCHA code is generated.

Table 1 was obtained using repeated CAPTCHA codes falling into four categories as shown in Figure 6 a - d. The four CAPTCHA designs are: English characters with numbers, English characters only, Accented characters with marks and Accented characters without marks.



Figure 6a. English characters with numbers



Figure 6b. English Characters only



Figure 6c. Accented Characters with mark



Figure 6d. Accented Characters without mark

The CAPTCHA designs in Figures 6 a-d were used in the usability tests. They were obtained from existing CAPTCHA authentication systems.

Table 1: Usability Results with different CAPTCHA designs

	CAPTCHA DESIGN			
	English characters with numbers	English characters only	Accented characters with marks	Accented characters without marks
Mean Time to Solve (seconds)	2.4622	2.3733	2.3200	2.3644
Standard Deviation	0.48114	0.40643	0.39692	0.51650
No of participants	45	45	45	45

There were 45 participants available for the usability testing as stated in Table 1. The time taken to solve the CAPTCHA code is about the same, even though the CAPTCHA with numbers took more time than the others. The accented character CAPTCHA without marks took more time to solve than the CAPTCHA with marks. This indicates that the presence of accented characters in the CAPTCHA did not affect the user’s ability to solve the code.

V. CONCLUSION

This research work has integrated accented characters into the CAPTCHA system for improved security and wider usage by non-traditional English language natives. The accented CAPTCHA system has been tested for usability in a controlled environment with positive results. Accented characters feature in a lot of human languages such as Yoruba, Latin, and French. The inclusion of accented characters will increase acceptability and diversification of the CAPTCHA authentication system. The development of audio capabilities for accented characters would be an enhancement to the work.

ACKNOWLEDGMENT

The 2015/2016 students of Department of Computer Science and Department of Science Laboratory Technology of Federal College of Animal Health & Production Technology, Moor Plantation, Ibadan, Nigeria are appreciated for testing the accented CAPTCHA system.

REFERENCES

- [1] G. Chaurasia, K.; Singhaniya, M.; Jain, S. and Sivaselvan, B. 2014. Novel CAPTCHA Design based on Cognitive Factors. Proc. of Int. Conf. on Advances in Communication, Network, and Computing, CNC pp789-795.
- [2] Ahmad El Ahmad; Jeff Yan and Wai-Yin Ng. 2012. CAPTCHA Design: Color, Usability, and Security. IEEE Internet Computing pp 44-51.
- [3] Yu, J.; Ma, X. and Han, T. 20XX. Four-Dimensional Usability Investigation of Image CAPTCHA
- [4] Nguyen, V. D., Chow, Y. & Susilo, W. 2014. A CAPTCHA scheme based on the identification of character locations. Lecture Notes in Computer Science, 8434 60-74.
- [5] Krol, K.; Parkin, S. and Sasse, M. A. 2016. Better the Devil You Know: A User Study of Two CAPTCHAs and a Possible Replacement Technology. USEC '16, 21 February 2016, San Diego, CA, USA Copyright 2016 Internet Society, ISBN 1-891562-42-8. <http://dx.doi.org/10.14722/usec.2016.23013>
- [6] Fidas, C.A. and A.G. Voyiatzis. 2013. An empirical study on efficiency and effectiveness of localized vs. Latin-based CAPTCHA challenges. ACM International Conference Proceeding Series.
- [7] Shirali-Shahreza, M. H. and Shirali-Shahreza, M. 2016. Persian/Arabic Baffletext CAPTCHA. J. UCS, 2006. 12(12): p. 1783-1796.
- [8] Yan, J. and Ahmad, A. S. E. 2008. Usability of CAPTCHAs or usability issues in CAPTCHA design, in Proceedings of the 4th symposium on Usable privacy and security. 2008, ACM: Pittsburgh, Pennsylvania, USA. p. 44-52.
- [9] Yang, T.C.; Ince, I. F. and Salman, Y. B. 2009. A Korean CAPTCHA Study: Defeating OCRs In a New CAPTCHA Context By Using Korean Syllables. International Journal of Contents, 2009. 5(3): p. 50-56