Accessibility Challenges with CAPTCHA Services in Digital Libraries

Todor Todorov[1, 2][0000-0002-2443-4618], Galina Bogdanova[1][0000-0002-5463-4274], Nikolay Noev[1][0000-0003-3290-1439]

[1] Institute of Mathematics and Informatics, Bulgarian Academy of Sciences, Sofia, Bulgaria
[2] St. St. Cyril and Methodius University of Veliko Tarnovo, Veliko Tarnovo, Bulgaria
`todor@math.bas.bg, galina@math.bas.bg, nickey@math.bas.bg`

Abstract. In the paper are discussed security problems related to online digital libraries and portals. Special attention is placed to one the methods to distinguish usage of online digital libraries from non-human users. The approaches known under the abbreviation CAPTCHA (Completely Automated Public Turing test to tell Computers and Humans Apart) are considered in the context of their accessibility to the users with different types of disabilities.

Keywords: Accessibility, CAPTCHA, Digital Libraries.

1   Introduction

Digital libraries and web sites in general are very important of contemporary digital life. There are different types of digital web service deeply incorporated in every day's life like email services, travel and shopping applications, social media etc. (Kaur, 2015). There are also some thematic digital libraries with specialized data and services included (Gospodinov & Gospodinova, 2019), (Grigorova, et al., 2012).

One of the common challenges to all these different web services is to prevent all the security issues that could arise during their usage (Todorova-Ekmekci, 2021).

In the paper is discussed one of the main security problems in Internet communication – to distinguish automatically if the other side of communication is a computer or a human. One of the possible approaches to solve this problem is CAPTCHA (Completely Automated Public Turing test to tell Computers and Humans Apart). CAPTCHA is an effective way to prevent automatic bot attacks but there are a lot of issues related to its usability. In the paper is researched the problem with accessibility in the sites that uses CAPTCHA and what are the possible alternatives.

More than one billion people or 13% of the population in the world have various kinds of disabilities. According to the research of the WHO - World Health Organization, the number of persons with a disability grows in time. One of the main aims of EU in the last years has been making valuable content and public services more accessible and with more accessibility for everyone – this includes not just having Internet access and computer to use digital media and content but also having this content in an

understandable and perceivable way for people with difficulties in perception and disabilities (Georgieva-Tsaneva & Subev, 2018), (Sonnenberg, 2020).

CAPTCHA methods have been studied in previous projects and research by the team due to the importance of this technology for people with visual impairments (Web Access, 2016). Their accessibility is included in the requirements of the WCAG web accessibility standard for people with disabilities (WCAG 2, 2005). The research of the team is continued in the project "Digital Accessibility for People with Special Needs: Methodology, Conceptual Models and Innovative EcoSystems".

In Section 2 is presented the main idea behind the CAPTCHA and its practical applications.

In Sections 3 are presented accessibility issues related to CAPTCHA usage and their solutions.

2   CAPTCHA

CAPTCHA is a technique that includes automated tasks that could be easily completed by a human but is difficult to be accomplished by a machine (Inaccessibility of CAPTCHA, 2021).

Character-Based CAPTCHA

The oldest type for human verification. It is based on text characters presented in distorted way as an image. The user is asked to enter the viewed sequence for verification. It is still used in some simple applications, but it is considered as already quite insecure and as we will discuss later it is also quite inaccessible.
An example of character-based CAPTCHA is presented on Figure 1.

With all the contemporary methods for OCR it appears quite easy for computer algorithms and bots to read and bypass this type of CAPTCHA (Tangmanee, 2016).

Fig. 1. Character-Based CAPTCHA.

Sound CAPTCHA

A set of characters or whole words are presented to the user as a sound file. The user should hear it and enter the character sequence. The sound could be played online or downloaded as a separate file.

General idea is to avoid automatic recognition of characters from text but in recent years algorithms for speech recognition also increases their capabilities. To make this approach more difficult for automatic recognition different types of noises are added to

original sound but this also leads to misunderstanding from the human users and need of many repetitions (Kulkarni & Fadewar, 2018).

Logic Puzzle
A set of logical challenges or games that should be easy to be solved by human but hard for a nonhuman user. There are a lot of challenges to implement such an algorithm bearing in mind randomness of puzzles and questions and dealing with free form answers.

An example of logic puzzle CAPTCHA is presented on Figure 2. There are also a lot of accessibility problems from different type that are related to this type of CAPTCHA (Moreno, González, & Martínez, 2014).
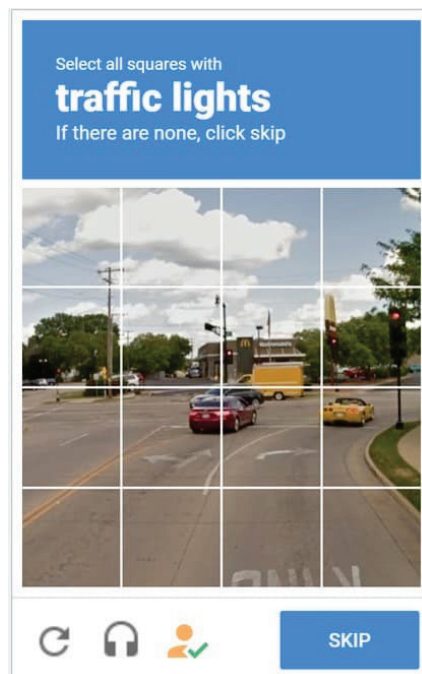

Fig. 2. Logic puzzle CAPTCHA.

Biometrics and Cryptography
Biometrics identifiers like fingerprint are very popular in web applications and especially in mobile applications. They are very effective but sometimes they could provide a false result. Also, there are anonymity problems that could be raised using this approach. That is why it is often used with additional cryptographic methods to guarantee the anonymity of the user (Web Authentication, 2019).

Non-Interactive Methods
These are sets of heuristic and algorithms that identify if the user is a human or a robot without direct interaction. One of the approaches is to restrict some functionalities from digital library or give the user less computational time until the algorithm verifies that the user has "good intentions". The algorithm tracks user interaction with the application, the amount of the data used etc. and decides about the suspicious actions.

Google reCAPTCHA
A PKI (Public key infrastructure) for CAPTCHA services that uses Java Script communication and Google specially developed AI algorithms to solve the Turing test (reCAPTCHA v3, 2021). In Version 2 of the API are used logical puzzles similar to those described in the previous sections. In Version 3 are use non-interactive methods. The AI algorithms generates a score for possibility of the user to be a robot based on its interaction with the application.

3 Accessibility Challenges

World Wide Web Consortium (W3C, World Wide Web Consortium) presents Web Accessibility Initiative (WAI, Web Accessibility Initiative) as guidelines and recommendations for web accessibility. WAI initiated the development of the Web Content Accessibility Guidelines (WCAG), which are now at version 2.1.

The guidelines were built on four principles of W3C:
- Perceivable - information and user interface components must be presentable to users in ways they can perceive;
- Operable - User interface components and navigation must be operable. The interface cannot require interaction that a user cannot perform;
- Understandable - users must be able to understand the information as well as the operation of the user interface;
- Robust - content must be robust enough that it can be interpreted reliably by a wide variety of user agents, including assistive technologies.

In Table 1. Accessibility of CAPTCHA. are presented observations about the accessibility of the CAPTCHA approaches presented in the previous section.

Table 1. Accessibility of CAPTCHA.

CAPTCHA	ACCESSIBILITY
Character-Based	Due to distorted symbols, it is inaccessible to: 1. People with visual disabilities due to distorted symbols; 2. People cognitive and learning disabilities (Gafni, 2016). In most cases not accessible to people that are not using English language.

CAPTCHA	ACCESSIBILITY
Sound	1. Inaccessible to deaf people; 2. Totally inaccessible to deaf-blind people because they also could not use image alternative; 3. People with cognitive disability may also have problems due to the additional noises added to the main sound; 4. Due to distorted speech screen readers could also experience problems.
Logic Puzzle	Inaccessible to people with language, learning or cognitive disabilities. There are disabilities like dyscalculia that causes difficulty in making arithmetical calculations even in simple cases.
Biometrics and Cryptography	In most cases inaccessibility is caused by the impossibility of biometrics information tracking due to some physical disabilities i.e., aging. The cryptography solves these problems to some extent.
Non-Interactive Methods	There are no serious accessibility problems since there is no interaction with the user. It should be considered that different types of disabilities lead to non-usual behaviour in the usage of web portals, and this could cause problems to CAPTCHA decision algorithm.

In general, many specialists advise not to use CAPTCHA without reason if we want to ensure the accessibility. That means to use it only in case when we expect an intensive traffic and large number of users (Digital Accessibility, 2021). Also, all potential disabilities should be addressed and not on concrete types like visual or audio disabilities. Web portals and digital libraries should avoid time limitations in all their functionalities including CAPTCHA services. Accessibility issues should be addressed to the whole application and not only to CAPTCHA technology.

4   Conclusions

The paper presents the essence of CAPTCHA an approach to extent the security of digital libraries and web portals. Different types of CAPTCHA methods are presented. Special attention is paid to web accessibility issues related to CAPTCHA usage.

Acknowledgements.

References

Digital Accessibility. (2021). Retrieved June 11, 2022, from How To Make CAPTCHA Accessible To Everyone: https://www.boia.org/blog/how-to-make-captcha-accessible-to-everyone

Dochkova-Todorova, J., Varbanova, S., & Stoeva, A. (2018). Technologies for mathematics and informatics students with vision loss. *Pedagogical Almanac, 6(2),* 231-219.

Gafni, R. (2016). *The Effect of CAPTCHA on User Experience among Users with and without Learning Disabilities.*

Georgieva-Tsaneva, G., & Subev, N. (2018). Technologies, standards, and approaches to ensure web accessibility for visually impaired people. *Digital Presentation and Preservation of Cultural and Scientific Heritage, Volume 8*, pp. 143 – 149.

Gospodinov, M., & Gospodinova, E. (2019). Implementing a web-based application for analysis and evaluation of heart rate variability using serverless architecture. *International Journal on Advanced Science, Engineering and Information Technology, Volume 9, Issue 6*, 1927 – 1935.

Grigorova, V., Sotirova, K., Naoumova, V., Sameva, A., Dobreva, M., Ivanova, K., & Stanchev, P. (2012). Digitization of Old Mathematical Periodicals Published by the Institute of Mathematics and Informatics, Bulgarian Academy of Sciences. *Digital Presentation and Preservation of Cultural and Scientific Heritage. Vol. 2*, pp. 222-227.

Inaccessibility of CAPTCHA. (2021). Retrieved June 11, 2022, from https://www.w3.org/TR/turingtest/

Kaur, H. (2015). Role of Digital Libraries in the Present Era: Challenges and Issues. (S. Thanuskodi, Ed.) *Handbook of Research on Inventive Digital Tools for Collection Management and Development in Modern Libraries*, pp. 86-102.

Kulkarni, S., & Fadewar, H. (2018). *Audio CAPTCHA Techniques: A Review.*

Moreno, L., González, M., & Martínez, P. (2014). *CAPTCHA and Accessibility - Is This the Best We Can Do?* WEBIST.

Paneva-Marinova, D., Scarpa, M., Riparante, M., Goynov, M., Luchev, D., Pavlova, L., & Zlatkov, L. (2022). Machine-readable Descriptive Structures to Study Medieval South Slavic Scriptoria and Scribes. *TEM journal, 11(2),* 675-682. doi:10.18421/TEM112-22.

Luchev, D., Paneva-Marinova, D., Pavlov, R., Pavlova, L., Márkus, Zs. L., Kaposi, G., Szkaliczki, T., & Veres, M. (2017). Presenting Bulgarian Cultural and Historical Sites with Panorama Pictures. *Digital Presentation and Preservation of Cultural and Scientific Heritage. Conference Proceedings, 7*, pp. 113-122.

reCAPTCHA v3. (2021). Retrieved June 11, 2022, from https://developers.google.com/recaptcha/docs/v3

Sonnenberg, C. (2020). E-Government and Social Media: The Impact on Accessibility. *Journal of Disability Policy Studies, 31(3)*, 181-191.

Tangmanee, C. (2016). Effects of Text Rotation, String Length, and Letter Format on Text-based CAPTCHA Robustness. *Journal of Applied Security Research, 11(3)*, 349-361.

Todorova-Ekmekci, M. (2021). Using Innovative Technologies, Digital Media and Site Tools for Presentation and Sustainable Preservation of Cultural Heritage. *ISMSIT 2021 - 5th International Symposium on Multidisciplinary Studies and Innovative Technologies, Proceedings*, pp. 135 – 140.

W3C, World Wide Web Consortium. (n.d.). Retrieved June 11, 2022, from https://www.w3.org

WAI, Web Accessibility Initiative. (n.d.). Retrieved June 11, 2022, from https://www.w3.org/WAI/

WCAG 2. (2005). Retrieved June 11, 2022, from https://www.w3.org/WAI/standards-guidelines/wcag/

Web Access. (2016). Retrieved June 11, 2022, from https://webaccess.horizonti.bg/aggregator/sources/1

Web Authentication. (2019). Retrieved June 11, 2022, from An API for accessing Public Key Credentials Level 1. W3C Recommendation: https://www.w3.org/TR/webauthn-1/