[1.] Mirko SAJIĆ, [2.]Dušanka BUNDALO, [3.]Dejan LALIĆ, [4.]Zlatko BUNDALO, [5.]Radmila BOJANIĆ, [6.]Luka SAJIĆ

USING SPEECH-TO-TEXT AI CLOUD TECHNOLOGY TO IMPLEMENT SPEECH CONTROL ON SELF-SERVICE DIGITAL DEVICES

[1]Independent University, Banja Luka, Faculty of Informatics, Banja Luka, BOSNIA & HERZEGOVINA
[2.]University of Banja Luka, Faculty of Philosophy, Banja Luka, BOSNIA & HERZEGOVINA
[3.]NLB Banka a.d. Banja Luka, Banja Luka, BOSNIA & HERZEGOVINA
[4.]University of Banja Luka, Faculty of Electrical Engineering, Banja Luka, BOSNIA & HERZEGOVINA
[5]Independent University, Banja Luka, Faculty of Economics, Banja Luka, BOSNIA & HERZEGOVINA
[6.]Apeiron University, Faculty of Information Technologies, Banja Luka, BOSNIA & HERZEGOVINA

Abstract: The purpose of this paper is to point out the possibilities of practical application of already created tools that use modern technology, API functions, artificial intelligence and cloud solutions, whose purpose is to convert speech into text. Using the specific Google Speech-to-text application, the Phyton programming language, its libraries and a multifunctional digital self-service device, a practical solution for using voice commands in working with the device is presented. The role of the aforementioned multi-functional digital self-service devices is to replace and fully automate the work of counter workers. With the additional option of recognizing voice commands, that work is further automated and makes it easier for the client to communicate with the device in an even more natural way. Also, this way of working is more hygienic, because there is no contact with the device, which could be a source of viruses and bacteria. This way of working is especially important during pandemics (covid, flu, etc.)

Keywords: API functions, artificial intelligence and cloud solutions, Google Speech-to-text application, Phyton programming language

INTRODUCTION

The dizzying rise of digital technology has led to the fact that many things can be automated, robotized and adapted to perform many jobs much more efficiently [1]. This especially applies to jobs that are cyclically repeated, and until now they were done by humans. Such daily jobs, where the worker did the same things every day and repeated them, led to an inevitable drop in concentration, irritability and dissatisfaction of those same workers, who hated those jobs over time, because they were monotonous. And it should not be emphasized that the possibility of error was great. The working hours of such jobs usually coincided with the working hours of other jobs, so it was very difficult for the users of those services to provide those other services during working hours, except to leave their workplaces and look for exits. This, in turn, contributes to the dissatisfaction of their managers and business owners.

Therefore, highly automated digital devices are welcome to solve these problems. Over time, such devices were introduced, which achieved an increasing degree of automation, as technologies advanced [2]. Now we live in an age where we can say that the degree of such devices has reached such a level that there are almost no so-called counter jobs, which cannot be replaced completely. We can also say that the jobs of counter workers will slowly almost die out in the sense that they are performed by a human being and will be replaced by machines in the coming period [3].

The advantages of working with automated and robotic devices are multiple:

- Their working hours are 24/7/365;
- Once the bugs are cleared, they no longer error and are able to loop the default actions ad infinitum (basically until something breaks);
- They are much more profitable and reduce the costs of the services they provide, because compared to humans, they do not ask for a salary, paid holidays, sick leave, and they work non-stop. Their investment pays off very quickly, and maintenance costs are significantly lower than when a human being is employed.
- The use of voice command significantly simplifies user operation and increases the speed of operation and speed of obtaining needed services.
- The use of new technologies, such as voice command, significantly increases the level of hygiene and protection against communicable diseases, because now the service user does not come into contact with the operator, a human being, as a potential

carrier of diseases. Especially with the use of voice command technology, the degree of hygiene and health care is further increased.

The specific use of the above-described multifunctional digital self-service device with a built-in option to work via voice command, voice recognition technique, is presented through the device used for the needs of the bank. The following services are implemented on that device:

- payment of bills by filling in payment forms;
- bill payment by scanning the bill;
- printing of statements for individuals and legal entities;
- printing of various bank receipts.

All the listed services can be performed using voice commands, by saying the commands offered on the screen and the numbers when filling out the payment forms. All this is possible in multiple languages. The number of languages for which the voice command can be implemented on this device depends exclusively on the number of supported languages of the API function of the AI Cloud application that will be implemented. In the specific case, it is about the implementation of the GOOGLE Speech-To-Text application, using the Phyton programming language and library support for Speech to text applications [4].

The presented multifunctional digital device is designed to be a smart device, and the possibility of installing a wide range of hardware components, i.e., its multifunctionality gives it the possibility of being used in various places and for various needs. With proper use and arrangement of adequate hardware components, the device can be easily adapted to provide various counter services, such as [5-8]:

- usage for issuing of various certificates issued by municipal institutions;
- counter services in banking;
- services in parking systems;
- services at hotel reception;
- services of telecommunications operators;
- services at petrol/gas stations;
- services sponsored by the city administration (for the needs of the tourist board, theatre, cinema and various other concert tickets, tickets for cultural events, etc.).

Adaptation for work with voice command is also possible for all of these listed services. The principles of such an implementation are explained in this paper.

DESCRIPTION OF MULTIFUNCTIONAL DIGITAL SELF-SERVICE DEVICE

The Figure 1 shows one of the variants of the physical appearance of the device [3]. The physical appearance of the device depends on the wishes of the customer, but also the need to arrange all the necessary hardware components in an ergonomically correct manner.

In the specific case, it is a device developed for the needs of a bank that operates on the market of Bosnia and Herzegovina. Complete autonomy of working with the device, in the sense of working remotely, without the need to touch the device's touch screen, is provided by an adequate selection of used and specially developed applications, which provide full support in work by issuing voice commands. In order for the distance to be fully respected, a built-in version of the POS device with contactless support is placed inside the device.

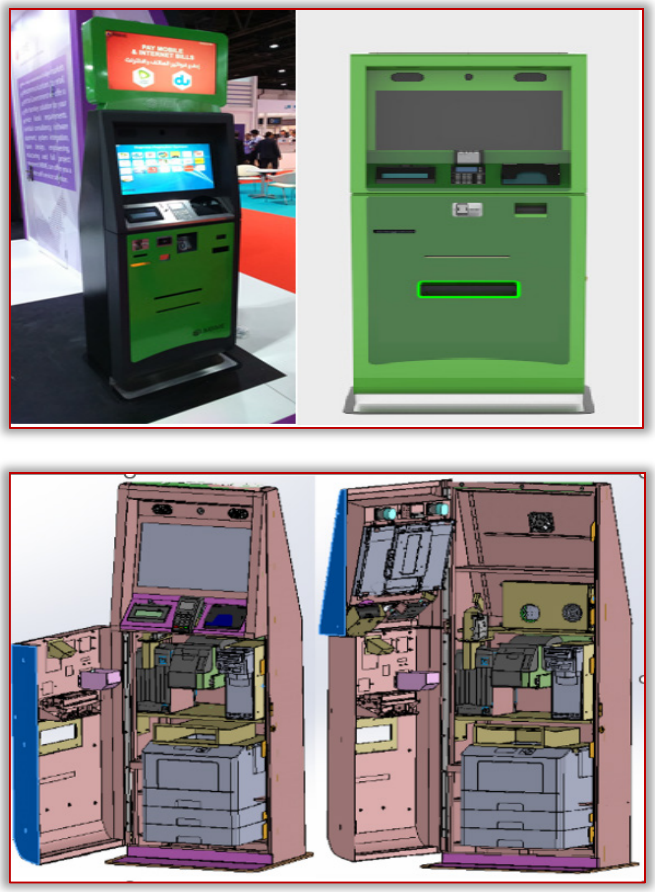


Figure 1. Standing variants of multifunctional digital self-service devices

The Figure 1 shows the basic versions of implemented devices, their standing variants. When using such devices, users stand in front of them and perform all activities to obtain needed service. The standing versions of the devices are intended for use in open areas of banks and their branches, that are permanently available to customers, in areas with operation on 24/7/365 principle.

Figure 2 shows a block diagram with listed hardware components that can be installed. From the selection of listed hardware components, one can sense the multifunctionality of the device, and thus the wide range of services that can be implemented. From the point of view of smart technologies, support for user authentication is particularly interesting [9]:

- Face recognition via embedded camera;
- Identification through the iris of the eye;
- Fingerprint identification;
- Identification through biometric documents (passport, personal documents);
- Identification via RFID, contactless, magnetic trace, chip, i.e., all represented technologies of electronic payment cards'
- Identification by voice recognition.

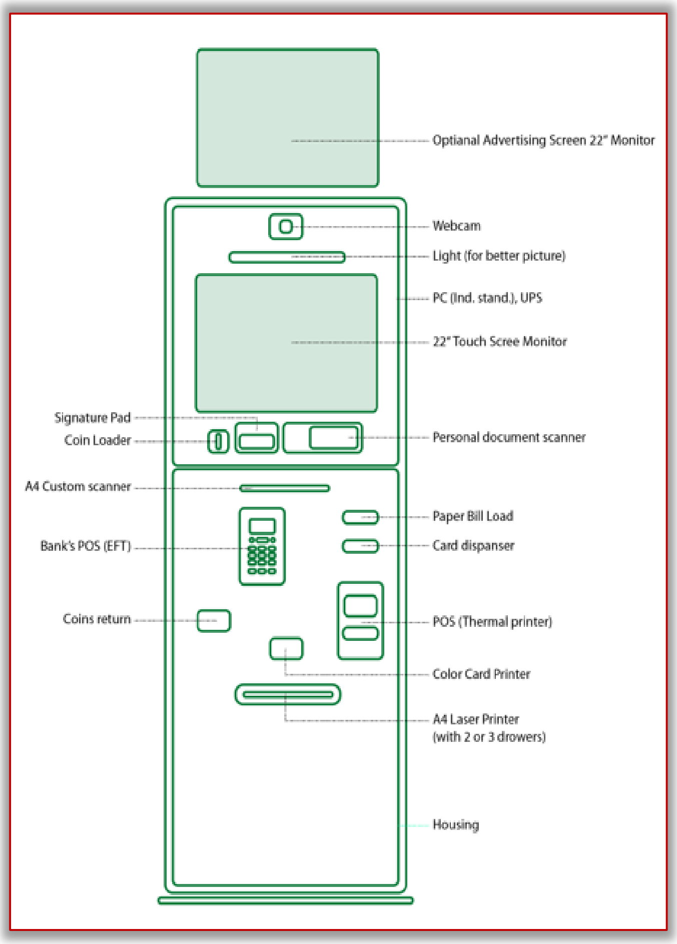


Figure 2. Block diagram of multifunctional digital self-service devices with hardware components

Regardless of the fact that the specific device was made for the needs of the Bosnia and Herzegovina market, it is clear from the multifunctionality of the device that it can be easily adapted for other markets as well.

IMPLEMENTATION OF THE FUNCTIONALITY OF VOICE COMMANDS ON THE PRESENTED DEVICE

The software package of the device consists of several components, made with various software solutions. For example. the support for the camera was done using the Phyton programming language, the support for the scanner for scanning biometric documents was done using the Java programming language, the adaptation of the API functions to support the A4 scanner for scanning receipts was done using the C# programming solution. The support of working with the banking POS device was done through the Phyton software solution. Various sensors for detecting the opening of the device and tampering with it were made using the C programming language. The main application and the connection with the banking Core Banking system and its database (in the specific case Oracle) was done through the SAP PowerBulder12 solution. Applications that monitor the operation of the complete system and the sending of certain messages were developed using the same solution.

The part related to the implementation of the voice command functionality was done using the Google Speech-to-Text AI Cloud solution, using the corresponding API function. More details about this solution can be found on the Google website [4].

For the communication between the device and the Google Speech-to-Text AI Cloud application, a specially designed application developed in Phyton with the support of libraries that enable work with Speech-to-Text types of applications is responsible for that purpose.

The communication is done via JSON format, and the Phyton application delivers the sound record of the spoken command to the Google Speech-to-Text application, by communicating through the API function provided by Google, and with the same function, the Phyton application downloads to the device's computer, in a previously defined folder, a text file, which represents a textual (symbolic) interpretation of a spoken voice command. The system is able to capture and interpret competent sentences with considerable accuracy, i.e., more of them. The degree of accuracy depends on whether the commands are spoken fluently or less fluently. In order to minimize the dependence on the degree of intelligibility of spoken commands in a language, carefully selected inscriptions on the displayed program buttons have been implemented, so that the texts of the displayed commands on the screen, which need to be spoken, have a minimum of coincidence. For example. Avoids the word "Print" appearing as a caption multiple times on the offered set of commands that can be executed at that moment.

Otherwise, the commands are pronounced as clearly as possible and with a moderate volume from a distance of approx. 50 cm from the device. The built-in computer with its built-in microphone records the spoken word on the computer. The mentioned Phyton application distributes that file, via an Internet connection to the Google Cloud Speech-to-Text application, via an API function. The Google app processes it and returns it as interpreted characters. Phyton app downloads those characters packed in a txt file to a previously defined folder. The main application, through its timer event, waits and when it receives the file, processes it in the sense of reading the sent characters. Based on them, it starts the part of the application that executes the interpreted voice command, as if the user activated the corresponding button via the touch screen. In the event that data entry fields are filled in, the program writes the read characters into those fields.

In order to achieve the most accurate interpretation of the spoken command, the program works so that it executes the command even based on a part of the recognized keyword. The list of words that replace the purely spoken and recognized word is written in the configuration file. If it happens that a word is not recognized, i.e., the complex of signs that marks it, it is recorded as unrecognized. At the same time, any touch screen command that follows is recorded, so it is enough to repeat it several times and the program learns to interpret that word as well. So, we can say that a certain degree of artificial intelligence is represented here as well in the code itself.

For example. If we work with the German language and we have the command "Zuruck", to go back one step, when we say it (especially someone whose native language is German), we will sometimes get this command processed by Google as "Zurich". Since we don't have any other command at that moment, which would give a similar combination of characters, we can safely conclude what the user wants us to execute the "Zuruck" command.

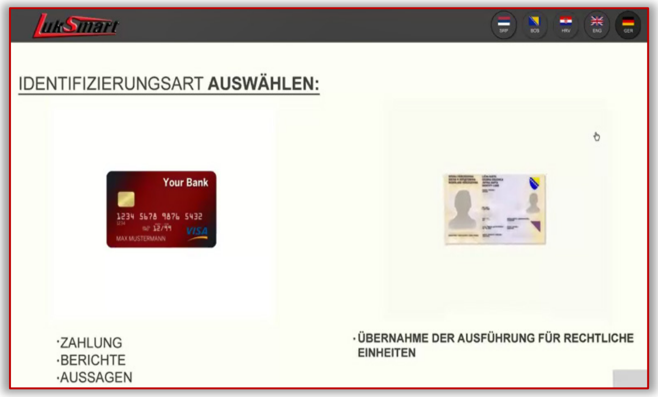
As time goes on and as Google receives an increasing number of different pronunciations of certain words in a language, processed in its BIG data, the quality of its speech recognition applications will also improve. Especially when it comes to well-known world languages, such as English. That is why it is recommended to use software solutions for converting speech to text by those who have a wide range of samples. Google Translate can serve as an example of great progress in quality lately.

When we talked about the Phyton app that communicates with the aforementioned device and the Google Speech-to-text app, we were talking in the singular. Actually, it is a series of adaptations, that is, one application for each spoken language. Which of them will the system, i.e. the main program to use, depends on the selected language from the menu? The language in which commands will be spoken can also be selected by voice command, by saying the name of that language in English (Serbian).

But before the user starts speaking voice commands, he needs to identify himself. It can do this in several ways, depending on the authentication implemented. Authentication through electronic payment cards, contactless, through biometric documents, iris, face recognition, ensure that there is no direct contact with the touch screen, which can be a source of bacteria, viruses, etc.

Saying the voice command "HELP" aloud, the user receives the necessary instructions for working with the device. Depending on the place in the program where the client is currently, the text that is spoken as help is adjusted. Of course, in the chosen language.

Figure 3 shows the beginning screen of the main application. To show multilingualism, the screen is in German, and the flags of the countries are visible on the top right, whose languages are represented on that device.



IDENTIFIZIERUNGSART **AUSWÄHLEN:**

·ZAHLUNG
·BERICHTE
·AUSSAGEN

·ÜBERNAHME DER AUSFÜHRUNG FÜR RECHTLICHE EINHEITEN

Figure 3. Selection of way of identification of client and language for voice commands

In the specific case, when you look at the content of the screen, you can see that the following are implemented as authentication:

- Electronic payment card (RFID, chip, magnetic trace, contactless);
- Biometric personal documents (passport, identity card).

Figure 4 shows the first choice that can be made by speaking the text inside the two buttons offered. The first slide of the image gives a choice whether it is a client in the role of a legal entity or a physical entity. After successful authentication, which is achieved through communication between the main program and CBS Bank, the program receives information whether it is only a physical entity, or whether that person is in the service of a legal entity or is a client of the bank both as a legal entity and as a natural person. Only in the latter case does this window and the possibility of selection appear. If the client is only one of those, there is no need for the program to ask him to decide in which capacity he wants the service, because he already knows. After identifying the client as a legal or physical entity, there is the next set of commands that the client can give by saying the offered text out loud inside the buttons, to choose.

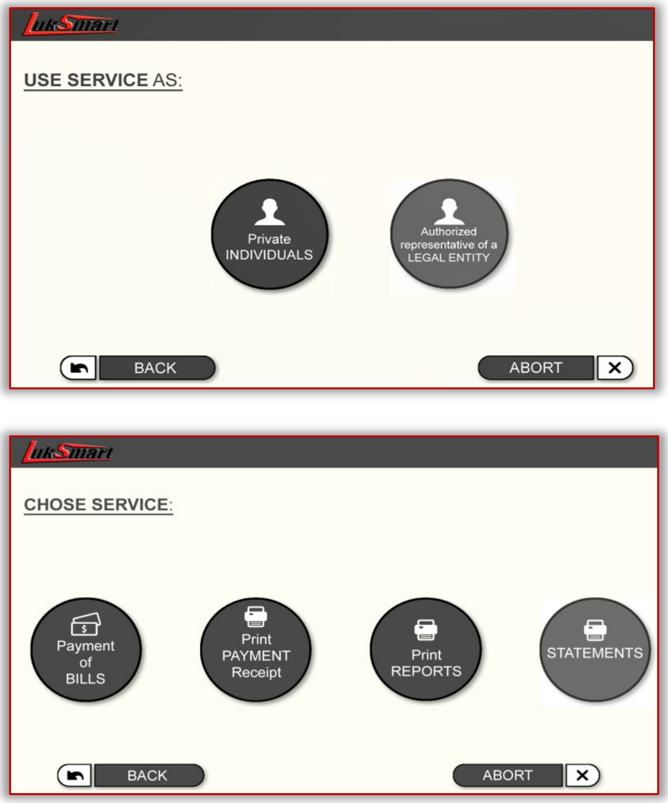
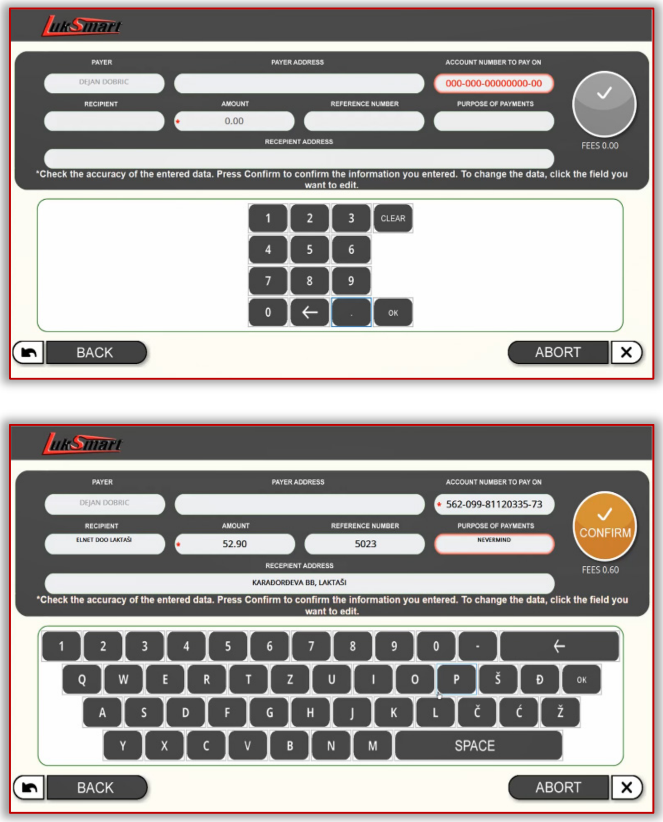


Figure 4. Service selection

Figure 5 shows the screen in case the client chooses to pay the bill by manually filling out the payment slip. However, in this case, instead of filling in the required fields via the touch screen, the client does so by saying the text aloud. The first field to be filled in is the number of the current account to which the payment is made. The client can say one number at a time (this is a recommendation), and can also say a set of numbers in sequence. After properly filling the number of the current account to which the payment is made, the program checks whether that current account exists, and if everything is in order, it transfers it to the field for filling in the payment amount. If an error has occurred, the program asks for re-filling. There is no need to fill in the fields concerning the person who pays, because the system has that information after successful authentication in the banking CBS.

In the lower part of Figure 5, the appearance of a properly filled payment slip is given. The image shows that the confirm button is now enabled and the client can:

- or to add another account or to access payment via the contactless variant by bringing the card to a sufficient distance from the built-in POS device,
- or by inserting the payment card into the POS and entering the pin.

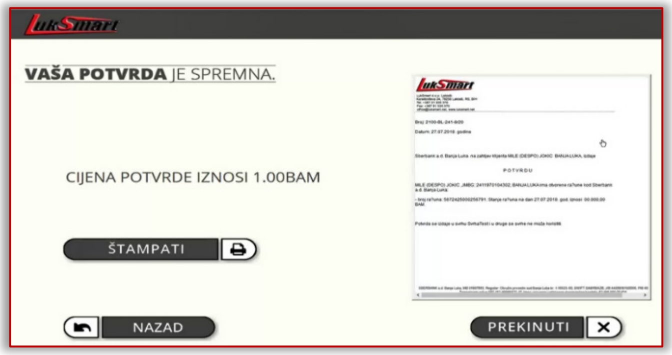

PAYER: DEJAN DOBRIC
PAYER ADDRESS:
ACCOUNT NUMBER TO PAY ON: 000-000-00000000-00
RECIPIENT:
AMOUNT: * 0.00
REFERENCE NUMBER:
PURPOSE OF PAYMENTS:
RECEPIENT ADDRESS:
FEES 0.00
*Check the accuracy of the entered data. Press Confirm to confirm the information you entered. To change the data, click the field you want to edit.

PAYER: DEJAN DOBRIC
PAYER ADDRESS:
ACCOUNT NUMBER TO PAY ON: * 562-099-81120335-73
RECIPIENT: ELNET DOO LAKTAŠI
AMOUNT: * 52.90
REFERENCE NUMBER: 5023
PURPOSE OF PAYMENTS: NEVERMIND
RECEPIENT ADDRESS: KARAĐORĐEVA BB, LAKTAŠI
FEES 0.60
*Check the accuracy of the entered data. Press Confirm to confirm the information you entered. To change the data, click the field you want to edit.

Figure 5. Filling payment slip

It depends on the amount, but in the future, everything will probably be done using the contactless method. For now, there is a limit to the amount that can be paid via the contactless method, due to security reasons.

The Figure 6 shows the appearance of the application in Serbian language in the case when the service of printing a specific bank certificate is provided.



VAŠA POTVRDA JE SPREMNA.

CIJENA POTVRDE IZNOSI 1.00BAM

Figure 6. Service - printing of the selected bank receipt -confirmation (Serbian language)

CONCLUSIONS

The use of ready-made AI Cloud solutions is on the rise and is strengthening in parallel with the improvement of the quality of these solutions. It is to be expected that large companies, which have the ability to work on a large number of samples, will reach satisfactory product quality sooner. We saw on the specific example of using the Speech-to-text application that there are still errors in interpretation, especially if words are pronounced "impurely" and in slang. But for some needs, such as the use on multifunctional smart devices that replace the jobs of counter workers, described in the paper, the quality of the speech-to-text application works at a quite satisfactory level. It is also clear that the quality of these and similar applications will grow day by day. At the same time, the quality of services provided by the described device will increase. The need to introduce voice commands in an age when various transitory diseases and pandemics appear, because they raise the level of hygiene, was also emphasized.

Also, by explaining the operation of the device, its multifunctionality, the possibility of adaptation when introducing new functionalities was shown. The mentioned device is in itself a smart type and has elements of the use of AI technology (it learns how to recognize voice commands and applies them over time, without the need to change the application source code).

Note: This paper was presented at ICAS 2023 – International Conference on Applied Sciences, organized by University Politehnica Timisoara (ROMANIA) and University of Banja Luka (BOSNIA & HERZEGOVINA), in Hunedoara, ROMANIA, in 24–27 May, 2023.

References

[1] ***Holzer H J January 19 2022 Understanding the impact of automation on workers, jobs, and wages, https://www.brookings.edu/articles/understanding-the-impact-of-automation-on-workers-jobs-and-wages/ An NCR Banking eBook, "Bringing personal teller experiences to self-service with ITMs"

[2] ***ncr.com 2020 DeliveryMktg_Banking_ITM_eBook_FNL, https://www.ncr.com/content/dam/web/documents/landing-pages/060220_DeliveryMktg_Banking_ITM_eBook_FNL.pdf

[3] Sajić M, Bundalo D, Vidović Ž, Bundalo Z, Lalić D and Sajić L 2021 Transformation of teller/counter services using modern mobile digital information technologies, Carpathian Journal of Electrical Engineering **15**(1) 203-212

[4] *** Cloud Speech-to-Text, 2023 Speech-to-Text, https://cloud.google.com/speech-to-text/

[5] Sajić M, Bundalo D, Kuzmić G, Bundalo Z and Lalić D 2019 Automation of teller/counter services in smart cities concept using universal digital devices, 27th Telecommunications forum TELFOR 2019, Belgrade, Serbia, November 26-27, pp 1-4

[6] ***replacedbyrobot.info, 2023 Will "Tellers" be Replaced by Robots? https://www.replacedbyrobot.info/70212/tellers

[7] ***Unified Communications Industry, Oct. 17, 2016 The Future of Banking: Interactive Teller Machines & Beyond, https://www.vyopta.com/blog/uc-industry/future-telebanking-interactive-teller-machines-beyond/

[8] ***The Financial Brand, M. Weber, 2023 5 Tips for Your Next Branch Transformation Project , https://thefinancialbrand.com/39641/bank-credit-union-branch-design-tip

[9] Sajić M, Bundalo D and Bundalo Z 2020 User identification and authentication in universal teller/counter digital devices, Conference of Transformative Technologies: Legal and Ethical Challenges of the 21st Century, Banja Luka, Bosnia and Herzegovina, February 7-8, pp 203-215

ISSN: 2067-3809
copyright © University POLITEHNICA Timisoara,
Faculty of Engineering Hunedoara,
5, Revolutiei, 331128, Hunedoara, ROMANIA
http://acta.fih.upt.ro