Software architecture model for a Competence Mapping Application – MapComp

Daniela TIMISICĂ, Răzvan-Alexandru DUȚESCU

National Institute for Research & Development in Informatics – ICI Bucharest

daniela.timisica@ici.ro, razvan.dutescu@ici.ro

Abstract: The MapComp Platform was designed to centralize employees' knowledge and to obtain an up-to-date status of their skills, abilities and certifications. This application is an extensive tool for examining and monitoring competences, which contributes to the efficient allocation of employees to ongoing projects within the Institute, taking into account their abilities. In the case of organisations where the activity is carried out on several projects, employees can be distributed and allocated according to their knowledge and to the needs of the ongoing projects. Competences are entered into the application by the user and are divided into several categories. For each employee, documents attesting the knowledge acquisition can be attached, in order to verify that the information provided is real. The platform offers a transparent way to quickly evaluate the employees' status and to identify, with few clicks, the suitable candidates for various ongoing projects within the Institute. Moreover, MapComp contributes to the employees' development, in accordance with the needs of the Institute, as well as to the identification of talents.

Keywords: Web development, Python, Django, competency mapping, natural language processing.

Model de arhitectură software pentru aplicația Harta Competențelor – MapComp

Rezumat: Platforma MapComp a fost concepută pentru a centraliza cunoștințele angajaților și pentru a obține o situație la zi a abilităților, a aptitudinilor și a certificatelor acestora. Această aplicație este un instrument utilizat pentru a examina și a monitoriza competențele, care contribuie la alocarea eficientă a angajaților pe proiectele aflate în derulare, în cadrul Institutului, luând în considerare abilitățile acestora. În cazul organizațiilor unde activitatea se desfășoară pe mai multe proiecte, angajații pot fi distribuiți și alocați în funcție de cunoștințele lor și de nevoile proiectului. Competențele sunt introduse în aplicație de către utilizator și sunt divizate pe mai multe categorii. Pentru fiecare angajat, documentele care atestă dobândirea de cunoștințe pot fi atașate pentru a verifica dacă informațiile furnizate sunt reale. Această platformă oferă o modalitate transparentă de a evalua rapid situația angajaților și de a identifica, prin câteva click-uri, angajații potriviți pentru diversele proiecte aflate în derulare, în cadrul Institutului. Mai mult, MapComp contribuie la dezvoltarea angajaților, în conformitate cu necesitățile Institutului, precum și la identificarea talentelor.

Cuvinte cheie: dezvoltare web, Python, Django, harta competențelor, prelucrarea limbajului natural.

1. Introduction

MapComp is an online digital platform meant to centralise all the skills of the research staff within the National Institute for Research & Development – ICI Bucharest. The primary purpose is to facilitate the bringing together of similarly skilled engineers and researchers in order to form teams or to find someone with a certain set of skills easily, without having to go door to door or sending e-mails. The secondary purpose is to find the areas where the employees of ICI are the most skilled and the areas where the skill set institute-wide is lacking, so that it can be rectified in the future.

The overall objectives of this application are the following:

- to allow the collection and organization of information regarding both the professional competences, technical skills, diplomas and certifications obtained in IT, as well as the communication skills and the areas of interest in IT, such as IoT, AI, ML;

https://doi.org/10.33436/v32i4y202205

- to allow the display of information on the employees' professional competences within the Institute;
- to enable a correlation between various areas of competence and/or technical skills in IT;
- to offer users the opportunity to access their own competency profile, performing operations such as updating/deleting information, including deleting their user account;
- to offer users the possibility to search or filter data, according to certain criteria of the displayed information;
- to offer the possibility of providing static data of interest.

2. Background & motivation

The European Commission has been monitoring the progress of member states in the field of digitisation since 2014, with their Digital Economy and Society Index (European Commission, 2022), which is a report coming out every subsequent year, Romania also being included in this report, as part of the European Union. Their latest report, DESI 2022, shows that most European countries have made significant strides towards the shared goal of a Digital Europe. However, not all the countries have done so, with Romania significantly lagging behind. Figure 1 shows all the European countries ranked and Romania's place in the said ranking. There are four DESI dimensions related to the Path to a Digital proposal. These four dimensions are: human capital, connectivity, integration of digital systems and digital public services.

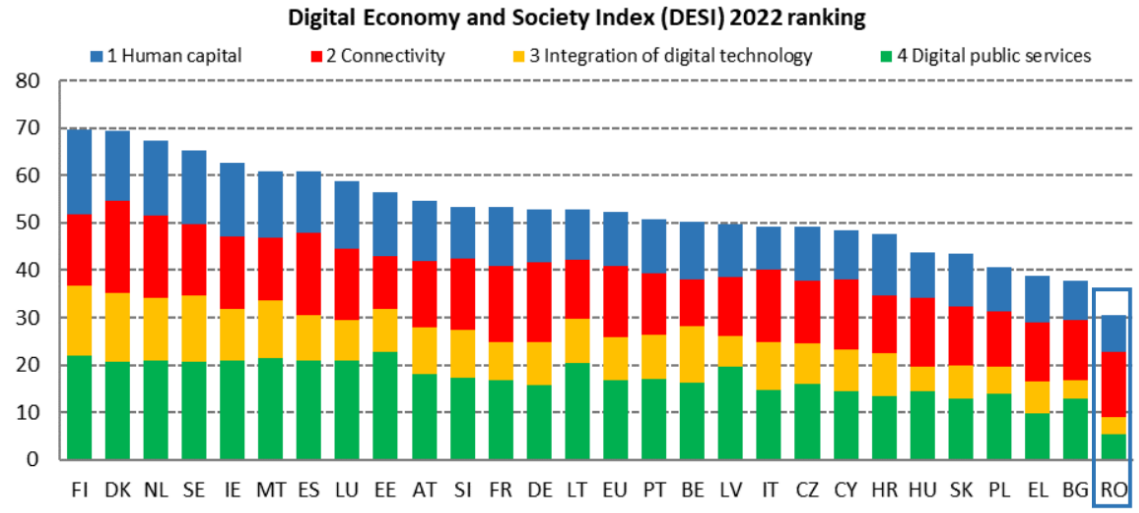

Figure 1. Digital Economy and Society Index 2022 ranking, showing Romania in the last place (European Commission, 2022)

Of the four, the human dimension will be analyzed more thoroughly. This includes personal attributes deemed useful in the production process such as health, level of education and, especially important to this application, knowledge and skills. While 87% of EU citizens have access to Internet, only 54% have basic digital skills. Romania is moving more slowly than its peers, with just over 30% of the population having basic digital skills. A strategy aimed at reducing the gap between Romania and other EU countries is represented by the use of money from NRRP (the National Recovery and Resilience Plan), 20% of which being set aside for digitisation (approximately 6 billion euros). Out of all the components related to digitisation, the largest is Component 7, the Digital Transformation, which pledges 1.8 billion euros towards the transformation of the digital sector. Two other significant components that are worth mentioning are Component 9 which grants 1 billion euros to business support and R&D, in the context of digitisation, and Component 15 which grants another 1,3 billion euros for the digitisation of education.

Digital competence is one of the eight key competences and refers to both accessing and using information and communication technologies, and to having the appropiate knowledge, skills

and competences in relation to them. This involves the confident and critical use of the entire Information and Communication Technology (ICT) range, in all the areas, for work, leisure and communication. The skills and knowledge in the field of ICT are essential for developing an efficient research and innovation system. In addition, ICT usage and development skills are new factors that stimulate employment and research and development in Europe (Smada & Cristescu, 2020).

Competences are a combination of knowledge, skills and attitudes appropriate to the context. Key competences are those which all individuals need for personal fulfilment and development, active citizenship, social inclusion and employment. Thus, the Reference Framework sets out eight key competences (EUR-Lex, 2018):

- Literacy competence;
- Multilingual competence;
- Mathematical competence and competence in science, technology and engineering;
- Digital competence;
- Personal, social and learning to learn competence;
- Citizenship competence;
- Entrepreneurship competence;
- Cultural awareness and expression competence.

The key competences are all considered equally important; each of them contributes to a successful life in society. They overlap and interlock; essential aspects of one domain will support competence in another. Skills such as critical thinking, problem solving, team working, communication and negotiation skills, analytical skills, creativity, and intercultural skills are embedded throughout the key competences.

In order to aggregate the total knowledge of a workplace, it has been concluded that there is a need for a centralised platform that can quantify the employees' skills within the ICT field. There have been some previous attempts at centralising knowledge, including an ontology-based approach for a medical e-Learning application. The implemented personalized concept is based on ontologies and it provides new solutions to several aspects such as: profile, knowledge, learning style, learning objective, training level, student' competences, training level assessment and feedback (Băjenaru et al., 2016).

Lastly, the architecture for a knowledge management system comprises of three levels: a user interface, tools that link the user interface and the database, such as a server, and a knowledge base which is stored using a standard database management system. The tools used to manage this system are applied in every process that the application performs, including storage, aggregation, retrival and processing of the data (Niculescu, 2006).

3. MapComp

MapComp is an online platform designed with two goals in mind. The first one arose from the need to find either people with similar skills or someone with a particular skill set within the Institute. Due to the large number of employees, it is difficult for someone to know all his collegues and all their related computer science skills. Having a centralised database where users input all their skills and their expertise level will make it significantly easier for the leaders to form their teams or for an employee to seek specialised help. The first point is particularly important in the research field where the ability to quickly form a team is of paramount significance, since team members would then have to compete for funds and grants for projects afterwards. MapComp expedites the team building process, extending the duration of the time-consuming process that comes thereafter. In addition to computer science related skills such as programming or domain specific software, the database will also include hardware. This will allow researchers, should they require a specific piece of hardware, to borrow it from a colleague, instead of buying a brand new one, bringing a financial relief for the project or for the Institute.

The second goal of the platform is for leadership to identify the strong areas in which the Institute can perform, as well as the weaknesses within the workforce. The backend has to be designed in such a way that it would facilitate making such a comprehensive report. However, this report isn't the only one produced, as other reports can be generated on the spot by the system, upon request. This can include a list of all the instances where an employee having a specific piece of hardware, regardless of generation, can obtain a normal distribution of the skill level that employees have within a specific programming language or can assess the most fringe skills that researchers have for future projects.

3.1. Front-end

In building this online platform, a simple and intuitive design was chosen. The reason why it was chosen is that users are not expected to spend much time with it, so having an unfriendly user interface may deter some potential users.

The web page is the main component of a web application or website that users interact with. However, useful and complex functionalities of web app or website should be exposed through well-structured web pages with an attractive design. Today, users have high expectations from any web page: to load quickly, to be easy to navigate, to adapt to any device (smartphone, tablet, netbook, laptop, desktop or TV) and to show better search engines performance and accessibility.

The online platform is a dynamic system, a collection of codes written in one or more specific programming languages, used to generate dynamic web pages. On the front-end side, specific web technologies are used, together with libraries and frameworks for the web pages. Also, the web pages are going to be accessed by users through a web browser, the communication between the user and the server application being realised through the HTTP protocol.

Component architectures are an important part of every front-end framework. The software front-end includes the major elements, such as navigation, mockup (text containers, graphic objects, buttons, links, input fields, etc.), graphics, text boxes and interactive forms (WebCase, 2018).

The technologies used by the MapComp online platform, on the front-end side, include the following:

- *HTML* – is the acronym of Hyper Text Markup Language. It is a descriptive language used to structure and describe the content of various documents, including web pages. This markup language represents the code used to structure a web page and its content. For example, the content can be structured within a set of paragraphs, a list of bulleted points, or by using images and data tables. In the case of the online digital platform proposed in this paper, each HTML page is dynamically generated using the Python programming language (MDN Web Docs, 2022b).
- *CSS* – is the acronym of Cascading Style Sheets. It represents the technology used for defining the appearance of the elements of a web page, in order to describe the structure and the content, so that they are as attractive as possible, from the design point of view. Thus, CSS styles were used to make the created web pages look good, not just contain information (MDN Web Docs, 2022a).
- *JavaScript* – is interpreted as a high-level object-oriented scripting language. It is used to create web pages, along with HTML and CSS. This programming language was used to add interactivity to the proposed online digital platform. This behaviour happenes in response to pressing buttons and entrering data into forms (MDN Web Docs, 2022c).
- *Bootstrap* – is a powerful open-source front-end framework. The newest version of Bootstrap was used, which is the most popular HTML, CSS, and JavaScript framework for creating responsive, mobile-first websites and web applications. Bootstrap makes it possible for a web page to detect the visitor's screen size and orientation and automatically adapt the display accordingly. The mobile-first approach

assumes smartphones, tablets and task-specific mobile apps are employees' primary tools used for getting work done. This front-end framework addresses the requirements of those technologies in design and includes UI (User Interface) components, layouts, JavaScript tools and the implementation framework (Zola, 2018).

3.2. Database

The default configuration of Django, namely Django Software Foundation (Django, 2022), includes an SQLite database. However, it will not be used as a database management system in the present work, in large part due to its drawbacks when it comes to scalability and performance. This platform is expected to evolve in the future, so a robust database is needed. SQLite has limited datatypes, making it a burden for developers seeking to add new features to this service. On the other hand, MySQL supports a wider variety of datatypes including date and time, multiple types of strings and numbers. Furthermore, SQLite can only handle a small dataset, with an increase in its size noticeably degrading the performance. In this case, as a large number of researchers with years of experience are expected to join the platform, filling up its database, thus switching to a new management system becomes essential, with MySQL becoming the clear alternative. The first change that was made to the default user was the usage of the email as an authenticator, meaning that the username field was removed. In addition, each user has a foreign key to a "Service", since all the members of the research staff are part of a service which, in its turn, is part of a department. Moreover, the first name and last name of the user will be stored. Figure 2 shows all the information stored within the user as a custom User.

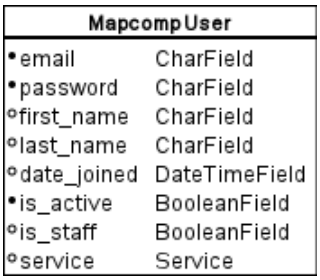

Figure 2. Database representation of the user

Because of the way Django handles databases, representing them in code as Python classes, a generic class for an internal structure which contains the name of the said structure can be defined. Then, this class can be extended to a department which, in addition to the name, has a code, and a service which also has a code and a foreign key to a department, as depicted in Figure 3. While the usefulness of having each employee assigned to a department within the database is limited when it comes to mapping competences, it can be invaluable when compiling statistics about the skill level of all the employees, as it allows team leaders to identify and address the present weaknesses.

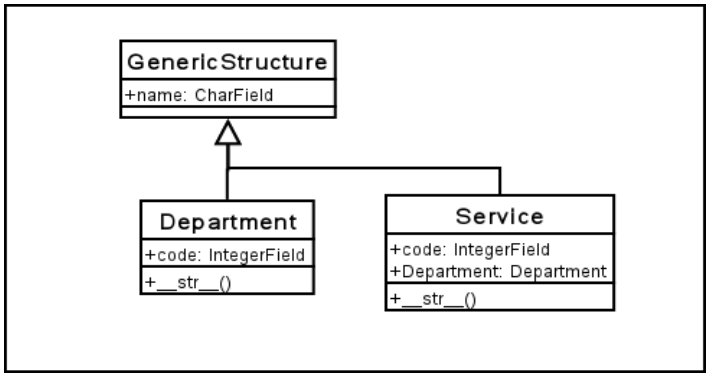

Figure 3. Generic structure class & the two classes inheriting and implementing it

Analogous to the internal structures, a generic competence is defined as a class containing the name of the competence in question and a text description which is flagged as being optional. For each type of competence, the generic class is extended to include more specific information.

For example, the programming competence also includes the users's expertise level, but that is not applicable to hardware competences, where the focus is on the number of units of a certain piece of hardware that the user has. The different types of competences and their respective pieces of information are shown in Figure 4.

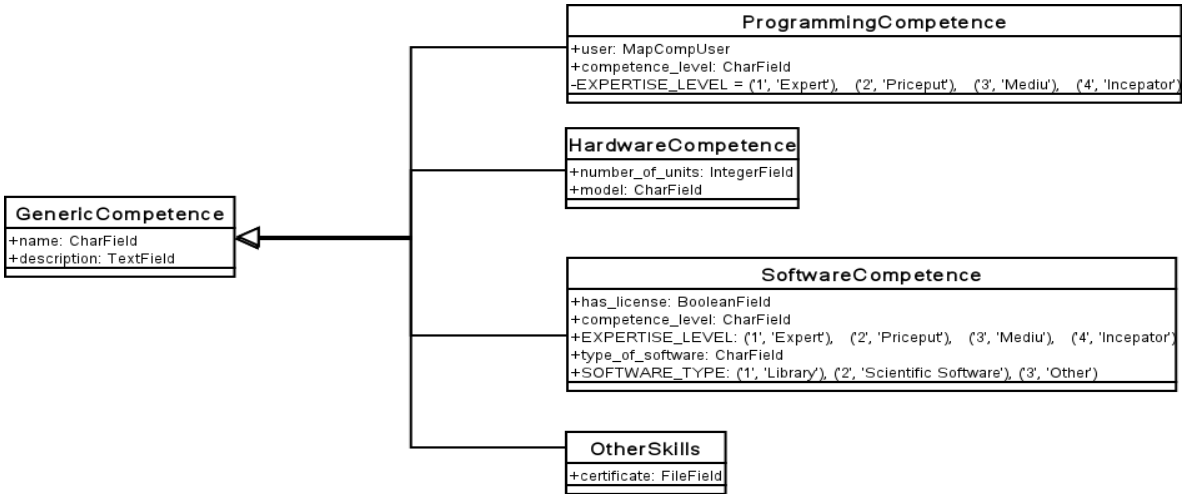

Figure 4. Generic competence class with different types of competences

3.3. General architecture

The MapComp platform was developed using the Django framework, programmed end-to-end using Python, using an MVC (Majeed & Rauf, 2018) architecture. Django was chosen as a base for developing this platform, due to the modularity of its code, resulting in fast development time.

By default, users defined by Django use a username and a password for logging in. Django also stores additional information such as the name and surname of the user, his/her email address and several Boolean values. In the case of MapComp, a custom user is defined to use the email for signing up and logging in. When creating an account, there is verification process to make sure that only ICI employees use the platform, by checking if the domain matches that of the Institute. This improves the employees' user experience, since there is no need to memorise another username. When creating an account, the other pieces of information that need to be provided are the employee's name and surname.

After creating an account, when first logging in, the user will be shown a wizard to input additional information about himself/herself (Figure 5). First, the department and service he/she is part of are introduced, then each individual competence. For each competence, the user first selects from a drop down menu the type of competence, which, in its turn, will unlock certain boxes to be completed. The only boxes that are universal for all the competences are the name and the description, with the other varying based on the user's selection. While this is a time consuming process, it is essential to be thorough in order to make sure that all the employees' skills are properly mapped and that the platform functions as intended.

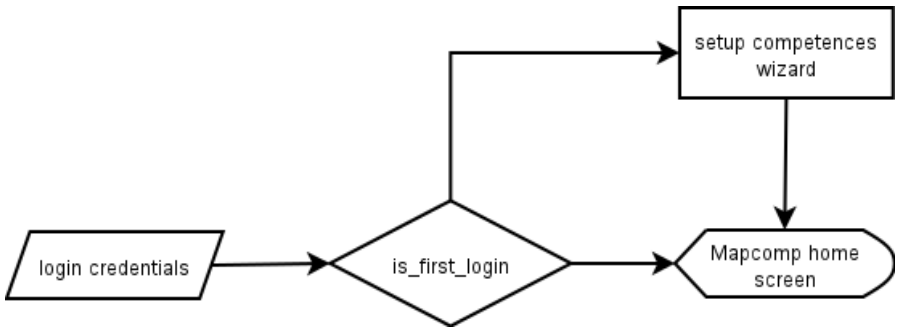

Figure 5. Flowchart of the initial login

Once the registration process is complete and all the skills have been added to the user's profile, the said user can now use the platform in order to find competences relevant to his/her tasks or projects. This can be done by creating a new query from the "create query menu". This prompts the user with a wizard similar to the one that greeted him/her when adding his/her own competences. After submitting it, all the employees who best match the given description will be shown, in order of their relevance. For each one, there is an option of contacting that person directly via email, in order to either ask for more details on a specific skill or to ask for help/advice.

In the likely event that a user acquires another skill after completing the initial registration, he/she can easily add it, by accessing the "add skill" in his/her user tab. He/she can also edit the existing skills, in case a mistake was made or if he/she has something to add to one of them.

Lastly, this platform also computes statistics about the level of knowledge existing within the research staff. This is done in order to monitor the employees' skill level and to find their strengths and weaknesses. On the landing page, after logging in, these statistics are randomly shown to the user. If the said user wants to view a certain statistic, he/she needs to go to the dedicated tab for this function. Then, he/she will fill out a form and wait for the system to output his/her specific request in a human readable format.

4. Competency mapping

The most important aspect of this platform is the competency mapping process. While for some human resources or hardware resources it can be easy to come up with a solution to find the right match for the user's query, with more complex/nuanced competences, the process becomes more difficult. This is the main drive behind the generic competence in the database having a description box. This will allow users to input more details about that specific skill, as well as extra details that can be given when inputting a query. This is especially useful when searching for a competence that can be used in different ways and in different fields. For example, knowing a certain programming language may not be enough information, as that programming languages tend to be used across multiple fields of computer science, from web development to data analysis or 3D graphics. Having an input box in which individual users provide details about their experience with the said competence, allows extra information to be encoded into that data. However, this description box is optional, as there are some scenarios where a competence does not need additional information.

The user querying the database does not have to input data in each field, but only fill in the relevant one for his/her inquiry. As a result, a prioritisation of the returned results has to be determined, which is achieved by ranking all the possible inputs. For some of them, such as name and expertise level, it's just simple matching. For others, such as the description of a competence mentioned above, there is a more complicated approach, in order to rank all the entries in the database.

In order to use the description box for competency mapping to it's fullest potential, a two step process is applied. First of all, this box can contain multiple unrelated sentences about the same competence name. In order to separate them, the user will use a new line between each phrase that is a different skill. Following that, the whole text will be split into those aforementioned new lines. In order to find whether any of these phrases matches the query, the semantically closest one has to be found. The easiest way to achieve this is to encode the meaning of the phrases into an array, in both the competence description and in the query. This method is used to find the closest phrases to the query, by employing the Universal Sentence Encoder by Tensorflow (Yang et al., 2021). This library takes the input and transforms it into a semantic vector with the embedding size of 512, as shown in Figure 6. There is no limit regarding the size of the input, but the longer the phrase is, the less information the semantic vector will hold. After transforming all the natural language input into vectors, it is trivial to calculate the similarity between them, to calculate a distance matrix between all the sentences and to keep track of which is a description and which is part of the query (Figure 7).

```
Message: Python
Embedding size: 512
Embedding: [-0.019093163311481476, -0.009827365167438984, -0.010161575861275196, ...]

Message: I used Python during university.
Embedding size: 512
Embedding: [0.02725698985159397, 0.027022460475564003, -0.02216530404984951, ...]

Message: I've been unsing Python for the past 15 years for web development using Django and Flask.
Embedding size: 512
Embedding: [0.015530381351709366, -0.01131885964423418, -0.019747929647564888, ...]
```

Figure 6. Example of three different sentences of varying lengths and their embeddings

Figure 7. Similarity between phrases

As seen in Figure 7, the first three are descriptions while the last two are queries. Note the similarity between the first description and first query as well as between the second description and second query.

The last step in competency mapping is sorting the relevant entries of the database, that have been previously retrieved. After converting all the text inputs into numerical data, it is now easier to operate on this data. By putting all the information into a table, the data can be sorted by using the most significant digit radix sort (Knuth, 1997). Only in this case, the most significant digit is the first column, the second most significant digit is the description and so on. The returned table is then formatted into a user-friendly list and displayed to the user in his/her browser page.

5. Conclusion

MapComp is an ambitious platform which aims to centralise the skills of all research staff within the Institute. It is built as a web application using Django, allowing a more pronounced agility in its development, and improving the ease of extending it further in the future. The data regarding the skills of all the research staff members has also been aggregated, allowing a thorough analysis of it and the identification of the areas where the Institute can improve. This project has highlighted some interesting limitations when it comes to human-computer interaction. It is easy for us, humans, to articulate our knowledge in a way that other humans can understand and match a set of skills with another set of skills. However, this task is not as straightforward for a computer. The challenge of this project has been twofold: finding the right representation for the data and finding a way of matching the said data. These two issues cannot be solved separately or one after the other, but rather at the same time, as they are not independent of one another. Furthermore, it is known that competence means the ability of an individual to achieve a task, but, for a computer, this term is very vague. Defining what competences are and how to efficiently store them is a laborious task that is domain-specific and scope-dependent.

Acknowledgement

The research reported in this work has been supported, in part, by the Ministry of Research, Innovation and Digitalization [project PN19370801 conducted at ICI Bucharest within the National Core Program].

REFERENCES

1. Băjenaru, L., Smeureanu, I. & Balog, A. (2016). An Ontology-Based E-Learning Framework for Healthcare Human Resource Management. *Studies in Informatics and Control,* 25(1), 99-108. DOI: 10.24846/v25i1y201611.
2. Django (2022). *Django Software Foundation.* Available at: https://www.djangoproject.com/foundation/ (Accessed: November 2022).
3. European Commission (2022). *Digital Economy and Society Index (DESI) 2022*. Available at: https://digital-strategy.ec.europa.eu/en/library/digital-economy-and-society-index-desi-2022 (Accessed: November 2022).
4. EUR-Lex (2018). *Council Recommendation of 22 May 2018 on key competences for lifelong learning.* Available at: https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32018H0604(01)&rid=7 (Accessed: November 2022).
5. Knuth, D. (1997). Sorting and Searching. Section 5.2.5: Sorting by Distribution, *The Art of Computer Programming*, volume 3, 3rd edition. 168-179. Addison-Wesley.
6. Majeed, A. & Rauf, I. (2018). MVC Architecture: A Detailed Insight to the Modern Web Applications Development, *Peer Review Journal of Solar & Photoenergy Systems,* 1(1), 1-7.
7. MDN Web Docs (2022a). *CSS basics*. Available at: https://developer.mozilla.org/en-US/docs/Learn/Getting_started_with_the_web/CSS_basics (Accessed: November 2022).
8. MDN Web Docs (2022b). *HTML basics*. Available at: https://developer.mozilla.org/en-US/docs/Learn/Getting_started_with_the_web/HTML_basics (Accessed: November 2022).
9. MDN Web Docs (2022c). *JavaScript basics*. Available at https://developer.mozilla.org/en-US/docs/Learn/Getting_started_with_the_web/JavaScript_basics (Accessed: November 2022).
10. Niculescu, C. (2006). Proiectarea unei platforme web dedicată centrelor de competență virtuale. *Revista Română de Informatică şi Automatică (Romanian Journal of Information Technology and Automatic Control)*, 16(1), 77-94.
11. Smada, D. & Cristescu, I. (2020). Assessing the needs for ICT skills and competences in public administration personnel. *Revista Română de Informatică şi Automatică (Romanian Journal of Information Technology and Automatic Control)*, 30(1), 49-60. DOI: 10.33436/v30i1y202004.

12. WebCase (2018). *Components of a Website*. Available at: https://webcase.studio/components-website/ (Accessed: November 2022).
13. Yang, Z., Yang, Y., Cer, D., Law, J. & Darve, E. (2021). Universal Sentence Representation Learning with Conditional Masked Language Model, *arXiv:2012.14388 [cs, CL]*. Available at: https://arxiv.org/abs/2012.14388 (Accessed: November 2022). DOI: 10.48550/arXiv.2012.14388.
14. Zola, A. (2022). *What is Bootstrap?*. Available at: https://www.techtarget.com/whatis/definition/bootstrap (Accessed: November 2022).

Daniela TIMISICĂ este absolventă a Facultății de Management din cadrul Academiei de Studii Economice din București. Are un masterat în Consultanță în Management (ASE București). În prezent, este Cercetător Științific în cadrul Institutului Național de Cercetare – Dezvoltare în Informatică – ICI București. Principalele sale domenii de interes și activitate sunt securitatea cibernetică, tehnologia Blockchain, IoT – Internet of Things și programare web. De asemenea, este interesată de studiul Inteligenței Artificiale, aplicarea tehnologiei 5G și implementarea acestora în noile ecosisteme digitale emergente.

Daniela TIMISICĂ is a graduate of the Faculty of Management of the Bucharest University of Economic Studies, and has a Master's degree in Management Consulting (ASE Bucharest). She is currently a Scientific Researcher at the National Institute for Research and Development in Informatics – ICI Bucharest. Her main areas of interest and activity are cybersecurity, Blockchain technology, IoT – Internet of Things and web development. She is also interested in the study of Artificial Intelligence, the application of 5G technology and their implementation in new emerging digital ecosystems.

Răzvan-Alexandru DUȚESCU a absolvit Universitatea din Manchester (Regatul Unit), cu o diplomă de licență în Informatică, și și-a finalizat masteratul în Inteligența Artificială, în cadrul Universității din Aberdeen (Regatul Unit), în 2019. Principalele arii de expertiză includ procesarea de imagini și învățarea automată nesupervizată. În prezent, lucrează la Institutul Național de Cercetare – Dezvoltare în Informatică – ICI București.

Răzvan-Alexandru DUȚESCU graduated from the University of Manchester (United Kingdom), with a BSc in Computer Science, and completed his MSc in Artificial Intelligence, at the University of Aberdeen (United Kingdom), in 2019. His main areas of expertise include image processing and unsupervised machine learning. He is currently working at the National Institute for Research & Development in Informatics – ICI Bucharest.