CareProfSys

Smart Career Profiler based on a Semantic Data Fusion Framework

Scientific report – phase II

Project financed from the state budget

PN-III-P1-1.1-TE-2021-1446

Contract NR. TE 151/ 2022

Start date: 13 mai 2022. Duration: 24 months

CONTENT

Executive summary

The CareProfSys project (http://careprofsys.upb.ro/) aims to develop, test and validate an intelligent career profile system by implementing it in an observant environment, the Career Counseling Center within the National University of Science and Polytechnic University of Technology Bucharest (UNSTPB). The created system will be able to provide career counseling using advanced user profile analysis, automatically extracted from various data sources. CareProfSys users will receive recommendations of professional occupations, based on this data, using ontological inferences from the "Classification of occupations in Romania" (COR) ontology (developed in the project), aligned with the European list of qualifications and classification algorithms specific to automatic learning. A conversational agent will provide personalized advice on recommended occupations and necessary steps for the future, while virtual 3D scenes will help users visualize activities connected to a recommended profession. The project has three phases: (1) Design of the system and development of the COR ontology – in 2022, (2) Development and testing of the system – in 2023, (3) Implementation of the CareProfSys system in the UPB-CCOC center – in 2024. Second phase is "System development and testing". The phase has only one main activity, but extremely important for the success of the project, "Development, testing and optimization of the CareProfSys system", with five specific tasks, described below.

The CareProfSys system has a tier-based architecture. The data required for the system is extracted through the web interface, the data being extracted from several sources, following user authentication: Europass CV, social media accounts, answers given to the questions in a form. A separate category of data is that required to create the user account. Login is done by implementing SpringBoot and Spring Security. All user information is saved in a MongoDB database. Data processing consisted in: eliminating contradictory information, coming from different data sources; identifying skills and interests useful for the recommendation process, according to the COR ontology; mapping data from natural language to structured data. The development of the services in the system consisted of the development of career recommendation services, access to scenarios in virtual reality on the Web (WebVR) and the conversational agent/chatbot CareerBot. Currently the recommendation mechanism contains two recommendation methods: ontological inference and recommendation by applying a machine learning algorithm. The professions recommended by both algorithms are offered as a result in the first positions, then, to allow the user to explore as many careers as possible, all other recommendations, originating from either of the two methods, are offered. Every time a user uses the CareProfSys system, answers questions and an electronic profile is attached to him in the system, an individual is also instantiated within the COR ontology, an individual who, with the help of the HermiT reasoner, will be classified as being of the type to a class that represents a profession from the COR ontology, which means that the respective user fits that profession. At the same time, using the K-Nearest Neighbors (KNN) algorithm from the sklearn Python library, we trained a machine learning model that allowed the recommendation of professions, based on 8 features extracted from filling out the form by the authenticated CareProfSys user. Since the development of 3D animated scenarios is not easy, we chose to develop scenarios for only six professions, within the project, to exemplify the concept of representing the recommended professions through VR, all of which have a lot of gamification elements: specialist in computer networks, civil, industrial and agricultural construction engineer, web and multimedia system designer, chemical engineer, university professor and similar, project manager. For the development of a virtual reality application that can be executed directly from a web browser, we used the game engine Unity Engine together with specific packages such as WebXR or VRTK Tilia. The virtual career advisor chatbot was developed using the Pandorabots platform and the AIML tag-based language. The Web platform is the user's access point to the CareProfSys system. For the frontend, HTML, CSS, Bootstrap, React technologies were used. In the implementation of the backend, a Maven project was created in which technologies such as Spring Boot, Apache Jena, OWL API were installed. During the development, we applied modular testing by developers, functional alpha and beta testing. The system was also tested by 48 university students and a group of 27 high school students from all over the country, participants in a summer school organized by UNTSPB. The feedback provided by the participants was obtained through questionnaires and was the starting point for the system optimization activities.

During the II phase, 1 deliverable was created: "Technical report on implementation and testing of the system". Project results were disseminated through 5 articles presented at international conferences (4 ISI), 1 oral presentation at a US conference, 1 updated ISI journal article from phase I, 5 new journal articles (of which 1 ISI Q2, 2 ISI , 2 BDI), 1 popularization intervention in the media, by updating the project's website, creating and popularizing the project's Facebook page, publishing relevant information on LinkedIn, participating in two relevant educational fairs.

Introduction

CareProfSys project description, objectives and phases

The project aims to provide career counseling using advanced user profile analysis, automatically extracted from various data sources. CareProfSys users will receive recommendations of professional occupations, based on this data, using ontological inferences from the "Classification of occupations in Romania" (COR) ontology (developed in the project), aligned with the European list of qualifications and classification algorithms specific to automatic learning. A conversational agent will provide personalized advice on recommended occupations and necessary steps for the future, while virtual 3D scenes will help users visualize activities connected to a recommended profession. Apart from the direct benefits brought to students and high school students, the system will bring indirect benefits to career counseling centers, higher education institutions, the government, employing companies. The system will integrate the latest technologies: semantic web and ontologies, machine learning, social network connectors, web virtual reality (WebVR), recommender or conversational agents and modern programming interfaces (APIs). The main objective of CareProfSys is to validate and test the concept of an intelligent career profile system by implementing it in an observant environment, UPB-CCOC – the career counseling center within the National University of Science and Technology POLITEHNICA Bucharest (UNSTPB).

The specific objectives of the project are: (a) building user profiles, in a fast and accurate manner - creating social connectors through current social network APIs, creating PDF extractors, web form processors and other modules for Web mining, combining and transforming all data sources into structured data, validating and analyzing this data with the help of expert psychologists; (b) creating the COR ontology and populating it with user profiles; (c) recommending occupations from the COR ontology to users (pupils, students and graduates who want to change careers), based on their profiles, through a customized algorithm; (d) providing additional information (based on text or 3D scenes) about recommended occupations; (e) providing support in advising users (students and pupils) through modern human-computer interaction tools, e.g. Web-VR based cameras or virtual cameras for interviews and conversational agents; (f) providing a flexible tool that can be easily customized to support career counselors and students in high schools and universities; (g) enabling the rapid development of several other modules in the future; (h) incorporating the product into a university ecology - pilot implementation in CCOC-UPB; (i) attracting potential users from other universities and high schools, through its intensive promotion; (j) providing an instrument of regional value.

The project has three phases: (1) Design of the system and development of the COR ontology - in 2022, (2) Development and testing of the system - in 2023, (3) Implementation of the CareProfSys system in the UPB-CCOC center - in 2024.

Objectives and activities of phase II

The second phase of the project consists in the Development and testing of the system and aims to achieve the specific objectives from (a) to (j). The phase has only one main activity, but extremely important for the success of the project: A2.1. Development, testing and

optimization of the CareProfSys system, with five concrete tasks, the results of which are described below.

Purpose of the document

This document contains the description of the scientific activities carried out within the second phase of the CareProfSys project, "Development and testing of the system", the results obtained, the ways of disseminating them and highlights the result indicators. The second phase had only one main activity, Activity 2.1. "Development, testing and optimization of the CareProfSys system", with several tasks, corresponding to specific objectives a-j.

Activity 2.1: Development, testing and optimization of the CareProfSys system

Within this activity, five main tasks were performed, with the aim of fulfilling objectives (a)-(j), corresponding to the "WP3: CareProfSys development, testing & optimization" package from the project proposal:

- T3.1. Development of the data extraction method
- T3.2. Development of the data processing module
- T3.3. Development of services in the system
- T3.4. Web platform development
- T3.5. Testing and optimization

Details about the scientific results of the activities carried out can be found in Deliverable 4 "Technical report on system implementation and testing", annex to this report, as follows: T3.1 in sub-chapter A "Data extraction and user authentication" of the chapter "Details on system development"; T3.2 in subchapter B "Data processing" of the same chapter; T3.3 in subchapter C "Development of the recommendation mechanism", subchapter D "Development and access to WebVR scenarios" and subchapter E "Development of the conversational agent"; T3.4 in subchapter E "Web Platform"; T3.5 in the chapter "System testing and optimization details". The results of the activities carried out at this phase were disseminated in three ISI journal articles (R6, R8, R9), two BDI journal articles (R5, R7), 6 international conferences (R10, R11, R12, R13, R14, R15) and 1 popularization intervention in the media (R16), all present in the Appendices of this report.

The proposed objectives for activities T3.1, T3.2, T3.3, T3.4 and T3.5 were 100% achieved. The system created through these activities is complex, having a tiered architecture (see Figure 1), communication between modules is done through HTTP requests.

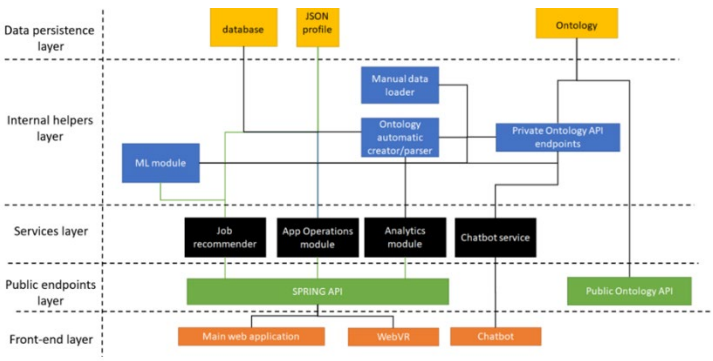

Figura 1. Architecture of the CareProfSys system

Development of the data extraction method

The data required for the system is extracted through the web interface, the data being extracted from several sources, following user authentication: Europass CV [1], social media accounts, answers given to the questions in a form. The questions in the form are similar to the Briggs Myers personality test [2], in that each user has to self-assess on a 5-point Likert scale regarding certain statements, e.g.: "I like to sit in an office when I work" , "I like to build and invent things." The statements also allow the scaling of the response for the 8 skill categories used by the ESCO matrices [3] to characterize the occupational profiles of the occupational standard The International Standard of Occupations (ISCO-08) [4]. For each type of data extraction, we made a controller. To extract data from Facebook and Instagram of the current user, we used Facebook Graph API [5] and Instagram Graph API [6], by implementing the OAuth 2.0 protocol [7] within the project, and to extract data from LinkedIn, we used web scraping . A separate category of data is that required to create the user account. Login is done by implementing SpringBoot and Spring Security [8]. All user information is saved in a MongoDB database [9].

Development of the data processing module

After extracting/gathering the data and building the user profile in JSON format, we will have both a static and a dynamic profile for an authenticated user. The static profile will contain the data from the CV and the test from the Web platform, and the dynamic profile will contain the data extracted through the social connectors, from LinkedIn, Facebook, Instragram. The phases of data processing are: elimination of contradictory information, coming from different data sources; the identification of useful skills and interests for the recommendation process, according to the COR ontology developed within the project, which contains all the professions from the Classification of Occupations in Romania [10]; mapping data from natural language to structured data.

Development of the services in the system

The main services are referral services, access to scenarios in WebVR and the conversational agent/chatbot CareerBot.

Development of the recommendation mechanism

Currently the recommendation mechanism contains two recommendation methods: ontological inference and recommendation by applying a machine learning algorithm. The professions recommended by both algorithms are offered as a result in the first positions, then, to allow the user to explore as many careers as possible, all other recommendations, originating from either of the two methods, are offered.

Both Apache Jena [11] and OWL API [12] are used in the implementation of the ontology-based recommendation module of CareProfSys, allowing the execution of inferences (automatic reasoning) based on the COR ontology, described in Deliverable 3 of the project. Every time a user uses the CareProfSys system, answers questions and an electronic profile is attached to the system, an individual is also instantiated within the ontology, of the Person type, which, with the help of the HermiT reasoner [13], will be classified as being of the type of a certain Job (which represents

a profession from the COR ontology), which means that that person fits that profession, taking into account the description of the person and the profession, e.g. skills, aptitudes, work style, occupational interests, important needs and values, etc. An example of the classification of a new user, careprofsys-user, within the Job "Chemical Engineer" can be seen in Figure 2 (image obtained with the help of the Protégé ontology editor [14]), which means that that user has received that profession as a recommendation.

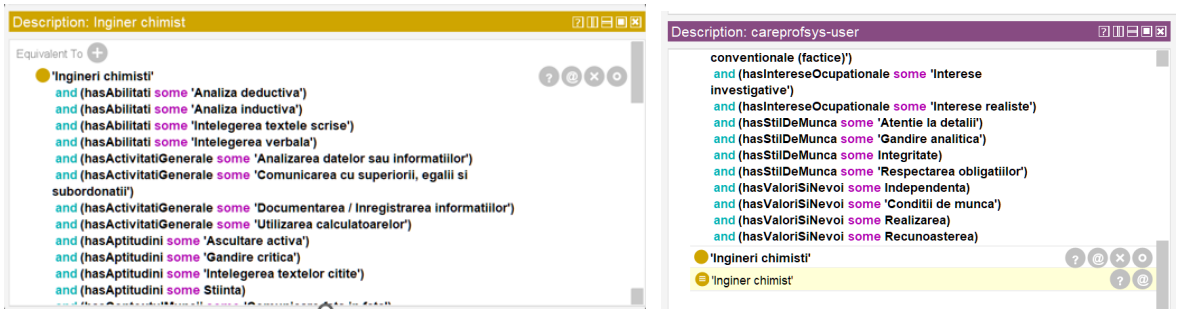

Figura 2. Career recommendations as inferred by the HermiT Reasoner within the Protégé ontology editor

HermiT is one of the reasoners used in ontology classification and outperforms others by using a mechanism based on a hypertable and other techniques to optimize and improve the inference computation, which should solve the problems that arise due to ontology dimensions [15] . The classification algorithm implemented by HermiT works by expanding the hypertable iteratively, with each new node and vertex representing the axioms of the ontology, until all the axioms have been found. In OWL, an axiom is the name given to restrictions or constraints about entities/concepts [16]. The classification algorithm of the HermiT reasoner starts by initializing the hypertable that will represent the entire OWL file. In this phase, the algorithm tries to avoid unnecessary tests so that the process is optimized [15]. For example, if a class A belongs to the superclass S and there are three other classes that are children of class A, then the algorithm decides that the three satisfy the constraint of A and also infers that they are members of S. Finally, this phase it should determine the initial knowledge of the relationships between the classes and be able to proceed to the next step, the automatic classification of an individual as a member of a class. As soon as the initialization phase is completed, the algorithm starts the classification. It iteratively expands and refines the relationships between classes, creating the class hierarchy based on all existing relationships [16]. The Reasoner is easily initialized by defining a new variable of the Reasoner class in Jena that takes the ontology itself as an attribute. The Reasoner will then apply the classification algorithm, check that it is valid, and then we can iterate through all the recommended classes of the individual for which we want to check the profiling results, saving each recommendation as a Job object.

After creating the user profile and creating the object with that profile, it is classified as being close to a maximum of 8 professions, according to the machine learning algorithm used, which has the following phases: data collection; data preprocessing and cleaning; feature extraction and engineering; machine learning model selection and training; evaluating and fine-tuning the machine learning model. For the training and validation of the machine learning algorithm, we used, as data sources: ESCO matrix tables [2] linking occupations from the International Standard of Occupations (ISCO-08) [3] to European ESCO competences and a newly completed Google Form survey by Romanian respondents (to give local color) and the LinkedIn type data collected in the first phase; each respondent was asked to declare their occupation

and to self-assess the skills necessary for the optimal exercise of that occupation, using the same eight characteristics used in the ESCO. According to the ESCO tables, each profession is characterized by 8 characteristics, which can have values between 0 and 1, depending on how important they are for the successful exercise of that profession. The eight characteristics considered important are: handling and movement; skills for working with information; working with the computer; building skills; managerial skills; working with specialized machinery and equipment; assistance and care; communication, collaboration and creativity. These features became the variables of the machine learning model. We used the K-Nearest Neighbors (KNN) algorithm from the sklearn Python library [17]: it predicts the values of new instances based on how close their features are to already known ones; therefore, professions are suggested based on how close their characteristics are to those in the user's profile. The model was trained with data from ESCO's professions and skills page and tested with data provided by our survey. The model's success was measured by how well it was able to discover jobs that match the preferences and skills of people who completed the form, with a current accuracy of 86.33%.

Developing and accessing WebVR scenarios

Once the names of suitable professions are received, following the recommendation mechanism, the user can try activities specific to the recommended professions by accessing some VR scenarios. Since the development of such scenarios is not easy, we chose to develop scenarios for only six professions, within the project, to exemplify the concept, all of which have a lot of gamification elements.

The scenario for Computer Network Specialist - COR 2523 is simulated in an office scene where the user has to replicate and configure various network schemes, using items such as PCs, servers, switches and routers. The virtual reality training process contains three levels of difficulty (easy, medium and hard), each of which has a different network to be replicated. Each level starts with a whiteboard (Figure 3) from which the user can choose the desired difficulty. Then the necessary buttons will be pressed to generate the correct number and type of elements according to the network diagram shown on the board. Items will automatically appear in the room, positioned as shown in the diagram. Various existing packages in Unity [18], customized for our needs, are used to speed up the implementation process. Such examples include a physical keyboard for entering IP and console commands (ping, ipconfig) - Figure 3. The user must press the appropriate keys using two sticks similar to those used for playing drums. Each character pressed by the corresponding key will appear on the screen in real time.

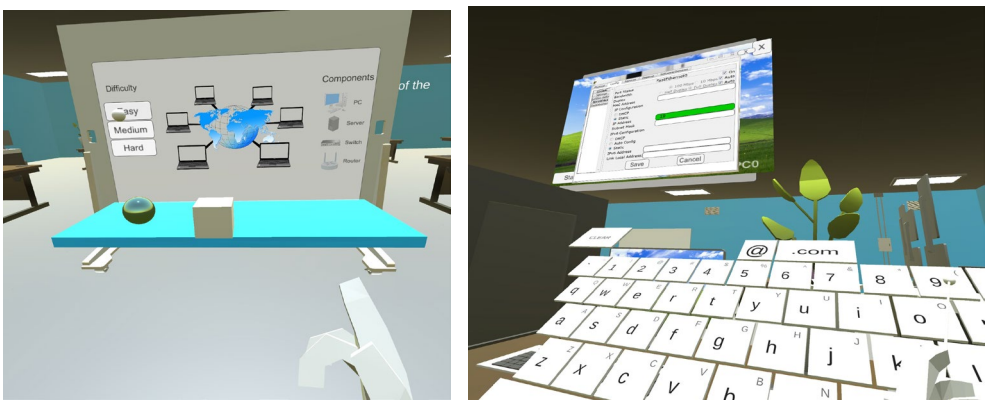

Figura 3. Scenes from the scenario for the profession "specialist in computer networks - COR 2523"

The following scenario replicated in virtual reality is intended to give an idea of some of the duties of a civil, industrial and agricultural construction engineer – COR 214201. The scenario focuses on the safety part of the workplace, within a working construction site. The goal of the "player" is to interact with as many workers as possible to activate certain interaction options that change the appearance of characters not properly equipped. This scenario is open-ended, allowing the player to interact with the workers in any order. The differences from one level to another consist in the increased number of actions that must be performed in the same time interval. Included in this scenario are 9 character types (see Figure 4), assigned to different positions to provide diversity. All workers are animated, having at least 3 types of animations actively used.

Figura 4. Characters from the scenes specific to the profession "Civil, industrial and agricultural construction engineer - COR 214201"

For the scenario involving the work of web and multimedia system designers - COR 2513, we chose to simulate the work on the front-end side, one of the possible duties of a web design specialist. The goal is to replicate the template provided in the left scene onto the artboard in the right scene in Figure 5.

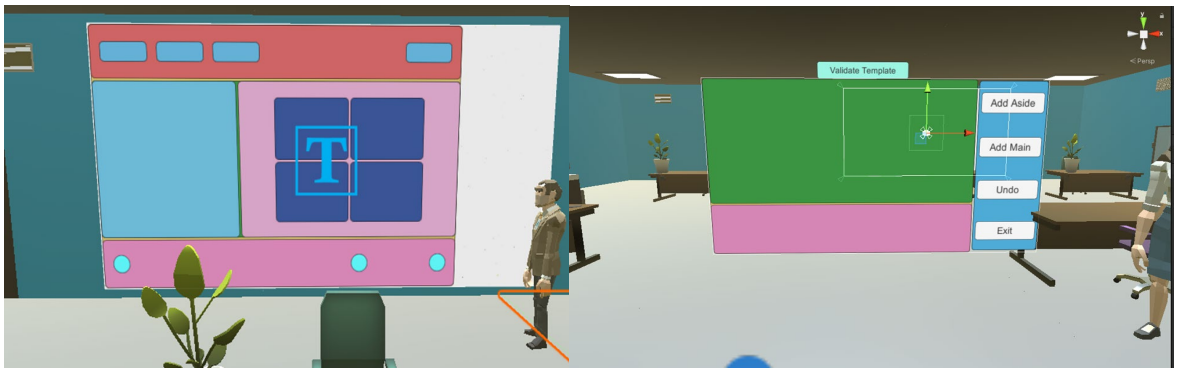

Figura 5. Plans for editing from the scenario assigned to the profession "Designers of web and multimedia systems - COR 2513

The Chemical Engineer Scenario - COR 2145 takes place in a hospital laboratory, where the user has to perform a series of chemical analyses, of varying complexity. Specific elements will be used, such as test tubes, pipettes, reagents, analyzer. The virtual reality training process includes 3 levels of difficulty (easy, medium, hard), each with different tasks to perform. On the wall of the laboratory is a panel on which the buttons for choosing the difficulty are displayed, as well as instructions related to the tasks to be performed performed (Figure 6). A particular element of this scene is the use of the particle system specific to the simulation of water flow. A particle system is used consisting of spheres (the drops) and including the "trail" that they leave, thus creating the sensation of leakage: see Figure 6.

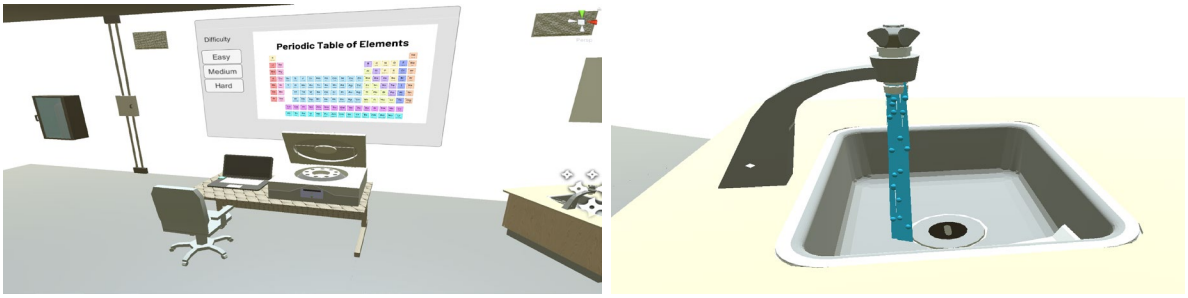

Figura 6. Scene din scenariul alocat profesiei "Inginer chimist - COR 2145"

The project manager scenario - COR 242101 is done in a conference room, the user has to make a Gantt chart, a Work Breakdown Structure(WBS) chart or both according to the chosen difficulty: see Figure 7.

Figura 7. Scenes from the scenario assigned to the profession "Project Manager - COR 242101"

The scenario for university and equivalent teachers - COR 2310 takes place in a classroom equipped with desks, a chair, computers, blackboard, projector: see Figure 8. Students are seated in desks and carry out different activities (programming, talking , raise your hand to ask questions). The teacher must select the action to be performed by the students in various situations, starting from being careful, to evacuating the room in case of fire, depending on the level. For the difficult level, an artificial intelligence script about the movement of students during the evacuation is used. Thus, the students have attached a NavMeshAgent type component that determines their automatic movement towards the door once the teacher is in his proximity.

Figura 8. Scenes from the scenario assigned to the profession "University and assimilated teachers - COR 2310 "

For the development of a virtual reality application that can be executed directly from a web browser, we used the game engine Unity Engine together with specific packages such as WebXR or VRTK Tilia. An application can be hosted on a web browser (see Figure 11) if the application's build type is WebGL. WebGL is a JavaScript API for rendering 3D graphics without the help of additional plugins. WebXR Exporter is a Unity package that allows the development of VR applications

in WebGL format, compatible with browsers such as Mozilla Firefox, Google Chrome, Microsoft Edge on Windows, Oculus Browser and Firefox Reality on Oculus Quest. Due to the format, WebGL, the application is compatible with several models of VR equipment, as tests were successfully carried out with HTC Vive Cosmos Elite, Oculus Rift and Meta Quest (1 and 2).

Development of the conversational agent

The CareerBot chatbot service can be accessed from the CareProfSys web platform to learn about COR professions: it serves as an experienced advisor to people seeking career guidance, useful for two distinct types of users, each with their own requirements specific. The first type targets aspiring learners, e.g. high school students or students who want to practice a job related to their field of study. The chatbot provides details on universities found in various cities across the country and admission requirements, helping users make informed choices about their educational path. The second type of users is those who want to make a career change. The provided conversational agent is an important source for these individuals, providing useful information on topics such as salaries and retraining requirements. In addition, it identifies firms that could hire people in the desired field and provides guidance on securing internships. Through relevant and useful guidance, the chatbot within CareProfSys supports these mature users in their professional development, as well as in the desired career change, in order to reach the professions recommended by the system. The chatbot has multilingual interaction capability, suitable for discussions in both Romanian and English.

The conversational agent built by us is based on the Pandorabots platform [19]: the user enters the chatbot platform, chooses the conversation language (English or Romanian) and engages in a conversation; the replies received from the bot conform to the rules previously created using the tag-based Artificial Intelligence Markup Language (AIML). A dialogue takes place according to the following flow: after entering some necessary information, the user asks questions, and the answers received from the chatbot are based on templates developed in AIML files on the basis of which the Pandorabots platform works. If the questions are not found in the templates, an exception is thrown and a friendly message is displayed asking the user to enter another question. Snippets of a possible user-chatbot discussion are available in Figure 9, as well as snippets of the AIML file behind the discussion.

The main AIML files of our chatbot are described below: “name.aiml” – contains the dialog flow for when a user introduces himself; it is also the part where the chatbot stores the user's name for later use; “career.aiml” – stores the dialog flow for when the user starts discussing a future career; "greetings.aiml" - is the beginning part where the user greets in English or Romanian; "salary.aiml" – stores information about salaries in Romania; “universities.aiml” – contains the dialogues for when the chatbot is asked about the option of universities and their requirements; “alreadyengineer.aiml” – the chatbot has information for users who are already working and want a career conversion; "generalinfos.aiml" - stores general information about professions; "ro.aiml" - has everything found in the files above, but translated into Romanian; all the files above (except the “ro.aiml” file) create the chatbot dialog flow in English.

AIML is based on pattern matching as the main mechanism and leverages the power of recursion. AIML files define templates to be found in the questions asked and the corresponding answers. These patterns act as triggers for the chatbot to identify user

inputs and appropriate responses are generated accordingly. AIML reproduces the natural writing style of humans. Recursion allows AIML to handle complicated conversational flows and generates responses dynamically. It empowers the chatbot to refer to its previous responses within a new response, leading to a continuous loop of pattern matching and template generation. This recursive approach improves AIML's ability to address various conversational scenarios while maintaining context throughout the interaction. Next, we exemplify some templates.

In Figure 9 we have fragments from the "greetings.aiml" file, where we frequently encounter the following labels or predefined tags: <category> - is a fundamental component of AIML; the main container where the model and corresponding templates are defined; contains the knowledge of the chatbot, it is used to understand a user's input and to provide an appropriate response; <pattern> - defines the input expected from a user; is the tag where, as the name suggests, pattern matching happens; the tag can include a keyword, expressions, metacharacters and their variations; <template> - contains the response that the chatbot will give if the value within the <pattern> tag is found in the user's reply; this tag can contain words, expressions, variables and conditionals; <srai> - is used within the <template> tag and supports recursion; recursive calls can be used in many ways, e.g. in Figure 9 it is used to respond to synonyms, "GREETINGS" being defined as a synonym of "HELLO", as you can see the result in the figure.

```
    </category>
<category>
    <pattern>HI</pattern>
    <template><srai>HELLO</srai></template>
</category>

<category>
    <pattern>HELLO THERE</pattern>
    <template><srai>HELLO</srai></template>
</category>

<category>
    <pattern>GREETINGS</pattern>
    <template><srai>HELLO</srai></template>
</category>
```

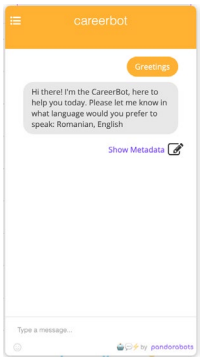

Figure 9. Fragments of the descriptive AIML files and Pandorabots interface for the CareerBot chatbot, within CareProfSys

For creation, testing and optimization, we used Pandorabots, but for the integration of the CareerBot chatbot into the CareProfSys system, we used the React JS and Node JS technology stack. Thus, we created a web interface to our chatbot (available in Figure 10), which can be accessed by the main CareProfSys web platform. To connect the web interface of the bot with the Pandorabots platform, we used a RESTful API.

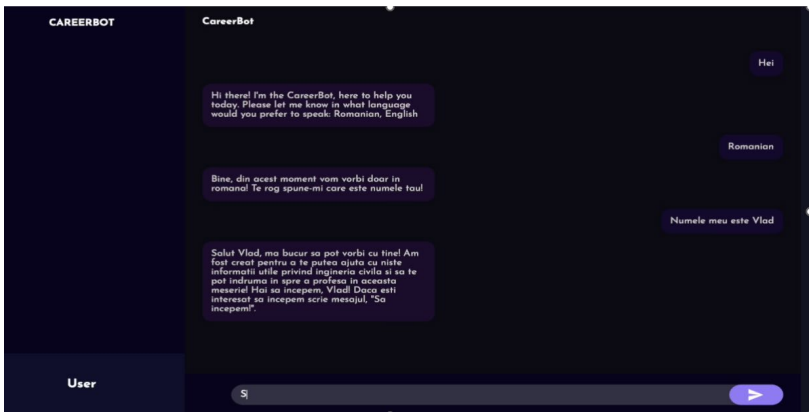

Figure 10. The CareerBot chatbot accessed from the Web platform of the CareProfSys system

Web platform development

The Web Platform is the user's access point to the CareProfSys system, which allows the user to create an account, complete a profile, get recommendations based on the profile, view WebVR scenarios for the recommended professions, and access the Career Bot chatbot for further advice (to see Figure 11). From a technical point of view, the platform follows a Model-View-Controller architecture (MVC).

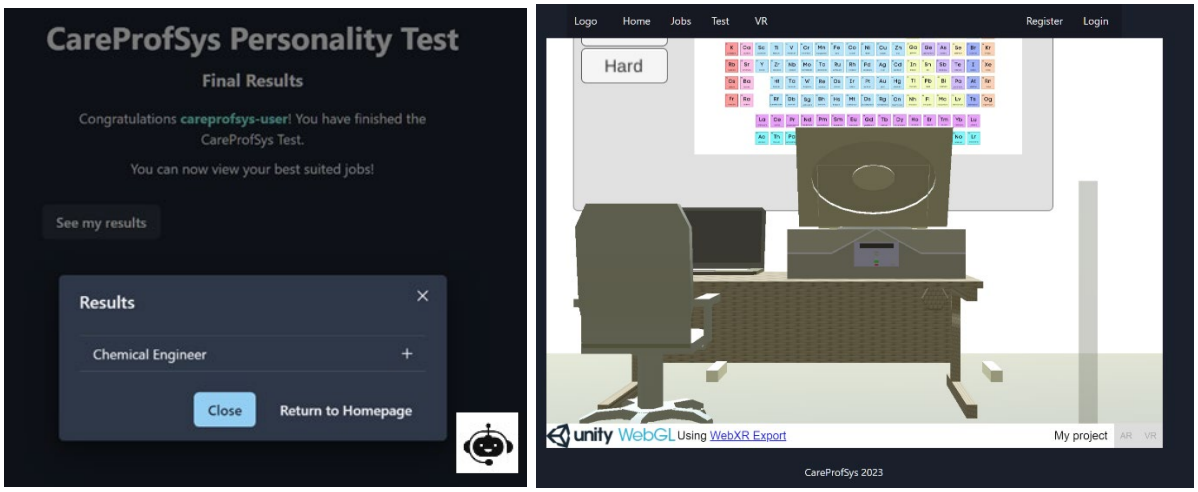

Figura 11. Screens from the CareProfSys Web platform

For the frontend, HTML, CSS, Bootstrap, React technologies were used. In the backend implementation, a Maven project was created in which technologies such as Spring Boot, Apache Jena, OWL API and HermiT reasoner were installed to enable application deployment. The project has been separated into different packages: the model, where the primitive classes are found; the service, where the implementation is written; the controller, where the RESTful interface is defined to allow data exchanges between backend and frontend, accessing the machine learning algorithm in Python, etc. As for the controller, it will create mappings for HTTP requests so that the ontology can be queried, managed and the classification accessible from a web interface.

Testing and optimization

During development, we applied modular testing, meaning each component was tested individually. In this regard, we used SwaggerUI [20] to test the various APIs in the system. To achieve rapid testing of WebVR environments, we included the OpenXR plugin [21] that allows testing directly from the Unity editor, without the need to create a web build after each change. We also maintain data persistence by using player preferences (PlayerPrefs) - a Unity module that ensures that small amounts of data relevant to game logic are saved. During development, we tested the chatbot directly in Pandorabots [19].

In addition to the testing performed by the main developers of the modules (after the implementation of each functionality), 6 students who did the practice in the university laboratories were involved in the alpha functional testing process. A black-box testing approach was thus used. The bug tracking process was facilitated by the use of the GitHub platform [22].

The system was tested by 48 students from the university (during 27/03/2023-20/10/2023) and a group of 27 high school students from all over the country, participants in a summer school organized by UNTSPB (during 15/07 /2023-30/07/2023), during beta functional testing. Both testing sessions took place in the laboratories of the CJ building in the Polytechnic Campus, being supervised by members of the project team, as can be seen in Figure 12.

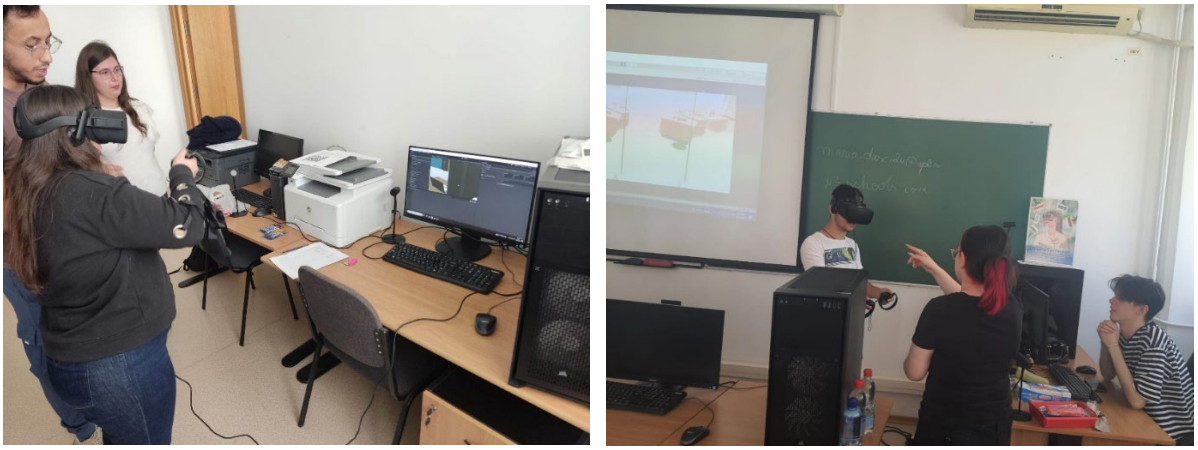

Figura 12. Functional testing of the CareProfSys system with students and high school students in the university laboratories

Testing with both groups had the following steps: (1) a member of the project team explained to the students, respectively, the idea of the project and how the system works; (2) the participants in the test sessions signed a document according to which they did not participate in another test session, they agree that the data obtained in the test sessions will be anonymized and will be used for research purposes, and that they have no known medical problems that prevent them from using VR technology; (3) participants used the system, still receiving help from the project team, in case of need; (4) participants were invited to complete a feedback questionnaire.

CareProfSys users found job recommendations to be a good match for their characteristics, including career preferences. They reported no difficulties in understanding the job recommendations provided by the system or any gender bias. They really liked the idea of experiencing the activities specific to the recommended professions in VR, as well as the fact that this experience was like a game. We noted that the scenarios in VR were accessible even to those who had not used this technology before. This is explained by the fact that our audience is young, very adaptable to new things. The scenario intended for network specialists is however difficult to navigate if users do not have prior knowledge, so we decided to simplify it as much as possible.

After testing, the VR scenes were especially optimized. One of the most complex scenarios, the networking one, has seen the most optimizations. Thus, interacting with canvas objects in the scene (containing IP addresses or other specific configuration elements of PCs or routers) was categorized as difficult by users, and since the same canvas object was used for each configuration, the scenario was prone to bugs (saving the state of some fields from one object to another). It was thus decided to modularize the approach and dynamically generate a new Canvas associated with each object to be configured. The virtual keyboard will be bound to it when the Canvas is generated and will facilitate typing input into the correct fields. The civil engineering scenario required some character interaction optimizations. The chemical engineering scenario is one of the most interactive, so it requires some dexterity/familiarity with the VR

system controllers. So that users do not find themselves unable to continue the level if they drop one of the objects (e.g. pipette, test tube), the automatic regeneration of a similar object at the moment of its contact with the floor was realized.

In terms of chatbots, a significant limitation of using AIML is its sensitivity to variations in customer input as well as typos. AIML relies heavily on predefined patterns and themes to recognize and react to specific individual questions or statements. To lessen this minus, we have defined as many synonyms as possible.

The results obtained regarding the recommendations were satisfactory. As threats to their validity, we note that it is difficult to calculate the real impact of the system in a short period of time and that the system is not yet mature, not all recommended professions had VR scenarios attached. Only in a few years will users be able to declare whether the recommendations received were really useful. However, the fact that they considered using the CareProfSys system a learning experience proves its usefulness for today's society. Indeed, in order to optimize the recommendation at the moment, we decided not to offer the professions among the results of both algorithms (the one of ontological inference and the one of automatic classification) as the result of the recommendation, but to offer the meeting of these results.

Summary of progress

Deliverable

During the II phase, 1 deliverable was created: Deliverable 4 - Technical report on the implementation and testing of the system, within the activity A 2.1, available in the annexes. The purpose of this deliverable is to describe the details regarding the development, testing and optimization of the CareProfSys system, which were based on Deliverables 1, 2 and 3, made in the previous phase. The document describes aspects regarding data extraction and processing, profession recommendation, using ontological inferences, but also a machine learning algorithm, aspects regarding the presentation of recommendations in the form of virtual reality scenarios and aspects regarding the development of a conversational agent that can support orientation in career. Both functional descriptions of the web platform introducing the recommendations, virtual reality scenes and chatbot to users, as well as detailed technical descriptions are provided.

Outcome indicators and dissemination of results

The result indicators of phase 2 are:

- • 1 delivered deliverable (available in the Appendices);
- • 1 ISI journal article made in the previous phase, updated, accepted and published in the current phase (available in the Appendices);
- • 5 new articles published or submitted for evaluation to journals, of which 2 articles published in ISI journals, 2 articles published in BDI journals (1 Scopus and 1 WorldCat), 1 article under evaluation in 1 ISI Q2 journal (all available in Appendices), exceeding the planned number of two articles submitted to journals for the current phase;

- • 6 participations in international conferences (see Figure 13), through 4 articles submitted and presented at international ISI conferences (2 IEEE Xplore indexed articles presented in Bucharest, 2 Springer Link articles presented in Västerås, Sweden, respectively Berlin, Germany), 1 article presented at an international Francophone conference in Bucharest and 1 oral presentation made at an international conference in the USA, published in the journal (all available in Appendices), exceeding the planned number of 3 articles sent to conferences for the current phase;
- • 1 popularization intervention in the media: the recording of CareProfSys from the show @UPB-Euronews Romania on the YouTube channel of the Faculty of Engineering in Foreign Languages, UNSTPB, https://www.youtube.com/watch?v=JwRik4zcUYk (description available in Annexes), exceeding the planned number of 0 popularization interventions in the media for the current phase.

Figura 13. CareProfSys at international conferences

The results were disseminated through papers published or to be published in the Proceedings of 5 prestigious international conferences (of which 4 Proceedings will be sent for evaluation and inclusion in Web of Science), 1 oral presentation at an international conference from the USA, 5 new articles published/under evaluation in journals (of which 3 are indexed by ISI, 1 Scopus, 1 WorldCat), 1 ISI journal article created in the previous phase, but published in the current phase, 1 popularization intervention in the media:

(R4) M.I. Dascalu, C.N.Bodea, I.V. Nemoianu, A.Hang, I.F. Puskás, I.C. Stanica, M. Dascalu, CareProfSys – AN ONTOLOGY FOR CAREER DEVELOPMENT IN ENGINEERING DESIGNED FOR THE ROMANIAN JOB MARKET, Rev. Roum. Sci. Techn.– Électrotechn. et Énerg. (RRST-EE), ISSN: 0035-4066, vol. 68 (2), WOS:001026628400016, DOI: https://doi.org/10.59277/RRST-EE.2023.68.2.16, pg. 212-217– diseminare rezultate din Activitatea 1.2, realizat și declarat în etapa I, actualizat și publicat în etapa a II-a, Article jurnal ISI

(R5) I.C. Stanica, I.A.Bratosin, D.A.Mitrea, C.N.Bodea, M.I. Dascalu, Electronic Profiling in CareProfSys System for Career Recommendation, Journal of Internet Social Networking & Virtual Communities, vol. 2023 (2023), Article ID 188953, 11 pages, ISSEN: 2166-0794, DOI: 10.5171/2023.188953 - diseminare rezultate din Activitatea 2.1, T3.1, Article jurnal BDI (WorldCat)

(R6) C.G. Dragomirescu, R.M. Ciuceanu, M.I. Dascalu, I.V. Nemoianu, THEORY OF CATASTROPHES REGARDING THE OPERATION OF A DC ELECTRIC MOTOR WITH SERIES EXCITATION, Rev. Roum. Sci. Techn.– Électrotechn. et Énerg. (RRST-EE), ISSN: 0035-4066, vol. 68 (1), WOS:000973414700017, DOI:

https://doi.org/10.59277/RRST-EE.2023.68.1.15, pg. 90-95 - diseminare rezultate din Activitatea 2.1, T3.2, Article jurnal ISI

(R7) M.I. Dascalu, A. Hang, I.F. Puskás, C.N. Bodea, CareProfSys : a job recommender system based on machine learning and ontology to support learners' employability at regional level, Issues in Information Systems, ISSN: 1529-7314, vol. 24(3), DOI: https://doi.org/10.48009/3_iis_2023_107, pg. 71-82 - diseminare rezultate din Activitatea 2.1, T3.2 și T3.3, Article jurnal BDI (Scopus)

(R8) M. Mitu, M. Dascalu, M.I. Dascalu, ROMANIAN TOPIC MODELING – AN EVALUATION OF PROBABILISTIC VERSUS TRANSFORMER-BASED TOPIC MODELING FOR DOMAIN CATEGORIZATION, Rev. Roum. Sci. Techn.– Électrotechn. et Énerg. (RRST-EE), ISSN: 0035-4066, vol. 68 (3), WOS:001087001200008, DOI: https://doi.org/10.59277/RRST-EE.2023.3.8, pg. 295-300 - diseminare rezultate din Activitatea 2.1, T3.2 și T3.3, Article jurnal ISI

(R9) C.N. Bodea, M. Paparic, R.I. Mogos, M.I. Dascalu, ARTIFICIAL INTELLIGENCE ADOPTION AT WORKPLACE AND ITS IMPACT ON THE UPSKILLING AND RESKILLING STRATEGIES: AN ETHNOGRAPHIC RESEARCH, Amfiteatru Economic, ISSN: 1582-9146, vol. 25 (65) (in curs de evaluare) - diseminare rezultate din Activitatea 2.1, T3.3 și T3.5, Article jurnal ISI Q2

(R10) I.V. Nemoianuc V. Manescu (Paltanea), Gh. Paltanea, M.I. Dascalu, R.M. Ciuceanu, Detailed Investigation of the Residual and Non-Symmetry Active and Reactive Power Flow for No-Neutral Three-Phase Nonlinear Circuits, The 13th International Symposium on ADVANCED TOPICS IN ELECTRICAL ENGINEERING (ATEE2023), IEEE, Bucharest (Romania), ISBN: 979-8-3503-3193-6, ISSN: 2159-3604, DOI: 10.1109/ATEE58038.2023.10108343, 23-24 March 2023 - diseminare rezultate din Activitatea 2.1, T3.2 și T3.3, Article conferință ISI

(R11) A. Hang, I. Puskas, M. Nitu, I.V. Nemoianu, M.I. Dascalu, CareProfSys Recommender for Modern Engineering Roles based on Emergent Technologies, The 13th International Symposium on ADVANCED TOPICS IN ELECTRICAL ENGINEERING (ATEE2023), IEEE, Bucharest (Romania), ISBN: 979-8-3503-3193-6, ISSN: 2159-3604, DOI: 10.1109/ATEE58038.2023.10108292, 23-24 March 2023 - diseminare rezultate din Activitatea 2.1, T3.1, T3.2 și T3.3, Article conferință ISI

(R12) M.I. Dascalu, A. Hang, I.F. Puskás, C.N. Bodea, CareProfSys : a job recommender system based on machine learning and ontology to support learners' employability at regional level, 63th IACIS Annual Conference, Clearwater Beach, Florida, SUA, 4-7 October 2023 - diseminare rezultate din Activitatea 2.1, T3.2 și T3.3, prezentare orala conferință

(R13) M.I. Dascalu, R. Birzaneanu, C.N. Bodea, An Ontology-based Recommendation Module for Optimal Career Choices, Proceedings of 2024 Future of Information and Communication Conference (FICC), Springer series "Lecture Notes in Networks and Systems", Future of Information and Communication Conference (FICC), Berlin (Germany), 4-5 April 2024 - diseminare rezultate din Activitatea 2.1, T3.2, T3.3, T3.4 și T3.5, Article conferință ISI

(R14) M.I. Dascalu, A.S. Bumbacea, I.A. Bratosin, I.C. Stanica, C.N. Bodea, CareProfSys - Combining Machine Learning and Virtual Reality to Build an Attractive Job Recommender System for Youth: Technical Details and Experimental Data, Engineering of Computer-Based Systems. ECBS 2023. Lecture Notes in Computer Science, vol 14390, ISBN: 978-3-031-49251-8, DOI: https://doi.org/10.1007/978, pg.

289–298, ECBS 2023: 8th International Conference on Engineering of Computer-based Systems, Västerås (Sweden), 16-19 Oct. 2023 - diseminare rezultate din Activitatea 2.1, T3.2, T3.3, T3.4 și T3.5, Article conferință ISI

(R15) M.I. Dascalu, V.A. Brîndușescu, I.C. Stanica, B.I. Uta, I.A. Bratosin, A. Mitrea, CHATBOT CAREPROFSYS POUR SOUTENIR LE CONSEIL ET L'ORIENTATION VERS UNE PROFESSION DANS UN DOMAINE D'INGENIERIE, La 2ème édition du Symposium de la recherche scientifique francophone en Europe centrale et orientale, Bucharest (Romania), 27-28 Nov. 2023 - diseminare rezultate din Activitatea 2.1, T3.3 și T3.5, Article conferință

(R16) Înregistrarea CareProfSys de la emisiunea @UPB-Euronews Romania pe canalul YouTube al Facultatii de Inginerie in Limbi Straine, UNSTPB, https://www.youtube.com/watch?v=JwRik4zcUYk - diseminare rezultate din Activitatea 2.1, T3.3, intervenție media

All articles contain project acknowledgment.

At the same time, the project was disseminated through its website (see Figure 14), continuously updated and available in two languages (English and Romanian), to ensure the visibility of the results (http://careprofsys.upb.ro/), by creating a social media account (Facebook) for the project (https://www.facebook.com/CareProfSys.UPB), publishing news related to the project on the social media profile (LinkedIn) of the project manager (https://www.linkedin.com/in/mariaiulianadascalu/), participation in the educational fairs POLIFEST 2023 and ROBOFEST 2023, organized by UNSTPB (see Figure 15).

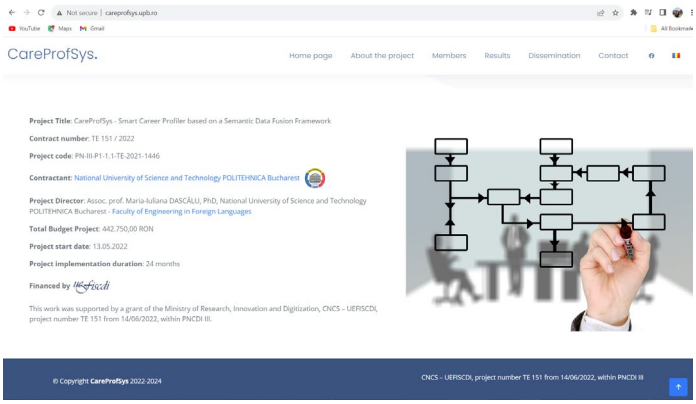

Figura 14. Website project CareProfSys: http://careprofsys.upb.ro/

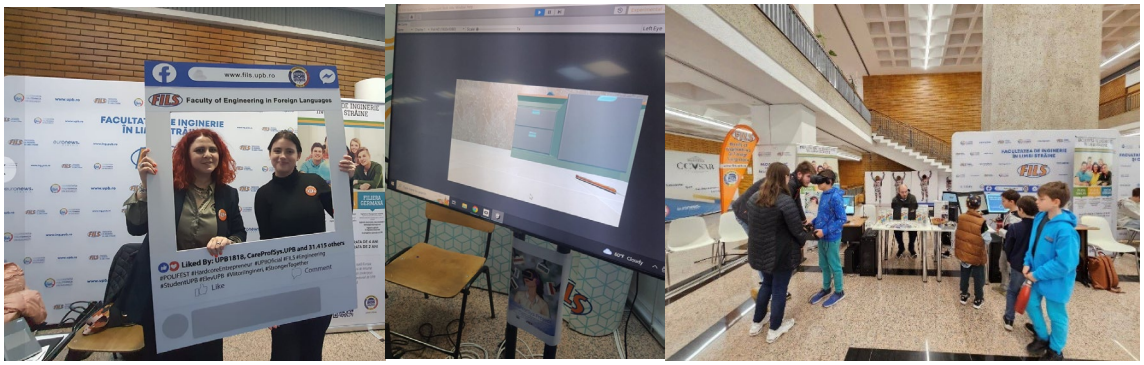

Figura 15. CareProfSys at UNSTPB's POLIFEST 2023 and ROBOFEST 2023 education fairs

Conclusion

The main purpose of any career guidance process is to support the client to identify and weigh options regarding the most suitable career path. The CareProfSys system aims to be a support for young people (high school students, students, professionals who want a professional reconversion) in finding the ideal profession and at the same time a help for counselors in career guidance centers. CareProfSys exploits today's emerging technologies, trying to successfully integrate them: artificial intelligence, machine learning, ontologies, virtual reality on the Web, social web, chatbot technologies. The system architecture is tiered and highly modular, heavily based on APIs. This report describes the main technical and functional details of the system, namely: extraction of user data from different sources and their authentication, data processing within CareProfSys, development of the recommendation mechanism, development and access to WebVR scenarios, development of the CareerBot conversational agent and the Web platform, which is users' access point to the system. Other aspects described in the report are: technical difficulties and how they were resolved, but also early positive results from functional tests and how they were used to optimize the system. The deliverable made at this phase, through the unitary way of presenting the development and testing details, through the clear functional and technical description, through the suggestive images, the easy-to-follow and understand tables, through the architecture and workflow diagrams, provides an eloquent picture of of the CareProfSys system and reflects the research activities carried out in 2023. At the same time, the dissemination activities and numerous results obtained this year within the project prove that the system is well received and considered useful by the scientific world and society in general.

ACKNOWLEDGMENT

This work was supported by a grant of the Ministry of Research, Innovation and Digitization, CNCS–UEFISCDI, project number TE 151 from 14/06/2022, within PNCDI III: “Smart Career Profiler based on a Semantic Data Fusion Framework”.

Bibliography

1. CV Europass, http://www.europass.eu/1.0
2. Test de personalitate Briggs Myers, https://www.truity.com/test/type-finder-personality-test-new
3. Tabele ESCO profesii-competențe, https://esco.ec.europa.eu/en/about-esco/data-science-and-esco/esco-skill-occupation-matrix-Tabeluls-linking-occupation-and-skill-groups
4. ISCO - International Standard Classification of Occupations https://www.ilo.org/public/english/bureau/stat/isco/
5. Facebook Graph API, https://developers.facebook.com/docs/graph-api/get-started
6. Instagram Graph API, https://developers.facebook.com/docs/instagram-api/
7. Mecanism de autentificare OAUTH2, https://oauth.net/2/
8. Spring Boot, https://www.bezkoder.com/spring-boot-jwt-auth-mongodb/
9. MongoDB, https://www.mongodb.com/
10. COR – Clasificarea ocupațiilor din România, https://mmuncii.ro/j33/index.php/ro/2014-domenii/munca/c-o-r
11. Apache Jena - Getting started with Apache Jena, https://jena.apache.org/getting_started/index.html, last accessed 2023/06/18
12. Horridge, M., Bechhofer, S.: The OWL API: A Java API for OWL Ontologies. Semantic Web Journal (2010), https://www.semantic-web-journal.net/content/owl-api-java-api-owl-ontologies
13. Glimm, B., Horrocks, I., Motik, B. et al. HermiT: An OWL 2 Reasoner. J Autom Reasoning 53, 245–269 (2014). https://doi.org/10.1007/s10817-014-9305-1
14. Protégé, https://protege.stanford.edu/
15. Glimm, B., Horrocks, I., Motik, B. et al. HermiT: An OWL 2 Reasoner. J Autom Reasoning 53, 245–269 (2014). https://doi.org/10.1007/s10817-014-9305-1

16. Horridge, M., Bechhofer, S.: The OWL API: A Java API for OWL Ontologies. Semantic Web Journal (2010), https://www.semantic-web-journal.net/content/owl-api-java-api-owl-ontologies
17. SCIKIT-Învățare automata în Python, https://scikit-learn.org/stable/
18. Unity, https://docs.unity3d.com/Packages/com.unity.entities@0.2/manual/ecs_core.html
19. Pandorabots, https://home.Pandorabots.com/home.html
20. SwaggerUI, https://swagger.io/tools/swagger-ui/
21. OpenXR, https://www.khronos.org/openxr/
22. GitHub, https://github.com/

Annexes

Annex 1. Deliverable 4 - Technical report on implementation and testing of the system

Annex 2. R4- Article in extenso updated from phase I – ISI Rev. Roum journal. Sci. Techn.– Électrotechn. et énerg.

Annex 3. R5 - Article in extenso – BDI journal Journal of Internet Social Networking & Virtual Communities

Annex 4. R6 - Article in extenso – ISI journal Rev. Roum. Sci. Techn.– Électrotechn. et Énerg.

Annex 5. R7 - Article in extenso –BDI journal Issues in Information Systems

Annex 6. R8 - Article in extenso –ISI journal Rev. Roum. Sci. Techn.– Électrotechn. et Énerg.

Annex 7. R9 - Article in extenso –ISI Q2 journal Amfiteatru economic

Annex 8. R10 - Article in extenso – conference IEEE ATEE

Annex 9. R11 - Article in extenso – conference IEEE ATEE

Annex 10. R12 – Prezentare orală – conference IACIS

Annex 11. R13 - Article in extenso – conference FICC

Annex 12. R14 - Article in extenso – conference ECBS

Annex 13. R15- Article in extenso – conference RSF

Annex 14. R16- Popularization article in media

PROJECT DIRECTOR,

DASCĂLU, MARIA-IULIANA