CareProfSys

Smart Career Profiler based on a Semantic Data Fusion Framework

Scientific report – phase I

Project financed from the state budget

PN-III-P1-1.1-TE-2021

Contract NR. TE 151/ 2022

Start date: 13 mai 2022. Duration: 24 months

CONTENT

Executive Summary

The main goal of **CareProfSys** project (http://careprofsys.upb.ro/) is the validation and testing of the intelligent career profile system concept by implementing it in an observed environment - the career development centre within the University POLITEHNICA of Bucharest. CareProfSys aims to provide career counselling to students and pupils using advanced analyses of user profiles, automatically extracted from various data sources, through occupation recommendations for people with similar profiles using ontological inferences and learning classification algorithms. The project has **three phases**: (1) System design and COR ontology development – in 2022, (2) System development and testing in 2023, (3) Implementation of the CareProfSys at the UPB-CCOC center in 2024. **The first phase** of the project consists of two main activities: (1) Design and technology choices, (2) Development of the COR ontology and exploitation tools.

During the **"Design and technology choices"** activity, an analysis was conducted to establish the functionalities of CareProfSys, which included detailed study of the literature and existing similar systems, as well as creating online surveys with 317 high school and university students to determine their interests. The results were then materialized into defining use cases and user requirements (visitors, authenticated users – pupils and students, and system administrators). Both traditional counseling practices (psychological or personality tests, one-on-one communication methods between counselor and client, etc.) and modern practices (use of recommendation systems based on content filtering, collaborative filtering, machine learning algorithms, ontologies) were identified. In Romania's educational system, the number of school/university counselors for career guidance is insufficient for the current need, thus new, modern solutions that can improve the counseling and career guidance process are necessary, and the survey responses highlighted young users' trust in automatic recommendation systems and social networks. After analyzing user requirements, we concluded that there are 2 main purposes for using the CareProfSys: (1) to receive personalized career recommendations by students or pupils; (2) to discover if a recommended career suits the user through virtual reality mode and chatbot, but only after a recommendation is provided. The functional requirements that need to be implemented are those that serve the above purposes in a user-friendly manner. Technically, a "microservices" type architecture will be used, where each module will run as a microservice in its own container, communicating with others through HTTP requests: data persistence layer, internal execution layer, services layer, public *endpoints* layer, *front-end* layer. In the design activity, technologies, algorithms were chosen, and software components were designed within the architecture. For the actual recommendation, collaborative filtering models as well as machine learning models (e.g., a transformer model - BERT, an unsupervised learning model - K-means) will be used.

In the **"Development of the COR ontology and exploitation tools"** activity, the following were analyzed: professional identity in industry 4.0, relevant professional bodies, taxonomies, and standards, the "Classification of Occupations in Romania (COR), examples of similar software artifacts - ontologies. COR represents a system for identifying and coding Romanian occupations regardless of their type and location, following the structure of the International Standards of Occupations. The purpose of the COR ontology is to describe the professional competencies required for an individual to access a job in the labor market. The taxonomy of occupations in Romania is the main pillar of the ontology. In addition to the occupation pillar, the ontology contains three more pillars: a) fields of study, thus connecting education with professions, b) characteristics related to occupation (general activities, work context, work style, values, and needs), and c) characteristics necessary to fulfill a certain position (abilities, aptitudes, and interests). The developed framework for exploiting the ontology is in the form of an API that will serve both the public users through public endpoints and the other microservices of the CareProfSys system through private endpoints. The public methods exposed by the API include: extracting the code of a specific job from COR; extracting all jobs from COR, along with their respective codes. The private methods include: adding a new triplet in the ontology, modifying a specific triplet in the ontology, deleting a specific triplet in the ontology, extracting all details related to a specific occupation.

In phase I, **3 deliverables** were produced: (1) Analysis report and user requirements of the system; (2) Report on the software requirements of the system and its design; (3) COR ontology and its description.

The project results were disseminated through 3 articles published in the Proceedings of international conferences (which will be ISI indexed) and **1 article in an ISI journal**, through the creation of the project website, and by publishing relevant information on ResearchGate.

Introduction

CareProfSys project description, objectives and Phases

The project proposes to provide career counseling using advanced user profile analysis, automatically extracted from various data sources: school record, CV, social media profiles or responses from web forms. CareProfSys users will receive job recommendations of people with similar profiles, using ontological inference and machine learning classification algorithms. The profile database will be completed with information about professionals who have occupations from the "Classification of occupations in Romania" (COR) ontology (developed by the project), aligned with the European list of qualifications. A conversational agent will provide personalized advice on recommended occupations, while virtual 3D scenes will help users visualize activities connected to a future profession. Apart from students and high school students, the system will also provide benefits to career counseling centers, higher education institutions, government, employing companies. The system will exploit the latest technologies: semantic web, ontologies, machine learning, social network connectors, virtual reality (VR) on the web, recommendation proposal and modern programming interfaces dedicated to big data analysis. Thus, the concept of an intelligent system for building the career profile will be validated and tested by implementing it in an observed environment - the career development center of a Romanian university. The main objective of CarProfSys is to validate and test the concept of an Intelligent Career Profile System by implementing it in an observed environment - CCOC UPB. The system is based on the analysis of several data sources (social networks, educational history, psychological tests), which are input for the algorithm for recommending the use of the ontology "Classification of Occupations in Romania" (COR) (developed within the project).

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.

Phase I objectives and activities

The first Phase of the project consists of the System Design and development of the COR ontology and aims to fulfill the specific objectives from (a) to (j). The Phase has two main activities: A1.1. Design and technological choices and A 1.2. Development of the COR ontology and exploitation tools, each with concrete tasks, the results of which are described below.

Purpose of the document

This document contains the description of the scientific activities carried out in the framework of Phase 1 of the CareProfSys project, the results obtained, the ways of disseminating them and outlines the result indicators.

Activity 1.1: Design and technology choices

Within this activity, **six main tasks** were performed, **with the aim of achieving objectives (a)-(j):**

- T1.1. Literature review on career guidance practices and support tools
- T1.2. Online interviews and surveys: conducting and analyzing the results
- T1.3. Identifying system user requirements
- T1.4. Selection of technologies and development of the necessary algorithms
- T1.5. Identifying system software requirements
- T1.6. Detailed system design

Details about the scientific results of activities T1.1, T1.2 and T1.3 can be found in Deliverable 1 "Analysis report and system user requirements", and their results were disseminated in the conference article (R1). The proposed objectives for activities T1.1, T1.2 and T1.3 were 100% achieved.

Details of the scientific results of activities T1.4, T1.5 and T1.6 can be found in Deliverable 2 "Report on the system software requirements and its design", and their results have been disseminated in the conference papers (R1), (R3) and the journal article (R4). The proposed objectives for activities T1.4, T1.5 and T1.6 were 100% achieved.

The analysis carried out to establish the functionalities of the CareProfSys system included both a thorough study of the specialized literature and existing similar systems, as well as the creation of online surveys with high school students and students (the main actors of the system) to determine their interests. The results obtained were then concretized in the definition of use cases and user requirements.

Literature review on career guidance practices and support tools

The literature review on career guidance practices and support tools consisted of two parts, depending on the degree of innovation of the existing work counselling and referral approaches - traditional practices and modern practices, respectively.

In **traditional practices**, the advisor-client relationship is paramount. Direct communication between them is a specific feature of the counseling process and depends on several characteristics such as the counselor's attitude, the characteristics of the interaction environment, the counselor's specific communication and support skills. Career guidance is a one-to-one professional relationship between a specialist and another person seeking specific help/support. The counseling process aims to provide the client with the opportunity to explore, discover and clarify ways of living their own life, creating a solid sense of well-being [1]. In the educational system in

Romania, the number of school/university counselors for career guidance is insufficient for the current need (one counselor ends up being responsible for 2400 students, according to the Romanian Ministry of Education) [2], so new solutions are needed, modern, which can improve the counseling and career guidance process. Traditional counseling tools include methods of collecting data about the client (e.g. psychological or personality tests, such as the Holland Questionnaire [3], MBTI [4], INEM [5], etc.), communication methods, methods of investigating labor market, personal branding, career planning methods [6].

Modern practices center on recommender systems, two main types being common: recommendations based on content filtering and recommendations based on collaborative filtering [7]. Systems based on content filtering use "text filters and feature similarity metrics" to users in order to provide the most suitable recommendations. Matching existing occupations and user profiles with different probabilities is the basis of this approach [7]. Recommender systems based on collaborative filtering are based on the user's preferences or those of similar users (jobs that appeared in their search history or similar) [8]. There are also advanced approaches, for example based on machine learning algorithms. Neural networks are often used to capture a candidate's preferences for specific jobs as they evolve/change over time [9] [10]. Supervised and unsupervised learning are also used in research. K-means is one of the simplest and most popular algorithms for unsupervised learning, which aims to divide data into groups (so-called clusters) based on their common characteristics [11]. In [12], user data is collected through surveys and includes items such as skills, expected salary, and desired geographic location, while job data is collected from social media posts. The aforementioned characteristics are given certain weights and are used to create job vacancy groups. Supervised learning refers to labeled data that can be used in machine learning to classify or predict the outcome. Decision trees are used in [13] to select meaningful attributes of job seekers. Information is searched recursively, starting at the root and working up to a leaf of the tree, grouping the information into classes. Among the criteria included in the decision trees are school results (high, average, low grades), age group, gender, studies (bachelor's, master's, doctorate), specialization, experience. One of the challenges facing personalized job recommendation systems is finding an appropriate format for representing the data required for recommendations. A commonly used method is represented by ontologies [14][15] which can formally represent data through classes, attributes, rules and relationships.

Online interviews and surveys: conducting and analyzing the results

In order to establish the relevant principles and functionalities for an occupation recommendation system, we conducted two surveys using the Google Forms platform, available in Deliverable 2. The target group included both high school students (for the first survey) and university students (for the second survey), with a total of 317 respondents, of which 209 high school students and 108 students. The surveys are similar for both categories, including three main sections: 1. socio-demographic characteristics and studies, 2. principles of choosing the current educational institution, and 3. support tools for receiving job/profession recommendations. Students have an optional additional section related to work experience, which is only completed by those who are currently or previously employed. Most of the questions include the possibility to give an additional answer (open answer) outside the predefined options, in order to get as much information as possible and leave some freedom to the respondents.

When asked to identify two main criteria for choosing their current educational institution, both pupils and students have similar top preferences, but in a different order - their preferred subjects of study (45.5% for pupils, 37% for students) and future job (36.4% for students, 43.5% for students). For students, the focus is on the possible employment opportunities that the college can offer (50%), an option that would not have been viable for high school students regarding the choice of high school. Both high school students and college students have similar preferences regarding the people they asked for advice when choosing their educational institution - parents, friends and teachers. As students age, they tend to be more independent and able to make their own decisions - 28.7% of university respondents said they looked online for information about the college they intended to attend. School counsellors are not popular among young respondents, they were consulted by no high school students and only 2 students.

Important information to consider for a good college/career choice should be based on interests, skills, and personality, in that order, for both high school and college students. Interests (what a person likes to do) dominate the statistics, chosen by 82.3% of high school students and 73.1% of college students. College students are more trusting of automated platforms than students, likely due to their more frequent study and interaction with technology, while trust in advisors decreases with age. There is a very low percentage of people who would not want to have access to the information needed for a good college/career recommendation (3.3% for high school and 2.8% for university).

Support tools seem to be considered useful by both students and high school students, and their perception of them is similar in several respects. When asked what information they consider useful from a job recommendation platform, the top 4 responses were identical for both categories, with the same order and similar percentages (1. Suitable job recommendations, 2. Information related to suitable schools or courses, 3. Job descriptions, 4. Information related to personality traits). The high trust of young people in technology is again underlined, as only 3.3% of high school students and 0.9% of students would not want to use a job recommendation platform. Another interesting aspect again shows the similarity of respondents' perception despite the age difference, this time related to the sources they consider relevant for extracting the information needed for job recommendations. The main choices include, with their percentages for high school and university level, resumes (62.7%, 70.4%), personality tests (59.3%, 66.7%), intelligence tests (40.7 %, 53.7%), emotional intelligence tests (40.7%, 51.9%). Students tend to view grades as less relevant these days (27.8% for female students vs. 18.5% for female students).

In terms of social networks, Instagram is by far the most popular network for all respondents (90% use it at least once a week), followed by TikTok for high school students (63.2%) and Facebook for college students (61 ,1%). In terms of usefulness for job referrals, Instagram remains the top choice for high school students (60.8%), but is replaced by LinkedIn for college students (65.7%). LinkedIn is the most professional social media platform, so it seems to be the obvious choice, but it is not very familiar to young people, hence the lower percentage of high school students who consider it appropriate (27.8%).

When asked to rate from 1 (least important) to 5 (most important) the features of a job referral platform, all respondents rated "Facilitating the identification of the right direction in choosing a school or profession" as the more relevant characteristic, with an average score of 3.92 and 3.89 out of 5 for high school students and students, respectively. On the other hand, "Replacing the need to turn to specialists to guide them in choosing a school or profession" was among the least popular (2.54 and 1.98 out of 5, respectively), which underlines the idea that young people have a strong opinion. about the need for human advisors in this matter.

Identifying system user requirements

As a result of the literature analysis regarding career guidance practices and support tools, as well as the analysis of surveys applied to high school students and students (possible users of the CareProfSys system), we have established the actors, the main use cases and the functionalities of the system, from the perspective the actors.

The actors identified are the administrator, the visitor, the registered user (pupil/student). Use cases include functionality related to: account management (account creation, login, reset forgotten password); profile and job recommendations management (profile creation – static and dynamic, receiving job recommendations, chatbot consultation, viewing jobs in virtual reality); system administration (user, profile administration) (see Figure 1).

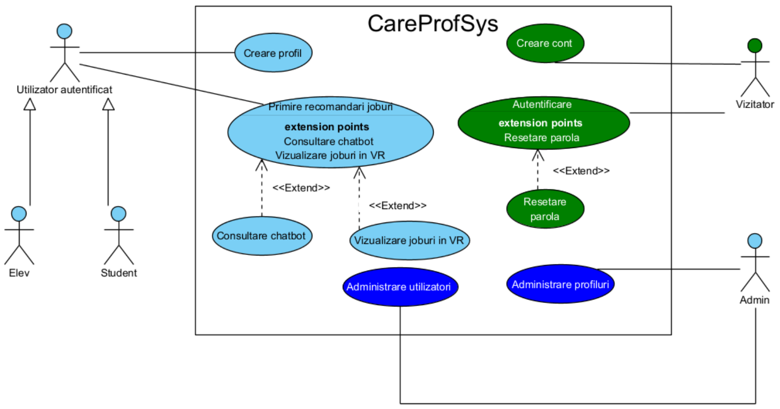

Figure 1. Use cases of the CareProfSys system

For the use cases, a brief summary is presented, preconditions and post-conditions, the steps required to fulfill the functionality, any exceptions that may occur within these steps. Each use case thus described has an associated user requirement, which includes a name, actors involved, description, attributes (importance – from critical to minor, stability – the likelihood of change during development, verification – the testing process), constraints (performance, interface, security). Critically important requirements can thus be easily identified (ex: account creation and authentication, profile creation, receiving job recommendations), requirements prone to change can be analyzed, as well as the constraints that must be considered when establishing the architecture, choosing technologies and starting the development process.

Selection of technologies and development of necessary algorithms

Data collection from various sources is done through worker services that will send the data in the format extracted from their sources in JSON format to the data collection microservice, which will process and store them in the database. Among the data collection services, we will have: collection of raw data (through the necessary APIs), collection of data from social networks (through scrappers), collection of data from forms (through special methods), collection of data from CVs (via CV parsing service). This microservice exposes endpoints for CRUD operations that can be performed based on the profile, only to authorized services that will authenticate to communicate with it. An Object-Relational Mapping (ORM) type framework will be used to work with the data. The collection of CV data will be done in a dedicated microservice, which is based on the Europass CV parser [16]. CareProfSys will recommend professions of individuals with a profile similar to that of the current user. Therefore, it is necessary to have as many profiles of professionals or professions in social networks as possible. For this, I developed scripts in Python for scrapping social recipes (LinkedIn, TikTok, Instagram - the most used by students and pupils, according to the questionnaires applied within the project), I used public data sets from Kaggle [17] and developed a questionnaire to collect data similar to those existing on the LinkedIn network.

The process of building the user's virtual profile is based on data collection and is very important in the implementation of the job recommendation algorithm. Based on the results of surveys with high school students and students from the deliverable "Analysis report and requirements of system users", as well as the data extracted from the chosen sources (social networks and Europass CV), we selected the essential characteristics needed to create the user profile and the occupational profile. Automatically extracted data is used to create the dynamic profile, while additional data that cannot be automatically extracted is manually entered by the user, creating the static profile. Static user profile data consists primarily of demographic information entered by the user when he first decides to use a career recommendation system. They include first name, last name, date of birth, email, nationality, city/country of residence and, optionally, preferred or avoided jobs. Tests that aim to reveal additional characteristics of the user (personality, temperament, IQ, EQ) are also included in the static profile, since they must always be updated and cannot be automatically extracted from third-party data sources. Dynamic profiling is primarily based on the person's Europass CV [16]. The information collected from social networks is intended either to add new content to the profile or to complete and validate the data already extracted from the CV. Such examples include LinkedIn (location of residence – optional, self-description, work experience, education – history, certificates, projects – optional, skills and abilities – both technical and social skills, languages spoken), Instagram and TikTok (about – self description, statistics – number of followers and posts, information about posts – description, topics). The choice of the 3 social networks (LinkedIn, Instagram and TikTok) was made as a result of the analysis of the surveys applied in the project. By combining both static and dynamic profiles, we get a complete user profile, while the same information obtained from several sources can help verify the result and avoid the use of misleading or false data. All merged data is structured into a user profile in JSON format.

As a result of the literature analysis regarding career guidance practices and support tools, presented in the deliverable "Analysis report and system user requirements", the following recommendation algorithm is proposed: (Step 1) Identification of the candidate's skills and interests; (Step 2) Mapping skills extracted from natural language to structured data; (Step 3) Obtaining the occupation recommendation using the COR ontology; (Step 4) Selecting the optimal recommendations obtained from the ontology using a collaborative filtering and machine learning algorithms; (Step 5) Presentation of user data. The hard parts are the implementation of step 2 and 3. We need to map the skills that are extracted from different data sources and saved in the JSON profile of the user to the standard skills that will be used later in the recommendation process, from within the COR ontology. To map natural language skills extracted from various sources into a common computer-understandable language we will need to run a natural language processing algorithm that will classify each given skill into a predefined code or standard structure. To build the algorithm, we will use a new machine learning model, a transformer - a neural network that learns context and therefore meaning by tracking relationships in sequential data, such as words in a sentence [18]. Among all available transformer models, we decided to choose BERT (Bidirectional Encoder Representations from Transformers) [19], mainly due to the rather large community of researchers supporting this initiative and the wide range of models that we can find online that can be adjusted for our needs. Unlike recent language representation models, BERT is designed to pre-train deep bidirectional representations from unlabelled text, jointly conditioning both left and right context across all layers. After all our data is structured, we can use the COR Ontology Mining API to retrieve all the occupations the candidate will be suitable for by querying the ontology/applying inferences based on their skills and areas of interest. The ontology will return us a large amount of data, which will be reduced to the most relevant occupations for the current user profile by applying collaborative filtering models combined with BERT and a K-means unsupervised learning model [20].

To implement the Virtual Reality (VR) scenes in the system, the Unity 3D engine [21] and the JavaScript-based API, WebGL [22] will be used. For 3D objects used in VR simulation we will use Blender when necessary [23]. We will have six initial scenarios in which we will expose activities from the occupations: chemist, project manager, university professor, network specialist, software developer and civil construction engineer. In the RV, an activity as attractive as possible from each job will be presented, which would last approximately 5-10 minutes. The duration of the simulation and the type of activity presented must also consider performance limitations when running on a browser and using a VR headset.

For the conversational agent part, the Pandora Bots framework [24] will be used, which is already used in education, offers us the necessary functionalities and features and supports the Romanian language. It uses AIML (Artificial Intelligence Markup Language), an open-source standard similar to XML.

System software requirements

After an analysis of user requirements, we have come to the conclusion that there are 2 main purposes of using the CareProfSys system: (1) using the system to receive personalized career recommendations by pupils or students; (2) using the system to discover whether a recommended career is to the user's liking via virtual reality mode and chatbot, but only after a recommendation is provided. The functional requirements to be implemented are those that serve the above purposes in a user-friendly manner. Regarding performance, the biggest challenge is the connection between the application itself and the ontology, which is why we decided to create the API for its exploitation, so that we can have a common entry point to the ontology, with which we can communicate through the methods known HTTPs. From a reliability point of view, the services will run within Docker images [25], which will be organized through Kubernetes [26]. As with any large-scale application, maintenance costs are not small, as each microservice requires a dedicated container for it. Github [27] is used for code collaboration and distribution, and AGILE [28] is used as an organization methodology.

Detailed system design

A "microservices" type architecture will be used, grouped by levels, in which each of the modules will run as a microservice in its own container, communicating with the others through HTTP requests: see Figure 2.

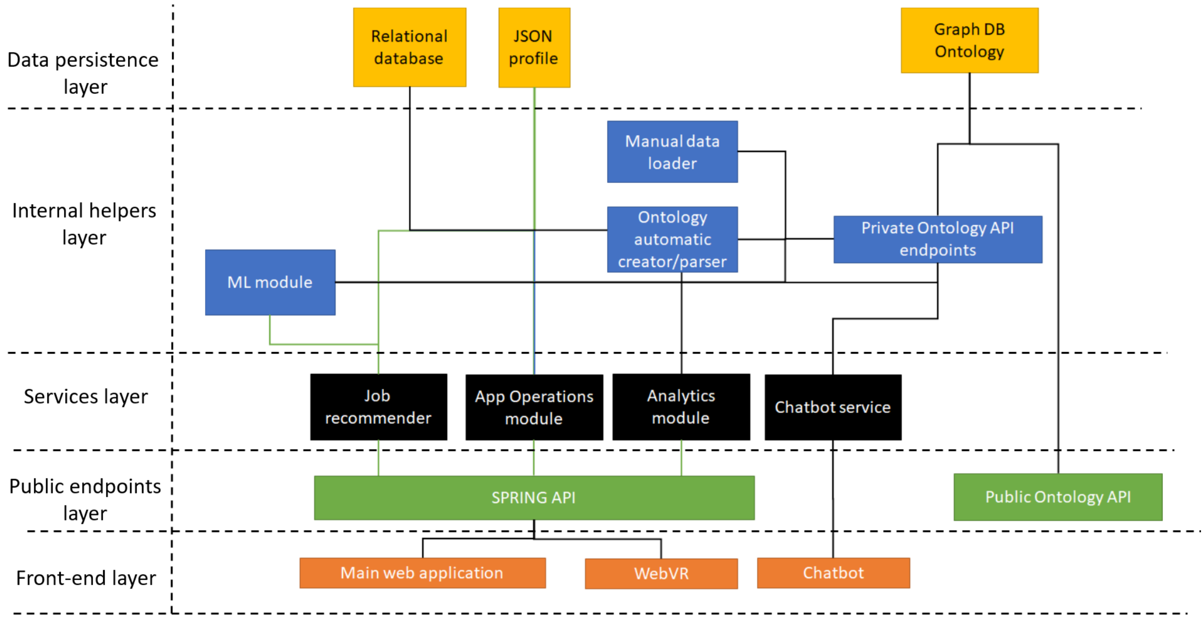

Figure 2. CareProfsys system architecture

The data persistence layer will contain the COR ontology that will run in a separate container using a GraphDB repository and will be accessed using the exploitation API, a non-relational database, NoSQL, MongoDB/ Firebase, which will be used to store data from a user's virtual profile in JSON format, and a relational database for the other types of data (e.g. user account). The internal helpers' layer will contain both the microservices that act as a repository for the ontology, as well as the services created by us to facilitate an easier manual or automatic loading of data into the application or machine learning module. The Services layer is where we will retrieve data from various sources and call certain internal services to successfully implement each use case. The services will serve both the main web application, as well as the chatbot module and the VR application. This is also where the recommendation algorithm and any semantic calculation based on the ontology will be implemented. The public endpoints layer is the connection between the front-end (the web application, the Web VR module and the chatbot) and the back-end (all the above services). On the front-end side, the architecture will be typical for modern applications, namely a single page web app, for which we will use the classic React + Redux stack [29], combined with Bootstrap [30]/ Material UI [31] as framework for style. In addition to the main application, we will have 2 more independent front-end modules, namely the VR application and the chatbot. The chatbot will benefit from a separate service on the back-end side, which can be connected either to the front-end created by us, or to a page on a certain social network, such as Facebook, Slack or Telegram.

In order to use the main system of the CareProfSys project (ontology access, obtaining job recommendations, accessing chatbot) no special hardware components are required, the system can be accessed from any laptop / PC connected to the Internet and equipped with a compatible browser. To use the VR module of the CareProfSys system, it is necessary to use a VR set, it is recommended to use the Oculus Quest 2 set and a desktop that allows this.

Activity 1.2: Development of COR ontology and exploitation tools

Within this activity, **two main tasks** were performed, **with the aim of achieving the objectives (b)** creating the COR ontology and populating it with professional profiles, **(f)** providing a flexible tool that can be easily customized to support career counselors and high school and university students and **(g)** enabling the rapid development of several other modules in the future:

- T2.1. Development of the COR ontology
- T2.2. Development of the ontology exploitation framework

Details about the scientific results of these activities can be found in Deliverable 3 "COR Ontology and its description", and the obtained results were disseminated in the conference article (R1) and the journal article (R4). The proposed objectives for T2.1 and T2.2 activities were 100% achieved.

Development of the COR ontology

In order to develop the ontology, aspects related to the professionalization of work were studied: professional identity in industry 4.0, relevant professional bodies, taxonomies and standards, the nomenclature "Classification of occupations in Romania (COR) was studied, then examples of software artifacts- similar ontologies. Industry 4.0, characterized by the extensive use of artificial intelligence (AI), has led to the emergence of the concept of digital transformation, which involves changes at the operational and organizational level, but also cultural changes through the integration of technologies in production processes and skill profiles at all levels and the functions of the organization [32]. The adoption of AI in the workplace can completely transform work and the way work is organized [33], leading to the emergence of new occupations. Occupations tend to turn into professions, which require professional qualification, systematic assessment of professional skills and compliance with the requirements of the code of professional conduct. The professionalization of work represents the process of transforming an occupation into a profession with a high degree of integrity and competence [34]. Professionalization implies the existence of professional qualification frameworks, professional associations that define and recommend to the members of the professional community the best practices, codes of professional conduct and professional certifications, to differentiate between competent and incompetent professionals. In order for all these processes to be aligned and performed properly, standards and regulations must be defined and the main actors in the labor market should comply with them. At the same time, these standards also lead to the shaping of the concept of professional identity: the shared perception of opportunities by the associated professional, including more than improving the level of income. The pillar of occupational regulation and organization is the international or national classification of occupations, which includes both professional and non-professional occupations. The International Standard Classification of Occupations (ISCO according to the English acronym), with ISCO-08 as the most recent edition, is a system for classifying and aggregating information about occupations present on the labor market [35]. The Occupational Classification of Romania (COR) is aligned with European standards [36].

The Romanian national classification of occupations (COR) is a nomenclature that appeared in 1995. The COR is a system for identifying and coding Romanian occupations regardless of their type and location. In COR, a six-digit code is assigned to each job position on the labor market in Romania. The classification is periodically revised and is updated with occupations not yet included in it. The last review and update was in 2022. Like many other national classifications of occupations, the Romanian classification follows the structure of the International Standard Classification of Occupations (ISCO-08, 2022). The taxonomy of occupations has 5 levels. The classes of the first four levels have an ISCO code correspondence. The fifth level includes the occupations carried out in the Romanian economy that are already included in the COR. The occupations do not have ISCO correspondence, they are specific to the labor market in Romania. The highest level is divided into nine main categories with the same names as in ISCO, not taking into account the tenth category, from the military field. For each occupation, the COR contains all the necessary details about the job, including general activities, work context, work style, values and needs attached to the job. It also specifies all the skills, abilities and knowledge that a person must have in order to fill that position. A significant number of occupations also have a job description attached, which further details all job characteristics and expectations from a candidate [36].

The desire to digitize the standards and taxonomies that support the concept of professionalization and professional identity in Industry 4.0 is reflected by the many examples of software artifacts [37][38]. The main software artifact used in knowledge modeling and work professionalization is the ontology. A formal and clear specification of a common conceptualization containing ideas, entities, characteristics, and qualities related to a particular domain. it is called an ontology. Using computational ontologies, one can explicitly represent the structure of a domain of knowledge, including the entities and pertinent relationships that result from observing it and serve the needs of the system in which the ontology is embedded. Ontologies help to analyze, share and reuse knowledge and reasoning.

The purpose of the COR ontology developed within the CareProfSys project is to describe the professional skills that an individual must have in order to access an occupation on the labor market, the ontology thus contributing to several fundamental aspects of human resource management: optimizing the professional path, learning and development continues, recruitment, strategic workforce planning, etc. This ontology will always evolve, as new occupations, fields, activities are always emerging to provide users with relevant and up-to-date information so that they can keep up with the needs of the labor market.

All data about occupations available in the labor market in Romania are stored in a conceptual system formally described by the COR ontology, whose content can be administered through a dedicated API, also built within the CareProfSys project. Users of the API have the ability to search the knowledge base by running a predefined SPARQL template.

The COR ontology was created using the RDF (Resource Description Framework) and RDFS (Resource Description Framework Schema) data models, but also the OWL (Web Ontology Language) language. The ontology editor that today offers the most functionality for creating and manipulating OWL ontologies available is Protégé. It has an open architecture and has a large number of extensions and plugins available and still growing. This fact and that it can be easily connected to an inference engine makes it the most suitable tool for COR ontology development. To make queries on the ontology, we used the classic SPARQL language (SPARQL Protocol and RDF Query Language). The COR ontology was created based on the content found at [36], considering the 9 major groups of occupations, except for the 10th one in the military field (see Figure 3). To automate the ontology creation process, each web page was automatically read by Python functions and saved to csv files, then each csv was parsed, and the classes/concepts created in OWL. At the same time, the rules for defining the classes that represent the occupations were created automatically, by running a machine learning algorithm based on transformers that will classify each given skill to a standard structure [39]. From all available transformer models, we decided to choose BERT [40], and to train a classification model for the tasks, we provided it with a training set consisting of a two-column CSV file: the representation in “natural language” of the skill taken from the job description in the web page and the structured output used in the OWL rule definition.

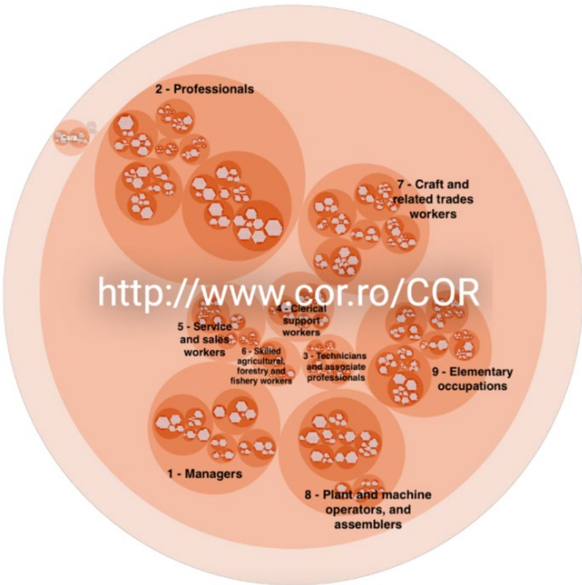

Figure 3. The main occupation groups of the COR ontology, made with the GraphDB tool.

The structure of the ontology

The taxonomy of Romanian occupations is the main pillar of the ontology. In addition to the pillar of occupations, the ontology contains three more pillars: a) fields of study, thus connecting education with professions, b) characteristics related to the occupation (general activities, work context, work style, values and needs) and c) characteristics necessary for to fulfill a certain position (ie skills, aptitudes and interests). All characteristics have a description, and the importance of the characteristic or the associated domain to be able to access a certain occupation is also indicated. This importance is labeled with one of the following values: indispensable with 81-100 points out of 100, very important with 61-80 points, quite important with 41-60 points, important with 21-40 points and less important with 1-20 points out of 100. These values are actually individuals of the Value class.

There is no inherent hierarchical relationship between any two of the four pillars. Any concept can become central depending on its contextual system in which the ontology is used: an organization providing vocational training for job seekers might focus on fields of study, but a recruitment company would focus more on places for work. An occupation can also be seen as a collection of skills, aptitudes, etc. but, at the same time, a single skill can be defined by all jobs where it is required.

Figure 4 shows the basic structure of the COR ontology with the four pillars ("Skills and interests", "Occupational features" in "Characteristics", "COR" and "Domains"). The last class in the list is “Value”, which represents the importance of each feature. For each occupation you need the COR code, a URI, its description and name and the correspondence in ISCO.

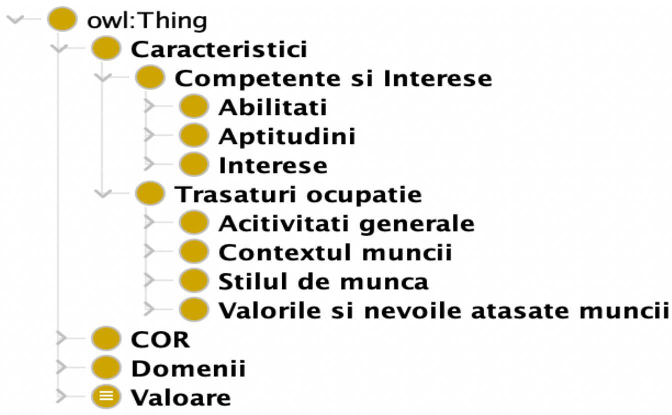

Figure 4. Basic structure of the COR ontology

In order to connect the domains, the characteristics of the occupation and the characteristics of the possible candidate for a specific job, object properties are developed, see Figure 5.

The purpose of the ontology is to correlate the skills of an individual's profile with those of an occupation, helping in several aspects: (1) Companies will find suitable candidates for vacancies; (2) Individuals can find those occupations for which they have all/sufficient necessary skills; (3) Individuals can find out what skills they still need to develop in order to apply for a certain position; (4). Education systems (public and private) will be aware of changes in the labor market.

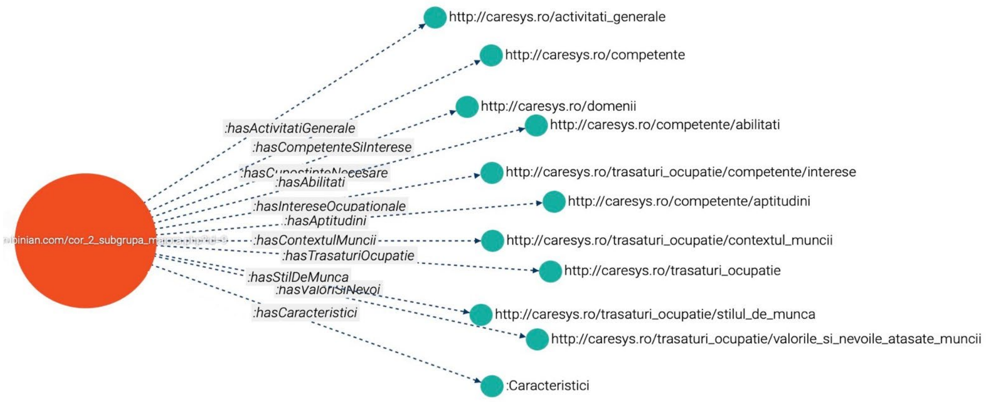

Figure 5. Relationships in the COR ontology

Development of the framework for ontology exploitation

The ontology exploitation framework is in the form of an API (Application Programming Interface), which will serve both public users through public endpoints and the other microservices of the CareProfSys system through private endpoints. For the implementation of the mining API, the classic 3-layer model was used: see the way the API works in Figure 6.

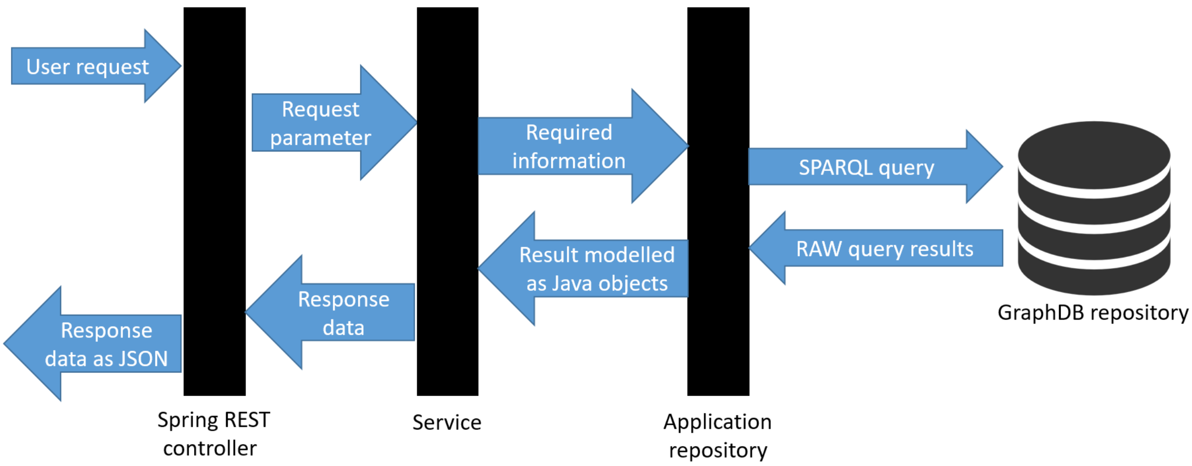

Figure 6. Workflow of the COR ontology exploitation API

The public methods exposed by the API are the following: extracting the code of a particular job from the COR; extracting all jobs from the COR, together with their related code. The private methods are as follows: adding a new triplet to the ontology, modifying a specific triplet from the ontology, deleting a specific triplet from the ontology, extracting all details related to a specific occupation. The API is developed in Java.

Summary of progress

Completed deliverables

During the first Phase, 3 deliverables were made: (1) Deliverable 1- Analysis report and system user requirements - within activity A1.1; (2) Deliverable 2 - Report on the software requirements of the system and its design - within the activity A1.1; (3) Deliverable 3 - COR Ontology and its description - within activity A1.2. All deliverables are available in the attachments.

The purpose of the first deliverable is to specify the results of the analysis performed to establish the functionalities of the CareProfSys system. This analysis included both a thorough study of the literature and existing similar systems, as well as the creation of online surveys with high school students and students (the main actors of the system) to determine their interests. The results obtained are then concretized in the definition of use cases and user requirements.

The purpose of the second deliverable is to present the technical details in order to properly implement the functionalities of the CareProfSys system. The technological choices for the creation of the web application, the ontology, the recommendation engine, the automatic extraction of user data, as well as the creation of the chatbot and virtual reality scenes will be justified. The document also reviews non-functional requirements, hardware and software components, and system architectural design. The descriptions are accompanied by schematics and suggestive diagrams to facilitate the understanding of the system by people with a technical *background*.

The purpose of the third deliverable is to identify relevant aspects regarding the professionalization of work, with an emphasis on the software artifacts used in this direction, to describe the development of the ontology that reflects the Classification of Occupations in Romania, within the CareProfSys project, as well as the API for its use.

Outcome indicators and dissemination of results

The result indicators of Phase 1 are: 3 delivered deliverables (available in the Appendices), 3 articles at conferences and 1 article under evaluation at an ISI journal (available in the Appendices) and the creation of the project website, thus exceeding the proposed indicators (3 deliverables, 1 scientific article submitted to a journal, 1 scientific article submitted to a conference and the project website).

The results were disseminated through papers published in the Proceedings of 3 prestigious international conferences (Proceedings that will be sent for evaluation and inclusion in the Web of Science) and 1 article under review in an ISI indexed journal.

(R1) M.I. Dascalu, I. Marin, I.V. Nemoianu, I.F. Puskás, A. Hang, AN ONTOLOGY FOR EDUCATIONAL AND CAREER PROFILING BASED ON THE ROMANIAN OCCUPATION CLASSIFICATION FRAMEWORK: DESCRIPTION AND SCENARIOS OF UTILISATION, 15th annual International Conference of Education, Research and Innovation, Sevilia (Spain), DOI: 10.21125/iceri.2022, ISBN: 978-84-09-45476-1, ISSN: 2340-1095, pg. 7386-7395. 7-9 noiembrie 2022 – diseminare rezultate din Activitatea 1.2.

(R2) I.C. Stanica, S.M. Hainagiu, S. Neagu, N. Litoiu, M.I. Dascalu, HOW TO CHOOSE ONE'S CAREER? A PROPOSAL FOR A SMART CAREER PROFILER SYSTEM TO IMPROVE PRACTICES FROM ROMANIAN EDUCATIONAL

INSTITUTIONS, ICERI2022, 15th annual International Conference of Education, Research and Innovation, Sevilia(Spain), DOI: 10.21125/iceri.2022, ISBN: 978-84-09-45476-1, ISSN: 2340-1095, pg. 7423-7432. 7-9 noiembrie 2022 – diseminare rezultate din Activitatea 1.1.

(R3) I.C. Stanica, I.A. Bratosin, D.A.Mitrea, C.N.Bodea, M.I. Dascalu, A.Hang, BUILDING RELEVANT ELECTRONIC PROFILING FOR AUTOMATED CAREER RECOMMENDATIONS, 40th IBIMA (International Business Information Management Association) Tech Conference 2022, Sevilia (Spain), ISBN: 979-8-9867719-1-5, ISSN: 2767-9640. 29-30 noiembrie 2022 – diseminare rezultate din Activitatea 1.1.

(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, ISI. dec 2022 (în evaluare) 2022 – diseminare rezultate din Activitatea 1.2.

All papers contain the acknowledgment of the project.

At the same time, the project was disseminated through its *website*, available in two languages (English and Romanian), to ensure the visibility of the results (http://careprofsys.upb.ro/)

and on ResearchGate (https://www.researchgate.net/project/Smart-Career-Profiler-based-on-a-Semantic-Data-Fusion-Framework).

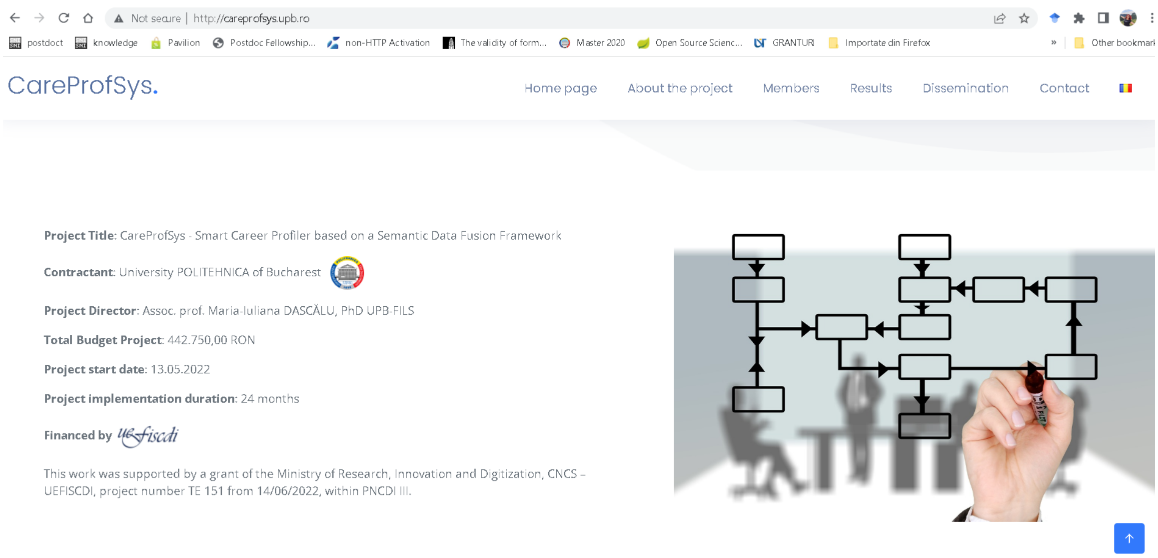

Figure 7. CareProfSys project website: http://careprofsys.upb.ro/

Conclusions

Interdisciplinary research projects have a high-risk probability and impact. In addition to the risks characteristic of such a project, the following risks can be determined for the CareProfSys project: (1) the combination of new technologies (semantic, social, virtual reality, conversational agents); (2) the performance of recommendation algorithms and big data analysis; (3) insufficient number of users/sufficient amount of data. Mitigation is achieved through: (1) the existence of an interdisciplinary team, with both technical experts and psychologists/career counselors, as well as continuous integration tests, as well as microservices-based architecture; (2) emphasis on experimentation before choosing the algorithms used in the final version of the system.; (3) in addition to web scrapping applied to social platforms, we also created a questionnaire to obtain the data needed to create LinkedIn-type professional profiles. We will also ensure data confidentiality by using appropriate means of storage and processing and by securing the hardware and software components, as well as the communication interfaces between them.

The main and ultimate objective of any online career recommendation platform is to identify the right direction in choosing the profession and, by implication, the required studies, therefore this should be the most relevant feature that CareProfSys is designed to fulfill. The deliverables from this Phase, through the clear and unified way of presenting the use cases, through tables that are easy to follow and understand, through the architecture diagrams and the technical details provided, constitute the working basis for the next Phase of system development.

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. R. Nelson-Jones, “Counselling manual”, Trei Publishing House, Bucharest, 2007
2. Romanian Ministry of Education, Official Website, www.edu.ro
3. Teoria personalitatii a lui Holland, https://www.scrigroup.com/educatie/psihologie-psihiatrie/TEORIA-PERSONALITATII-A-LUI-HO62718.php
4. Chestionar MBTI, https://www.jobtestsuccess.com/mbti-test/
5. Chestionar INEM, https://www.psihoprofile.ro/Content/Questionnaires/1/41/RaportdeconsilierelNEM2022.pdf
6. M. Jigau, “Career counselling. Compendium of methods and techniques”, *AFIR Publishing House*, Bucharest, 2007
7. R. Burke, "Hybrid Recommender Systems:Survey and Experiments," User Model. User-Adapt. Interact, vol. [12], no. 4, pp. 331–370, 2002.
8. N. D. Almalis, G. A. Tsihrintzis, N. Karagiannis and A. D. Strati, "FoDRA — A new content-based job recommendation algorithm for job seeking and recruiting", *2015 6th International Conference on Information, Intelligence, Systems and Applications (IISA)*, pp. 1-7, 2015.
9. R. Mishra, S. Rathi, “Enhanced DSSM (deep semantic structure modelling) technique for job recommendation”, *Journal of King Saud University - Computer and Information Sciences*, 2021.
10. A. Nigam, A. Roy, H. Singh and H. Waila, "Job Recommendation through Progression of Job Selection," *2019 IEEE 6th International Conference on Cloud Computing and Intelligence Systems (CCIS)*, pp. 212-216, 2019
11. I. Mohamad and D. Usman, “Standardization and Its Effects on K- Means Clustering Algorithm,” Res. J. Appl. Sci. Eng. Technol., vol. 6, pp. 3299–3303, Sep. 2013, doi: 10.19026/rjaset.6.3638.
12. B. D. Puspasari, L. L. Damayanti, A. Pramono, A. K. Darmawan, "Implementation K-Means Clustering Method in Job Recommendation System," *2021 7th International Conference on Electrical, Electronics and Information Engineering (ICEEIE)*, pp. 1-6, 2021.
13. A. Gupta, D. Garg, "Applying data mining techniques in job recommender system for considering candidate job preferences," *2014 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, pp. 1458-1465, 2014.
14. S. R. Rimitha, V. Abburu, A. Kiranmai and K. Chandrasekaran, "Ontologies to Model User Profiles in Personalized Job Recommendation," 2018 IEEE Distributed Computing, VLSI, Electrical Circuits and Robotics (DISCOVER), 2018, pp. 98-103, doi: 10.1109/DISCOVER.2018.8674084
15. May Fern Koh, Yew Choong Chew, Intelligent Job Matching with Self-learning Recommendation Engine, Procedia Manufacturing, Volume 3, 2015, Pages 1959-1965
16. Parser CV Europass, https://github.com/europass/europasscv-parser-js
17. Kaggle, https://www.kaggle.com/
18. Model transformator, https://blogs.nvidia.com/blog/2022/03/25/what-is-a-transformer-model/
19. Model BERT, https://huggingface.co/docs/transformers/model_doc/bert

20. I. Mohamad and D. Usman, “Standardization and Its Effects on K- Means Clustering Algorithm,” Res. J. Appl. Sci. Eng. Technol., vol. 6, pp. 3299–3303, Sep. 2013, doi: 10.19026/rjaset.6.3638.
21. Unity, https://unity.com/
22. WebGLhttps://developer.mozilla.org/en-US/docs/Web/API/WebGL_API
23. Blender, https://all3dp.com/2/blender-vr-simply-explained/
24. Pandorabots, https://home.pandorabots.com/
25. Docker, https://www.docker.com/
26. Kubernets, https://kubernetes.io/
27. Github, https://github.com/
28. Agile, https://www.atlassian.com/agile
29. React-Redux, https://react-redux.js.org/
30. BootStrap, https://getbootstrap.com/
31. Material-UI, https://mui.com/material-ui/getting-started/overview/
32. C.N. Bodea, C. Mitea, C., O. Stanciu, *Artificial intelligence adoption in project management: main drivers, barriers and estimated impact*, in: A.M. Dima (ed) Innovative Models to Revive the Global Economy, Proceedings of the 3rd International Conference on Economics and Social Sciences, Sciendo Proceedings (2020).
33. I-SCOOP, *Digitization, digitalization, digital and transformation: the differences*. *Online guide to digital transformation*, available at: https://www.i-scoop.eu/digital-transformation/digitization-digitalization-digital-transformation-disruption/ (2013).
34. F. Garcia-Lopera, JM Santos-Jaén, M. Palacios-Manzano, D. Ruiz-Palomo, *Exploring the effect of professionalization, risk-taking and technological innovation on business performance*. PLoS ONE 17(2) (2022).
35. MMSS, Clasificarea Ocupatiilor din Romania, Ministerul Muncii si Solidaritatii Sociale (2022).
36 COR 2022 Clasificarea ocupatiilor din Romania, https://www.rubinian.com/cor_1_grupa_majora.php (2022).
37. S. Pryimaa, O. Strokana, J. Rogushinab, A. Gladunc, A. Mozgovenko, *Ontology-Based Methods and Tools for Validation of Non-Formal Learning Outcomes*, in: Scientific Journal Problems of Programming, Volume 2, pp. 50-60 (2020).
38. E. Ilkou, H. Abu-Rasheed, M. Tavakoli, S. Hakimov, G. Kismihók, S. Auer, W. Nejdl, *EduCOR: An Educational and Career-Oriented Recommendation Ontology*, in: Lecture Notes in Computer Science, Volume 12922, pp. 546-562 (2021).
39 A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, I. Polosukhin, *Attention is all you need*. In 31st Conf. on Neural Information Processing Systems (NIPS 2017) (pp. 5998–6008). Long Beach, CA, USA (2017).
40 M. Masala, S. Ruseti, M. Dascalu, M. RoBERT – A Romanian BERT Model. In 28th Int. Conf. on Computational Linguistics (COLING) (pp. 6626–6637). online: ACL. (2020).

Annexes

Annex 1. Deliverable 1 - Analysis report and system user requirements

Annex 2. Deliverable 2 - System Software Requirements and Design Report

Annex 3. Deliverable 3 - COR Ontology and its description

Annex 4. R1 - Article in extenso - ICERI conference

Annex 5. R2 - Article in extenso - ICERI conference

Annex 6. R3 - Article in extenso - IBIMA conference

Annex 7. R4 - Article in extenso - ISI journal Rev. Rome. Sci. Techn. – Electrotechn. et Énerg.

DIRECTOR PROIECT,

DASCĂLU, MARIA-IULIANA