

CareProfSys

Sistem inteligent de profil de carieră bazat pe o platforma de fuziune semantica a datelor

Raport științific – etapa I

Proiect finanțat de la bugetul de stat

PN-III-P1-1.1-TE-2021

Contract NR. TE 151/2022

Data început: 13 mai 2022. Durata: 24 luni

CUPRINS

| | |
|--|----|
| Rezumat executiv | 3 |
| Introducere | 4 |
| Descriere, obiective și etape proiect CareProfSys | 4 |
| Obiectivele și activitățile etapei I | 5 |
| Scopul documentului | 5 |
| Activitate 1.1: Proiectare și alegeri tehnologice | 5 |
| Analiza literaturii cu privire la practicile de orientare în carieră și la instrumentele de support | 5 |
| Identificarea cerințele utilizatorilor sistemului | 8 |
| Selecția tehnologiilor și dezvoltarea algoritmilor necesari | 9 |
| Cerințele software ale sistemului | 11 |
| Proiectarea detaliată a sistemului | 11 |
| Activitate 1.2: Dezvoltarea ontologiei COR și a instrumentelor de exploatare | 12 |
| Dezvoltarea ontologiei COR | 12 |
| Dezvoltarea framework-ului de exploatare a ontologiei | 16 |
| Sumar al progresului | 17 |
| Livrabile realizate | 17 |
| Indicatori de rezultat și diseminarea rezultatelor | 17 |
| Concluzii | 18 |
| Bibliografie | 19 |
| Anexe | 20 |
| Anexa 1. Livrabil 1 - Raport de analiză și cerințele utilizatorilor sistemului | 20 |
| Anexa 2. Livrabil 2 - Raport privind cerințele software ale sistemului și proiectarea acestuia | 20 |
| Anexa 3. Livrabil 3 - Ontologia COR și descrierea acesteia | 20 |
| Anexa 4. R1 - Articol in extenso - conferința ICERI | 20 |
| Anexa 5. R2 - Articol in extenso - conferința ICERI | 20 |
| Anexa 6. R3 - Articol in extenso – conferința IBIMA | 20 |
| Anexa 7. R4 - Articol in extenso – jurnalul ISI Rev. Roum. Sci. Techn.– Électrotechn. et Énerg. | 20 |

Rezumat executiv

Proiectul **CareProfSys** (<http://careprofsys.upb.ro/>) are ca **obiectiv principal** validarea și testarea conceptului de sistem inteligent de profil de carieră prin implementarea lui într-un mediu observat - centrul de dezvoltare a carierei din cadrul Universității POLITEHNICA din București. CareProfSys își propune să ofere elevilor și studenților consiliere în carieră folosind analize avansate ale profilului utilizatorilor, extras automat din diverse surse de date, prin recomandări de ocupații ale unor persoane cu profiluri similare, folosind inferențe ontologice și algoritmi de clasificare a învățării. Proiectul are **trei etape**: (1) Proiectarea sistemului și dezvoltarea ontologiei COR – în 2022, (2) Dezvoltarea, și testarea sistemului – în 2023, (3) Implementarea sistemului CareProfSys în centrul UPB-CCOC – în 2024. **Prima etapă** a proiectului constă în două activități principale: (1) Proiectare și alegeri tehnologice și (2) Dezvoltarea ontologiei COR și a instrumentelor de exploatare.

În cadrul activității de “**Proiectare și alegeri tehnologice**”, s-a efectuat analiza pentru stabilirea funcționalităților sistemului CareProfSys, care a inclus atât studierea amănunțită a literaturii de specialitate și a sistemelor asemănătoare existente, cât și crearea de sondaje online cu 317 liceeni și studenți pentru a stabili interesele acestora. Rezultatele obținute au fost apoi concretizate în definirea cazurilor de utilizare și a cerințelor utilizatorilor (vizitatori, utilizatori autentificați – elevi și studenți și administratori sistem). S-au identificat practici tradiționale de consiliere (teste psihologice sau de personalitate, metode de comunicare unu-la-unu între consilier și client etc) și practici moderne (utilizarea sistemelor de recomandare bazate filtrarea conținutului, filtrarea colaborativă, algoritmi de învățare automată, ontologii). În sistemul educațional din România, numărul de consilieri școlari/universitari pentru orientare în carieră este insuficient pentru nevoia actuală, astfel încât sunt necesare noi soluții, moderne, ce pot îmbunătăți procesul de consiliere și orientare în carieră, iar răspunsurile la chestionare au subliniat încrederea utilizatorilor tineri în sistemele de recomandare automată și în rețelele sociale. După o analiză a cerințelor utilizatorilor, am ajuns la concluzia că există 2 scopuri principale ale utilizării sistemului CareProfSys: (1) utilizarea sistemului pentru a primi recomandări de carieră personalizate de către elevi sau studenți; (2) utilizarea sistemului pentru a descoperi dacă o carieră recomandată este pe placul utilizatorului prin intermediul modului de realitate virtuală și a chatbot-ului, dar doar după ce o recomandare este oferită. Cerințele funcționale care trebuie implementate sunt cele care deservește scopurile de mai sus, într-o manieră prietenoasă pentru utilizator. Din punct de vedere tehnic, se va folosi o arhitectură de tip “microservicii”, grupate pe niveluri, în care fiecare dintre module va rula ca un microserviciu în propriul său container, comunicând cu celelalte prin intermediul unor cereri HTTP: nivelul de persistență a datelor, nivelul intern de execuție, nivelul de servicii, nivelul de *endpoint-uri* publice, nivelul de *front-end*. În cadrul activității de proiectare, s-au stabilit tehnologiile, algoritmii și s-au proiectat componentele software din cadrul arhitecturii mai sus menționate. Pentru recomandarea propriu-zisă, se vor folosi modele de filtrare colaborativă, dar și de învățare automată (e.g. un model de transformator - BERT, un model de învățare nesupervizată K-means).

În cadrul activității de “**Dezvoltarea ontologiei COR și a instrumentelor de exploatare**” s-au analizat: identitatea profesională în industria 4.0, organisme profesionale relevante, taxonomii și standarde, nomenclatorul “Clasificarea ocupațiilor din România (COR), exemple de artefacte software- ontologii similare. COR-ul reprezintă un sistem de identificare și codificare a ocupațiilor românești indiferent de tipul și locul lor, care urmează structura din Standardele Internaționale a Ocupațiilor. Scopul ontologiei COR este descrierea competențelor profesionale pe care trebuie să le aibă un individ pentru a accesa o ocupație de pe piața muncii. Taxonomia ocupațiilor din România este pilonul principal al ontologiei. Pe lângă pilonul ocupațiilor, ontologia conține încă trei piloni: a) domeniile de studiu, conectând astfel educația cu profesiile, b) caracteristicile legate de ocupație (activități generale, contextul muncii, stilul de lucru, valori și nevoi) și c) caracteristicile necesare pentru a îndeplini o anumită poziție (abilități, aptitudini și interese). Framework-ul dezvoltat pentru exploatarea ontologiei este sub forma unui API ce va servi atât utilizatorii publici prin intermediul endpoint-urilor publice, cât și celelalte microservicii ale sistemului CareProfSys prin intermediul endpoint-urilor private. Metodele publice expuse de către API sunt următoarele: extragerea codului unui anumit job din COR; extragerea tuturor job-urile din COR, împreună cu codul lor aferent. Metodele private sunt următoarele: adăugarea unui nou triplet în ontologie, modificarea unui anumit triplet din ontologie, ștergerea unui anumit triplet din ontologie, extragerea tuturor detaliilor legate de o anumită ocupație.

În cadrul etapei I au fost realizate **3 livrabile**: (1) Raport de analiză și cerințele utilizatorilor sistemului; (2) Raport privind cerințele software ale sistemului și proiectarea acestuia; (3) Ontologia COR și descrierea acesteia. **Rezultatele proiectului au fost diseminate prin 3 articole publicate în Proceedings-urile unor conferințe internaționale** (care vor fi indexate ISI) și **1 articol într-un jurnal ISI**, prin realizarea **website**-ului proiectului și prin publicarea informațiilor relevante pe ResearchGate.

Introducere

Descriere, obiective și etape proiect CareProfSys

Proiectul își propune să ofere consiliere în carieră folosind analize avansate ale profilului utilizatorilor, extras automat din diverse surse de date: situația școlară, CV, profiluri de social media sau răspunsuri din formulare web. Utilizatorii CareProfSys vor primi recomandări de ocupații ale unor persoane cu profiluri similare, folosind inferențe ontologice și algoritmi de clasificare a învățării automate. Baza de date de profiluri va fi completată cu informații despre profesioniști care au ocupații din ontologia „Clasificarea ocupațiilor din România” (COR) (dezvoltată de proiect), aliniată cu lista europeană a calificărilor. Un agent conversațional va oferi sfaturi personalizate despre ocupațiile recomandate, în timp ce scenele 3D virtuale vor ajuta utilizatorii să vizualizeze activitățile conectate cu o profesie viitoare. În afara studenților și liceenilor, sistemul va oferi beneficii și centrelor de consiliere în carieră, instituțiilor de învățământ superior, guvernului, companiilor angajatoare. Sistemul va exploata cele mai noi tehnologii: web semantic, ontologii, învățare automată, conectori de rețea socială, realitate virtuală (VR) pe web, agenți de recomandare și interfețe de programare moderne dedicate analizei datelor de mari dimensiuni. Astfel, se va valida și testa conceptul de sistem inteligent pentru construirea profilului de carieră prin implementarea acestuia într-un mediu observat - centrul de dezvoltare a carierei dintr-o universitate românească. Obiectivul principal al CarProfSys este de a valida și testa conceptul unui Sistem inteligent de profil de carieră prin implementarea lui într-un mediu observat - CCOC UPB. Sistemul se bazează pe analiza mai multor surse de date (rețele sociale, istoric educațional, teste psihologice), care sunt input pentru algoritmul de recomandare ce utilizează ontologia "Clasificarea Ocupațiilor din România" (COR) (dezvoltată în cadrul proiectului).

Obiectivele specifice ale proiectului sunt: (a) construirea de profiluri de utilizatori, într-o manieră rapidă și precisă - crearea de conectori sociali prin intermediul API-urilor actuale de rețele sociale, crearea de extractoare PDF, procesoare de formulare web și alte module de minare Web, combinarea și transformarea tuturor surselor de date în date structurate, validarea și analiza acestor date cu ajutorul psihologilor experți; (b) crearea ontologiei COR și popularea ei cu profilurile utilizatorilor; (c) recomandarea ocupațiilor din ontologia COR utilizatorilor (elevi, studenți și absolvenți care doresc să-și schimbe cariera), pe baza profilurilor acestora, printr-un algoritm personalizat; (d) oferirea de informații suplimentare (pe bază de text sau scene 3D) despre ocupațiile recomandate; (e) oferirea suportului în consilierea utilizatorilor (studenți și elevi) prin instrumente moderne de interacțiune om-calculator, de ex. camere bazate pe Web-VR sau camere virtuale pentru interviuri și agenți conversaționali; (f) oferirea unui instrument flexibil care poate fi personalizat cu ușurință pentru a sprijini consilierii de carieră și pe cei școlari din licee și universități; (g) oferirea posibilității dezvoltării rapide a mai multor alte module în viitor; (h) încorporarea produsului într-o ecologie universitară - implementarea ca pilot în CCOC-UPB; (i) atragerea potențialilor utilizatori din alte universități și licee, prin promovarea intensă a acestuia; (j) oferirea unui instrument cu valoare regională.

Proiectul are **trei etape**: (1) Proiectarea sistemului și dezvoltarea ontologiei COR – în 2022, (2) Dezvoltarea, și testarea sistemului – în 2023, (3) Implementarea sistemului CareProfSys în centrul UPB-CCOC – în 2024.

Obiectivele și activitățile etapei I

Prima etapă a proiectului constă în *Proiectarea sistemului și dezvoltarea ontologiei COR* și își propune îndeplinirea obiectivelor specifice de la (a) la (j). Etapa are două activități principale: A1.1. Proiectare și alegeri tehnologice și A 1.2. Dezvoltarea ontologiei COR și a instrumentelor de exploatare, fiecare cu sarcini concrete, ale căror rezultate sunt descrise în continuare.

Scopul documentului

Prezentul document conține descrierea activităților științifice desfășurate în cadrul etapei 1 a proiectului CareProfSys, a rezultatelor obținute, a modalităților de diseminare a acestora și subliniază indicatorii de rezultat.

Activitate 1.1: Proiectare și alegeri tehnologice

În cadrul acestei activități, s-au executat **șase sarcini principale, cu scopul îndeplinirii obiectivelor (a)-(j):**

- T1.1. Analiza literaturii cu privire la practicile de orientare în carieră și la instrumentele de suport
- T1.2. Interviuri și sondaje online: realizare și analiza rezultatelor
- T1.3. Identificarea cerințelor utilizatorilor sistemului
- T1.4. Selecția tehnologiilor și dezvoltarea algoritmilor necesari
- T1.5. Identificarea cerințelor software ale sistemului
- T1.6. Proiectarea detaliată a sistemului

Detalii despre rezultatele științifice ale activităților T1.1, T1.2 și T1.3 se găsesc în Livrabilul 1 “Raport de analiză și cerințele utilizatorilor sistemului”, iar rezultatele acestora au fost diseminate în articolul de conferință (R1). Obiectivele propuse pentru activitățile T1.1, T1.2 și T1.3 au fost 100% atinse.

Detalii despre rezultatele științifice ale activităților T1.4, T1.5 și T1.6 se găsesc în Livrabilul 2 “Raport privind cerințele software ale sistemului și proiectarea acestuia”, iar rezultatele acestora au fost diseminate în articolele de conferință (R1), (R3) și articolul de jurnal (R4). Obiectivele propuse pentru activitățile T1.4, T1.5 și T1.6 au fost 100% atinse.

Analiza efectuată pentru stabilirea funcționalităților sistemului CareProfSys a inclus atât studierea amănunțită a literaturii de specialitate și a sistemelor asemănătoare existente, cât și crearea de sondaje online cu liceeni și studenți (principalii actori ai sistemului) pentru a stabili interesele acestora. Rezultatele obținute au fost apoi concretizate în definirea cazurilor de utilizare și a cerințelor utilizatorilor.

Analiza literaturii cu privire la practicile de orientare în carieră și la instrumentele de suport

Analiza literaturii cu privire la practicile de orientare în carieră și la instrumentele de suport a constatat în două părți, în funcție de gradul de inovație al abordărilor existente de consiliere și recomandare a muncii - practici tradiționale și respectiv practici moderne.

În **practicile tradiționale**, primordială este relația consilier-client. Comunicarea directă dintre aceștia este o trăsătură specifică a procesului de consiliere și depinde de o serie de caracteristici precum atitudinea consilierului, caracteristicile mediului de interacțiune, abilitățile specifice de comunicare și suport ale consilierului. Orientarea în carieră este o relație profesională de tip „unu la unu” între un specialist și o altă

persoană care caută un ajutor/ sprijin specific. Procesul de consiliere își propune să ofere clientului oportunitatea de a explora, descoperi și clarifica modalități de a-și trăi propria viață, creând un sentiment solid de bunăstare [1]. În sistemul educațional din România, numărul de consilieri școlari/universitari pentru orientare în carieră este insuficient pentru nevoia actuală (un consilier ajunge să fie responsabil pentru 2400 de elevi, conform Ministerului Educației din România) [2], astfel încât este nevoie de noi soluții, moderne, ce pot îmbunătăți procesul de consiliere și orientare în carieră. Instrumentele tradiționale de consiliere includ metode de colectare a datelor despre client (de exemplu, teste psihologice sau de personalitate, precum chestionarul Holland [3], MBTI [4], INEM [5] etc.), metode de comunicare, metode de investigare a pieței muncii, branding personal, metode de planificare a carierei [6].

Practicile moderne au în centru sistemele de recomandare, uzuale fiind două tipuri principale: recomandări bazate pe filtrarea conținutului și recomandări bazate pe filtrarea colaborativă [7]. Sistemele bazate pe filtrarea conținutului folosesc „filtre de text și măsurători de similaritate a caracteristicilor” utilizatorilor, în vederea oferirii de recomandări cât mai potrivite. Potrivirea dintre ocupațiile existente și profilurile utilizatorilor, cu diferite probabilități, stă la baza acestei abordări [7]. Sistemele de recomandare bazate pe filtrare colaborativă se bazează pe preferințele utilizatorului sau pe cele ale utilizatorilor similari (joburi ce au apărut în istoricul lor de căutare sau similare cu acestea) [8]. Există de asemenea abordări avansate, de exemplu bazate pe algoritmi de învățare automată. Rețelele neuronale sunt adesea utilizate pentru a capta preferințele unui candidat pentru anumite locuri de muncă, pe măsură ce acestea evoluează/se schimbă în timp [9] [10]. Învățarea supervizată sau cea nesupervizată sunt și ele utilizate în cercetări. K-means este unul dintre cei mai simpli și populari algoritmi pentru învățarea nesupervizată, care are scopul de a împărți datele în grupuri (așa-numite clustere) pe baza caracteristicilor lor comune [11]. În [12], datele despre utilizatori sunt colectate prin sondaje și includ elemente precum abilitățile, salariul așteptat și poziția geografică dorită, în timp ce datele despre locurile de muncă sunt colectate din postările de pe rețelele sociale. Caracteristicile menționate anterior primesc anumite ponderi și sunt utilizate pentru crearea de grupuri de locuri de muncă vacante. Învățarea supervizată se referă la datele etichetate care pot fi utilizate în învățarea automată pentru a clasifica sau prezice rezultatul. Arborii de decizie sunt utilizați în [13] pentru selectarea atributelor semnificative ale utilizatorilor aflați în căutarea unui job. Informațiile sunt căutate recursiv, începând de la rădăcină și avansând până la o frunză a arborelui, grupând informațiile în clase. Printre criteriile incluse în arborii de decizie se remarcă rezultatele școlare (note mari, medii, scăzute), grupa de vârstă, genul, studiile (licență, master, doctorat), specializarea, experiența. Una dintre provocările cu care se confruntă sistemele personalizate de recomandare a locurilor de muncă este găsirea unui format adecvat de reprezentare a datelor necesare pentru recomandări. O metodă frecvent utilizată este reprezentată de ontologii [14][15] care pot reprezenta în mod formal date prin clase, atribute, reguli și relații.

Interviuri și sondaje online: realizare și analiza rezultatelor

Pentru a stabili principiile și funcționalitățile relevante pentru un sistem de recomandare de ocupații, am realizat două sondaje folosind platforma Google Forms, disponibile în Livrabilul 2. Grupul țintă a inclus atât elevi de liceu (pentru primul sondaj), cât și studenți (pentru al doilea sondaj), cu un total de 317 respondenți, dintre care 209 liceeni și 108 studenți. Sondajele sunt similare pentru ambele categorii, incluzând trei secțiuni principale: 1. caracteristici socio-demografice și studii, 2. principii de

alegere a instituției de învățământ actuale și 3. instrumente de sprijin pentru primirea de recomandări de ocupații/ profesii. Studenții au o secțiune suplimentară opțională legată de experiența profesională, secțiune completată doar de cei care au în prezent sau au avut în prealabil un loc de muncă. Majoritatea întrebărilor includ posibilitatea de a oferi un răspuns suplimentar (răspuns deschis) în afara opțiunilor predefinite, pentru a obține cât mai multe informații și a lăsa o anumită libertate respondenților.

Când au fost rugați să identifice două criterii principale de alegere a instituției de învățământ curente, atât elevii cât și studenții au preferințe similare în top, dar într-o ordine diferită - disciplinele lor preferate de studiu (45,5% pentru elevi, 37% pentru studenți) și viitorul loc de muncă (36,4% pentru elevi, 43,5% pentru studenți). Pentru studenți, focusul este pe posibilele oportunități de angajare pe care facultatea le poate oferi (50%), opțiuni care nu ar fi fost viabilă pentru liceeni în privința alegerii liceului. Atât elevii de liceu, cât și studenții au preferințe similare în ceea ce privește persoanele cărora le-au cerut sfaturi atunci când și-au ales instituția de învățământ - părinții, prietenii și profesorii. Odată cu vârsta, studenții tind să fie mai independenți și mai capabili să ia propriile decizii - 28,7% dintre respondenții universităților au declarat că au căutat online informații despre facultatea la care intenționau să meargă. Consilierii școlari nu sunt populari printre tinerii respondenți, ei nu au fost consultați de niciun licean și doar de 2 studenți.

Informațiile importante care trebuie luate în considerare pentru o bună alegere a facultății/ocupației profesionale trebuie să se bazeze pe interese, abilități și personalitate, în această ordine, atât pentru elevii de liceu, cât și pentru studenți. Interesele (ceea ce îi place unui om să facă) domină statisticile, alese de 82,3% dintre liceeni și 73,1% dintre studenți. Studenții sunt mai încrezători în platformele automate decât elevii, probabil datorită studiului și interacțiunii lor mai frecvente cu tehnologia, în timp ce încrederea în consilieri scade odată cu vârsta. Există un procent foarte scăzut de persoane care nu ar dori să aibă acces la informațiile necesare pentru o bună recomandare de facultate/ ocupație profesională (3,3% pentru liceu și 2,8% pentru universitate).

Instrumentele de sprijin par a fi considerate utile atât de către studenți, cât și de către liceeni, iar percepția lor despre acestea este similară în mai multe privințe. Când au fost întrebați despre ce informații consideră ca fiind utile dintre cele oferite de o platformă de recomandare de locuri de muncă, primele 4 răspunsuri au fost identice pentru ambele categorii, cu aceeași ordine și procente similare (1. Recomandări de locuri de muncă potrivite, 2. Informații legate de școli sau cursuri potrivite, 3. Descrierile posturilor, 4. Informații legate de trăsături de personalitate). Încrederea ridicată a tinerilor în tehnologie este din nou subliniată, întrucât doar 3,3% dintre liceeni și 0,9% dintre studenți nu ar dori să folosească o platformă de recomandare a locurilor de muncă. Un alt aspect interesant arată din nou asemănarea percepției respondenților în ciuda diferenței de vârstă, de data aceasta legată de sursele pe care le consideră relevante pentru extragerea informațiilor necesare recomandărilor de locuri de muncă. Opțiunile principale includ, cu procentele lor pentru nivel liceal și universitar, CV-uri (62,7%, 70,4%), teste de personalitate (59,3%, 66,7%), teste de inteligență (40,7%, 53,7%), teste de inteligență emoțională (40,7%, 51,9%). Studenții tind să considere notele ca fiind mai puțin relevante în zilele noastre (27,8% pentru elevi față de 18,5% pentru elevi).

În ceea ce privește rețelele sociale, Instagram este de departe cea mai populară rețea pentru toți respondenții (90% o folosesc cel puțin o dată pe săptămână), urmată de

TikTok pentru elevii de liceu (63,2%) și Facebook pentru studenți (61,1%). În ceea ce privește utilitatea pentru recomandarea de locuri de muncă, Instagram rămâne alegerea de top pentru liceeni (60,8%), dar este înlocuit de LinkedIn pentru studenți (65,7%). LinkedIn este cea mai profesionistă platformă de socializare, așa că pare a fi alegerea evidentă, însă ea nu este foarte familiară tinerilor, de unde și procentul mai mic de liceeni ce o consideră adecvată (27,8%).

Când li s-a cerut să noteze de la 1 (cel mai puțin important) la 5 (cel mai important) caracteristicile unei platforme de recomandare de locuri de muncă, toți respondenții au considerat că „Facilitatea identificării direcției corecte în alegerea școlii sau profesiei” este cea mai relevantă caracteristică, cu nota medie de 3,92, respectiv 3,89 din 5 pentru liceeni și respectiv studenți. Pe de altă parte, „Înlocuirea necesității de a apela la specialiști care să-i ghideze în alegerea unei școli sau profesii” a fost printre cele mai puțin populare (2,54 și, respectiv, 1,98 din 5), fapt care subliniază ideea că tinerii au o părere puternică, despre necesitatea consilierilor umani în această chestiune.

Identificarea cerințele utilizatorilor sistemului

Ca urmare a analizei literaturii cu privire la practicile de orientare în carieră și la instrumentele de suport, precum și a analizei sondajelor aplicate liceenilor și studenților (posibili utilizatori ai sistemului CareProfSys), am stabilit actorii, principalele cazuri de utilizare și funcționalitățile sistemului, din perspectiva actorilor.

Actorii identificați sunt: administratorul, vizitatorul, utilizatorul înregistrat (elev/student). Cazurile de utilizare includ funcționalități legate de: gestionarea contului (creare cont, autentificare, resetare parolă uitată); gestionarea profilului și a recomandărilor de joburi (creare profil – static și dinamic, primire recomandări joburi, consultare chatbot, vizualizare joburi în realitate virtuală); administrarea sistemului (administrarea utilizatorilor, a profilurilor) (a se vedea Figura 1).

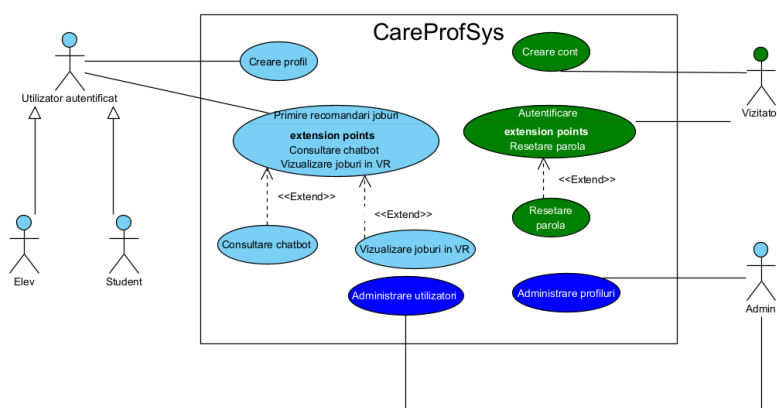


Figura 1. Cazuri de utilizare ale sistemului CareProfSys

Pentru cazurile de utilizare, se prezintă un scurt sumar, precondiții și postcondiții, pașii necesari pentru îndeplinirea funcționalității, eventuale excepții ce pot apărea în cadrul acestor pași. Fiecare caz de utilizare astfel descris are o cerință utilizator asociată, ce presupune o denumire, actorii implicați, descriere, atribute (importanță – de la critică la mică, stabilitate – probabilitatea de schimbare pe parcursul dezvoltării, verificarea – procesul de testare), constrângeri (performanță, interfață, securitate). Se pot identifica astfel cu ușurință cerințele cu importanță critică (ex: crearea contului și autentificarea, crearea profilului, primirea recomandărilor de ocupații), se pot analiza

cerințele predispuse la schimbare, precum și constrângerile ce trebuie avute în vedere în momentul stabilirii arhitecturii, alegerii tehnologiilor și demarării procesului de dezvoltare..

Selecția tehnologiilor și dezvoltarea algoritmilor necesari

Colectarea datelor din diverse surse se face prin intermediul unor *worker services* ce vor trimite datele în formatul extras din sursele lor în format JSON către microserviciul de colectare a datelor, care le va prelucra și le va stoca în baza de date. Printre serviciile de colectare a datelor, vom avea: colectare a datelor brute (prin intermediul API-urilor necesare), colectare a datelor din rețele sociale (prin intermediul unor *scrapere*), colectare a datelor din formulare (prin metode speciale), colectare a datelor din CV-uri (prin serviciu de parsare a CV-urilor). Acest microserviciu expune *endpoint-urile* pentru operațiile CRUD ce pot fi efectuate pe baza profilului, doar către serviciile autorizate ce se vor autentifica pentru a putea comunica cu acesta. Se va folosi un framework de tip Object-Relational Mapping (ORM) pentru lucrul cu datele. Colectarea datelor din CV se va face într-un microserviciu dedicat, ce se bazează pe parser-ul de CV-uri Europass [16]. CareProfSys va recomanda profesii ale unor indivizi cu profil similar celui deținut de utilizatorul curent. Prin urmare, este necesară existența cât mai multor profile de profesioniști sau profile de profesii din rețelele sociale. Pentru acest lucru, am dezvoltat scripturi în Python pentru *scraping* de rețelele sociale (LinkedIn, TikTok, Instagram-cele mai utilizate de elevi și studenți, conform chestionarelor aplicate în cadrul proiectului), am folosit seturi de date publice de pe Kaggle[17] și dezvoltat un chestionar care să colecteze date similare cu cele existente pe rețeaua LinkedIn.

Procesul de construire a profilului virtual al utilizatorilor are la bază colectarea datelor și este foarte important în implementarea algoritmului de recomandare de ocupații profesionale. Pe baza rezultatelor sondajelor cu liceenii și studenții din livrabilul “Raport de analiză și cerințele utilizatorilor sistemului”, precum și a datelor extrase din sursele alese (rețele sociale și CV europass), am selectat caracteristicile esențiale necesare pentru crearea profilului utilizatorului și a profilului ocupațional. Datele extrase automat sunt folosite pentru realizarea profilului dinamic, în timp ce datele suplimentare care nu pot fi extrase automat sunt introduse manual de către utilizator, creând profilul static. Datele statice ale profilului utilizatorului constau în principal din informații demografice introduse de utilizator atunci când decide să utilizeze pentru prima dată un sistem de recomandare a carierei. Ele includ nume, prenume, data nașterii, e-mail, naționalitate, oraș/țara de reședință și, opțional, locurile de muncă preferate sau de evitat. Testele care au scopul de a dezvălui caracteristici suplimentare ale utilizatorului (personalitate, temperament, IQ, EQ) sunt incluse și în profilul static, deoarece acestea trebuie să fie mereu actualizate și nu pot fi extrase automat din surse de date terțe. Profilarea dinamică se bazează în primul rând pe CV-ul Europass [16] al persoanei. Informațiile colectate de pe rețelele sociale au scopul fie de a adăuga conținut nou la profil, fie de a completa și valida datele extrase deja din CV. Astfel de exemple includ LinkedIn (locație de reședință – opțional, descrierea propriei persoane, experiență de muncă, educație – istoric, certificate, proiecte – opțional, abilități și aptitudini – atât abilități tehnice, cât și sociale, limbi vorbite), Instagram și TikTok (despre – descrierea propriei persoane, statistici – numărul de urmăritori și postări, informații despre postări – descriere, subiecte). Alegerea celor 3 rețele sociale (LinkedIn, Instagram și TikTok) s-a făcut ca urmare a analizei sondajelor aplicate în proiect. Combinând atât profilul static, cât și cel dinamic, obținem un profil de utilizator complet, în timp ce aceleași informații obținute din mai multe surse pot

contribui la verificarea rezultatului și pot evita utilizarea de date înșelătoare sau false. Toate datele îmbinate sunt structurate într-un profil de utilizator în format JSON.

Ca urmare a analizei literaturii cu privire la practicile de orientare în carieră și la instrumentele de suport, prezentată în livrabilul “Raport de analiză și cerințele utilizatorilor sistemului”, se propune următorul algoritm de recomandare: (Pas 1) Identificarea abilităților și intereselor candidatului; (Pas 2) Maparea abilităților extrase din limbaj natural la date structurate; (Pas 3) Obținerea recomandării de ocupație folosind ontologia COR; (Pas 4) Selectarea recomandărilor optime obținute din ontologie folosind o filtrare colaborativă și algoritmi de învățare automată; (Pas 5) Prezentarea datelor utilizatorului. Elementele dificile sunt implementarea pasului 2 și 3. Trebuie să mapăm abilitățile care sunt extrase din diferite surse de date și salvate în profilul JSON al utilizatorului la abilitățile standard care vor fi utilizate ulterior în procesul de recomandare, din cadrul ontologiei COR. Pentru a mapa abilitățile din limbajul natural extras din diverse surse într-un limbaj comun care poate fi înțeles de calculator va trebui să rulăm un algoritm de procesare a limbajului natural care va clasifica fiecare abilitate dată într-un cod predefinit sau o structură standard. Pentru construirea algoritmului, vom folosi un model nou de învățare automată, un transformator - o rețea neuronală care învață contextul și, prin urmare, sensul urmărind relațiile în date secvențiale, cum ar fi cuvintele din propoziție [18]. Dintre toate modelele de transformatoare disponibile, am decis să alegem BERT (Bidirectional Encoder Representations from Transformers) [19], în principal datorită comunității destul de mari de cercetători care susțin această inițiativă și pentru gama largă de modele pe care le putem găsi online și care pot fi ajustate pentru nevoile noastre. Spre deosebire de modelele recente de reprezentare a limbajului, BERT este proiectat pentru a antrena prealabil reprezentările bidirecționale profunde din text neetichetat, condiționând în comun atât contextul stânga cât și cel din dreapta în toate straturile. După ce toate datele noastre sunt structurate, putem folosi API-ul de exploatare a ontologiei COR pentru a prelua toate ocupațiile pentru care candidatul se va potrivi, interogând ontologia/ aplicând inferențe în funcție de abilitățile și domeniile sale de interes. Ontologia ne va returna o cantitate destul de mare de date, care va fi diminuată la ocupațiile cele mai relevante pentru profilul curent de utilizator prin aplicarea modelelor de filtrare colaborativă combinată cu BERT și a unui model de învățare nesupervizată K-means [20]

Pentru implementarea scenelor de Realitatea Virtuală (VR) din sistem, se va folosi motorul Unity 3D [21] și API-ul bazat pe JavaScript, WebGL [22]. Pentru obiectele 3D utilizate în simularea VR vom folosi Blender atunci când este necesar [23]. Vom avea șase scenarii inițiale în care vom expune activități din ocupațiile: chimist, proiect manager, profesor universitar, specialist în rețelistică, dezvoltator software și inginer în construcții civile. În RV se va prezenta o activitate cât mai atractivă din partea fiecărui job ce ar dura aproximativ 5-10 minute. Durata simulării și tipul de activitate prezentat trebuie să țină cont de asemenea de limitările de performanță în cazul rulării pe un browser și a utilizării unei căști VR.

Pentru partea de agent conversational, se va folosi framework-ul Pandora Bots [24], care deja este folosit în educație, ne oferă funcționalitățile și caracteristicile necesare și acceptă limba română. Acesta folosește AIML (Artificial Intelligence Markup Language - Limbaj de Markup Inteligență Artificială), un standard open-source similar cu XML.

Cerințele software ale sistemului

După o analiză a cerințelor utilizatorilor, am ajuns la concluzia că există 2 scopuri principale ale utilizării sistemului CareProfSys: (1) utilizarea sistemului pentru a primi recomandări de carieră personalizate de către elevi sau studenți; (2) utilizarea sistemului pentru a descoperi dacă o carieră recomandată este pe placul utilizatorului prin intermediul modului de realitate virtuală și a *chatbot-ului*, dar doar după ce o recomandare este oferită. Cerințele funcționale care trebuie implementate sunt cele care deservește scopurile de mai sus, într-o manieră prietenoasă pentru utilizator. Legat de performanță, cea mai mare provocare o reprezintă conexiunea între aplicația propriu-zisă și ontologie, motiv pentru care am decis crearea API-ului de exploatare a acesteia, ca să putem avea un *entry point* comun spre ontologie, cu care putem comunica prin metodele HTTP cunoscute. Din punct de vedere fiabilitate, serviciile vor rula în cadrul unor imagini de Docker [25], ce vor fi organizate prin intermediul Kubernetes [26]. Ca în orice aplicație de dimensiuni mari, costurile de întreținere nu sunt mici, deoarece fiecare microserviciu necesită un *container* dedicat acestuia. Github [27] este folosit pentru colaborare și distribuirea codului, iar ca metodologie de organizare, se folosește AGILE [28].

Proiectarea detaliată a sistemului

Se va folosi o arhitectură de tip “microservicii”, grupate pe niveluri, în care fiecare dintre module va rula ca un microserviciu în propriul său container, comunicand cu celelalte prin intermediul unor cereri HTTP: a se vedea Figura 2.

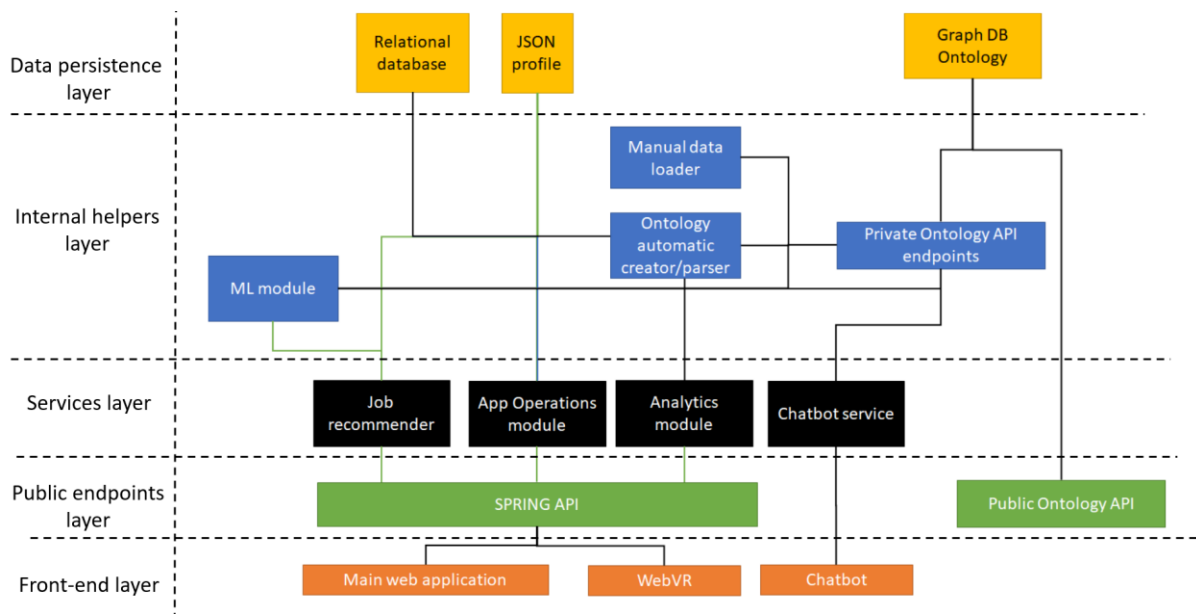


Figura 2. Arhitectura sistemului CareProfSys

Data persistence layer (Nivelul de persistență a datelor) va conține ontologia COR care va rula într-un container separat, folosind un *repository* GraphDB și va fi accesată folosind API-ul de exploatare, o bază de date non-relațională, NoSQL, MongoDB/Firebase, care va fi folosită pentru stocarea datelor din profilul virtual al unui utilizator în formatul JSON și o bază de date relațională pentru celelalte tipuri de date (e.g. cont utilizator). Internal helpers layer (Nivelul intern de execuție) va conține atât microserviciile ce acționează ca *repository* pentru ontologie, cât și serviciile create de către noi pentru a facilita o mai ușoară încărcare manuală sau automată a datelor în aplicație sau modulul de învățare automată. Services layer (Nivelul de servicii) este

locul în care vom prelua datele din diverse surse, și vom apela anumite servicii interne pentru a implementa cu succes fiecare caz de utilizare. Serviciile vor deservi atât aplicația web principală, cât și modulul de chatbot și aplicația VR. Tot aici se va implementa și algoritmul de recomandare și orice calcul semantic pe baza ontologiei. Public endpoints layer (Nivelul de endpoint-uri publice) reprezintă conexiunea dintre front-end (aplicația web, modulul Web VR și chatbot-ul) și back-end (toate serviciile de mai sus). Pe partea de front-end, arhitectura va fi una tipică pentru aplicațiile moderne, și anume un *single page web app*, pentru care vom folosi *stack-ul* clasic React + Redux [29], combinat cu Bootstrap [30]/ Material UI [31] ca și *framework* pentru stil. Pe lângă aplicația principală, vom mai avea 2 module de *front-end* independente și anume aplicația VR și *chatbot-ul*. *Chatbot-ul* va beneficia de un serviciu separat pe partea de *back-end*, ce va putea fi conectat ori la *front-end-ul* creat de către noi, ori la o pagină pe o anumită rețea socială, cum ar fi Facebook, Slack sau Telegram.

Pentru utilizarea sistemului principal al proiectului CareProfSys (accesare ontologie, obținere recomandări de ocupații, accesare chatbot) nu sunt necesare componente hardware speciale, sistemul putând fi accesat de pe orice laptop / PC conectat la Internet și dotat cu un browser compatibil. Pentru utilizarea modulului de VR a sistemului CareProfSys este necesară utilizarea unui set VR, recomandată fiind utilizarea setului Oculus Quest 2 și a unui desktop care permite acest lucru.

Activitate 1.2: Dezvoltarea ontologiei COR și a instrumentelor de exploatare

În cadrul acestei activități, s-au executat **două sarcini principale, cu scopul îndeplinirii obiectivelor (b)** crearea ontologiei COR și popularea ei cu profiluri de profesioniști, **(f)** oferirea unui instrument flexibil care poate fi personalizat cu ușurință pentru a sprijini consilierii de carieră și pe cei școlari din licee și universități și **(g)** oferirea posibilității dezvoltării rapide a mai multor alte module în viitor:

- T2.1. Dezvoltarea ontologiei COR
- T2.2. Dezvoltarea framework-ului de exploatare a ontologiei

Detalii despre rezultatele științifice ale acestor activități se găsesc în Livrabilul 3 “Ontologia COR și descrierea acesteia”, iar rezultatele obținute au fost diseminate în articolul de conferință (R1) și articolul de jurnal (R4). Obiectivele propuse pentru activitățile T2.1 și T2.2 au fost 100% atinse.

Dezvoltarea ontologiei COR

Pentru a realiza dezvoltarea ontologiei, s-au studiat aspecte ce țin de profesionalizarea muncii: identitatea profesională în industria 4.0, organisme profesionale relevante, taxonomii și standarde, s-a studiat nomenclatorul “Clasificarea ocupațiilor din România (COR), apoi s-au analizat exemple de artefacte *software-ontologii* similare. Industria 4.0, caracterizată prin utilizarea extensivă a inteligenței artificiale (IA), a condus la apariția conceptului de transformare digitală, care implică schimbări la nivel operațional și organizațional, dar și schimbări culturale prin integrarea tehnologiilor în procesele de producție și profilurile de competențe la toate nivelurile și funcțiile organizației [32]. Adoptarea IA la locul de muncă poate transforma complet munca și modul în care este organizată munca [33], ducând la apariția de noi ocupații. Ocupațiile tind să se transforme în profesii, care necesită calificarea profesională, evaluarea sistematică a competențelor profesionale și respectarea cerințelor codului de conduită profesională. *Profesionalizarea muncii* reprezintă

procesul de transformare a unei ocupații într-o profesie cu un grad ridicat de integritate și competență [34]. Profesionalizarea presupune existența unor cadre de calificare profesională, asociații profesionale care definesc și recomandă membrilor comunității profesionale cele mai bune practici, coduri de conduită profesională și certificări profesionale, pentru a face diferența între profesioniști competenți și incompetenți. Pentru ca toate aceste procese să fie aliniate și realizate corespunzător, trebuie definite standarde și reglementări, iar actorii principali de pe piața muncii ar trebui să le respecte. Totodată, aceste standarde conduc și la conturarea conceptului de *identitatea profesională*: percepția comună a oportunităților de către profesionistul asociat, incluzând mai mult decât îmbunătățirea nivelului de venit. Pilonul de reglementare și organizare a ocupațiilor este clasificarea internațională sau națională a ocupațiilor, care include atât ocupațiile profesionalizate, cât și cele neprofesionalizate. Clasificarea Internațională Standard a Ocupațiilor (ISCO conform acronimului englezesc), cu ISCO-08 ca cea mai recentă ediție, reprezintă un sistem de clasificare și agregare a informațiilor referitoare la ocupațiile prezente pe piața muncii [35]. Clasificarea Ocupațională a României (COR) este aliniată la standardele europene [36].

Clasificarea națională a ocupațiilor (COR) românească este un nomenclator care a apărut în anul 1995. COR-ul reprezintă un sistem de identificare și codificare a ocupațiilor românești indiferent de tipul și locul lor. În COR, pentru fiecare post de muncă pe piața muncii din România este alocat un cod din șase cifre. Clasificarea este revizuită periodic și este actualizată cu ocupații neincluse încă în ea. Ultima revizuire și actualizare a fost în anul 2022. La fel ca multe alte clasificări naționale de ocupații, clasificarea românească urmează structura Clasificării Standarde Internaționale a Ocupațiilor (ISCO-08, 2022). Taxonomia ocupațiilor are 5 niveluri. Clasele primelor patru niveluri au o corespondență de cod ISCO. Al cincilea nivel cuprinde ocupațiile desfășurate în economia românească care deja sunt incluse în COR. Ocupațiile nu au corespondențe ISCO, ele fiind specifice pieței muncii din România. Cel mai înalt nivel este împărțit în nouă categorii principale cu aceleași nume ca în ISCO, neluând în considerare a zecea categorie, din domeniul militar. Pentru fiecare ocupație, COR conține toate detaliile necesare despre post, inclusiv activitățile generale, contextul muncii, stilul de lucru, valorile și nevoile atașate postului respectiv. De asemenea, se precizează toate abilitățile, aptitudinile și cunoștințele pe care trebuie să le dețină o persoană pentru a putea ocupa postul respectiv. Un număr semnificativ de ocupații are conectată și fișa postului, care detaliază în continuare toate caracteristicile postului și așteptările de la un candidat [36].

Dorința de a digitaliza standardele și taxonomiile care susțin conceptul de profesionalizare și identitate profesională în industria 4.0 se reflectă prin numeroasele exemple de artefacte software [37][38]. Principalul artefact software utilizat în modelarea cunoștințelor și profesionalizarea muncii este ontologia. O specificare formală și clară a unei conceptualizări comune, care conține idei, entități, caracteristici și calități legate de un anumit domeniu. se numește *ontologie*. Folosind ontologii computaționale, se poate reprezenta în mod explicit structura unui domeniu de cunoștințe, inclusiv entitățile și relațiile pertinente care rezultă din observarea acestuia și servesc nevoilor sistemului în care ontologia este inclusă. Ontologiile ajută la analiza, partajarea și reutilizarea cunoștințelor și raționamentelor.

Scopul ontologiei COR dezvoltată în cadrul proiectului CareProfSys este descrierea competențelor profesionale pe care trebuie să le aibă un individ pentru a accesa o ocupație de pe piața muncii, ontologia contribuind astfel la mai multe aspecte

fundamentale ale managementului resurselor umane: optimizarea parcursului profesional, învățarea și dezvoltarea continuă, recrutarea, planificarea strategică a forței de muncă etc. Această ontologie va evolua mereu, deoarece mereu apar noi ocupații, domenii, activități pentru a oferi utilizatorilor informații relevante și actuale ca ei să poată ține pasul cu nevoile pieței muncii.

Toate datele despre ocupațiile disponibile în piața muncii din România sunt stocate într-un sistem conceptual descris formal de ontologia COR, a cărei conținut poate fi administrat printr-un API dedicat, construit tot în cadrul proiectului CareProfSys. Utilizatorii API-ului au posibilitatea de a căuta în baza de cunoștințe prin rularea unui șablon predefinit SPARQL.

Ontologia COR a fost creată folosind modelele de date RDF (Resource Description Framework) și RDFS (Resource Description Framework Schema), dar și limbajul OWL (Web Ontology Language). Editorul de ontologii care astăzi oferă cele mai multe funcționalități pentru crearea și manipularea ontologiilor OWL din cele disponibile este Protégé. Are o arhitectură deschisă și are un număr mare de extensii și plugin-uri disponibile și în continuare este în creștere. Acest fapt și că poate fi conectat simplu la un motor de inferențe îl face cel mai potrivit instrument pentru dezvoltarea ontologiei COR. Pentru realizarea interogărilor asupra ontologiei, am folosit clasicul limbaj SPARQL (SPARQL Protocol and RDF Query Language). Ontologia COR a fost creată pe baza conținutului aflat la adresa [36], considerându-se cele 9 grupe majore de ocupații, excepție cea de-a X-a din domeniul militar (a se vedea Figura 3). Pentru a automatiza procesul de creare a ontologiei, fiecare pagină web a fost citită automat prin funcții Python și salvată în fișiere csv, apoi fiecare csv parsat și creat clasele/conceptele în OWL. Totodată, regulile de definire a claselor ce reprezintă ocupațiile au fost create automat, prin rularea unui algoritm de învățare automată bazat pe transformatoare care vor clasifica fiecare abilitate dată la o structură standard [39]. Din toate modelele de transformatoare disponibile, am decis să alegem BERT [40], iar pentru a antrena un model de clasificare pentru sarcini, i-am furnizat un set de antrenament format dintr-un fișier CSV cu două coloane: reprezentarea în „limbaj natural” a abilității luată din descrierea ocupației din pagina web și rezultatul structurat, folosit în definirea regulei din OWL.

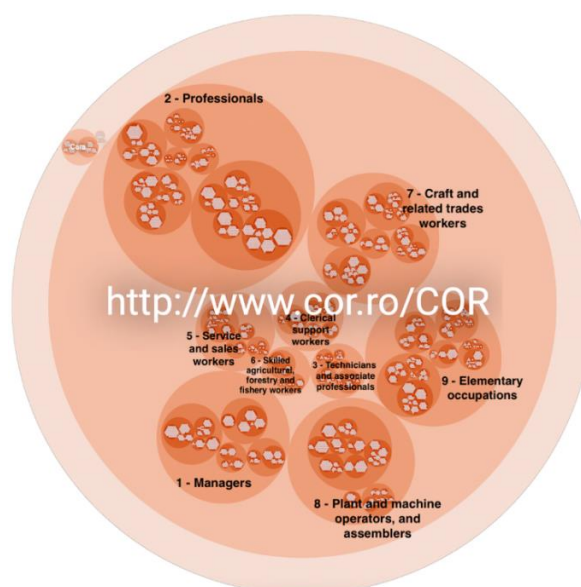


Figura 3. Principalele grupe de ocupații ale ontologiei COR, realizate cu instrumentul GraphDB.

Structura ontologiei

Taxonomia ocupațiilor din România este pilonul principal al ontologiei. Pe lângă pilonul ocupațiilor, ontologia conține încă trei piloni: a) domeniile de studiu, conectând astfel educația cu profesiile, b) caracteristicile legate de ocupație (activități generale, contextul muncii, stilul de lucru, valori și nevoi) și c) caracteristicile necesare pentru a îndeplini o anumită poziție (adică abilități, aptitudini și interese). Toate caracteristicile au o descriere, iar importanța caracteristicii sau a domeniului asociat pentru a putea accede la o anumită ocupație este de-asemenea indicată. Această importanță este etichetată cu una dintre următoarele valori: *indispensabil* cu 81-100 de puncte din 100, *foarte important* cu 61-80 de puncte, *destul de important* cu 41-60 de puncte, *important* cu 21-40 de puncte și *mai puțin important* cu 1-20 de puncte din 100. Aceste valori sunt de fapt indivizi ai clasei *Valoare*.

Nu există o relație ierarhică inerentă între oricare doi piloni dintre cei patru. Oricare concept poate deveni central în funcție de sistemul său contextual în care este folosită ontologia: o organizație care oferă training profesional pentru indivizi care caută job s-ar putea concentra pe domenii de studii, dar o companie de recrutare ar pune mai mult accent pe locuri de muncă. O ocupație poate fi privită și ca o colecție de abilități, aptitudini, etc. dar, în același timp, o singură abilitate poate fi definită prin toate locurile de muncă unde aceasta e necesară.

În Figura 4 este prezentată structura de bază a ontologiei COR cu cei patru piloni ("Competențe și interese", "Trăsături ocupație" în "Caracteristici", "COR" și "Domenii"). Ultima clasă în listă este "Valoare", ceea ce reprezintă importanța fiecărei caracteristici. Pentru fiecare ocupație este nevoie de codul COR, un URI, descrierea și denumirea acesteia și corespondența în ISCO.



Figura 4. Structura de bază a ontologiei COR

Pentru a conecta domeniile, caracteristicile ocupației și caracteristicile posibilului candidat pentru un post specific, sunt dezvoltate proprietăți de obiecte, a se vedea Figura 5.

Scopul ontologiei este corelarea competențelor profilului unui individ cu cele ale unei ocupații, ajutând la mai multe aspecte: (1) Companiile vor găsi candidații potriviți pentru locurile libere; (2) Indivizii pot găsi acele ocupații pentru care au toate/ suficiente competențele necesare; (3) Indivizii pot afla ce aptitudini trebuie să își mai dezvolte pentru a aplica pentru o anumită poziție; (4). Sistemele de educație (publice și private) vor fi conștiente de schimbările pe piața muncii.

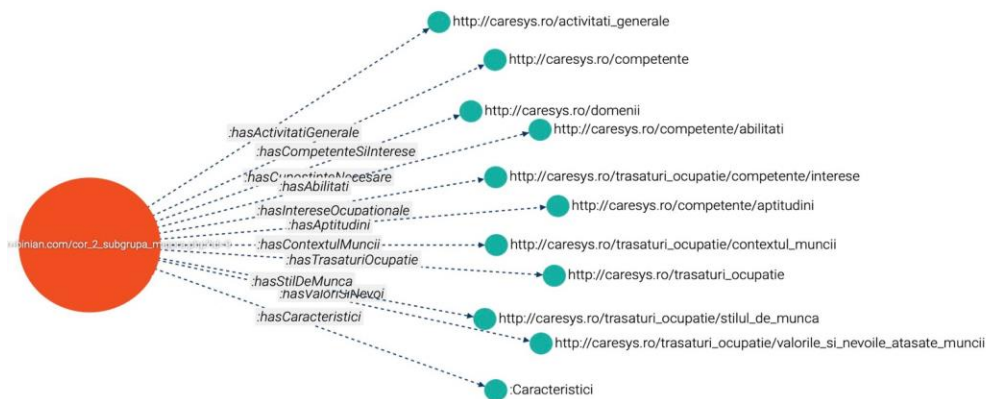


Figura 5. Relații în ontologia COR

Dezvoltarea framework-ului de exploatare a ontologiei

Framework-ul de exploatare a ontologiei este sub forma unui API (Application Programming Interface/ Interfață de programare a aplicațiilor), ce va servi atât utilizatorii publici prin intermediul *endpoint-urilor* publice, cât și celelalte microservicii ale sistemului CareProfSys prin intermediul *endpoint-urilor* private. Pentru implementarea API-ului de exploatare s-a folosit modelul clasic pe 3 straturi: a se vedea modalitatea de funcționare a API-ului în Figura 6.

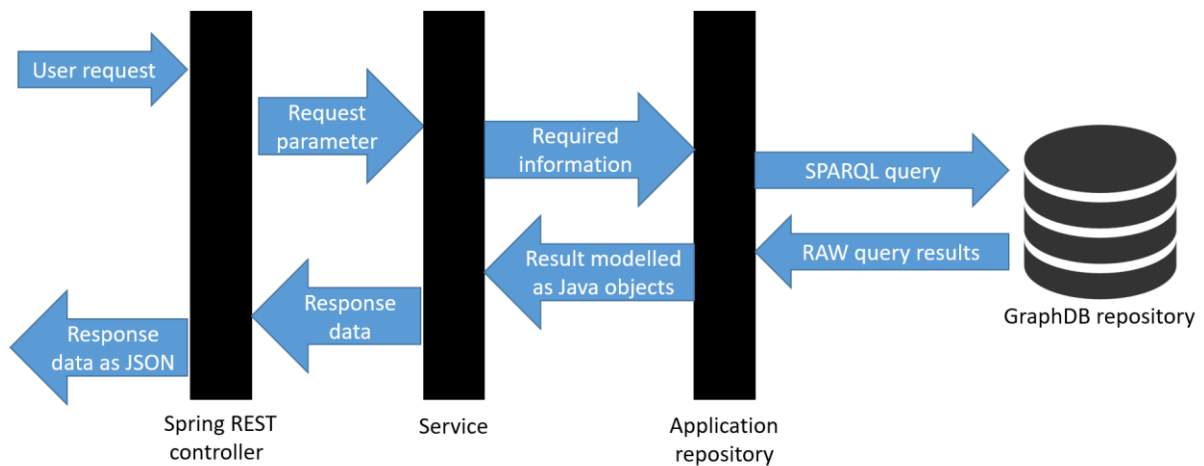


Figura 6. Workflow al API-ului de exploatare a ontologiei COR

Metodele publice expuse de către API sunt următoarele: extragerea codului unui anumit job din COR; extragerea tuturor job-urile din COR, împreună cu codul lor aferent. Metodele private sunt următoarele: adăugarea unui nou triplet în ontologie, modificarea unui anumit triplet din ontologie, ștergerea unui anumit triplet din ontologie, extragerea tuturor detaliilor legate de o anumită ocupație. API-ul este dezvoltat în Java.

Sumar al progresului

Livrabile realizate

În cadrul primei etape au fost realizate 3 livrabile: (1) Livrabil 1- Raport de analiză și cerințele utilizatorilor sistemului - în cadrul activității A1.1; (2) Livrabil 2 - Raport privind cerințele software ale sistemului și proiectarea acestuia - în cadrul activității A1.1; (3) Livrabil 3 - Ontologia COR și descrierea acesteia - în cadrul activității A1.2. Toate livrabilele sunt disponibile în anexe.

Scopul primului livrabil este de a preciza rezultatele analizei efectuate pentru stabilirea funcționalităților sistemului CareProfSys. Această analiză a inclus atât studierea amănunțită a literaturii de specialitate și a sistemelor asemănătoare existente, cât și crearea de sondaje online cu liceeni și studenți (principalii actori ai sistemului) pentru a stabili interesele acestora. Rezultatele obținute sunt apoi concretizate în definirea cazurilor de utilizare și a cerințelor utilizatorilor.

Scopul celui de al doilea livrabil este de a prezenta detaliile tehnice în vederea bunei implementări a funcționalităților sistemului CareProfSys. Se vor justifica alegerile tehnologice pentru crearea aplicației web, a ontologiei, a motorului de recomandare, a extracției automate a datelor utilizatorilor, precum și crearea *chatbot-ului* și a scenelor de realitate virtuală. Documentul realizează și trecerea în revistă a cerințelor nonfuncționale, a componentelor hardware și software, precum și proiectarea arhitecturală a sistemului. Descrierile sunt însoțite de scheme și diagrame sugestive, pentru a facilita înțelegerea sistemului de către persoane cu *background* tehnic.

Scopul livrabilului al treilea este de a identifica aspecte relevante privind profesionalizarea muncii, cu accent pe artefactele software folosite în această direcție, de a descrie dezvoltarea ontologiei ce reflectă Clasificarea Ocupațiilor din România, din cadrul proiectului CareProfSys, precum și API-ul de utilizare a acesteia.

Indicatori de rezultat și diseminarea rezultatelor

Indicatorii de rezultat al etapei 1 sunt: 3 livrabile realizate (disponibile în Anexe), 3 articole la conferințe și 1 articol aflat în evaluare la un jurnal ISI (disponibile în Anexe) și realizarea website-ului proiectului, depășindu-se astfel indicatorii propuși (3 livrabile, 1 articol științific transmis către un jurnal, 1 articol științific transmis către o conferință și website-ul proiectului).

Rezultatele au fost diseminate prin lucrări publicate în *Proceedings-urile* a 3 conferințe internaționale de prestigiu (*Proceedings-uri* care vor fi trimise spre evaluare și incluse în Web of Science) și 1 articol în evaluare în revistă indexată ISI.

(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, Sevilla (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, Sevilla(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, Sevilla (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.

Toate lucrările conțin *acknowledgment-ul* proiectului.

Totodată, proiectul a fost diseminat prin *website-ul* acestuia, disponibil în două limbi (engleză și română), pentru a asigura vizibilitatea rezultatelor (<http://careprofsys.upb.ro/>)

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

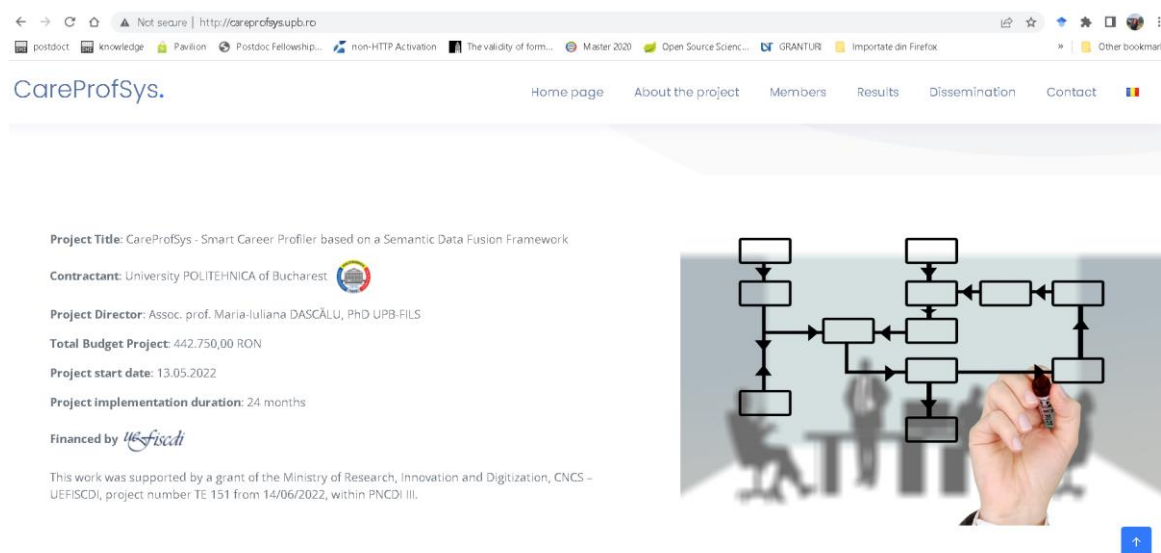


Figura 7. Website proiect CareProfSys: <http://careprofsys.upb.ro/>

Concluzii

Proiectele de cercetare interdisciplinare au o probabilitate și un impact ridicat de risc. Pe lângă riscurile caracteristice unui astfel de proiect, se pot determina următoarele riscuri pentru proiectul CareProfSys: (1) combinarea noilor tehnologii (semantică, socială, realitate virtuală, agenți conversaționali); (2) performanța algoritmilor de recomandare și analiză a datelor mari; (3) număr insuficient de utilizatori/ cantitate suficientă de date. Atenuarea se realizează prin: (1) existența unei echipe interdisciplinare, atât cu experți tehnici, cât și psihologi/consilieri de carieră, precum și prin testele de integrare continua, cât și arhitectura bazată pe microservicii; (2) accent pe experimentare înainte de alegerea algoritmilor utilizați în varianta finală a sistemului.; (3) pe lângă web scrapping-ul aplicat platformelor sociale, am creat și un

chestionar pentru obținerea datelor necesare creării profilurilor profesionale de tip LinkedIn. De-asemenea, vom asigura confidențialitatea datelor prin utilizarea mijloacelor adecvate de stocare și procesare și prin securizarea componentelor hardware și software, precum și a interfețelor de comunicare dintre acestea.

Obiectivul principal și suprem al oricărei platforme online de recomandare a ocupațiilor profesionale este de a identifica direcția corectă în alegerea profesiei și, implicit, a studiilor necesare, prin urmare, aceasta ar trebui să fie cea mai relevantă caracteristică pe care CareProfSys este conceput să o îndeplinească. Livrabilele din această etapă, prin maniera clară și unitară de prezentare cazurilor de utilizare, prin tabele ușor de urmărit și înțeles, prin diagramele de arhitectură și detaliile tehnice oferite constituie baza de lucru pentru etapa următoare, de dezvoltare a sistemului.

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”.

Bibliografie

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.scripgroup.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/RaportdeconsiliereINEM2022.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 transformer, <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. WebGL https://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. Kubernetes, <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. Nejdli, *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).

Anexe

Anexa 1. Livrabil 1 - Raport de analiză și cerințele utilizatorilor sistemului

Anexa 2. Livrabil 2 - Raport privind cerințele software ale sistemului și proiectarea acestuia

Anexa 3. Livrabil 3 - Ontologia COR și descrierea acesteia

Anexa 4. R1 - Articol in extenso - conferința ICERI

Anexa 5. R2 - Articol in extenso - conferința ICERI

Anexa 6. R3 - Articol in extenso – conferința IBIMA

**Anexa 7. R4 - Articol in extenso – jurnalul ISI Rev. Roum. Sci. Techn.–
Électrotechn. et Énerg.**

DIRECTOR PROIECT,
DASCĂLU, MARIA-IULIANA

