Electronic Profiling in CareProfSys System for Career Recommendation

Iulia-Cristina STANICA¹, Ioan-Alexandru BRATOSIN², Dan-Alexandru MITREA³, Constanța-Nicoleta BODEA⁴ and Maria-Iuliana DASCALU⁵

¹,²,³,⁵Department of Engineering in Foreign Languages, University Politehnica of Bucharest, Splaiul, Bucharest, Romania
⁴The Bucharest University of Economic Studies, Department of Economic Informatics and Cybernetics, Calea Dorobanți, Bucharest, Romania

Correspondence should be addressed to: Iulia-Cristina STANICA; iulia.stanica@upb.ro

Received date: 4 November 2022; Accepted date: 7 March 2023; published date: 27 March 2023

Academic Editor: Mircea Georgescu

Copyright © 2023. Iulia-Cristina STANICA, Ioan-Alexandru BRATOSIN, Dan-Alexandru MITREA, Constanța-Nicoleta BODEA and Maria-Iuliana DASCALU. Distributed under Creative Commons Attribution 4.0 International CC-BY 4.0

Abstract

Recommender systems represent tools capable of providing suggestions based on large datasets or collections of objects in the user’s area of interest. Career recommendations have become more and more popular, as they are able to offer guidance and help people build their most suitable future professional path. Without any doubt, electronic profiling can improve the automatization and accuracy of the recommendation process. Our paper presents the process of electronic profiling used in the CareProfSys research project; a smart career profiler based on semantic data fusion. The research methodology includes the analysis of a job survey filled in by university students, containing similar information as that included in a LinkedIn user profile. In a detailed case study, we further perform the analysis of the web and multimedia systems developer job which is of interest to most respondent students. All results and conclusions will be used in improving the construction of the user profile and the recommendation process of the CareProfSys.

Keywords: career recommendation, electronic profiling, ontology, data scraping, web development

Introduction

In the last few years, recent graduates have benefited from the rapid growth of recruitment platforms making their career choice more complex. Job seekers are required to spend more and more time searching and navigating to find the right job because there is an excess of information online and this task has become boring for many applicants. Even more concerning is the fact that many graduates do not really comprehend their skills in terms of advantages and limits, especially since these skills will have to be improved in the future thanks to artificial intelligence. Therefore, the way each employee works will also evolve. It is possible that, in the future, some jobs will require skills that are not taught in schools which is why acquiring new skills through lifelong learning (Pew Research Center, 2017) is essential to cope with in a modern world, both at work and in private life (Laal, 2011).

Our project, Smart Career Profiler based on a Semantic Data Fusion Framework (CareProfSys) uses ontology inference, conversational agents, and virtual reality to recommend career options to users using advanced user profile analysis. The system will extract and process data from CVs, studies, social network profiles or psychological tests. The data are grouped into nine major groups of occupations according to the COR occupation ontology (Romanian occupation classification) and ESCO (European classification of skills, competences, qualifications, and occupations). Therefore, job recommendation systems play an increasingly important role because they can offer personalized advice with high precision.

The current paper presents the process of electronic profiling suitable for job recommendations. The following two sections present state of the art related to the creation of profiles and systems for career recommendations. We will further focus on the CareProfSys job recommendation system, emphasizing the process of automatic data extraction for electronic profiling. Finally, we will detail the results of a job survey conducted with university students, having the purpose of clarifying our view on their approach to writing a personal profile.

Profiling as a Tool in Career Recommendations

The employee’s profile has become more difficult to create as social networks have started to restrict access to certain information from the user’s profile. As the data are accessible through APIs, each platform has its own rules for accessing information. Burke et al. propose a method of extracting skills from text documents to generate a list of professional skills with the role of hiring and managing employees in a company. By filtering the content, posts are analyzed to identify the information that is of interest, however, they do not take into account whether the user’s preferences will change in the future (Kivimäki, 2013).

There are four different ways through which recommender systems suggest the best match for a user: content-based, collaborative, demographic, and hybrid filtering (Mohamed et al., 2013). Research by Deshpande and Karypis (2004) is just an example that user-based collaborative filtering is widely used for building commercial recommender systems, but, for this, the system must be scalable for very large numbers of users. As shown by Burke (2007), after the user profile modules are determined, the hybrid recommender system will combine multiple hybridization techniques to achieve a synergy between them. Job referral systems are significantly different from traditional referral systems, as
job posting targets an employee or a few employees, while traditional systems recommend a large number of users (Kamehkhosh et al., 2017).

**Recommender Systems**

The job recommendation system offers personalized recommendations based on classification rules. Moreover, the job recommendation system helps companies to recruit online employees who meet the requirements of a job seeker; thus, candidates are recommended to recruiters. Almalis et al. (2015) proposed a job recommendation algorithm that extends and updates the Minkowski distance to match people to jobs based on content analysis of job postings and candidate resumes. Lee and Brusilovsky (2007) proposed an adaptive job recommendation system that includes four different types of recommendations for the user. Their system dynamically updates its database by collecting new information on the basis of which two ontologies are generated, one for jobs and another for companies. Puspasari et al. (2021) used the K-Means Clustering method to group applicants and vacancies based on skills, salary, location and other skills, and each variable has an assigned weight. Their job recommendation method will display job vacancies in descending order based on a percentage that is calculated based on the level of match. Appadoo et al. (2020) developed a job recommendation system using Machine Learning with data collected in the past to make a prediction for a candidate if he is good enough for a job. Their system provides the HR department with a sorted list of all candidates based on a calculated score. On the other hand, Zarandi and Fox (2009) presented a hybrid ontology-based approach for matching job applicants and job advertisements. They show that this method can also be used to improve skills management systems within an enterprise. Chou and Yu (2020) propose a job recommendation system based on machine learning, text mining and big data to analyze personal competitiveness and personality traits according to CVs submitted by candidates. All this aims to help candidates to identify the most suitable job, and for companies the recommendation system creates a list of possible talents among the candidates.

Recommender systems have gained ground over traditional methods because the application process and participating in a job search are time-consuming for both the applicant and the job seeker. A recommendation system is used to reduce a large amount of information by generating personalized suggestions (Malinowski et al., 2006). Most researchers propose recommendation algorithms for finding a job, but few consider that a graduate needs to be guided towards a career path that fits their profile (Stanica et al., 2022), (Dascalu et al., 2022). Nonetheless, recommendation systems become effective if they can provide actionable feedback and guide through the improvement processes of a graduate who aims to achieve their career goals (Ghosh et al., 2020).

Looking for a job can create disappointment for the new generation in case of a failure and can mark them for the following interviews. For this reason, it is important that a career recommendation system has a high accuracy. Generation Z is more capable of interacting with technology, but it is also a more sensitive generation, being concerned with social and environmental issues (Wong, 2021). Systems that recommend a career path must also consider millennials as potential users who want to have a balance between work and private life in order to consider a job attractive (Firfray and Mayo, 2017).

**CareProfSys System - data extraction for profiling and job recommendations**

The CareProfSys System has the purpose of creating a job recommendation platform which uses a personalized user profile to
identify suitable career paths. Data are extracted automatically from social media or CVs as part of dynamic profiling, while custom information introduced by the user (in surveys, dedicated fields or personality tests) creates the static profile. Recommendation algorithms are then applied to match the user's profile information to the data extracted from job postings. The relevant data sources and functionalities of the platform have been established after conducting a survey with over 300 high school and university students. In order to cover a wide range of occupations and domains and create the proof of concept of our system, we selected 6 jobs from the official Classification of Romanian Occupations, which will also be represented in virtual reality, in order to offer a realistic simulation of the job's tasks to its possible future candidate. These jobs include: chemist, project manager, web developer, university professor, computer networks architect and civil engineer.

**Data collection - user survey**

A user job survey was created with the scope of providing more data for our matching algorithms. The main reason for this is the limitations imposed by LinkedIn on the number of profiles that can be accessed in a certain timeframe which eventually leads to a shortage of LinkedIn user profiles that can be extracted for our system. To allow easy merging, the survey was created with a similar composition as that of LinkedIn. First questions involve information regarding the user's current workplace and their future career path to see if they are aligned to the first 6 types of jobs we decided to take in consideration for the proof of concept of our system. Next questions are related to education, technical and non-technical skills, work experience, known foreign languages, certifications and projects. All these questions are similar to the sections provided by LinkedIn's user page. The main advantage of using a survey is that we will obtain data in an easy to interpret format with some of the items already displayed in statistics that will give us relevant information of the users interested in that domain. This information will be further used alongside the data extracted from LinkedIn, Europass CV, TikTok and Instagram. The results of this survey are further described and interpreted in the "Data analysis" chapter.

**Automatic data extraction from CV and social media**

For the data extraction process, we use multiple sources such as LinkedIn, TikTok, Instagram, the previously presented survey and Europass CV. LinkedIn is one of the primary sources because it holds user information pertaining to the construction of a user profile, including: current workplace, previous experience, education, skills, projects. The data extraction process is accomplished with Python scripts that use “Selenium Webdriver” and “BeautifulSoup4” libraries. These 2 tools are intended for the creation of web automation testing tools as they behave as a user that locates specific elements and accesses that information contained in the specific tag or activates the element in case of button interactions. The main method of finding the required elements containing multiple similar types of data is by using a combination of tag identifiers and their attributes; in case of single specific elements, the first method is with the use of "xpath" that specifies the exact location of the element on the webpage. This approach works as intended on platforms such as TikTok or Instagram, but, in some cases, they are unusable on LinkedIn as it has a responsive web design and each refresh changes the "xpath" of the element. To address this problem, we use a combination of multiple tag identifiers and children locators until we obtain a single result per element as intended. The collected data are then printed in a readable format as a CSV file.
The collected data are necessary for enhancing the search or the matching algorithms between experience, education and skills that were obtained from the scraping process. From the first platform, LinkedIn, the script extracts user defining information: name, current workplace, location and saves it in three separate columns; we then select the most relevant recent work experience, skills and experience, each section having separate columns in the CSV for easier classification; the last part pertains miscellaneous data like projects, “about” section or interests all together in one column as it is low priority for our project’s purpose. Constant improvements of the extraction process are being done, especially on the navigation part; the script was made for a stable Internet connection, in case the connection has high latency, the script might try to execute a command before the page is loaded that would result in an exception management case or a full stop. To prevent this issue, we optimize the code to wait for the page load event to make sure all required elements are displayed on the page before trying to access them, instead of using an arbitrary time wait value.

Our tool extracts data regarding job offers and company data from LinkedIn as well. The script base code was adapted for this requirement as the location, format and way, to access the data changed completely when compared to the user page. For the user scraping process, it first saved all the profile links found on the current page then opened each one of them in a new tab; for jobs, this is not required as they have a secondary display window on the right side when an interaction is triggered with the section of the webpage that contains the job. To extract the data of the jobs, we use a function to focus the view on the current item and locate the items by tags or class name; we have 4 columns in the CSV that store the respective data: “job_title”, “job_location”, “job_type”, “job_details”. We first take the title of the position then the geographical area where this position is located, if it’s a full time or part time job, if they are actively recruiting or not and finally all job details that include skills, tasks involved, responsibilities and benefits. These are required for matching the user profile skills and previous experience or education with the available job offerings to suggest an appropriate job.

Data extraction for TikTok and Instagram is not as difficult as for LinkedIn as the amount of relevant information is significantly lower. The script extracts the username, user title, description, follower count, post count and posts’ descriptions; this information is then stored in separate columns in a CSV file. TikTok data are handled in a similar fashion as the Instagram one since the 2 social media platforms have similar structure. The generated file contains user title, sub-title, following, followers, likes count, description and “TikToks” descriptions in different columns. The main parameters for the data scraping are hashtags and the number of pages to extract data from.

Multiple tests were done on all social platforms. A test case for LinkedIn consists in the extraction of the first 50 “chemist” profiles. For this we need to insert the keyword “#chemist people”, and select the number of pages, in this case 5, after this step the script automatically opens a new browser tab, logs in with predefined credentials and uses the specific keyword to start the search, it then takes all profile links from each page and opens them individually to extract the profile data and save them in a CSV. When the program receives an input to search for jobs, for instance “#chemist jobs”, it will search for job opportunities in a similar fashion; the only difference is that instead of saving the links for each job, it opens them directly in a preview sidescreen, which leads to an easier data extraction with less time consumption per processed item. For TikTok and Instagram, we
do not use pages as parameters because these social platforms are being scrolled down continuously so we can use "screen heights" as a loop condition. Using this method, we take all profiles currently displayed then we scroll down until the last link analyzed becomes the first being displayed, we exclude it and take the new list on screen to take the data.

For the Europass CV information extraction is similarly done with Python programming language and two libraries: one meant for text extraction and another for table data extraction. The CV data can be compared with the LinkedIn data so that we fill in the gaps or vice-versa to obtain a complete user profile and infer different relations between the user knowledge, languages, and work-experience. The data will be compared as well with the one obtained from the previous form that has a structure close to the one of LinkedIn, so that more data for profile creation are provided.

**Job Recommendation Algorithm**

The recommendation process of the CareProfSys tool includes several distinct phases:

- the extraction of the users' skills and interests (manual completion, CV parsing, social media scraping)
- mapping content expressed through natural language to structured data representing skills
- using our custom job ontology based on COR to find suitable professions for the candidate
- optimize ontology matches by including collaborative filtering techniques (only when the database is mature enough, so that similar user profiles can be identified)
- show the recommended jobs to the user (in an attractive manner, including virtual reality scenes with gamified job tasks' simulations).

The recommendation algorithm is in constant improvement, since users of different ages can have different styles of filling a social media profile, different technical and social skills, and, therefore, need an adapted recommendation process.

**Data Analysis**

In order to get information about the manner in which students fill in professional information, we decided to conduct a survey similar in structure to a LinkedIn user profile. The job survey was filled in by 50 university students - 78% of them are currently following their bachelor studies, 18% their master's studies and 2% are studying to get their PhD. The main goal is to analyze their education and work experience, in connection with the 6 jobs selected as proof of concept for the CareProfSys project. The one job getting the most responses (both in terms of current workplace or envisioned workplace) will be further analyzed in the Case study subsection.

**Survey Results**

When asked about their current occupation, the majority of the respondents declared that they are students and do not have a job (40%), followed closely by those working in the software development field (38%). Other jobs with low percentages include computer networks specialist (10%) or data analyst (4%) (Figure 1). Some of the companies where the respondents have their current workplace are: Huawei, IBM, Info World, Goodyear, Dell, Finastra, Societe Generale or Cegedim.
When asked about their preferred future career path, once again, the majority of people (28%) chose the web and multimedia systems developer job. The answers (Figure 2) cover this time a wider range of areas, including jobs such as construction engineer, project manager, computer networks specialist, chemist, university teacher. It is notable that 34 out of 50 students (68%) will see themselves working in a software-related field in the future. The question also offered the option of mentioning if they did not want to change their job in the future. Only 7 respondents mentioned they would like to keep their current job, which represents 23.3% of all students that currently have an employment. These statistics emphasize the importance of a job recommender system, since most students are not entirely content with their current workplace. Respondents that desire to change their current job usually want to remain in a technical specialty related to computer science (from software development to cybersecurity or AI development), but there are also situations when they would like to change completely their path, to work in education (university teacher), project management or even entertainment (music / media).
As it can be observed in figures 1 and 2, the web and multimedia systems developer job is the most frequent one among the 6 ones selected as case study, both as the students’ current job and their envisioned one. Since it is of interest to most of the students of a technical faculty, we decided to analyze it in more detail. The current subsection will focus on analyzing the survey answers given by students interested in this particular job.

In order to collect similar information as that included in a LinkedIn user profile, students filling in the survey were asked to describe themselves in a few words (similar to an “About” section), write details about their education, previous work experience, hard and soft skills, foreign languages, licenses and certifications, volunteering and previous projects they worked on. It is worth mentioning that all these were open questions, leaving a huge liberty to the respondents. The About section offered various results, showing that people have very different ways of characterizing themselves when no indications are given. Software-centered students described themselves as ambitious and hardworking, mentioned extracurricular activities and summer internships, talked about their technical skills and passions (cybersecurity, game or mobile development, foreign languages). It is worth mentioning a particular student who was proud that they managed to gain some relevant IT experience, by balancing both university studies and employment, despite being discouraged and disregarded by their highschool teachers. On the opposite side, another student mentioned that despite developing their technical skills in the computer science domain and even having a software job at the moment, they figured out that this is not their right career path and tried to develop their artistic skills in order to have a career reprofiling in the future. These two opposite examples prove that students need better guidance and career orientation during their youth in order to make sure that their chosen path is the right one, matching their skills, personality and abilities.

Students were asked to fill answers related to their education - the last finished education cycle (high school baccalaureate, bachelor, masters or PhD), as well as their current one. It is worth mentioning that open questions do not work well when precise information is required. Students’ answers had the tendency of containing less details than required and usually included more details in the case of their current faculty and specialization. Among the students who mentioned their interest in having a web and multimedia systems developer-related job, 85.3% are currently studying to get their bachelor degree, 11.8% are master’s students and only 2.9% PhD ones.

In the Work experience question, students were asked to fill in their entire work experience, including company name, job title, duration, period and location. Most answers include one or two jobs per student, suggesting that most of them are at the beginning of their professional journey and that they prefer stability at the workplace over frequent job changes. In terms of Technical Skills, the wide range of abilities they acquire at the faculty or during jobs and internships are being portrayed: programming languages, OS, database tools, algorithms or various software working environments (NodeJS, C++, C#, PHP, Java, Linux, SQL, Dart, Typescript, JavaScript, Swift, Python, Solidity, Spring Boot, Flutter, Angular, React, Docker, Kubernetes, Android, Firebase, Firestore, MongoDB, Elasticsearch, AWS Mobile Development, Hybrid cloud, Unity, Azure, Matlab, Rest API, Autocad, Catia), version control and project tracking tools (Git, Jira). For the Soft Skills,
people mentioned good communication skills, effective team working, team management and coordination, good capacity of integrating and working in multicultural environments, task prioritization, punctuality, desire to learn, seriousness, motivation, creativity, leadership, responsibility, curiosity, optimism, perseverance, time management, problem solving, using resources in a practical way, multitasking, ability to work under pressure. It is worth mentioning that there are only 6 people out of 34 that did not mention good teamwork or collaboration as one of their soft skills, a fact which is not surprising as students are often asked to work in teams for various university projects or extracurricular activities. 47% of the respondents interested in a software career mentioned they possess at least a communication-related soft skill.

When asked to fill in their spoken foreign Languages, all respondents mentioned they know English, as it is generally required for working in the software field. 67.65% also know French, while other languages are known by low percentages of students (German - 14.7%, Spanish - 11.37%, Italian - 8.82%, Arabic - 2.94%).

The last three questions were optional, so they include less information collected. For Licenses and certifications, multiple students hold a language certificate (Cambridge, IELTS or Delf), various technical certifications (Microsoft Certifications, European Computer Driving Licence - ECDL, International Software Testing Qualifications Board - ISTQB, Cisco Certified Network Associate - CCNA), Erasmus study mobilities and youth exchange programmes. Volunteering activity is less represented, including only a few students mentioning their actions, yet with a wide range of actions: members in students' associations, Red Cross, teaching primary school students with disabilities, Waste Awareness Youth Exchange, Reaching out Romania, organizers in music festivals or other cultural events, organizers and promoters of the Hardcore Entrepreneur programming competition, waste collectors, blood donors. Finally, students were asked to mention relevant research or development Projects they worked on. Only a few personal projects were mentioned (Sustainable Web Server and Cloud solution on a Raspberry Pi 4, OpenChess, various e-commerce web applications), yet without any details or relevant description which could be useful for a job recommendation or a recruiter.

Conclusions

In this paper we presented the approach used in CareProfSys for creating a complete user electronic profile usable in a job recommendation process. Our research will help students find their most suitable future jobs, matching their skills, personality and interests.

A survey collecting data similar to a LinkedIn user profile was conducted with technical students in order to see the manner in which they fill in their professional information, as well as their interest for the 6 jobs selected as a case study.

The next step of our research will involve the improvement of the students' survey, so that its collected data can be merged with the data automatically extracted by the scraping tools. A survey following the same structure as a user’s LinkedIn profile can provide useful information when analyzed as a case study on a low number of respondents. However, open questions and optionality can lead to data sparsity and inconsistency - people tend to give shorter answers and often miss some details that they were asked about (the most conclusive example being the question requiring all details related to students' work experience, duration, date, yet most people resumed writing their job title and company,
as they only considered those to be important. Structured data collection (lists with options, checkboxes), compulsoriness of all fields and data validation are thus preferred, in order to ensure a high correspondence between manually collected and automatically extracted data, and, thus, create a complete user profile for a career recommendation platform.

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".

References