

Study Report of using Web Crawling, Scraping, and Text Mining, Investigated the Medical Informatics Labour

Rajia Khatun*

Department of Bioinformatics, Islamic International Medical College, Islamabad, Pakistan

Abstract

Employability is a major goal of higher education, according to the European University Association (EUA). As a result, a competency-based approach to education is critical. The representation of a standardized job profile in the field of medical informatics, based on the most common labour market criteria, is critical for identifying and communicating the learning goals associated with these capabilities. We took numerous job adverts from the job marketplace to find out what the most prevalent needs were. We used web crawling, web extraction, and text mining as part of a programme we wrote in R using the “rvest” library. Many job adverts remained after we removed duplicates and filtered for occupations that required a bachelor’s degree, from which we retrieved qualification terms. Professional

competence, soft skills, teamwork, procedures, learning, and problem-solving abilities were divided into six categories. Employee soft skills studies have yielded comparable findings. Programming, experience, project, and server were the most often used terms. The value of experience is our second significant conclusion, emphasizing the importance of practical skills. Surveys and narrative accounts were employed in previous investigations. This is the first-time web crawling, web extraction, and text mining have been used in a study. Soft skills and specialist expertise are given equal weight in our studies. The findings of this study could be useful in the development of medical informatics curriculum.

Keywords

Medical Informatics, Education, Challenges, CBME

Correspondence to:

Rajia Khatun,

Department of Bioinformatics,
Islamic International Medical College,
Islamabad, Pakistan
Email: khatun.ra2@gmail.com

Citation: Khatun R (2022). Study Report of using Web Crawling, Scraping, and Text Mining, Investigated the Medical Informatics Labour. *EJBI*. 18(3):30-31.

DOI: 10.24105/ejbi.2022.18.3.30-31

Received: 05-Mar-2022, Manuscript No. ejbi-22-57645;

Editor assigned: 07-Mar-2022, PreQC No. ejbi-22-57645(PQ);

Reviewed: 21-Mar-2022, QC No. ejbi-22-57645;

Revised: 24-Mar-2022, Manuscript No. ejbi-22-57645(R);

Published: 31-Mar-2022

1. Introduction

The European Higher Education Area (EHEA) is defined by the European University Association (EUA) as a space for global cooperation in higher education. Its members’ principal objectives are major changes and the exchange of educational tools. As educational aims, easier incorporation into the labour market, mobility within in the EU, and knowledge have been established. A continual improvement strategy has been created, as well as criteria and rules, to ensure the quality of specific courses [1].

As industries, society, and the environment become more complicated, the study reveals that graduates need skills in systemic thinking and problem solving, critical thinking, group collaboration, multidisciplinary work, planning/realization of creative initiatives, communication, and media use. CBME (Competence-Based Medical Education) is a collection of principles and methodologies for teaching medical students that has been thoroughly examined. CBME, when implemented efficiently and dynamically, has the potential to improve all

medical training programmers, allowing medical practitioners to better serve their patients. Despite the fact that many research have sought to establish competency frameworks, few have focused on competence-based education in computer sciences, particularly in medical informatics [2].

Web data is becoming a more valuable source of information as the quality and amount of data accessible for automated retrieval improves. Using the rvest and tidy text packages, we constructed a script to extract the data for our investigation. We employed a combination of proven methods of content analysis and text mining for document and web pages, respectively, to determine if employers’ expectations are reflected in job postings. Preparation, organization, and reporting are the three stages of a content analysis [3].

The gathering of representative data that can aid in answering the research question is handled in the preparation phase. Employment listings from various online job portals are a good data source, according to us. We gathered suitable job adverts by applying web crawling. The process was repeated three times, yielding three distinct total job posts each time. To begin, we

used the job IDs in the adverts to remove several duplicates within the extracted timeframe. We then removed approximately around hundred entries for jobs that did not require a bachelor's or master's degree, leaving a final group of around hundred fifty job posting documents with an HTML document structure containing some data fields, such as Unique job ID, Company name, Title, The applicant's qualifications, Job description, and so forth [4].

We used a word cloud containing the top organizations looking for staff to run a frequency table of company names to validate the data. To guarantee that the samples contained only employment adverts from the healthcare industry, we employed an inductive content analysis approach. We created the following business area categories by identifying relevant employers using the company description: Hospital, Foundation/Institute, Medical Science Manufacturer, Employer, Care Home, Pharmaceuticals, Operating system Manufacturer, Academy, Consultancy, Healthcare, Health Insurer, and so on. We constructed a competency matrix based on research that assessed the criteria of the healthcare information labour market during the organizational phase of our deductive methodology. A substantial amount of research has been done on the relationship between skills and employability [5].

2. Conclusion

The goal of our study was to look at the terminology connected to employability in medical informatics as part of the Bologna Process' goal of aligning learning objectives with labour market demands. Regular assessments and adjustments of curricula are required as healthcare 4.0 refashions what skills and knowledge are relevant. Our research proposes a semi-automated technique for web crawling job adverts, extracting database systems with text mining, and categorizing skills with content analyses, all culminating in a tree map graph that clearly illustrates the

required job skills. The study adds to the existing knowledge in recommendations on education in medical informatics" by proposing to broaden the existing strategic framework and employment conditions to include soft skills, by to provide current information to institutions of higher learning on how to update their school curriculum, by demonstrating the utility of the text mining approach for exploratory analysis of job advertisements, and by revealing that hospitals are the primary target group of a job advertisement. More research is needed to see how these recommendations align with the IMIA body of knowledge and the evolution of medical informatics curriculum in other countries.

3. References

1. Klemencic M. From student engagement to student agency: Conceptual considerations of European policies on student-centered learning in higher education. *Higher Edu Pol.* 2017;30(1):69-85.
2. Prisacariu A. New perspectives of quality assurance in European higher education. *Procedia-Social Behav Sci.* 2015;180:119-26.
3. Rieckmann M. (2012). Future-oriented higher education: Which key competencies should be fostered through university teaching and learning?. *Futures.* 2012;44(2):127-35.
4. Holmboe ES, Sherbino J, Englander R, Snell L, Frank JR. A call to action: the controversy of and rationale for competency-based medical education. *Med Teach.* 2017;39(6):574-81.
5. Chamberlain S, Gonzalez N, Dobiesz V, Edison M, Lin J, Weine S. A global health capstone: an innovative educational approach in a competency-based curriculum for medical students. *BMC Medical Education.* 2020;20(1):1-8.