UNIVERSITY OF AMSTERDAM

UvA-DARE (Digital Academic Repository)

Labour market information driven, personalized, OER recommendation system for lifelong learners

Tavakoli, M.; Mol, S.T.; Kismihók, G.

DOI
10.5220/0009420300960104

Publication date
2020

Document Version
Final published version

Published in
CSEDU 2020

License
CC BY-NC-ND

Link to publication

Citation for published version (APA):
Tavakoli, M., Mol, S. T., & Kismihók, G. (2020). Labour market information driven, personalized, OER recommendation system for lifelong learners. In H. C. Lane, S. Zvacek, & J. Uhomoibhi (Eds.), *CSEDU 2020: Proceedings of the 12th International Conference on Computer Supported Education : May 2-4, 2020* (Vol. 2, pp. 96-104). SciTePress. https://doi.org/10.5220/0009420300960104

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: https://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, P.O. Box 19185, 1000 GD Amsterdam, The Netherlands. You will be contacted as soon as possible.

UvA-DARE is a service provided by the library of the University of Amsterdam (https://dare.uva.nl)

Download date:09 Nov 2026

Labour Market Information Driven, Personalized, OER Recommendation System for Lifelong Learners

Mohammadreza Tavakoli[1][a], Stefan T. Mol[2][b] and Gábor Kismihók[1][c]
[1]*TIB Leibniz Information Centre for Science and Technology, Hannover, Germany*
[2]*University of Amsterdam, Amsterdam, The Netherlands*

Keywords: Lifelong Learning, Open Education Resources, Recommender Systems, Labour Market Intelligence, Machine Learning, Text Mining.

Abstract: In this paper, we suggest a novel method to aid lifelong learners to access relevant OER based learning content to master skills demanded on the labour market. Our software prototype 1) applies Text Classification and Text Mining methods on vacancy announcements to decompose jobs into meaningful skills components, which lifelong learners should target; and 2) creates a hybrid OER Recommender System to suggest personalized learning content for learners to progress towards their skill targets. For the first evaluation of this prototype we focused on two job areas: *Data Scientist*, and *Mechanical Engineer*. We applied our skill extractor approach and provided OER recommendations for learners targeting these jobs. We conducted in-depth, semi-structured interviews with 12 subject matter experts to learn how our prototype performs in terms of its objectives, logic, and contribution to learning. More than 150 recommendations were generated, and 76.9% of these recommendations were treated as useful by the interviewees. Interviews revealed that a personalized OER recommender system, based on skills demanded by labour market, has the potential to improve the learning experience of lifelong learners.

1 INTRODUCTION

The worlds of work and employment are changing rapidly in our post-industrial societies. As a consequence, matching processes between skill demand and supply are getting more and more complicated as skills dynamically evolve through an uncontrollable process (Colombo et al., 2018; Castello et al., 2014). These dramatic changes lead to a number of educational problems in relation to the gap between (dynamic) skills that job markets demand and the training that education programs offer (Smith and Ali, 2014; Wowczko, 2015; McGill, 2009). Furthermore, being up to date about actual job market skills has significant importance for individuals to remain employed or climb workplace hierarchy during active times of employment (Colombo et al., 2018; Khobreh et al., 2015). Notably, in order to mitigate mismatches between education and labour markets, we need to 1) understand the dynamic nature of labour markets, which requires the deconstruction of jobs into required skills, and 2) match those skills to relevant learning content.

In order to tackle the first problem, governments and international organizations have created a number of occupational and skill taxonomies to provide structure for job-seekers and employers about skill components of jobs (e.g. ESCO, ISCO, O*NET). However, there are obstacles limiting the usefulness of these taxonomies, such as keeping their information updated (Djumalieva and Sleeman, 2018). At the same time, researchers attempt to build ontologies to provide accurate representations of jobs and skills (e.g. (Sibarani et al., 2017)), and machine learning models to capture information from rich, text based, labour market data sources, like job vacancy announcements (Colace et al., 2019; Boselli et al., 2018b; Boselli et al., 2018a; Kobayashi et al., 2018).

To address the second issue, educational services should be tailored to the needs of individual lifelong learners. In this respect, open education become a key facilitator in many areas, including personal skills development (Kanwar and Mishra, 2018). Open Educational Resources (OERs) are also gaining popular-

[a] https://orcid.org/0000-0002-7368-0794
[b] https://orcid.org/0000-0002-9375-3516
[c] https://orcid.org/0000-0003-3758-5455

Tavakoli, M., Mol, S. and Kismihók, G.
Labour Market Information Driven, Personalized, OER Recommendation System for Lifelong Learners.
DOI: 10.5220/0009420300960104
In *Proceedings of the 12th International Conference on Computer Supported Education (CSEDU 2020) - Volume 2*, pages 96-104
ISBN: 978-989-758-417-6
Copyright © 2020 by SCITEPRESS – Science and Technology Publications, Lda. All rights reserved

ity as content sources for open education (Ha et al., 2011). Major OER repositories have large amount of regularly updated learning content in wide range of content areas. Therefore it is surprising that despite their growing capacity, OER platforms still under perform when offering personalised learning services. As an example, OER users must consult and search through several OER repositories (with different interfaces) manually in order to find appropriate learning content. Only few, initial efforts are reported, which attempt to build OER recommendation algorithms. These are done by collecting properties of users and OERs using various approaches such as building (or reusing existing) ontologies (Wan and Niu, 2018; Sun et al., 2017), conducting user behavior analysis in social networks (Lopez-Vargas et al., 2014), or applying Text Mining techniques to identify similar OER-Documents (Duffin et al., 2007). Nevertheless, due to the lack of personalized services like high quality search and recommendation, the popularity of OERs has been limited in most user groups (typically educators or lifelong learners) (Sun et al., 2018; Ruiz-Iniesta et al., 2014; Chicaiza et al., 2015; Ha et al., 2011; Chicaiza et al., 2017).

In this paper we address the above mentioned challenges and report on the prototype building of a personalized OER recommender system, which helps lifelong learners 1) to be informed about necessary skills required by their current or future jobs and 2) recommend them OERs to facilitate their progress towards mastering those skills. In this paper, after depicting the current state of the art, we reveal the methods and data we used to build up our skill classifier and OER recommender algorithms. Subsequently, we showcase the validation of our first prototype for two jobs (*Data Scientist* and *Mechanical Engineer*), using semi-structured interviews with domain experts. At the end of the paper we conclude our experiences and suggest further research directions.

2 STATE OF THE ART

2.1 Matching between Jobs and Skills

Having access to reliable labour market information on skills and jobs is not easy. Currently, only several governments or inter-governmental organizations (the most prominent actors are the US Government, European Commission or Singapore) attempt to build skill inventories and occupational taxonomies (such as ESCO, ISCO or O*NET). Although these taxonomy building efforts have created a stable basis for basic skill analytics (inter-skill relationships, high level matching to competences and occupations), most of these resources are created and maintained by human experts in several time-consuming steps, which makes them expensive and also susceptible to out-dating (Djumalieva and Sleeman, 2018). It is therefore not surprising that more and more commercial and research attempts target new ways to obtain real-time labour market information about skills, using and analysing alternative data sets like job vacancy announcement text, resume text, or social media data. These attempts can be clustered into the following three main categories:

2.1.1 Semantic-based Methods

This approach builds on ontologies to reveal and organise components of jobs (e.g. skills, tasks) (Sibarani et al., 2017; Castello et al., 2014; Khobreh et al., 2015). These methods provide meaningful information for stakeholders (i.e. structure of existing jobs, skills and their relationships), however, their dynamicity is limited, since building and maintaining ontologies to cover a wide range of occupations and skills, are currently done manually (by subject matter experts), which is a very costly and time-consuming exercise (Hepp, 2007).

2.1.2 Text Mining and Machine Learning Methods

A number of studies analyze online vacancy announcements to classify job components (e.g. skills, tasks) according to existing, static taxonomies (e.g. ESCO). This is done to update taxonomies and provide fresh information about labour markets. Most of these papers try to extract features from the vacancies by applying embedding techniques (e.g. word2vec and doc2vec)(Colombo et al., 2018), Topic Modeling techniques (e.g. LDA) (Colace et al., 2019; Colombo et al., 2018), TFIDF (Karakatsanis et al., 2017) and use classification techniques such as Logistic Regression, SVM, and Random Forest (Boselli et al., 2018b; Boselli et al., 2018a) or calculate distance (Karakatsanis et al., 2017) to assign job vacancies to the their closest job class. Furthermore, a number of papers are focusing on using Text Mining and clustering techniques to find relationships between skills and jobs, and to calculate similarity measures (Djumalieva and Sleeman, 2018; Wowczko, 2015). These papers build vectors for skills using embedding techniques, Bag of Words (Djumalieva and Sleeman, 2018), and apply clustering techniques such as K-Means (Djumalieva and Sleeman, 2018) to find the structure of related skills and jobs. Contrary to the Ontology-based systems, given that a powerful model is constructed,

these methods can automatically extract the required information form job vacancies. However, the identification of such general models remains challenging.

2.1.3 Content Analysis

Several papers focus on specific job areas, and collect related job vacancies from various sources (e.g. job boards, newspapers). Subsequently, they apply content analysis techniques such as counting the number of skills occurrence and skills co-occurrence in order to provide insights about skills in the investigated job area (Verma et al., 2019; Gardiner et al., 2018; Maer-Matei et al., 2019). Although, these methods are successful when finding and identifying required skills in a given job area, in most cases, they cannot scale. The reason is that mostly these studies use static lists of jobs and skills in their focus areas, which results in a "tunnel vision" and fail to detect new, emerging job components.

2.2 OER Recommendation

The area of OER recommendation systems has enormous development potential. The available literature on OER based content recommendations to learners is currently limited (Chicaiza et al., 2017) and there is no signal that factors related to typical lifelong learning goals (skills, jobs) play any role here. To structure recent developments, we clustered available studies into the following four categories:

2.2.1 Heuristic Method

(Sun et al., 2018) examines the Cold Start problem (Lam et al., 2008) in the case of new micro OERs. The paper defines rules, based on recommended sequences of learning objects (e.g. some learning objects should be learnt before others) using an existing ontology and calculates a Violation Degree according to the rules. The more a learning path violates the rules, the higher the Validation Degree is. Subsequently, the system recommends and adds new OERs into users' learning paths, based on minimizing the violation degree.

2.2.2 Semantic and Ontology based Methods

(Wan and Niu, 2018) builds an ontology for learners, learning objects, and their environments to establish similarity measures between learning objects. This is done in order to update learning objects' properties and provide diverse and adaptive recommendations. Some studies make use of ontologies and open source RDF data to leverage semantic content, and define recommendation algorithms suitable for linked data (Chicaiza et al., 2017; Ruiz-Iniesta et al., 2014; Sun et al., 2017). Moreover, (Chicaiza et al., 2015) tries to define an open linked vocabulary to describe user profiles, in order to facilitate recommendations.

2.2.3 Social Network Analysis

(Lopez-Vargas et al., 2014) uses social networks to build graphs of OERs and learners. Therefore, it finds tweets which have valid urls, and builds a graph, based on the co-occurrences of tweets' hashtags. They also build a similar graph with users, based on their *mentions* and *retweet*. Finally, they recognize important and influential hashtags, and use density and centrality measures from the graphs to provide recommendations.

2.2.4 Machine Learning

(Sun et al., 2017) attempts to classify users (and their demographic features) with the help of Decision Trees and Naive Bayes algorithms to recommend them OERs. Furthermore, (Duffin et al., 2007) uses Document Clustering and LSA in order to find similar OERs and use them for recommendations.

2.3 Research Question

Based on the state of the art, it is clear that 1) it is worthwhile and timely to consider labour market information to define learning goals; 2) Efforts to decompose jobs into components suitable for educational purposes are still in their infancy, and 3) the area of OER recommendation systems is an under-researched area, with a number of challenges from a technical (e.g. available algorithms, data integration, scalability) perspective. For these reasons the main research questions and objectives of this paper are:

- Empower lifelong learners to construct their own learning trajectories on the basis of labour market information and OER based learning content
- Create and evaluate a hybrid OER recommendation system prototype, relying on labour market information, learner and OER properties
- Create an algorithm to decompose jobs into unique skills and translate those skills into learning objectives
- Develop algorithms to match skills (learning objectives) with learning content available in OERs on the basis of learner and OER properties
- Conduct an initial evaluation of our hybrid OER recommendation system prototype against the

general project objectives, the applied logic, and its potential contribution to lifelong learning.

In general, with this work we expect to advance the potential of OERs to handle the increasing need for learning content and instruction (Ha et al., 2011), through personalized services for learners, based on labor market data.

3 METHODS

In this section, we detail the data and methods we used to identify required skills and their importance levels for jobs, followed by our OER recommendation algorithms. Finally, we illustrate our prototype system, which provides personalized OER recommendations to learners based on individual skill targets.

3.1 Data Collection

For the prototyping, we used a crawled sample data-set from Monster.com containing 22,000 job vacancies [1]. We used 80% of our dataset for training and cross validation and 20% of them as our test set. Moreover, for our OER recommendation, we have used APIs, provided by the following OER providers: SkillsCommons [2] and Wisc-Online [3].

3.2 Labour Market Intelligence (LMI)

3.2.1 Extracting Skills from Job Vacancies

Since our aim was to avoid any dependency on existing taxonomies (which are updated slowly), we put existing methods classifying jobs and skills into predefined classes aside, and created a dynamic job-skill matching mechanism to detect skill changes in jobs quickly. As the first step, we constructed a model to find skill related sentences in job vacancies. After an exploratory analysis, we concluded that large number of vacancies do not contain a "Required Skills" section. Therefore, in order to build our model, we selected vacancies with an explicit "Required Skills" section and run the following preprocessing procedure on each of those vacancies:

- Deletion of unimportant characters, punctuations and bullet-points
- Removal of irrelevant stop words
- Removal of conjunctions, articles, and prepositions
- Sentence Tokenization
- Lowercase Conversion
- Lematization

Altogether we obtained more than 60,000 sentences with this method. This corpus included both sentences, which were mentioned in a "Required Skills" section (we set their label to 1), and also sentences mentioned in other sections in vacancies (we set their label to 0). As a result, we got around 15,000 sentences related to "Required Skills" and labelled as 1 and around 45,000 sentences not related to "Required Skills" and labelled as 0. Subsequently, we applied embedding techniques on word-level n-grams, and built sentence vectors with averaging word/n-gram embeddings and using Multinomial Logistic Regression model to minimize the classification error [4]. It should be mentioned that word-level n-gram applies the n-gram concept on character level and find the most common sequences of characters. Therefore, vectors are created for each of the extracted sequence of characters and it helps us build vectors for new words (skills), based on our existing vector for the new word's sequences of characters (e.g. building an initial vector for *Mechatronic* based on existing vectors which are extracted from *Elecronic* and *Mechanic*). Applying our model on the test data-set resulted in the detection of 88.7% balance-accuracy (including precision and recall) of skill-related sentences. Finally, we used TFIDF weighting to detect skill terms in skill related sentences. It should be mentioned that we used *Minimum Document Frequency* of 3 as cut-off point in order to handle typing errors and remove rare words.

3.2.2 Calculating Skills' Importance for Jobs

To calculate the importance of particular skills associated with jobs in a specific geographical location, we calculated the rate of skill occurrence in the previous 6 months at the given job location. After normalizing the rates, we use a simple decay function to compute the new importance score, which combines the previous importance scores and the new rates with more weight on the new rates.

[1] The data-set is accessible from: https://www.kaggle.com/PromptCloudHQ/us-jobs-on-monstercom

[2] https://www.skillscommons.org/

[3] https://www.wisc-online.com/

Table 1: User Properties.

Property	Values	Note
Selected Job	Existing Jobs	selected by users
Skills-Levels	[0..100] for Skills	determined by users
Personal Information	Location Gender Education	entered by users
Pref_Resources	[0..100] for Resources	higher tendency →higher value
Pref_Length	[0..100]	*preferred_long* and *preferred_short*
Pref_Check	[0..100]	prefer assured →closer to 100
Pref_Accessibility	[0..100]	prefer higher accessibility →closer to 100

3.3 Recommending OERs

3.3.1 Method for Initializing Learners' Properties

Table 1 depicts learners' properties in our OER recommender prototype. During the initialization of a new user, we capture known properties entered by users (i.e. Personal Information, Skill Level List, and Selected Job), and also a number of properties without values (i.e. Resource scores, Length scores, Quality scores, and Accessibility scores). To set an initial value for these unknown properties, we sample similar users, based on the known properties and use weighted average (based on similarity) of their properties as initial values for unknown properties. This strategy scaffolds the cold start problem of new users. To sample similar users, we use (1) to compute the similarity between user i and j where the Similarity Effect function for user i and j in property k is calculated as (2).

$$similarity(i,j) = \frac{\sum_{k=known_properties} sim_effect(i,j,k)}{100} \quad (1)$$

$$sim_effect(i,j,k) = \begin{cases} equal_val(k), & \text{same k for i\&j} \\ 0, & \text{otherwise} \end{cases} \quad (2)$$

Furthermore, the equality value of property k (equal_val(k)), showing the effect of variable k on similar behaviour (rating) by users, is calculated through the following process:

1. We collect user pairs who gave exactly the same ratings for the same OER in the period
2. Compute the ratio of the number of pairs having exactly the same value in property k to the number of all pairs
3. Normalizing the ratios in a way that sum of all the ratios becomes equal to 100 and the normalized ratio of k is the Equality Value of k

This process is executed regularly, after defining a time period (e.g. after every month).

3.3.2 Method for Updating Learners' Properties

Since we aim to capture learners' preferences quickly and provide relevant OERs according to the changes and improvements in learners' property values, we decided to update user properties after each rating action on any of the recommended OERs. This is done by using a real-time updating process that, according to the rating score and the properties of the recommended OERs (i.e. length, quality, accessibility), updates the properties of the users. As a consequence, if a learner is satisfied (dissatisfied) by a given OER, we will encourage (discourage) the properties (see details in the next section) of that particular OER for that learner. For instance, if a user is dissatisfied by a long OER (e.g. it takes 10 week to complete), we will update the *Preferred Long* property of the user and decrease its value in order to provide shorter OERs in the future. Along the same line, with assigning positive ratings to accessible OERs, learners can enhance their accessibility criterion and increase their *Preferred Accessibility* value to receive content with accessibility support (critical for instance for visually impaired learners (Elias et al., 2017)).

3.3.3 OER Properties

Table 2 shows OER properties. Based on existing literature, we selected *Level*, *Length*, *Quality*, and *Accessibility* as important properties of OERs (Piedra et al., 2015; Atenas and Havemann, 2013; Elias et al., 2018). When assigning a value to a particular OER property, first we extract and order all existing values assigned to that property, then classify them, and count the number of classes. Based on the number of classes, we assign a value between 0 and 100 to that property. For instance, we take property *Level*, we extract 3 values (beginner, intermediate and advanced - 3 classes), and as a result we set the value for beginner OERs to 0, intermediate OERs to 50, and advanced OERs to 100.

[4]We used FastText Library in Python for our classification task (Joulin et al., 2016)

Table 2: OER Properties.

Property	Values	Note
Resource	Repositories	e.g. SkillCommons, Wisc-Online
Skill	Existing Skills	based on subjects
Author	Full Name	the provider
URL	URL	web address of OERs
Length	[0..100]	how_long and how_short
Level	[0..100]	higher level →closer to 100
Quality	[0..100]	more quality assurance →closer to 100
Accessibility	[0..100]	more accessibility →closer to 100
Relevance	[0..100]	decreased if defined *Irrelevant*

3.3.4 Method for Initializing OER Properties

For each OER, we attempt to identify similar OERs, based on its known properties. For instance, if we know *Skill* and *Author* of a new OER, we identify all other OERs provided by the same author and the same skill target, compute their average values, and set the initial property values accordingly.

3.3.5 Method for Updating OER Properties

Detecting OER properties is a slow process in the beginning, since change happens when users alter their rating pattern. This happens usually when they are confronted with new OERs. Therefore, we run the updating process after a specific time period (e.g. once each month). To adjust the properties (except Relevance) of each OER at first, we collect all related users and their ratings in the given time period. Afterwards, we compute the property values for the OER as X in order to minimize (3) using Gradient Descend, where θ_i is the property vector of user i and Y_i is the satisfaction rate of user i.

$$LossFunction = \sum_{i=users} |(\theta_i^T * X) - Y_i| \qquad (3)$$

This strategy of using all recent ratings in updating OER properties, enhances the diversity in our recommendations. All learners contribute to calculating these OER properties (for each OER they studied) through their individual evaluations. Users can also rate OERs as irrelevant. As a consequence, the *Relevance* property of an OER o is calculated as (4) where the *total_recom(o)* shows the number of times that OER o has been recommended to users and *irrelev_count(o)* is the number of times that o has been determined as *Irrelevant*. Finally, OERs with a Relevance Value less than the average in relation to a specific skill, are marked as Irrelevant (for that skill only), and therefore will not be recommended (for that skill) anymore.

$$relevancy(o) = \frac{total_recom(o) - irrelev_count(o)}{total_recom(o)} \qquad (4)$$

3.3.6 Recommendation Algorithm

For recommending an OER to a learner, we calculate Cosine Similarity between the properties of candidate OERs (which are related to the skill-level of any user) and the properties of the user. The system will recommend an OER with the lowest distance between those two. Since we update user properties in a real-time process and update OER properties after a predefined period, for recommending the best match for a user, we only need to find an OER, which has the closest properties to the user. Furthermore, Rating Sparsity problem (i.e. users rate only a few number of OERs) is one of the most important issue when building recommender systems. In our case users and OERs have mutual contribution to calculating properties, which intends to eliminate the effects of Rating Sparsity. Even if an OER has limited amount of ratings, we can rely on the properties of the learners. On the basis of their ratings on other (similar) OERs, we calculate the properties for OERs suffering from Rating Sparsity.

3.4 OER Recommender Prototype

Learners were confronted with a prototype of our recommender system in a form of a dashboard [5]. Through this dashboard learners can search for their current or desired job, display the list of required skills and set their level of expertise for each skills. Subsequently, on the learning tab, the dashboard shows the current expertise levels of the learner, and the links to the recommended OERs. OERs are ordered according to the importance of skills for the selected job. In case a learner thinks that a recommended OER is not related (Irrelevant) or does not find the content engaging, a new recommendation could be generated, without changing the expertise level of the learner. After consulting (learning) a recommended OER, learners are asked to rate their sat-

[5] You can find demo of our prototype from: https://github.com/rezatavakoli/CSEDU2020

isfaction with that OER. Finally, the dashboard updates the learner's expertise level and provide an updated recommendation based on the new rating. This is done until the learner masters all required skills on the highest level. Figure 1 depicts the building blocks of our proposed approach.

4 VALIDATION

To validate our proposed approach, we conducted semi-structured interviews with subject matter experts in the job areas of *Data Science* and *Mechanical Engineering*. We focused on jobs, which are related to these areas, and randomly selected 100 job vacancies for Data Scientists and 100 job vacancies for Mechanical Engineers from August 2019. Afterwards, we applied our skill extraction and importance detection model to select the most important skills in both occupations. To evaluate recommendations, we invited four university instructors with at least 12 years of teaching and 13 years of industrial experience and eight PhD students with a minimum teaching experience of 1 year and a minimum industrial experience of 2 years for a semi-structured interview[6]. Participants gave feedback on our prototype with regards to its general objectives, logic, and potential contribution to individual learning. Each interviewee had to go through the following protocol:

1. Learning about the research problem and the proposed approach - 15 minutes
2. Work with our prototype - 15 minutes
3. Going through a semi-structured interview with the help of a qualitative questionnaire[7] - 30 minutes

During working sessions with our prototype, participants generated more than 150 OER recommendations. 76.9% of these recommendations were useful and relevant to participants' skill levels and properties. 8.2% of the recommended OERs were signalled as irrelevant, and in 14.9% of the cases participants decided to change the recommended OERs. The results of the interviews are summarised in the following three sections.

4.1 Objectives

Interviewees confirmed that there is a potential value in building a labour market information driven OER recommender system. Both instructors and PhD students thought that there are several useful and high quality OERs available on the Internet, but finding them are complicated and time-consuming. Regarding the skill extraction, participants recommended to use alternative data sources, besides vacancy announcements. *Student_2* for example suggested that "you should also use other data sources related to labor market like CVs and available data about salaries". Moreover, interviewees thought that this approach is extremely useful for job-seekers, job-holders, and people who have clear ideas about their preferred occupation. However, they were skeptical about those learners, who want to focus their attention on a specific skill only.

4.2 Logic

Participants confirmed that our method to calculate the importance of particular skills in recent job vacancies can potentially help learners to focus on the most important elements of their current or future job. However, as it was also suggested by *Student_1*, a more intelligent decay function, to combine recent and previous skill important values might be desirable. Regarding the self assessment of learners to set their initial level of expertise, *Instructor_1* suggested to "introduce basic assessment in a form of technical or non-technical questions" for each targeted skill.

4.3 Contribution to Learning

Participants emphasized that interacting with learners in order to recognize their preferences (e.g. recommending OERs based on their previous ratings) is one of the most important, novel and engaging component of our proposed approach. *Student_5* recommended to include more properties: "You should capture more learners' properties such as language preferences or type of OERs (e.g. presentation, video)." Moreover, interviewees were convinced that setting specific and personalized goals for each skill in our prototype system has a strong and positive effect on the learning process.

5 CONCLUSION AND FUTURE WORK

In this paper, we showcased a hybrid OER Recommender system prototype to support individual skill development, targeting concrete, labour market oriented skills and jobs. For this prototype a skill extraction mechanism has been constructed, which cap-

[6]Detailed profiles of our interview participants are available on: https://github.com/rezatavakoli/CSEDU2020

[7]The questionnaire is available on: https://github.com/rezatavakoli/CSEDU2020

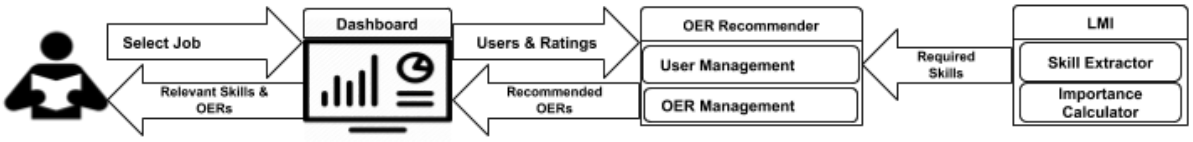

Figure 1: Components of our Labour Market Intelligence (LMI) based OER recommender.

tures skill related sentences in vacancy announcements with balanced accuracy of 88.7%. These dynamically generated skills became individual learning objectives and were connected to OER based learning contents. Recommendations were generated through a dashboard, with combining OER and learner properties. The system prototype was validated with semi-structured interviews. The initial results showed that our proposed approach has the potential to aid lifelong learners to construct their individual learning pathways and progress towards their desired job related skills. Moreover, participants valued that user properties were critical, when formulating recommendations.

We consider this study as an important first step (and a promising positive feedback) on our ongoing research project to empower lifelong learners on the basis of accurate labour market information. We believe that by confronting learners with labour market information, we also support them to develop critical transferable skills such as the awareness of their own learning needs, continuous reflection on their individual learning goals, capacities to (re)design personalised curricula, or measurement of learning achievements. Of course this prototype comes with a number of limitations (e.g. only two jobs were covered; content was only received from two OER repositories, the number of properties for the recommendation were limited), but we believe it is worthwhile to invest further effort in this area. As the next step we plan to expand the context of our investigation with adding more OER repositories to our system, together with extracting more properties from users to provide better recommendations. Moreover, accurate skill decomposition is another key problem to improve, in order to get better assessment about users' expertise level, and to construct more suitable learning pathways for lifelong learners. Finally, we plan to use (quasi-)experimental designs for further developing and validating our prototype in a number of use cases.

REFERENCES

Atenas, J. and Havemann, L. (2013). Quality assurance in the open: an evaluation of oer repositories. *INNOQUAL: The International Journal for Innovation and Quality in Learning*, 1(2):22–34.

Boselli, R., Cesarini, M., Marrara, S., Mercorio, F., Mezzanzanica, M., Pasi, G., and Viviani, M. (2018a). Wolmis: a labor market intelligence system for classifying web job vacancies. *Journal of Intelligent Information Systems*, 51(3):477–502.

Boselli, R., Cesarini, M., Mercorio, F., and Mezzanzanica, M. (2018b). Classifying online job advertisements through machine learning. *Future Generation Computer Systems*, 86:319–328.

Castello, V., Flores, E., Gabor, M., Guerrero, J., Guspini, M., Luna, J., Mahajan, L., McGartland, K., Szabo, I., and Ramos, F. (2014). Promoting dynamic skills matching: challenges and evidences from the smart project. In *INTED2014 Proceedings*, pages 2430–2438. Citeseer.

Chicaiza, J., Piedra, N., Lopez-Vargas, J., and Tovar-Caro, E. (2015). A user profile definition in context of recommendation of open educational resources. an approach based on linked open vocabularies. In *IEEE Frontiers in Education Conference*, pages 1–7. IEEE.

Chicaiza, J., Piedra, N., Lopez-Vargas, J., and Tovar-Caro, E. (2017). Recommendation of open educational resources. an approach based on linked open data. In *Global Engineering Education Conference*, pages 1316–1321. IEEE.

Colace, F., De Santo, M., Lombardi, M., Mercorio, F., Mezzanzanica, M., and Pascale, F. (2019). Towards labour market intelligence through topic modelling. In *Proceedings of the 52nd Hawaii International Conference on System Sciences*.

Colombo, E., Mercorio, F., and Mezzanzanica, M. (2018). Applying machine learning tools on web vacancies for labour market and skill analysis.

Djumalieva, J. and Sleeman, C. (2018). An open and data-driven taxonomy of skills extracted from online job adverts. *Developing Skills in a Changing World of Work: Concepts, Measurement and Data Applied in Regional and Local Labour Market Monitoring Across Europe*, page 425.

Duffin, J., Muramatsu, B., and Henson Johnson, S. (2007). Oer recommender: A recommendation system for open educational resources and the national science digital library. *White paper funded by the Andrew W. Mellon Foundation for the Folksemantic. org project*.

Elias, M., James, A., Lohmann, S., Auer, S., and Wald, M. (2018). Towards an open authoring tool for accessible slide presentations. In *International Conference on Computers Helping People with Special Needs*, pages 172–180. Springer.

Elias, M., Lohmann, S., and Auer, S. (2017). Ontology-based representation of learner profiles for accessible opencourseware systems. In *International Con-*

ference on Knowledge Engineering and the Semantic Web, pages 279–294. Springer.

Gardiner, A., Aasheim, C., Rutner, P., and Williams, S. (2018). Skill requirements in big data: A content analysis of job advertisements. *Journal of Computer Information Systems*, 58(4):374–384.

Ha, K.-H., Niemann, K., Schwertel, U., Holtkamp, P., Pirkkalainen, H., Boerner, D., Kalz, M., Pitsilis, V., Vidalis, A., Pappa, D., et al. (2011). A novel approach towards skill-based search and services of open educational resources. In *Research Conference on Metadata and Semantic Research*, pages 312–323. Springer.

Hepp, M. (2007). Possible ontologies: How reality constrains the development of relevant ontologies. *IEEE Internet Computing*, 11(1):90–96.

Joulin, A., Grave, E., Bojanowski, P., and Mikolov, T. (2016). Bag of tricks for efficient text classification. *arXiv preprint arXiv:1607.01759*.

Kanwar, A. and Mishra, S. (2018). Global trends in oer: What is the future?

Karakatsanis, I., AlKhader, W., MacCrory, F., Alibasic, A., Omar, M. A., Aung, Z., and Woon, W. L. (2017). Data mining approach to monitoring the requirements of the job market: A case study. *Information Systems*, 65:1–6.

Khobreh, M., Ansari, F., Fathi, M., Vas, R., Mol, S. T., Berkers, H. A., and Varga, K. (2015). An ontology-based approach for the semantic representation of job knowledge. *IEEE Transactions on Emerging Topics in Computing*, 4(3):462–473.

Kobayashi, V. B., Mol, S. T., Berkers, H. A., Kismihók, G., and Den Hartog, D. N. (2018). Text mining in organizational research. *Organizational research methods*, 21(3):733–765.

Lam, X. N., Vu, T., Le, T. D., and Duong, A. D. (2008). Addressing cold-start problem in recommendation systems. In *Proceedings of the 2nd international conference on Ubiquitous information management and communication*, pages 208–211. ACM.

Lopez-Vargas, J., Piedra, N., Chicaiza, J., and Tovar, E. (2014). Recommendation of oers shared in social media based-on social networks analysis approach. In *IEEE Frontiers in Education Conference*, pages 1–8. IEEE.

Maer-Matei, M. M., Mocanu, C., Zamfir, A.-M., and Georgescu, T. M. (2019). Skill needs for early career researchers—a text mining approach. *Sustainability*, 11(10):2789.

McGill, M. M. (2009). Defining the expectation gap: a comparison of industry needs and existing game development curriculum. In *Proceedings of the 4th International Conference on Foundations of Digital Games*, pages 129–136. ACM.

Piedra, N., Chicaiza, J., López-Vargas, J., and Caro, E. T. (2015). Seeking open educational resources to compose massive open online courses in engineering education an approach based on linked open data. *J. UCS*, 21(5):679–711.

Ruiz-Iniesta, A., Jimenez-Diaz, G., and Gomez-Albarran, M. (2014). A semantically enriched context-aware oer recommendation strategy and its application to a computer science oer repository. *IEEE Transactions on Education*, 57(4):255–260.

Sibarani, E. M., Scerri, S., Morales, C., Auer, S., and Collarana, D. (2017). Ontology-guided job market demand analysis: a cross-sectional study for the data science field. In *Proceedings of the 13th International Conference on Semantic Systems*, pages 25–32. ACM.

Smith, D. and Ali, A. (2014). Analyzing computer programming job trend using web data mining. *Issues in Informing Science and Information Technology*, 11(1):203–214.

Sun, G., Cui, T., Beydoun, G., Chen, S., Dong, F., Xu, D., and Shen, J. (2017). Towards massive data and sparse data in adaptive micro open educational resource recommendation: a study on semantic knowledge base construction and cold start problem. *Sustainability*, 9(6):898.

Sun, G., Cui, T., Xu, D., Shen, J., and Chen, S. (2018). A heuristic approach for new-item cold start problem in recommendation of micro open education resources. In *International conference on intelligent tutoring systems*, pages 212–222. Springer.

Verma, A., Yurov, K. M., Lane, P. L., and Yurova, Y. V. (2019). An investigation of skill requirements for business and data analytics positions: A content analysis of job advertisements. *Journal of Education for Business*, 94(4):243–250.

Wan, S. and Niu, Z. (2018). An e-learning recommendation approach based on the self-organization of learning resource. *Knowledge-Based Systems*, 160:71–87.

Wowczko, I. (2015). Skills and vacancy analysis with data mining techniques. In *Informatics*, volume 2, pages 31–49. Multidisciplinary Digital Publishing Institute.