Implementing “Rethinking Education”: Matching Skills Profiles with Open Courses through Linked Open Data technologies

Maria Zotou, Agis Papantoniou, Karel Kremer, Vassilios Peristeras, and Efthimios Tambouris

***Abstract*—The "Rethinking Education" EC strategy highlights the importance of learning outcomes in education. The so-called Massive Open Online Courses (MOOC) is becoming an increasingly important part of education. In this paper, we present a novel approach and proof-of-concept application to enable automatically matching MOOCs learning outcomes with learners' needs related to specific skills and competencies. For this purpose, we exploit the ESCO classification as well as Linked Open Data (LOD) technologies.**

***Index Terms*—ESCO, LOD, rethinking education, skills, competences, MOOC, Open Courses, Semantic Web**

I. Introduction

The "Rethinking Education" Strategy [1] from the European Commission (EC) aims at a significant shift in education, providing more focus on 'learning outcomes', i.e. all knowledge, skills and competencies acquired during learning which people can use as qualification evidence. An important relevant EU initiative is the EURES. The EU EURES Job Mobility Portal [2] provides information on worker mobility and access to more than two million job vacancies published by European public employment services. One of the main objectives of this portal is to enable competence-based job matching. For this purpose, a comparison between individuals competence profiles and employers needs is performed to find the best match.

From the education perspective, there is also a shift towards innovative approaches on learning technologies [3] and Open Courses (OC). The EC Working Document entitled “Analysis and mapping of innovative teaching and learning for all through new Technologies and Open Educational Resources in Europe” [4] clearly points out that the EC sees Techno and Open Educational Resources as opportunities to res education in the European Member States.

In this respect, a rapidly increasing number of Massive Open Online Courses (MOOCs) are becoming available. These MOOCs open up education by providing accessible, flexible, affordable and fast-track completion of courses for free. The European MOOCs Scoreboard [5] illustrates this propagation of MOOCs throughout Europe.

With such an ambition to focus on learning outcomes empowered by an EC agenda to enhance competence-based matching and with MOOCs flourishing, the problem of formalizing a dynamic lifelong learning process becomes apparent. The question that one may logically ask "how can I find open courses that match my profile, so that I can further enhance my skills and competences?” makes the problem much clearer.

Such matching typically imposes the key challenge of having software "systems" talking to each other in a common language supported by semantics that are understood from both sides. This is clearly a semantic interoperability challenge [6]. The European Interoperability Framework [7] provides a set of general recommendations to enable such communication. The European Skills/Competences, Qualifications and Occupations multilingual (ESCO) classification [8] provides the necessary common vocabulary to facilitate exchange of information related to skills and qualifications.

From the technology perspective, we argue that the adoption of the Linked Open Data (LOD) paradigm [9] and its related semantic technologies enables semantic interoperability.

Based on the above, this paper investigates the problem of formalizing the lifelong learning process in a dynamic way. The investigation is supported by matching skills profiles (like the ones found in a CV) with courses openly provided through the Internet. More specifically, it proposes a methodology on:

- How to transform skills profiles using a semantic interoperability hub and link them with the web of data
- How to enable candidate profiles to interact with open digital learning resources and learning outcomes and
- How to match skills profiles with open courses' learning outcomes.

The methodology is being applied for the implementation of a pilot, proof-of-concept application.

The paper is structured as follows. Section 2 provides a brief description of ESCO, while section 3 presents the proposed methodology. Section 4 describes the

M. Zotou is with the Department of Applied Informatics of the University of Macedonia, Greece; e-mail: mzotou@uom.gr.
A. Papantoniou is with TenForce BVBA e-mail: agis.papantoniou@tenforce.com.
K. Kremer is with TenForce BVBA, e-mail: karel.kremer@tenforce.com.
V. Peristeras is with the European Commission, Interoperability Solutions for European Public Administration (ISA) Unit, e-mail: Vassilios.PERISTERAS@ec.europa.eu
E. Tambouris is with the Department of Applied Informatics of the University of Macedonia, Greece; e-mail: tambouris@ uom.gr

implementation of a pilot system, while section 5 discusses our results and findings. Finally section 6 concludes and presents the next steps.

II. What is ESCO

To achieve and foster semantic interoperability in the area of job search, the European Commission is developing the multilingual European Skills/Competences, Qualifications, and Occupations (ESCO) classification. ESCO offers a reference vocabulary for the labor market and for the education and training sectors bringing together job offers supply and demand. The classification's first release, ESCO v0, which went live in October 2013, is a fully functional pilot version and can be accessed at https://ec.europa.eu/esco.

ESCO organizes available knowledge about the European labor market and the training and education sectors in three main pillars: occupations, skills/competences, and qualifications. Each pillar has its own hierarchical structure, making it easy for users to browse through the classification and find what they are looking for.

ESCO organizes terminology into *concepts* and *terms*. Each occupation, skill/competence, and qualification is a concept and is uniquely identified by a uniform resource identifier (URI). Multiple terms are linked to concepts and are used to name a concept in a human-readable way, depending on the language and the gender desired by the user. ESCO is being built on existing standards. A concept has one preferred term per language, but can have multiple non-preferred terms as well. It is based on the EURES taxonomy and also makes use of existing standardized classifications like ISCO-08, NACE [10], and FoET (Fields of Education and Training) [11].

Finally, the ESCO data model is represented in Simple Knowledge Organization System (SKOS) [12] which is the World Wide Web Consortium's (W3C) recommendation for representing any type of controlled vocabulary including taxonomies, ontologies, thesauri and classification schemes. SKOS is built on top of the Resource Description Framework (RDF; www.w3.org/TR/rdf-primer) and one of its main objectives is to enable the production, publication, and use of controlled vocabularies as linked data.

III. Methodology

Our goal has been to allow candidates to match their professional profiles with digital learning resources and receive access to available content regarding their skills and external relevant knowledge from multiple courseware providers, capitulating on the linked data approach. In order to achieve the above objective, we applied a multi-step research approach, resulting in a pilot implementation as a proof-of-concept. The steps are as follows:

Step 1: Identify a controlled vocabulary that appropriately describes skills, occupations and qualifications. As already discussed, this vocabulary is the ESCO taxonomy, which can be used as a semantic interoperability hub, linking skills with available learning resources.

Step 2: Enrich the description of open courseware metadata by linking their learning outcomes with ESCO concepts. We chose the Coursera MOOC, which provides an API [15] for accessing information about its courses. More specifically, we retrieved information on two exemplary courses. We then used the RDFazer tool [14] to tag the courses' metadata with corresponding ESCO skills and extracted the generated RDF triples.

Step 3: Set up an RDF store to host all the RDF data created during the previous steps to enable discoverability. We chose the Virtuoso triple store and installed it on a web server for storing all our data.

Step 4: Locate additional MOOC providers and present supplementary external open courseware suggestions relevant to the candidate's skills. We chose the Udacity MOOC which also provides an API [16] for accessing course information. We then communicated directly with Udacity's API to retrieve courses that foster the skills in question.

Step 5: Build a simple and intuitive interface to visualize the process of matching skills profiles with open courses through Linked Open Data technologies and the results. We developed a pilot implementation using the Eclipse IDE and languages such as Java, SPARQL, JSON, HTML and Javascript.

IV. Pilot Implementation

We have developed a pilot application based on linked data and semantic technologies, in an effort to test and validate our approach. The application aims to address and put into practice the objectives originally set, i.e. enhancing employability via learning, and following the LOD paradigm in order to bridge learning outcomes with skills, occupations and qualifications, and thus provide formal, informal and lifelong learners the opportunity to be linked with the web of data.

We examine the case scenario where a candidate wishes to create a professional CV and add match his/her profile with the corresponding ESCO skills as well as with open digital courses available. The system's architecture diagram is

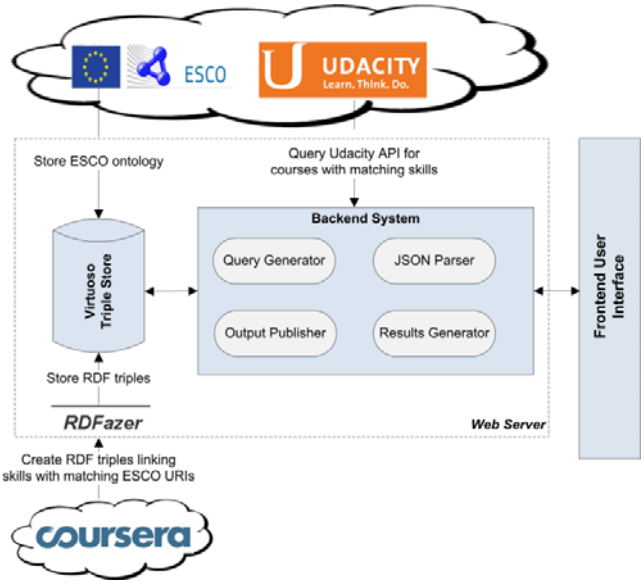

Fig. 1. Pilot software architecture diagram

presented in Fig. 1. The approach is that the candidate will have at his disposal ESCO concepts in order to do that, therefore the ESCO URIs will be injected in the CV amplifying the information regarding each skill.

The following Figure shows the user interface where the candidate can provide some initial personal information and add skills to be added to his/her CV. In this example, the user submits four skills, i.e. Data analysis, Python, Web design and Java.

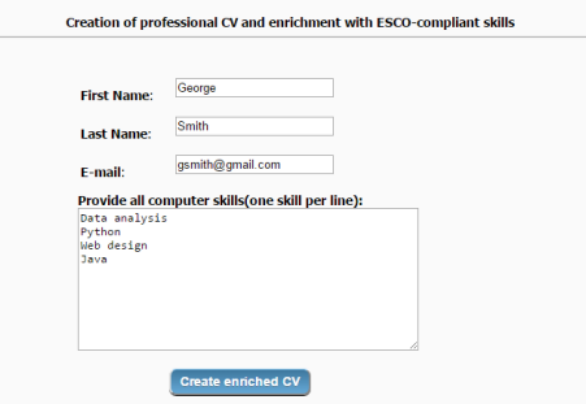

Fig. 2. Pilot's User Interface

Initially, we downloaded the ESCO ontology in its SKOS/RDF form and stored it in a Virtuoso triple store. This provided us the opportunity to query and retrieve 2,523,582 triples of information regarding skills, occupations and qualifications.

Next, we chose two courses available in Coursera as exemplary cases, namely "Programming for Everybody (Python)" and "An Introduction to Interactive Programming in Python". Using the Coursera API, we retrieved information about each course in JSON format, accessible via a web browser. More specifically, we retrieved the course's ID, title and section "About the course", which includes additional information on the curriculum, learning objectives and learning outcomes.

The next step was to enrich the courses' metadata by annotating each course's HTML snippet with the URIs of ESCO skills that matched the context.

This process was carried out using the RDFazer plugin for Google Chrome. We configured RDFazer to communicate with our Virtuoso store, and we queried the triples in order to tag the HTML page with RDFa concepts. As shown in Fig. 3, we searched for concepts in the page that indicated a specific skill, we searched for that skill in the triple store, and then we linked the matching ESCO skill URI with the HTML concept found in the course's description. Fig. 3 shows the annotation process for the first course, where we tagged the word "Python" with the ESCO skill "Python (programming language)", the phrase "data analysis" with the ESCO skill "Data analysis", the phrase "web design" with the ESCO skill "Web design" and the phrase "web browser" with the skill "Use internet browsers", while the second course was annotated with the skill "Python (programming language)". The RDF triples that were derived from this annotation process were stored in Virtuoso.

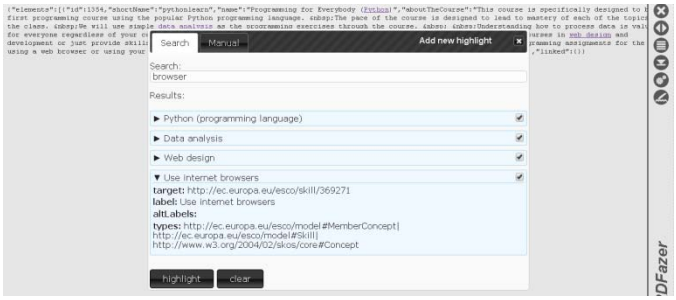

Fig. 3. Annotation of course skills with matching ESCO URIs

The backend system includes four components, as shown in Fig. 1. The Query Generator populates dynamic queries depending on the skills the candidate adds on the frontend and searches the store for the corresponding results. The queries involve searching:

1) ESCO skills URIs and labels that match or are similar to each candidate's skill. This matching is carried out automatically by the software. More specifically, the system searches in the triple store for ESCO skills that contain the given skill's name (e.g. python*). If any matches are found, the result viewed by the user in the web interface is the label of each skill linked to the ESCO website so that the user can retrieve additional information.
2) Coursera courses that have been linked with ESCO skills via RDFazer and stored in Virtuoso. The matching is again carried out by the tool, which queries the triple store for ESCO skills that have been linked as RDFa to content of the course's description. If any matches are found, the result viewed by the user is the title of each course linked to the course's page on Coursera. For example, if a user provides the skill "Python", the queries will return the courses "Programming for Everybody (Python)" and "An Introduction to Interactive Programming in Python", because both RDFa-zed courses available in the triple store include the word "python" in their descriptions. It should be noted that there is no review process included in these matches. The linking of candidate skills to ESCO skills and of ESCO skills with skills included in a Coursera course is realized through text-based matching.

The JSON parser requests information from the API provided by the Udacity MOOC on the courses available, then checks whether any courses are related with the given skills. If any matches are found, the result is the title of each course linked to the course's page on Udacity.

The Results Generator retrieves the aforementioned information per skill while the Output Publisher organizes and presents these results in an informative structure to the candidate. More specifically, a tab is dynamically created per given skill, which holds all information derived from the backend system, as shown in Fig.4.

Fig. 4. Pilot's Results page

In our example, the output in the tab for the skill "Python" includes a) the ESCO URI of the skill that matches the given attribute, b) a list of the courses available in Coursera (in this case both courses used in our example since they were linked to the skill) and c) a list of the courses available in Udacity that are related to the skill.

The pilot application can be accessed through the following URL: http://egov.it.uom.gr:8080/ESCO_impl/.

V. Discussion

The web of data provides abundant accessible and retrievable knowledge that can enhance learners' competencies in all professional fields. Learners are not always aware of the courses that exist and that they should attend in order to develop specific skills. This was the exact reason of the development of our simple pilot application; to show how the lifelong learning process can be formalized in a dynamic way

The EURES mobility portal provides the end-user with the possibility of preparing his CV online (following the Europass CV multilingual formats). Our intention was to simulate in a simplistic manner this process and show how annotation through a common and controlled vocabulary can support the process of matching skills (described in a CV) with the learning outcomes described in MOOCs that are supported by APIs. In this way the EURES portal could also be enhanced by providing information on existing MOOCs.

The overall objective is to have the user directly describe his/her skills in the CV by using ESCO URIs. If the same happens on the side of the MOOC providers, while preparing their course metadata, one can understand that there can be perfect matches.

ESCO will soon be published as Linked Open Data. Actually the whole SKOS/RDF file is already available for download from the ESCO Portal, but this isn't enough as developers using the taxonomy will want to receive metadata of specific ESCO concepts and in various formats like RDF, HTML, XML, CSV. This functionality, according to the ESCO roadmap [13] will be implemented in 2015.

What about Open Courses and Linked Data? As the pilot showed the main considerations in order to publish them as LOD are the following:

- Courses and course metadata need to be supported by an API
- ESCO should be used to annotate the content.
- Courses need to explicitly provide metadata about their learning outcomes, which currently and in most cases are only implicitly described in sections like "About the course".

VI. Conclusions

Within the context of the paper we aimed to investigate how the process of matching digital learning resources with professional profiles and competencies can be achieved. To this end, we exploited novel developments published by the European Commission such as the ESCO taxonomy as well as emerging paradigms such as Linked Open Data.

As a proof-of-concept, we developed a pilot application where candidates can create their professional profile by submitting their skills and enrich their profile by receiving supplementary content from the web of data. This pilot application demonstrates the feasibility and opportunities that derive from linking concepts from the ESCO taxonomy with skills on candidate profiles and learning outcomes of open digital resources. More specifically, the use of such an application can significantly benefit formal, informal and lifelong learners in developing appropriate competences that will increase their qualifications.

Future work will include an extension of the pilot software application to include more MOOC providers, as well as the improvement of the user interface with additional features when preparing skills profiles, a richer selection of sources, the ability to download the results in various formats etc.

References

[1] European Commission, "Rethinking Education initiative" http://ec.europa.eu/languages/policy/strategic-framework/rethinking-education_en.htm
[2] EURES, The European Jobs Network. https://ec.europa.eu/eures/
[3] European Commission, "Opening up education through new technologies", http://ec.europa.eu/education/policy/strategic-framework/education-technology_en.htm
[4] European Commission, "Opening up Education: Innovative teaching and learning for all through new Technologies and Open Educational Resources", http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex:52013DC0654
[5] Open Education Europa, "European MOOCS Scoreboard" http://www.openeducationeuropa.eu/en/european_scoreboard_moocs
[6] Wikipedia, "Semantic Interoperability", http://en.wikipedia.org/wiki/Semantic_interoperability
[7] Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of Regions 'Towards interoperability for European public services', "European Interoperability Framework (EIF) for European public services" http://ec.europa.eu/isa/documents/isa_annex_ii_eif_en.pdf
[8] European Commission, "ESCO – European Skills, Competencies, Qualifications and Occupations", https://ec.europa.eu/esco
[9] C. Bizer, T. Heath, and T. Berners-Lee, "Linked Data—The Story So Far," Int'l J. Semantic Web and Information Systems, vol. 5, no. 3, 2009, pp. 1–22
[10] NACE Rev. 2 - Statistical classification of economic activities. http://epp.eurostat.ec.europa.eu/portal/page/portal/nace_rev2/introduction
[11] UNESCO Institute for Statistics, "Global Consultation for the revised ISCED Fields of Education and Training", http://www.uis.unesco.org/Education/Documents/isced-fos-consultation-draft-2013-en.pdf
[12] A. Miles and Bechhofer, SKOS Simple Knowledge Organization System Reference, World Wide Web Consortium (W3C) recommendation, 18 Aug. 2009; www.w3.org/TR/skos-reference
[13] European Commission, "ESCOpedia", https://ec.europa.eu/esco/escopedia
[14] RDFazer, https://github.com/Rahien/RDFazer
[15] Coursera Catalog APIs, https://tech.coursera.org/app-platform/catalog/
[16] Udacity Course Catalog API, https://www.udacity.com/catalog-api