X5GON: Connecting OER Repositories

Jasna Urbančič
Jožef Stefan Institute
Ljubljana, Slovenia
jasna.urbancic@ijs.si

Erik Novak
Jožef Stefan Institute
Jožef Stefan International Postgraduate School
Ljubljana, Slovenia
erik.novak@ijs.si

ABSTRACT

The X5GON project aims to connect Open Educational Resources (OER) found across Europe and the globe as part of this project. In this showcase, we show how we acquire, process and enrich the OER, as well as how a user can then access these resources through a recommendation engine - both on the X5GON platform material search as well as an OER repository page.

CCS CONCEPTS

• **Applied computing** → **Education**; • **Information systems** → *Information retrieval*;

KEYWORDS

Open Educational Resources, Education, Recommender Engine, Information Retrieval, Network

ACM Reference Format:
Jasna Urbančič and Erik Novak. 2019. X5GON: Connecting OER Repositories. In *Proceedings of KDD 2019 Conference (KDD2019).* ACM, New York, NY, USA, Article 4, 2 pages. https://doi.org/10.475/123_4

1 INTRODUCTION

Open Educational Resources (OER), as defined by the UNESCO,[1] are teaching, learning and research materials in any medium – digital or otherwise – that reside in the public domain or have been released under an open license that permits no-cost access, use, adaptation and redistribution by others with no or limited restrictions. As such, digital OER have many advantages over traditional learning materials, namely OER reduce the costs for the students, the dissemination of information is faster, and the resources can be accessed from everywhere. However, there are also some reservations, such as quality and reliability of the materials, and intellectual property rights ownership. Additionally, an OER user faces a very fragmented landscape of repositories containing OER, which makes finding relevant OER a difficult task for both students and teachers.

[1]Definition adopted from https://en.unesco.org/themes/building-knowledge-societies/oer.

KDD2019, August 2019, Anchorage, Alaska, USA

ACM ISBN 123-4567-24-567/08/06.
https://doi.org/10.475/123_4

2 PROJECT DESCRIPTION

The X5GON project[2] is an innovation action project funded within the H2020 research and innovation program of the European Commission (ICT-19-2017 – Media and content convergence). The X5GON consortium consists of eight partners from five European countries (Slovenia, United Kingdom, France, Germany, and Spain). Partners represent universities (University College London, Universitat Politècnica de València, Université de Nantes and Universität Osnabrück), research institutions (Jožef Stefan Institute) non-profit (Knowledge 4 all Foundation LBG) and for-profit (Post of Slovenia) organizations, as well as, government bodies (Ministry of Education, Science and Sport of Slovenia). With the project beginning in September 2017, we are now at the two-thirds point of its three year duration.

The project aims to connect the scattered OER by enriching the material with additional semantic information, automatic transcription, and machine translation, as well as providing services for cross-site recommendations. With the transcriptions and translations of audible material we aim to include minority groups to the OER community, while the cross-site recommendations are meant to make finding the appropriate OER easier and make OER more accessible by connecting different OER repositories across the world. More specifically, the goals of the project include:

(1) creating a homogeneous network of OER sites by combining content understanding, user modelling, and quality assurance methods and tools; and
(2) creating a self-sustainable and growing mode of operation based on bottom-up OER sites collaboration.

3 PRESENTED INNOVATION

Currently, the OER landscape paints a picture of highly specialized repositories in terms of scientific domains, type of content, level of education, and language. From the perspective of a student this means that in the worst case the student has to search OER in different repositories for each class. Not only that such search is inefficient and time consuming, it also leads to sub-optimal search results and a negative user experience with OER. To bridge the gaps between OER repositories the project has developed:

(1) a system for gathering, indexing, and enriching OER;
(2) a search engine for OER; and
(3) a recommender engine for content-based recommendations.

Developed systems are available on Github.[3]

[2]X5GON — Cross Modal, Cross Cultural, Cross Lingual, Cross Domain, and Cross Site Global OER Network (https://www.x5gon.org). The project has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement No 761758.
[3]X5GON on Github (https://github.com/JozefStefanInstitute/x5gon).

3.1 OER Network

To expand the collection of the indexed OER we do this in two different ways. The first is with crawling, which we use for large structured repositories. During this process we scrape the websites for metadata of the learning material, such as the date of creation or publishing, author or speaker, description, license, type of material, etc. The second is through a REST API endpoint, where registered repositories can submit the URLs of their OER along with the resource metadata. Since downloading and storing materials would take up a lot of computer resources, we only collect and store material metadata. With the use of the transcription and translation platform developed by Universitat Politècnica de València[4] we extract raw text from the audible resources, whereas for textual documents we use a text extraction tool. Afterwards, we use Wikifier[5] to enrich the extracted raw text with semantic annotations, namely Wikipedia concepts. The Wikipedia concepts not only enrich the OER with additional information, but also enables us to process OER in various languages as the concepts in one language are linked to a concept in another language through the Wikipedia graph. Furthermore, this supports search and recommendation functionalities in various languages.

So far we have collected, enriched, and indexed over 32k OER consisting of more than 88k files. The numbers of OER per repository are shown in Figure 1. The discrepancy between the number of OER and the number of files comes from the structure of OER. For example, a digital OER may include a video of a lecture and the slides in textual format that were used in the lecture. We show the number of OER per type in Figure 2. The OER associated with more than one type is labelled as multimodal. The counts of OER per language is shown in Figure 3.

3.2 User Interaction with the OER Network

Currently, a user is able to interact with the global OER network in two ways. The first is through the X5GON platforms material search.[6] The user can browse the collection by entering a search query into the text box. The search engine then transforms the query into vector using TFIDF and finds the nearest neighbors among OER to the query vector in terms of cosine similarity. Afterwards, the user can filter the results based on the type of the OER.

The other entry point to the network is through one of the partnering OER repositories with the embedded X5GON recommendations, for example VideoLectures.NET.[7] We have developed an recommender engine API, which as a query takes URL to the OER in the network and suggest similar OER that are present in the network. Such suggestions enable users to move between different repositories, types of content, and languages. Recommendations are computed based on cosine similarity between sparse vectors of Wikipedia concepts extracted from the resources.

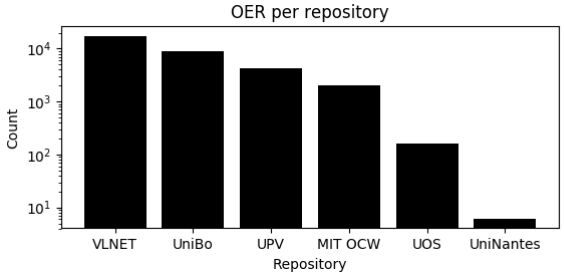

Figure 1: The number of OER materials per repository. Labels represent the following OER repositories: VLNET — VideoLectures.NET; UniBo — University of Bologna; UPV — Universitat Politècnica de València; MIT OCW — MIT OpenCourseWare; UOS — Universität Osnabrück; and UniNantes — Université de Nantes.

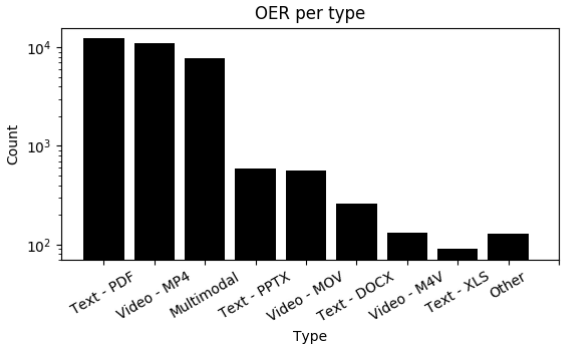

Figure 2: The number of OER materials per type. Multimodal OER is associated with more than one type of materials.

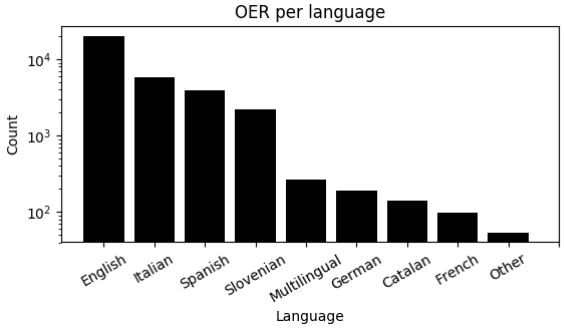

Figure 3: The number of OER materials per language. Multilingual OER is composed of learning materials that are in more than one language.

In addition, we provide REST API endpoints for OER providers in X5GON network to retrieve the OER metadata, its transcriptions and translations, as well as full documentation on using these endpoints.[8]

4 PROPOSED DEMONSTRATION

As a live demonstration we intend to show how the user can interact with the network by accessing different OER through the X5GON platform material search and the embedded recommendation list in the VideoLectures.NET repository.

[4] Available at https://ttp.mllp.upv.es.
[5] Available at http://wikifier.org.
[6] X5GON platform material search (https://platform.x5gon.org/search).
[7] Available at http://videolectures.net/.
[8] X5GON API documentation (https://platform.x5gon.org/documentation).