X Modal
X Cultural
X Lingual
X Domain
X Site
Global OER Network

Grant Agreement Number: 761758
Project Acronym: X5GON
Project title: X5gon: Cross Modal, Cross Cultural, Cross Lingual, Cross Domain, and Cross Site Global OER Network
Project Date: 2017-09-01 to 2020-08-31
Project Duration: 36 months
Document Title: D4.5 – Prototype of cross-site recommendation engine
Author(s): Jasna Urbančič, Ayşe Saliha Sunar, Erik Novak (IJS)
Contributing partners: JSI
Date:
Approved by:
Type: P
Status: Final
Contact: Erik Novak (erik.novak@ijs.si)

Dissemination Level		
PU	Public	x
PP	Restricted to other programme participants (including the Commission Services)	
RE	Restricted to a group specified by the consortium (including the Commission Services)	
CO	Confidential, only for members of the consortium (including the Commission Services)	

Revision

Date	Lead Author(s)	Comments
21/06/2019	Jasna Urbančič	Initial draft
15/08/2019	Jasna Urbančič, Erik Novak, Ayşe Saliha Sunar	Added data analysis
27/08/2019	Colin de la Higuera	Draft review
30/08/2019	Jasna Urbančič, Erik Novak	Final version

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS

Abbreviation	Definition
OER	Open Educational Resource

ABSTRACT

In this document we present the prototype of the cross-site recommendation engine. We provide the reasoning and aspects used in the cross-site recommendation design, and present the recommendation plugin – an easy to integrate solution of providing recommendations to the OER repositories. Furthermore, we perform an analysis of the user recommendation action dataset and provide preliminary results and insight on the recommendation list.

1. INTRODUCTION

The main objectives of this project are to create a cross-modal, cross-cultural, cross-lingual, cross-domain, and cross-site global OER network by connecting several OER repositories. To achieve this, we developed a recommendation engine which is capable of providing recommendations from different OER sources. At the time of this writing we had acquired more than 90k items of OER material from six different repositories. We used this data to develop a content-based recommendation engine prototype which generates recommendations from different OER sites. We direct the readers to access the following deliverable for more information:

- *D2.2 – Final Server Side Platform* for the data acquisition process,
- *D4.1 – Initial Prototype of User Modelling Architecture* for detailed description of user activity data and how we use the data, and
- *D4.3 – Early prototype of Recommendation Engine* for explanation how we combine both types of data.

In this report we present the prototype of the cross-site recommender engine. The report includes a brief overview of technical challenges of the cross-site recommender systems, a short description of our engine with examples, and the list of collaborating repositories. In addition, we present preliminary results of the cross-site recommendations – showing the user selection on the recommendation statistics.

The document is structured as follows. Section 2 describes the technical requirements and describes the solutions used in providing cross-site recommendations. Next, preliminary results on the cross-site recommendations is presented in Section 3. Finally, we conclude the report in Section 4.

2. CROSS-SITE RECOMMENDATION ENGINE

In this section we discuss the challenges we are facing when providing cross-site recommendations, how we solve them, and how successful are we when it comes to attracting other OER sites.

2.1. TECHNICAL ASPECTS OF CROSS-SITE RECOMMENDATIONS

We have identified several critical aspects of the cross-site recommendations and have described them in the following paragraphs.

Data harmonization. In deliverable *2.2 – Final Server-Side Platform* we already stated that each OER repository structures their material in a different way – meaning there is not clear definition of what an individual OER is and what additional metadata is associated with it. In addition, we find that OER materials are usually embedded inside a particular webpage on an OER repository. When discussing with the repository maintainers, we found that OER providers prefer recommendations that direct the users towards a webpage which embeds the material instead of the actual material. Therefore, we have created a recommendation model which supports content-based recommendations on OER webpages (defined as *bundles*). Since a webpage can contain multiple materials, we represent the bundle as an aggregate of the embedded materials. Each bundle in the recommendation list is described with a minimum set of information – the title of the OER material, the OER provider of the material, the URL used to access the OER, as well as, the type and language of the material. Further description of the model is found in deliverable *4.4 – Final Prototype of Recommendation Engine*.

Easy-to-use services. In our experience, the OER repositories are not willing to use services that are not easy to add to their web pages. Often the repositories do not have extra personnel they can allocate to adoption the new services. We found that embed-ready HTML objects work best as they can be integrated into web pages easily. To this end, we have found a solution which employs the iframe HTML tag for integrating the recommender list into the website.

Transparent services. The OER repositories also feel uneasy about using the services that they do not understand or do not deem transparent. In the case of the project this mostly applies to personalization through user modelling and profiling – as the repositories are fairly concerned about sharing the user data. Therefore, in cross-site recommendation engine we focus on content-based cross-site recommendations as these algorithms do not require any user data.

Detailed documentation. We prepared a detailed documentation of the services and their use. With this, we hope it will encourage the OER repositories to adopt the developed services. The full list of possible recommender models is available at [1], whereas the documentation for embed-ready recommendation list is at [2]. Both approaches use the same algorithms, data, and the infrastructure in the back. The only difference is in the structure of the output – one provides a JSON object of the recommended results, while the other provides an embed ready recommender list.

2.2. EMBED-READY LIST OF RECOMMENDATIONS

As a way to appeal to the OER repositories we have designed an embed-ready list of recommendations. It uses the same infrastructure in the backend as the regular X5GON recommender engine as the data is already harmonized. However, instead of a JSON object of recommendations it returns an HTML snippet in the desired size. The snippet can be then embedded into an iframe, which is a well-known HTML tag used to embed documents into HTML. The snippet must be configured to provide the

required recommendations – this is done by setting the query parameters within the URL. An example of the embedded recommendations is show in Figure 1.

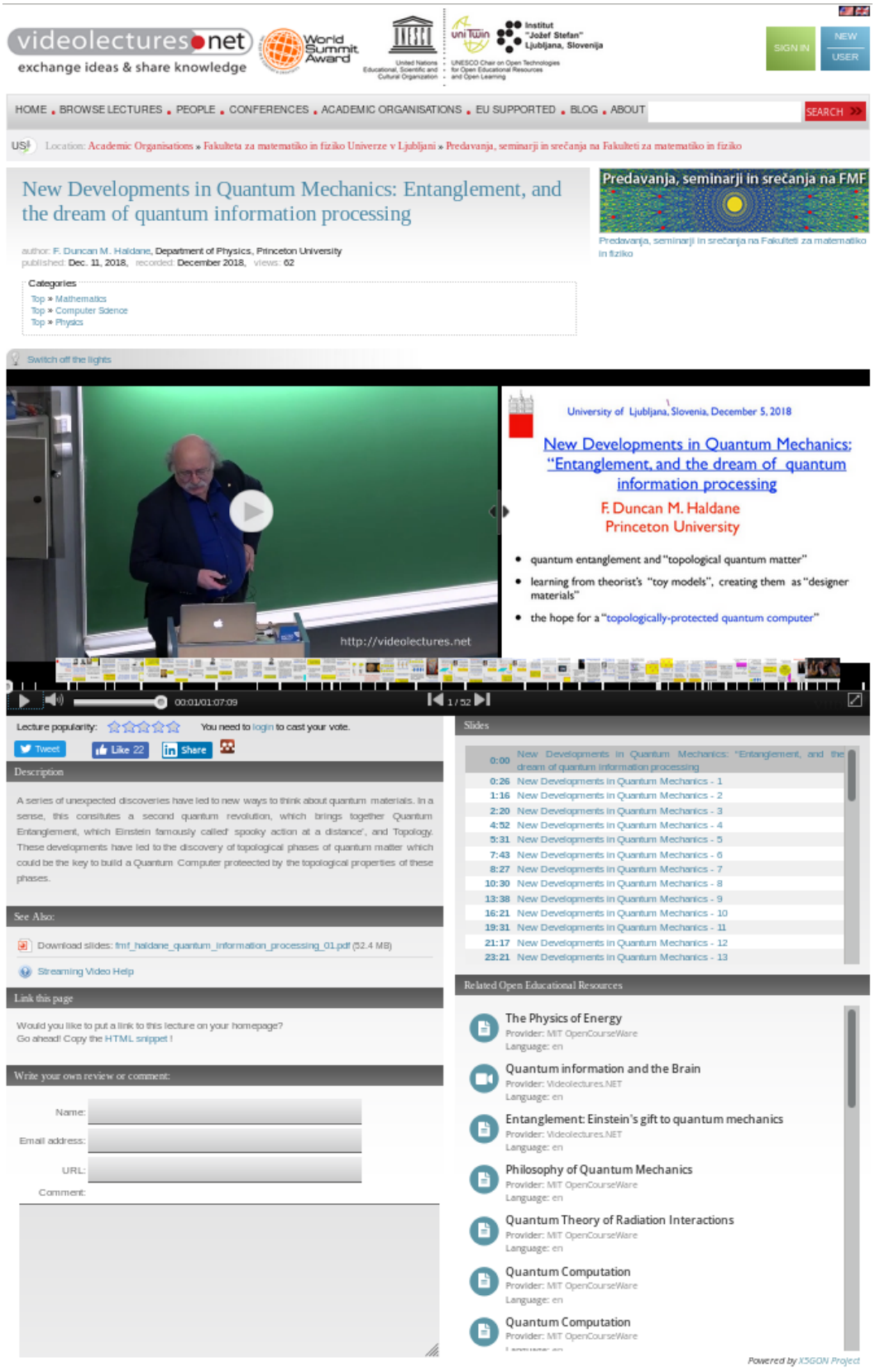

Figure 1: *Embed-ready recommendations in VideoLectures.NET in red frame.*

The example shows a lecture on the Videolectures.NET [3] repository. The embedded recommendations were configured to provide OER bundles that are similar to the current one. By clicking on one of the recommendation items, the user is redirected to the webpage associated with the selected item. This allows the user to transition from one bundle to another, enabling them to consume materials from different sources.

2.3. COLLABORATING OER REPOSITORIES

When inviting OER repositories to join the X5GON OER network we found that there are different levels of collaborations – some are prepared to integrate either the Connect Service or the Embed-Ready Recommendation Plugin, some only provide the OER material metadata, and some are willing to full collaborate. To this end, we established three levels of repository collaboration – full collaboration, partial collaboration and OER material only. The full integration conditions include:

1. Providing a way of collecting the OER material metadata,
2. Integrating the Connect Service, and
3. Including the Recommendation List into their webpages (either with the embed-ready plugin or with the use of the recommendation REST API).

Partial integration includes the collection of OER material metadata and adoption of only one of the two services – the Connect Service or the Recommendation List. The OER material only level states that a repository has only provided the OER material metadata at the moment – since the materials are openly available, we have acquired the materials using different ingestion scripts, as described in deliverable *2.2 – Final Server-Side Platform*. The current participating OER repositories and their level of commitment is shown in Table 1.

Full Collaboration	Videolectures.NET, Universitat Politècnica de València
Partial Collaboration	Université de Nantes, Unversität Osnabrück, eUčbeniki
OER Material Only	MIT OpenCourseWare, University of Bologna

Table 1: *An overview of collaborative repositories.*

We have fully recruited the Videolectures.NET and Universitat Politècnica de València repositories, both of which are partnering repositories. While most of the fully and partial collaborative repositories are maintained by the project partners, we are starting to recruit other OER repositories – one of which is eUčbeniki, a Slovenian repository containing textbooks for primary school and high school programs in Slovenia. In the future, we will continue recruiting repositories to join the X5GON OER Network.

3. PRELIMINARY RESULTS OF THE CROSS-SITE RECOMMENDATION ENGINE

Both Videolectures.NET and Universitat Politècnica de València repositories have embedded the embed-ready recommendation plugin. Through the plugin, we are able to collect data of the user material selection. Currently, we have collected above 233k user recommendation actions. Each action consists of the following information:

- The URL of the source webpage,
- The URL of the chosen material webpage,
- The list of recommended material webpages (bundles), and
- The position of the selected item.

In this section, we present the preliminary analysis results of the cross-site recommender engine using the collected data.

3.1. USER CLICK POSITION DISTRIBUTION

Using the user recommendation actions dataset, we first checked which of the material positions were most clicked. Due to the nature of the generated recommendations – having the most relevant materials at the top of the list – we expected the users will mostly click on the first couple of materials in the list. *Figure 2* shows the distribution of the number of clicks per material position in the recommended list.

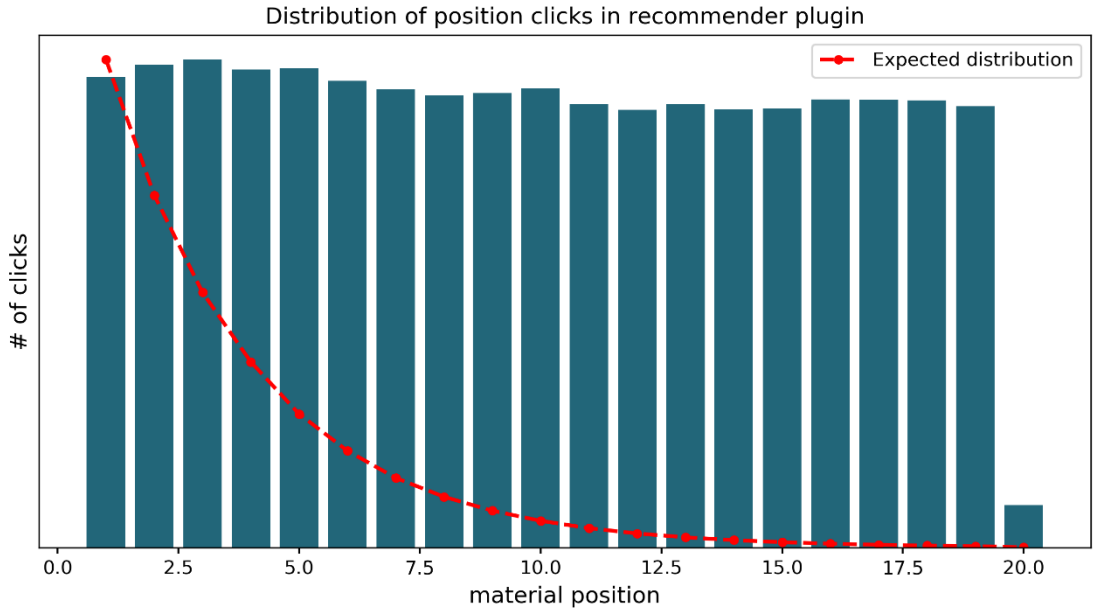

Figure 2: *The distribution of the number of users clicks per material position in the list. The red line shows the expected distribution of clicks.*

The distribution of clicks is close-to-uniformly distributed across the material positions, which signifies that users might not be interested in the most relevant materials but some other features. These features might be the provider, the language or the type of the material. Currently, we do not have any concrete answers which is it and will leave it for future work.

3.2. RECOMMENDED MATERIAL DOMAIN TRANSITION ANALYSIS

In addition to the user clicks, we have also analysed how frequent are users transitioning from one provider to another via the recommended materials. Since there

are only two repositories – Videolectures.NET and Universitat Politècnica de València – with the integrated recommendation plugin, there will be only two source providers, while the target domains are all providers for which we have acquired the material metadata. Figure 3 shows the user transition flow between different providers.

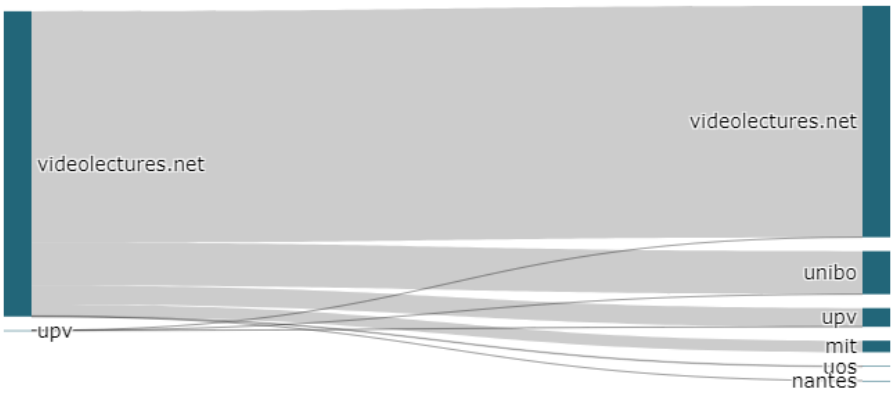

Figure 3: *The flow of user transitions among the providers through the recommended list.*

When users use the recommender engine, they are usually interested in materials provided by same source providers, e.g. if a user viewing a material on the Videolectures.NET repository and use the recommender engine, they will most likely select a material that is hosted by Videolectures.NET. This can be because of different reasons:

- The user is loyal to the repository and does not wish to learn from other repositories.
- The list of recommendations does not include materials from other domains.

Since we do not have access to the users to provide them a questionnaire about their preferences, we cannot evaluate the first reason. For the second, we have counted how many times a material from a particular provider has been found in the recommender list. Figure 4 shows the distributions of providers in the recommended list given that the source provider is Videolectures.NET or Universitat Politècnica de València, respectively.

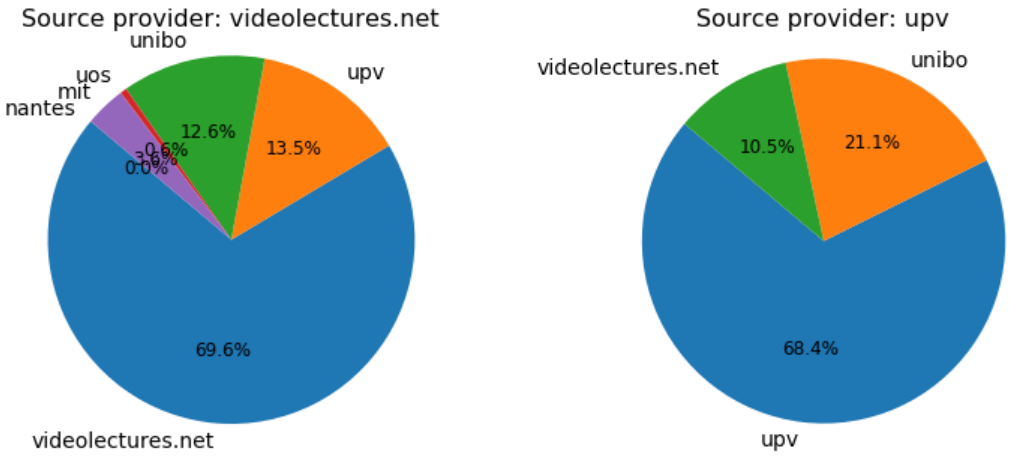

Figure 4: *The distributions of providers in the recommended list given the source provider is Videolectures.NET (left) and Universitat Politècnica de València (right), respectively.*

For both distributions, the most frequent provider in the recommended lists is the actual source provider – which is reasonable, since the providers have a tendency of providing materials that are on a similar topic (i.e. conferences, courses). To have a thorough overview of the provider presence in the recommendations, we will create bundle recommendations for all bundles that are in the X5GON platform and calculate the distributions of providers presence. Similar analysis will be performed in the future.

4. CONCLUSION

In this report we described the prototype of the cross-site recommendation engine. We present the technical aspects that were considered when developing the cross-site recommendation plugin. We describe the function of the recommendation plugin and show which of the OER providers have agreed to collaborate and join the X5GON OER Network.

We also presented preliminary results of the cross-site recommendation list. We have found that users are using the recommendation plugin and that they are selecting materials from all over the list. In addition, users tend to select materials that are provided by the same OER repository. This might be because the recommendations tend to mostly contain materials of the source OER provider.

In the future, we will continue to perform analysis on the usage of the recommendation plugin to better understand the users needs. The results will also provide insight into the usability of the recommendations, which will influence the overall design of the recommendation engine – both the models and recommendation plugin.

REFERENCES

[1] Jožef Stefan Institute, “API Documentation | X5GON Platform - Recommender REST API,” 14 August 2019. [Online]. Available: https://platform.x5gon.org/documentation#recommender-rest-api.

[2] Jožef Stefan Institute, “API Documentation | X5GON Platform - Connect Recommender,” [Online]. Available: https://platform.x5gon.org/documentation#connect-recommender. [Accessed 14 August 2019].

[3] “Videolectures.NET - Videolectures.NET,” Videolectures.Net, [Online]. Available: http://videolectures.net/. [Accessed 23 07 2018].