X5GON Project Showcase

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

Blaž Fortuna
Jožef Stefan Institute
Ljubljana, Slovenia
blaz.fortuna@ijs.si

Primož Škraba
Jožef Stefan Institute
Ljubljana, Slovenia
primoz.skraba@ijs.si

ABSTRACT

The X5GON project aims to connect Open Educational Resources found across Europe and the globe as part of this project. In this showcase, we present the Infominer service, which is used to explore the resources, group them together based on different criteria and create OER ontologies.

CCS CONCEPTS

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

KEYWORDS

Open Educational Resources, Learning Analytics, Machine Learning, Exploration Tool, Ontology Creation, Visualization, Statistics

ACM Reference Format:
Erik Novak, Blaž Fortuna, and Primož Škraba. 2018. X5GON Project Showcase. In *Proceedings of KDD 2018 Conference (KDD2018).* ACM, New York, NY, USA, Article 4, 2 pages. https://doi.org/10.475/123_4

1 INTRODUCTION

Open Educational Resources (OER) are teaching materials which are freely accessible and open license. These resources are gaining popularity within the education sector where anyone can access it and use it for teaching. As show in [4] they are found scattered across Europe and the globe and come in a variety of modalities, e.g. text, video, audio, presentations and other, and a variety of languages.

Due to its sparsity and variety, finding user relevant OER is difficult. In this process, we need to take into account users' interests, prior knowledge, and learning patterns, as well as, the quality, modality, and language of the OER material.

The X5GON project [1] is an H2020 EU project that began in August 2017. It aims to converge the scattered OER, enrich it, and provide valuable recommendations to the user. Additionally, it will also allow users to explore the enriched OER and create their own learning pathways. This will make finding the appropriate OER easier and make it more accessible by connecting different OER repositories found across the globe.

KDD2018, August 2018, London, United Kingdom

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

2 PROJECT PARTNERS

The X5GON consortium is made up of 8 partners from 5 European countries (Slovenia, United Kingdom, France, Germany and Spain). The partners represent universities (University College London, Universitat Politecnica de Valencia, Université de Nantes and Universitaet Osnabrueck), research institutions (Jožef Stefan Institute) non-profit (Knowledge 4 all Fooundation LBG) and profit (Posta Slovenije) organizations, as well as, government bodies (the Ministry of Education of Slovenia).

3 OER EXPLORATION

One of the project objectives is to acquire OER metadata and make it available to teachers, students, and other learning personnel, as well as, give them a way of exploring the metadata using machine learning methods. To this end, we are developing an online exploratory service called Infominer, which will enable users get basic statistics on the OER metadata, such as word importance and view count, as well as use machine learning methods to find similar OER and create ontologies. This will enable the users to

- find OER that match their interests,
- group OER using different machine learning methods to create learning material bundles, and
- create OER ontologies highlighting how the learning material is connected.

In the following sections, we present the functionalities of the Infominer service. For the showcase, we have used OER material provided by Videolectures.NET [5], a United Nations award-winning free and open access educational video lectures repository. The materials consist of the title, description, and author of the lecture, as well as, the language in which the lecture is presented, the number of views, and its duration.

3.1 Basic Statistics and Query

When a dataset is uploaded, the service automatically calculates some basic statistics, such as document count, word importance within a field, and number of word occurrences within the field. Different visualizations, e.g. wordclouds and histograms, are used to present the resulting statistics, see Figure 1.

The user can then browse through and query for resources of interest. They are presented in a table which contain all of the resource information. Both visualizations and the table give the user a structured view of the resources found within the dataset.

3.2 Analysis Selection

The main functionality of the Infominer service is to provide an easy way of applying machine learning methods to analyze resources.

Figure 1: Wordclouds generated based on a lecture's title and description TFIDF weight, respectively. The size and the color represent the word relevance within the shown fields.

This is done through the analysis selection panel, shown in Figure 2, enabling the user to select a clustering method and its parameters. The results are shown using visualizations presented in Section 3.1.

The user can then select a different analysis method and use it on the whole dataset or on one of the clusters of resources given in the previous results. The clustering provides a convenient way for the user to select subsets of the data and perform a more in-depth analysis. The user's exploration of the data creates a personal resource ontology representing how different subsets of the data are interconnected. These ontologies increase the understanding of the dataset and can help create new material bundles.

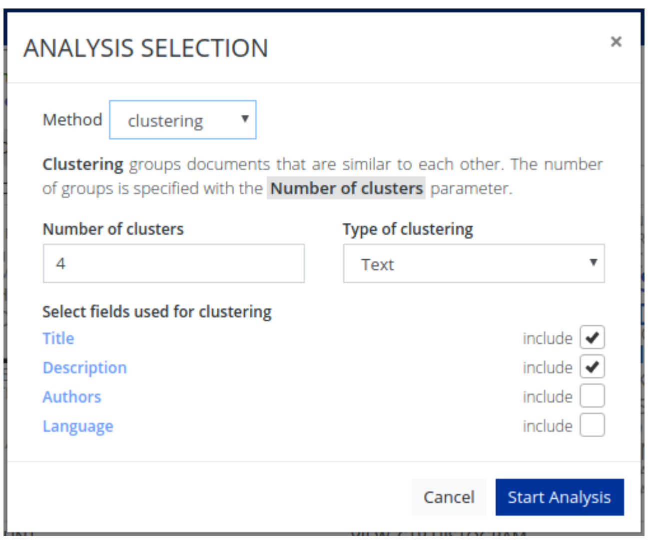

Figure 2: The analysis selection panel. It's used to select the method and the criteria by which the analysis should follow.

With this analysis the user will create his or her own resource ontology reflecting the exploration process and showing how are the subsets connected with each other. This will increase the understanding of the dataset and create learning material bundles.

3.3 Dataset Visualizations

Here we describe how the user defined ontology is presented within the service. Two separate components show the ontology, as seen in Figure 3.

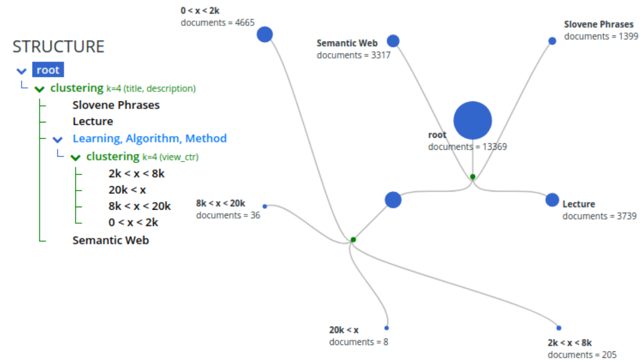

Figure 3: Ontology representation components. Left the hierarchical structure of the analysis process. Right the tree structure showing the number of documents in each of the subsets.

The first component is the hierarchical structure which shows the analysis process as it is. It shows which method was used on which subset of resources and what are the results. This component is always present next to the result page and serves as a navigation between different resource subsets.

The second component is the tree structure showing how are the resources clustered together. It shows the number of resources within the subset and how it is split in smaller subsets.

4 METHODOLOGY

Data is analyzed using QMiner [6], an analytics platform for processing structured and unstructured data. It provides clustering algorithms, i.e. kmeans clustering [3], and feature extractors used for creating subsets as well as calculating the statistics of the subsets. The analysis results are then used for visualizations described in the previous sections.

5 DEMONSTRATION

The Infominer service is available at [2]. It's use is not limited to OER material and can be tested by providing a dataset in an appropriate format - the first line of the dataset file contains field names separated by one of the following delimiters: comma (,), semicolon (;), tab (\t), pipe (|) or tilde (~). The remaining lines of the dataset are the resource attributes separated by the delimiter.

At the conference, we plan to show how is the service used to analyze educational material and explain the technical details of the service.

REFERENCES

[1] [n. d.]. Cross Modal, Cross Cultural, Cross Lingual, Cross Domain, and Cross Site Global OER Network. http://x5gon.org/
[2] [n. d.]. Infominer - Artificial Intelligence for the rest of us. https://infominer.ijs.si/
[3] [n. d.]. k-means clustering - Wikipedia. https://en.wikipedia.org/wiki/K-means_clustering
[4] [n. d.]. OER World Map. https://oerworldmap.org/
[5] [n. d.]. Videolectures.NET - Videolectures.NET. http://videolectures.net/
[6] Blaz Fortuna, J Rupnik, J Brank, C Fortuna, V Jovanoski, M Karlovcec, B Kazic, K Kenda, G Leban, A Muhic, et al. 2014. » QMiner: Data Analytics Platform for Processing Streams of Structured and Unstructured Data «, Software Engineering for Machine Learning Workshop. In *Neural Information Processing Systems.*