



New Ontology and Knowledge Graph for University Curriculum Recommendation

Nicolas Hubert, Armelle Brun, Davy Monticolo

► To cite this version:

Nicolas Hubert, Armelle Brun, Davy Monticolo. New Ontology and Knowledge Graph for University Curriculum Recommendation. ISWC 2022 - The 21st International Semantic Web Conference, Oct 2022, Hangzhou / Virtual, China. hal-03768154v2

HAL Id: hal-03768154

<https://hal.science/hal-03768154v2>

Submitted on 25 Sep 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

New Ontology and Knowledge Graph for University Curriculum Recommendation

Nicolas Hubert^{1,2,*}, Armelle Brun² and Davy Monticolo¹

¹Université de Lorraine, ERPI, France

²Université de Lorraine, CNRS, LORIA, France

Abstract

Education is a complex domain where students must make their curricula choices carefully. To model the intricacies of the educational domain, ontologies have successfully been leveraged in the past. However, no available ontology or dataset directly addresses the critical transition from high school to university from a decision-making perspective. Therefore, the contribution of our on-going work is twofold. Firstly, we introduce EducOnto – an ontology that aims at modeling university curricula and students' profiles. Secondly, we introduce EduKG – a knowledge graph inheriting the semantics of EducOnto and instantiated with data about French students and curricula.

Keywords

Knowledge Graph, Education, Ontology, Recommender System

1. Introduction

Higher education is an intricate domain: it features a lot of distinct concepts, relationships and structural constraints. Whenever a given application domain is too complex to be grasped directly, ontologies established themselves as the reference approach to model the richness of this domain. As such, they have naturally been leveraged in education, where there is a real need for structuring information [1, 2]. However, there is a lack of ontologies specifically concerned with the critical period ranging from the end of high school to the first years of university, in which students commit for their future.

Besides, owing to the vast amount of provided information on university curricula, students can find it tough to make the most appropriate decision regarding their educational pathways [3]. Recommender Systems (RSs) have successfully been deployed to reduce the overhead associated with decision-making by sifting through massive amounts of choices and suggesting only the most relevant items to the active user [4]. RSs have proved to be useful for recommending curricula to students [1]. Nevertheless, most educational datasets related to curriculum recommendation are university-specific and remain private [5, 6].

International Semantic Web Conference (ISWC) 2022: Posters, Demos, and Industry Tracks, October 23-27, 2022

*Corresponding author.

✉ nicolas.hubert@univ-lorraine.fr (N. Hubert); armelle.brun@loria.fr (A. Brun); davy.monticolo@univ-lorraine.fr (D. Monticolo)

🌐 <https://nicolas-hbt.github.io/> (N. Hubert)

🆔 0000-0002-4682-422X (N. Hubert); 0000-0002-9876-6906 (A. Brun); 0000-0002-4244-684X (D. Monticolo)



© 2022 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

CEUR Workshop Proceedings (CEUR-WS.org)

Hence, facing the need for public resources related to curriculum recommendation, we present an on-going work related to the development and provision of both a new ontology and a new dataset, referred to as EducOnto and EduKG, respectively. More specifically, EducOnto focuses on modeling the pivotal period bridging the last year of high school and the first years of university, while EduKG is a Knowledge Graph (KG) encompassing students' profiles and academic training information. EduKG is built on the basis of EducOnto, from which it inherits its semantics and structural constraints.

2. EducOnto: An Ontology for High School to University

In this section, we elaborate on the methodology used to design EducOnto. We follow best practices guidelines presented in [7]. We distinguish three main phases that will respectively form the following paragraphs: (1) Domain and purpose definitions; (2) Ontology building process; (3) Ontology evaluation. EducOnto is publicly available at the following link: purl.org/eduonto.

2.1. Domain and Purpose Definitions

In the educational domain, several ontologies are publicly available¹. However, none of them directly models the period ranging from the end of high school to the first years of university. EducOnto aims at modeling such a period. Together with education experts, we defined competency questions whose purpose is to frame the scope of the ontology and act as requirements that EducOnto should meet. The exhaustive list of competency questions is made available². Let us introduce two of them as an example:

CQ1. What are the recommended academic backgrounds for a specific university curriculum?

CQ2. What is the most popular university curriculum following a given high school major?

2.2. Ontology Building Process

With education experts, we first enumerate all the terms related to our application domain and the expected purpose of EducOnto. Two main concepts stand out: *student* and *curriculum*. Both concepts are related to other intermediate concepts such as *major* and *specialty*. Then, adopting a top-down development process, we end up with 71 classes and 30 object properties. A general overview of EducOnto that contains the main classes and properties is presented in Figure 1.

2.3. Ontology Evaluation

EducOnto can be considered as successfully designed if it helps to fulfill downstream tasks related to university curricula information retrieval and recommendation. According to [8], we rely on a task-based evaluation. To do so, we first take a general use case, after which we seek to answer the previously mentioned competency questions by querying EducOnto using SPARQL. The answers to the queries are retrieved from the data contained in EduKG. Below, we

¹<https://lov.linkeddata.es/dataset/lov/vocabs?&tag=Academy>

²<https://nicolas-hbt.github.io/educ-ontokg/competencyquestions/>

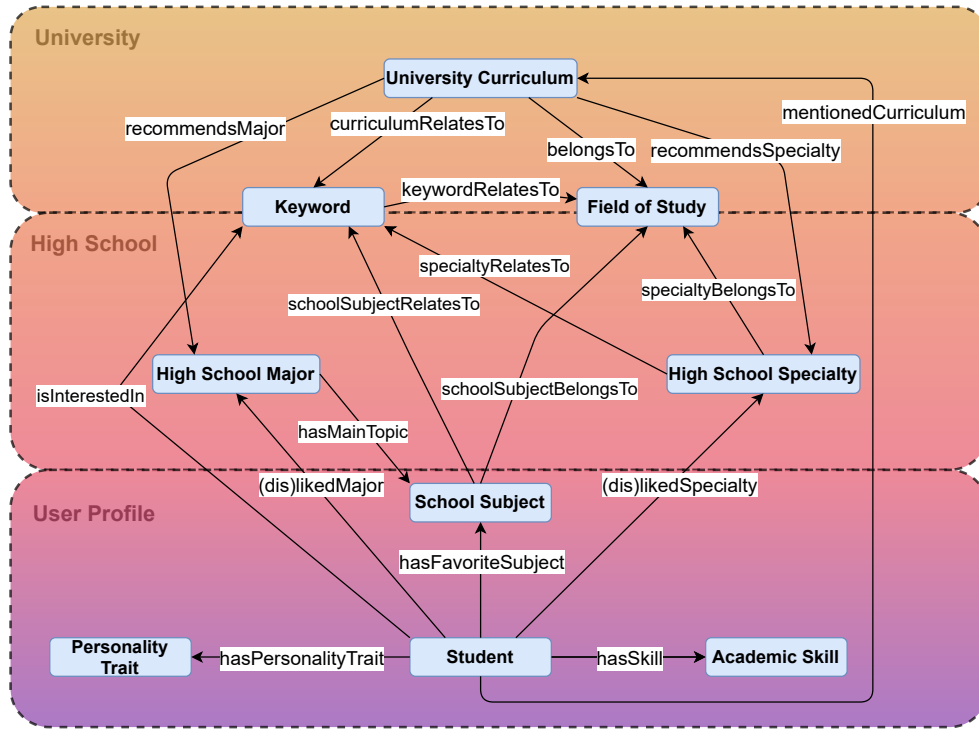


Figure 1: A conceptual overview of EducOnto. The lower, middle and upper parts of the diagram depict the classes and properties related to the User Profile, High School and University, respectively.

present the query for CQ1. More precisely, we retrieve the recommended high school majors to start a Bachelor's degree in Computer Science (hereinafter referred to as *curriculum:lg_cs*).

Listing 1: SPARQL query to address CQ1

```
PREFIX ed: <http://purl.org/eduonto/>
PREFIX major: <http://purl.org/edukg/major/>
PREFIX curriculum: <http://purl.org/edukg/curriculum/>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>

SELECT ?major
WHERE { ?major rdf:type/rdfs:subClassOf* ed:HighSchoolMajor .
        curriculum:lg_cs ed:recommendsHighSchoolMajor ?major . }
```

2.4. Directions for Future Work

We will enrich EducOnto by linking it with other ontologies. As previously said, few ontologies model the period from the end of high school to the first years of university. However, some ontologies may be partially reused, e.g. EduProgression Ontology³.

³<https://lov.linkeddata.es/dataset/lov/vocabs/edupro>

In addition, we intend to develop collaborative work with other European institutions and education experts to broaden the scope of EducOnto to countries belonging to the European Higher Education Area⁴.

3. EduKG: An Educational Knowledge Graph

In this section, we detail the whole pipeline for the construction of EduKG. EduKG is publicly available at the following link: purl.org/edukg.

3.1. Data Collection

Data were collected in tabular format through a self-administered online survey whose content has been developed with the education experts who helped define the competency questions (see Section 2.1). Several French high schools and universities were asked to disseminate the survey among their students. A wide spectrum of geographical areas was targeted to ensure representativeness among survey respondents. From January 2022 to April 2022, 3,583 respondents have fully answered the survey and are part of EduKG.

3.2. Construction and Description

Before constructing EduKG, some modeling choices were made, such as:

- Students rated their curriculum on a 0 to 5 scale. Ratings were binarized so as to lead to the predicates *likedCurriculum* (rating ≥ 3) and *dislikedCurriculum* (rating < 3);
- Students were allowed to provide keywords related to their educational interests. Because they were given a freeform text field to fill, consistency among respondents was needed. For instance, keywords that appeared too little were removed, and too precise keywords were replaced by broader concepts;
- Publicly available datasets provided by government agencies were used to enrich the knowledge graph with additional curriculum information.

Then we use Python 3.9 to create RDF triples meeting the conditions specified by EducOnto. All these triples forms our knowledge graph EduKG. A comprehensive view of EduKG is provided in Table 1: the number of distinct entities for each intermediate class (Table 1.a), information about the number of Student-Curriculum interactions and the relative sparsity of EduKG (Table 1.b) and the total amount of entities, relations and triplets of EduKG (Table 1.c).

3.3. Directions for Future Work

We will enrich EduKG by mapping EduKG keywords to DBpedia or Wikidata entities. This would give more genericity to EduKG and facilitate its use in downstream tasks. Similarly to EducOnto, we intend to extend EduKG by adding instances from other European countries, although its feasibility will depend on privacy-related and data-sharing concerns.

⁴https://en.wikipedia.org/wiki/Bologna_Process

Table 1
Statistics of EduKG

(a)	#Fields Of Study	#Keywords	#School Subjects	#Majors	#Specialties
	12	321	15	92	13
(b)	#Users (Students)	#Items (Curricula)	#Interactions		
	3,583	286	7,021		
(c)	#Entities	#Relations	#KG Triples		
	5,452	27	36,301		

On a technical side, a public SPARQL endpoint will be provided, so that information can be retrieved by running SPARQL queries against EduKG.

Finally, recall that the ultimate goal for building EduKG was to provide a public dataset for curriculum recommendation. Thus, the natural next step of this work is to design a RS that specifically relies on EduKG, by refining the work initiated in [9].

Acknowledgments

This work is supported by the AILES PIA3 project (<https://www.projetailes.com/>) that seeks to support French students in their educational choices. We thank all the project partners: the University of Reims Champagne-Ardenne, the University of Lorraine, the Technological University of Troyes and the two Academies of Reims and Nancy-Metz.

References

- [1] M. E. Ibrahim, Y. Yang, D. L. Ndzi, G. Yang, M. Al-Maliki, Ontology-based personalized course recommendation framework, *IEEE Access* 7 (2019) 5180–5199.
- [2] E. Ilkou, H. Abu-Rasheed, M. Tavakoli, S. Hakimov, G. Kismihók, S. Auer, W. Nejdl, Educor: An educational and career-oriented recommendation ontology, *LNCS* (2021) 546–562.
- [3] N. Hubert, A. Brun, D. Monticolo, Vers un système de recommandation explicable pour l’orientation scolaire, in: *Workshop EXPLAIN’AI - EGC*, Blois, France, 2022.
- [4] F. Ricci, L. Rokach, B. Shapira, *Introduction to Recommender Systems Handbook*, Springer US, Boston, MA, 2011, pp. 1–35.
- [5] B. Ma, Y. Taniguchi, S. Konomi, Course recommendation for university environment, in: *EDM*, 2020.
- [6] Z. A. Pardos, W. Jiang, Designing for serendipity in a university course recommendation system, in: *Proc. 10th ACM LAK, Association for Computing Machinery*, 2020, p. 350–359.
- [7] N. F. Noy, D. L. McGuinness, *Ontology development 101: A guide to creating your first ontology*, 2001.
- [8] K. Dellschaft, S. Staab, Strategies for the evaluation of ontology learning, in: *Proc. of the Conf. on Ontology Learning and Population*, IOS Press, 2008, p. 253–272.
- [9] N. Hubert, P. Monnin, A. Brun, D. Monticolo, New strategies for learning knowledge graph embeddings: The recommendation case, in: *EKAU - 23rd International Conf. on Knowledge Engineering and Knowledge Management*, Springer, 2022, pp. 66–80.