A Formalization Of Standard Knowledge And Skills For The French Elementary School Curricula

Oscar Rodríguez Rocha, Catherine Faron Zucker

▶ **To cite this version:**

Oscar Rodríguez Rocha, Catherine Faron Zucker. A Formalization Of Standard Knowledge And Skills For The French Elementary School Curricula. EKM 2016 - Second International Workshop on Educational Knowledge Management, Nov 2016, Bologne, Italy. hal-01393641

HAL Id: hal-01393641
https://inria.hal.science/hal-01393641

Submitted on 7 Nov 2016

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.

A Formalization Of Standard Knowledge And Skills For The French Elementary School Curricula

Oscar Rodríguez Rocha[1], Catherine Faron Zucker[2]

[1] INRIA Sophia Antipolis Méditerranée, Sophia Antipolis, France.
oscar.rodriguez-rocha@inria.fr
[2] Univ. Nice Sophia Antipolis, CNRS, I3S, UMR 7271, Sophia Antipolis, France
faron@unice.fr

Abstract. In the education field, in order to achieve learning goals, it is necessary to define learning paths that foresee a gradual and incremental acquisition of certain knowledge and skills that students should acquire. In this paper we analyze the educational progressions of the French educational system, we show how to formalize them through a web ontology and how to perform knowledge extraction from the official texts describing them to automate the population of such an ontology.

Keywords: e-Education, Ontologies, Semantic Web, Knowledge Extraction, Knowledge Representation

1 Introduction

In the education field, in order to achieve learning goals, it is necessary to define learning paths that foresee a gradual and incremental acquisition of certain knowledge and skills that students should acquire. Such educational programs are usually defined by a body (a ministry or department of education) and then implemented by the schools or institutes that are part of the educational system. Nowadays it is not possible to find them in a standardized format that can be accessible and queried automatically.

In this context, this paper proposes an answer to the following research questions:

1. How to model educational programs into a formal machine understandable ontology that can be processed and queried?
2. How to extract the knowledge and skills that an educational program requires from the official texts describing them?
3. How to automate the population of the ontology based on the extracted knowledge?
4. How can we exploit these formal representations?

Our scientific contribution is first, a web ontology that describes and represents an official standard of knowledge and skills, in this case the French educational system. It is formalized in the OWL Web Ontology Language, which

enables to process and query it with SPARQL, the standard query language for the Web of data. Second, we propose a process to extract knowledge and skills from the official texts describing the French educational program. Finally, we propose a process to automatically populate our ontology with the knowledge extracted from the official texts which we further enrich with the web of data.

The remainder of this paper is structured as follows: Section 2 presents state of the art approaches of ontology-based modeling in e-Education. Section 3 introduces to the French common base of knowledge and skills for students as well as the official educational progressions. In Section 4 we propose EduProgression: an ontology to model the educational progressions of the French educational system. We show, in Section 5, how to exploit the data of the progressions modeled through EduProgression and its thesaurus. Section 6 presents our method to automate the population of this ontology by extracting elements of knowledge and skills from the official texts and Section 7 describes how to link them to relevant resources available on the Web of Data. Finally, the conclusions and future work are presented in Section 8.

2 Related work

The term of "educational ontology" covers very different types of ontologies, modeling different kinds of knowledge related to e-Education. In [7], the authors provide a literature survey of the development and use of ontologies in the domain of e-learning systems. According to the classification proposed in this survey, our work falls within the "curriculum modeling and management" category. Among such works we have CURONTO [1, 2], an ontological model designed for the management of a curriculum and to facilite program review and assessment". Compared to it, our ontology is not focused on giving the means to conduct a review of the program nor an assessment; it focuses on the description and representation of the knowledge and skills required by an educational system and it provides the means to link knowledge and skills with resources from the web of data.

In the same category, Gescur [4] is a tool dedicated to the management and evaluation of the implementation of the curriculum, which facilitates the curriculum management process". It relies on an ontology of concepts relevant to curriculum management in Secondary Schools, such as teachers, departments, objectives, subjects, modules, tasks, documents, policies, activities, learning objects, quality criteria, etc. While this ontology focuses on the administration of the curriculum ours focuses on the management of the knowledge and skills required in each school year in the French educational system. In this sense, both ontologies could be complementary.

Besides the above cited literature review, the educational semantic web ontology proposed in [3] focuses on representing higher education concepts and assisting specialized e-learning systems. In contrast to this work, we propose an ontology that focuses on representing the elements of knowledge and skills of an educational system.

Finally, the OntoEdu educational platform architecture for e-Learning proposed in [5], relies on an activity ontology that describes all the education activities and the relations among them, and a material ontology which describes the educational content organization. The ontology is not available online, therefore it is not possible neither to analyze its classes and properties nor to determine the feasibility of use it in other scenarios outside the OntoEdu platform. Additionally their ontology does not contemplate the use of resources of the web of data to enrich the educational material.

3 The French educational model

In France, there is an official common base of knowledge and skills created by the Ministry of Education and published in the Official Journal; it is called the *"Socle commun de connaissances et de compétences"*[3]. It states that, compulsory education should at least guarantee to each student the necessary means to acquire a common base consisting of a set of knowledge and skills that are essential to master in order to successfully complete her education, build her personal and professional future and succeed in her life in society. The French common base is organized according to seven skills:

- Mastery of the French language
- Practice of a foreign language
- Basic math skills and scientific and technological culture
- Mastery of the common information and communication techniques
- Humanistic culture
- Social and civic competences
- Autonomy and initiative of the students

In order to organize a gradual acquisition of the knowledge and skills of the French common base, some official learning progressions *"Progressions pour l'ecole elmentaire"*[4] for the educational areas of the French elementary school have been defined and published (official bulletin of January 5, 2012)[5]. For each elementary school cycle, the progressions are organized by area of education and strictly follow the program labels to offer a comprehensive approach to knowledge, skills and attitudes to be mastered by the students. Our study is based on the analysis of these progressions (which are based on the French common base of knowledge and skills). By generalizing them, we built a standard model of the knowledge and skills necessary for the students through their educational path and a method to automate the population of this model by extracting knowledge from the texts describing these progressions. Being a standard model, it is also

[3] http://www.legifrance.gouv.fr/WAspad/UnTexteDeJorf?numjo=MENE0601554D

[4] http://eduscol.education.fr/cid58402/progressions-pour-l-ecole-elementaire.html

[5] http://media.education.gouv.fr/file/1/58/7/programmes_ecole-primaire_203587.pdf

possible to model progressions from other educational systems in the world, such as the U.S. academic standards[6] defined by the Common Core State Standards (CCSS)[7].

4 The EduProgression ontology and associated thesaurus

4.1 The EduProgression ontology

We created EduProgression, an ontological model for educational progressions, formalized in the standard Ontology Web Language (OWL). Its main classes are:

- **Element of Knowledge and Skills (EKS).** It is the key concept of our model. It is represented by the EKS class, which is the central class of the EduProgression ontology. An element of knowledge and skills is associated to a set of knowledge pieces (class `Knowledge`) and skills (class `Skill`) for a specific French school cycle (class `Course`) containing reference points (class `PointOfReference`) and also a vocabulary of associated terms (class `Vocabulary`). More precisely:
  - `Knowledge`. Instances of class `Knowledge` are also instances of `skos:Concept` and each one belongs to a `skos:ConceptScheme` that contains all the knowledge pieces of a given progression. An instance of `EKS` is related to an instance of `Knowledge` through the property `hasKnowledge`.
  - `Skill`. An instance of `EKS` is related to an instance of `Skill` through property `hasSkill`.
  - `Course`. In the French educational system, skills that students are expected to develop, are defined by cycle and each cycle is organized into courses. For example, "the consolidation cycle" includes "the second year elementary course" (CE2), "the first-year intermediate course" (CM1) and "the second year intermediate course" (CM2). In this context, an instance of class `Course` represents a course in a cycle. An instance of `EKS` is related to an instance of `Course` through property `hasCourse`.
  - `PointOfReference`. An instance of class `PointOfReference` represents an educational reference element on a specific element of knowledge and skills (an instance of `EKS`). An instance of `EKS` is related to an instance of class `PointOfReference` through property `hasPointOfReference`.
  - `VocabularyItem`. Each element of knowledge and skills has vocabulary items. An instance of `EKS` is related to an instance of class `VocabularyItem` through property `hasVocabularyItem`. A vocabulary item is also an instance of `skos:Concept` and it is related to an instance of `skos:ConceptScheme` which gathers the concepts of the thesaurus of the progression.

[6] http://www.corestandards.org/read-the-standards/
[7] http://www.corestandards.org/

– **Progression.** A progression is represented by an instance of class `Progression`. It can be associated to an existing learning domain (through property `hasLearningDomain`) and to one or several EKSs (through property `hasEKS`).
– **Learning domain.** A learning domain is represented by an instance of the `LearningDomain` class, and it is also an instance of `skos:Concept` that is part of (only) one `skos:ConceptScheme` containing the only learning domains of a progression. Also, as they are SKOS concepts, learning domains are hierarchically organized by using the `skos:broader` and/or `skos:narrower` properties. A learning domain can be associated to a `Progression` or an `EKS`.

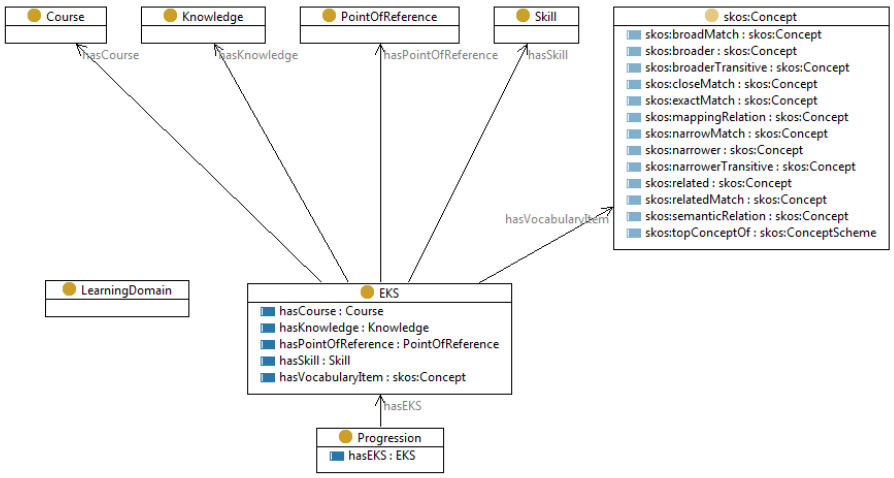

Fig. 1. Classes and Properties of the EduProgression Ontology

The EduProgression ontology is freely accessible online.[8] It is composed of 7 OWL classes and 8 OWL object properties. The relationships between them can be seen in Figure 1.

The two following excerpt is the description of the History progression and a related *EKS* representing the "european construction" based on EduProgression. We highlight with three different colors the three different types of resources related to specific thesaurus described in the following (see Section 4.2):

```
@prefix edp: <http://ns.inria.fr/semed/eduprogression> .
edp:Histoire a edp:Progression ; rdfs:label "History"@en, "Histoire"@fr ;
    edp:hasEKS edp:La_construction_europeenne .
edp:La_construction_europeenne a edp:EKS ;
    rdfs:label "The european construction"@en ;
    rdfs:label "La construction européenne"@fr ;
    edp:hasKnowledge edp:Le_role_de_Robert_Schuman ;
    edp:hasLearningDomain edp:Le_XXe_siecle_et_notre_epoque ;
```

[8] http://ns.inria.fr/semed/eduprogression

```
    edp:hasVocabularyItem edp:Euro .
    edp:hasCourse edp:CM2 ;
    edp:hasPointOfReference edp:Euro_monnaie_europeenne ;
    edp:hasRelatedResource dbpedia-fr:Union_européenne ;
    edp:hasSkill edp:Principales_etapes_de_la_construction_europeenne ;
```

4.2 Three thesaurus associated to each educational progression

Each formalized progression is associated to three thesaurus: one for its knowledge pieces, one for its vocabulary items and another one containing its learning domains. As a result, *knowledge pieces*, *vocabulary items* and *learning domains* are represented by instance classes of the EduProgression ontology which also are SKOS concepts belonging to one of these thesaurus. Grouping these elements in thesaurus allows, among other benefits, to manage in a simple and structured way the knowledge of each progression [6].

The following excerpt shows the structure of the three thesaurus of the progression of History (in blue the thesaurus of knowledge pieces, in green the thesaurus of learning domains and in red the thesaurus of vocabulary items).

```
edp:ElementsConnaissanceHistoire a skos:ConceptScheme ;
     skos:prefLabel "Knowledge pieces of History"@en .
     skos:prefLabel "Elements de connaissance de l'Histoire"@fr .
edp:Le_role_de_Robert_Schuman a edp:Knowledge, skos:Concept ;
     skos:inScheme edp:ElementsConnaissanceHistoire ;
     skos:prefLabel "Know the role of Robert Schuman in European construction"@en .
     skos:prefLabel "Connaître le rôle de Robert Schuman dans la construction européenne"@fr
          .
edp:DomainesApprentissageHistoire a skos:ConceptScheme ;
     skos:prefLabel "Learning domains of History"@en .
     skos:prefLabel "Domaines d'apprentissage de l'Histoire"@fr .
edp:Le_XXe_siecle_et_notre_epoque a edp:LearningDomain, skos:Concept ;
     skos:inScheme edp:DomainesApprentissageHistoire ;
     skos:prefLabel "Elements of knowledge and skills on the twentieth century and our
          time"@en ;
     skos:prefLabel "Éléments de connaissances et de compétences sur le XXe siècle et notre
          époque"@fr .
edp:VocabulaireHistoire a skos:ConceptScheme ;
     skos:prefLabel "Vocabulary of History"@en ;
     skos:prefLabel "Vocabulaire de l'Histoire"@fr .
edp:Euro a skos:Concept ;
     skos:inScheme edp:VocabulaireHistoire ;
     skos:prefLabel "Euro"@en, "Euro"@fr .
```

To give a clearer view of the elements of the History progression described in the previous excerpts, Figure 2 presents it as a graph.

At the time of writing, we have created the thesaurus of the vocabulary, knowledge pieces and learning domains relative to three progressions from the French curricula: History, Geography and Experimental sciences and technology, and we have populated the EduProgression ontology with the concepts in these thesaurus. For the progression on History, we identified 29 elements of knowledge and skills (instances of edp:EKS), the thesaurus of vocabulary items comprises 131 concepts, the thesaurus of knowledge pieces 45 concepts and the thesaurus of learning domains 7 concepts. For the progression of Geography, 32 elements of knowledge and skills were identified; its thesaurus of vocabulary items is composed of 131 concepts; its thesaurus of knowledge pieces and learning domains

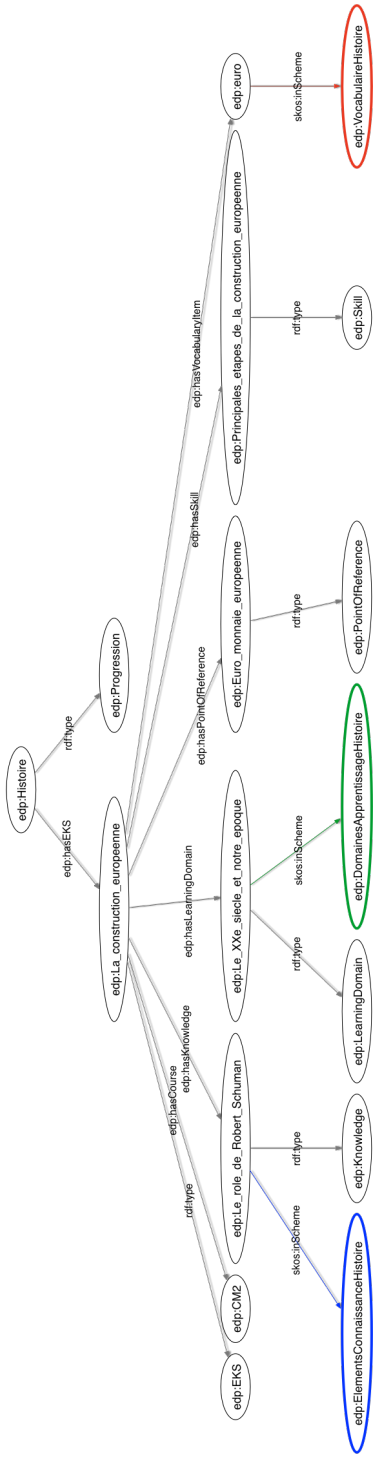

Fig. 2. Graph of the History progression excerpt

contaisn 34 and 6 concepts respectively. Finally, the pogression of Experimental sciences and technology was populated with a total of 50 elements of knowledge and skills; It has the largest vocabulary since it comprises 265 elements. Its thesaurus of knowledge pieces was populated with 40 concepts while its thesaurus of learning domains with 10 concepts. These thesaurus can be freely accessed online: `http://ns.inria.fr/semed/eduprogression/thesaurus`.

5 Exploitation of EduProgression and the associated thesaurus

As the EduProgression ontology has been populated, we can exploit it with SPARQL queries. For example, to retrieve the *EKS* required for the students attending the *second year intermediate course* (CM2) of *History*, we can use the following query:

```
PREFIX edp: <http://ns.inria.fr/semed/eduprogression#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
SELECT ?eks
WHERE {
  ?history_progression a edp:Progression .
  ?history_progression rdfs:label "History"@en .
  ?history_progression edp:hasEKS ?eks .
  ?eks edp:hasCourse edp:CM2 . }
```

Similarly, we can retrieve the required vocabulary for *the first-year intermediate course* (CM1) of *Geography*, within the learning domain of "Elements of knowledge and skills on the territories at different scales".

```
PREFIX edp: <http://ns.inria.fr/semed/eduprogression#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
SELECT ?vocabulary_item
WHERE {
  ?geography_progression a edp:Progression .
  ?geography_progression rdfs:label "Geography"@en .
  ?geography_progression edp:hasEKS ?eks .
  ?eks edp:hasCourse edp:CM1 .
  ?eks edp:hasLearningDomain edp:
       ECC_sur_les_territoires_a_differentes_echelles .
  ?eks edp:hasVocabularyItem ?vocabulary_item . }
```

Finally, the following query allows to retrieve the necessary knowledge pieces of *History* and *Experimental sciences and technology* for *the second year elementary course* (CE2).

```
PREFIX edp: <http://ns.inria.fr/semed/eduprogression#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
SELECT ?knowledge
WHERE {
  ?progression a edp:Progression .
  {?progression rdfs:label "Geography"@en }
  UNION
  {?progression rdfs:label "Experimental sciences and technology"@en }
  ?progression edp:hasEKS ?eks .
  ?eks edp:hasCourse edp:CE2 .
  ?eks edp:hasKnowledge ?knowledge . }
```

It is important to note that thanks to the EduProgression ontology it is possible to make queries to the existing progressions filtered by specific *courses*.

This is the basis for the implementation of educational applications adapted to each student profile.

6 Knowledge extraction from texts and formalization

For the extraction of the information contained in the progressions for elementary school we had to consider first, that the progressions are only available in PDF format, that is, a PDF document for each learning domain. The data contained in each PDF, was inside tables without a standard format, therefore, when we tried to use common techniques for PDF data extraction[9] such as OCR or data extraction from tables[10], the resulting extracted text was mixed among various tables and/or incomplete.

This first led us to perform a preliminary manual extraction of knowledge from text. The first process that we experimented aimed at directly extracting RDF data from PDF texts. This involved to copy textual data directly from each PDF file, and then to paste it into an RDF description as literal values of some RDF triples (also manually written). Not only this process was time-consuming but also it was prone to human errors in the format of the data and the resulting RDF triples.

To solve these problems, we split the data extraction process in two steps, the first one consisting in manually filling a JSON intermediate data structure, the second one consisting in automatically transforming JSON data into RDF data. We designed an intermediate data structure in JSON format in which we represented each progression as a JSON object containing the following information:

- The name of the progression
- An array of the different learning sub domains involved in the progression
- A dictionary with the French scholar courses covered
- An array of the elements of knowledge and skills (EKS) involved in the progression. Each EKS itself contains:
  - a title and a description in French
  - the scholar cycle covered by this EKS
  - the learning domain covered by this EKS
  - a dictionary of the knowledge pieces (titles and descriptions in French) required by this EKS
  - a dictionary containing the required skills (titles and descriptions in French) required by this EKS
  - an array of the vocabulary items required by this EKS
  - an array of points of reference required by this EKS

[9] `https://en.wikipedia.org/wiki/Optical_character_recognition`
[10] `http://tabula.technology/`

This allowed us to manually fill the data structure in a much faster and much more reliable way. In addition, in order to validate the format of the data, we defined a JSON schema[11] representing these aforementioned structural constraints. This enabled us to automatically validated the JSON objects representing a progression manually entered.

As a second step, we defined and performed an automatic conversion process of JSON objects into RDF graphs. It uses the EduProgression ontology to automatically generate an RDF graph.

The following excerpt is the intermediate JSON representation of the EKS on the Age of Discovery, which conversion in RDF is presented in the excerpt of section 4.2, which illustrates the use of the EduProgression ontology.

```
"EKSs": [{
    "short_label": [{"fr": "Les Grandes Découvertes", "en": "Age of Discovery
        "}],
    "label": [{"fr": "Les Grandes Découvertes", "en": "Age of Discovery"}],
    "course": "CE2",
    "learningDomain": "ECC sur les Temps Modernes",
    "knowledge": [{
            "short_label": [{"fr": "Les Européens se lancent à la découverte
                du monde"}],
            "label": [{"fr": "Savoir que les Européens se lancent à la d
                ecouverte du monde grâce notamment aux progrès de la
                cartographie et de la navigation"}]
    }],
    "skills": [{
            "short_label": [{"fr": "Raconter le voyage de Christophe Colomb
                "}],
            "label": [{"fr": "Être capable de raconter le voyage de
                Christophe Colomb et la découverte d'un nouveau monde"}]
    }],
    "points_of_reference": []
},...
]
```

7 Entity Linking

We enriched the EduProgression ontology and its related thesaurus by conducting an entity linking process. It allows us to link the EKS instances of our already populated progressions, with DBpedia resources related to the textual values of the properties describing EKS instances. The information provided by these linked resources may be further exploited directly as learning material or to discover related learning resources.

To make this possible, we have created an automated process that first extracts literal values of the properties of each EKS and builds a textual representation of the latter by concatenating these literal values, and then retrieves the related DBpedia resources by using the DBpedia spotlight web service. The process is as follows:

1. All the instances of *EKS* are retrieved from the populated progressions. This is the result of a `SELECT` *SPARQL* query processed by a **SPARQL engine** on our local triple store.

[11] `http://json-schema.org/`

2. For each retrieved *EKS* instance, the textual values of its properties are extracted and concatenated into a string which will be considered as a whole as the **EKS text representation**.
3. The process takes each **EKS text representation** and analyzes it by using the DBPedia Spotlight web service.
4. Each related resource obtained from the previous step is linked to the *EKS* by using the `edp:hasRelatedResource` property of EduProgression. This is done with an `INSERT` *SPARQL* query applied to our triple store.

The whole process is shown in Figure 3.

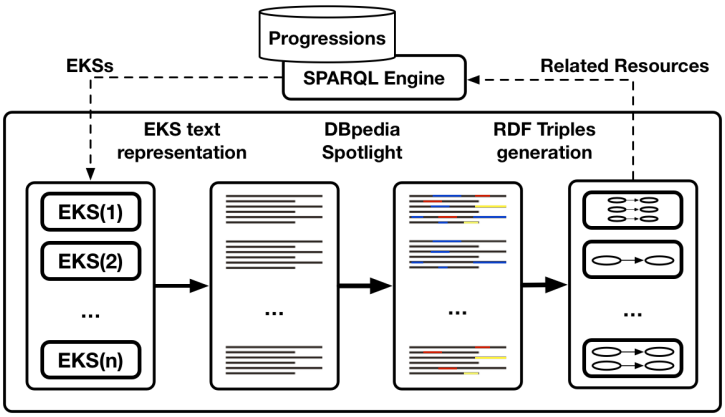

Fig. 3. Process to link EKS instances of a given progression to DBpedia resources

Currently, our process creates links between EKS instances and related resources in DBpedia. We chose to link EKS since they are conceived in our model as the main concept that represents skills and knowledge. Eventually, for other case studies, the EduProgression ontology could be adapted in such a way that the property `hasRelatedResource` can be applied to other types of instances, among which skills, knowledge pieces and vocabulary items, making it possible to link their instances with related resources from the web of data.

As a result of our process, we have obtained 187 resources related to the EKSs of the Geography progression with a precision of 94%, 205 for the History progression with a precision of 95% and 220 for the progression of Experimental Sciences and technology with 94% of precision. For these experiments DBpedia was configured with a confidence of 0.15 and a support of 20.

8 Conclusions and future work

In this paper, we have shown how to model an educational program by formalizing it as a Web ontology in OWL that is understandable by machines and

that can be queried and processed. In particular, we have taken as a case study the French education system, from which we have extracted and formalized educational progressions in an ontology called EduProgression. We sketched how this ontology can be used through some example SPARQL queries. We have described a process for populating the EduProgression ontology, by extracting the knowledge elements and skills involved in the educational progressions from the official texts describing these progressions. Additionally, we showed how we managed to automatically link the ontology instances to Web resources.

At the time of writing, the current "Socle commun de connaissances et de compétences" in France (which constitutes the basis of this work) is that of 2006. As of September 2016, an update to such document will be made[12], although it will involve minor changes, this scientific work will be adapted accordingly.

Future work includes the use of the EduProgression ontology and its related thesaurus for the indexation of educational resources to the French curricula through our ontology, with the aim of recommending them to students in their learning path. This activity will be performed in collaboration with our two industrial partners. In addition, we look forward to apply this work to other educational systems in the world as suggested in Section 3.

References

1. Al-Yahya, M., Al-Faries, A., George, R.: Curonto: An ontological model for curriculum representation. In: Proceedings of the 18th ACM Conference on Innovation and Technology in Computer Science Education. pp. 358–358. ITiCSE '13, ACM, New York, NY, USA (2013), http://doi.acm.org/10.1145/2462476.2465602
2. Alfaries, A., Al-Yahya, M., Chorfi, H., George, R.P.: Curonto: A semantic model of the curriculum for program assessment and improvement. INTERNATIONAL JOURNAL OF ENGINEERING EDUCATION 30(5), 1083–1094 (2014)
3. Bucos, M., Dragulescu, B., Veltan, M.: Designing a semantic web ontology for e-learning in higher education. In: Electronics and Telecommunications (ISETC), 2010 9th International Symposium on. pp. 415–418 (Nov 2010)
4. Dexter, H., Davies, I.: An ontology-based curriculum knowledgebase for managing complexity and change. 2014 IEEE 14th International Conference on Advanced Learning Technologies 0, 136–140 (2009)
5. Guangzuo, C., Fei, C., Hu, C., Shufang, L.: Ontoedu: a case study of ontology-based education grid system for e-learning. In: GCCCE International conference, Hong Kong (2004)
6. Pastor-Sánchez, J.A., Martínez Méndez, F.J., Rodríguez-Muñoz, J.V.: Advantages of thesaurus representation using the simple knowledge organization system (skos) compared with proposed alternatives. Information Research: An International Electronic Journal 14(4), n4 (2009)
7. Yahya, M.A., George, R., Alfaries, A.: Ontologies in e-learning: Review of the literature. International Journal of Software Engineering and Its Applications 9(2), 67–84 (2015)

[12] http://cache.media.education.gouv.fr/file/17/45/6/Socle_commun_de_connaissances,_de_competences_et_de_culture_415456.pdf