View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by INRIA a CCSD electronic archive server

OntoSIDES: Ontology-based student progress monitoring on the national evaluation system of French Medical Schools

Olivier Palombi, Fabrice Jouanot, Nafissetou Nziengam, Behrooz Omidvar-Tehrani, Marie-Christine Rousset, Adam Sanchez

To cite this version:

Olivier Palombi, Fabrice Jouanot, Nafissetou Nziengam, Behrooz Omidvar-Tehrani, Marie-Christine Rousset, et al.. OntoSIDES: Ontology-based student progress monitoring on the national evaluation system of French Medical Schools. Artificial Intelligence in Medicine, Elsevier, 2019, 96, pp.59-67. 10.1016/j.artmed.2019.03.006 . hal-02089318

HAL Id: hal-02089318
https://hal.archives-ouvertes.fr/hal-02089318

Submitted on 17 Apr 2019

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.

OntoSIDES: Ontology-based student progress monitoring on the national evaluation system of French Medical Schools.☆

Olivier Palombi[a,b], Fabrice Jouanot[c], Nafissetou Nziengam[c,d], Behrooz Omidvar-Tehrani[c], Marie-Christine Rousset[c,e], Adam Sanchez[c]

[a] *Université Grenoble Alpes, LADAF, CHU Grenoble Alpes, F-38000 Grenoble, France*
[b] *Université Grenoble Alpes, Grenoble INP, Inria, CNRS, LJK, F-38000, Grenoble, France*
[c] *Université Grenoble Alpes, Grenoble INP, CNRS, Inria, LIG, F-38000, Grenoble, France*
[d] *Yaoundé University, Cameroun*
[e] *Institut universitaire de France, 75005, Paris, France*

Abstract

We introduce OntoSIDES, the core of an ontology-based learning management system in Medicine, in which the educational content, the traces of students' activities and the correction of exams are linked and related to items of an official reference program in a unified RDF data model. OntoSIDES is an RDF knowledge base comprised of a lightweight *domain ontology* that serves as a pivot high-level vocabulary of the query interface with users, and of a *dataset* made of factual statements relating individual entities to classes and properties of the ontology. Thanks to an automatic mapping-based data materialization and rule-based data saturation, OntoSIDES contains around 8 millions triples to date, and provides an integrated access to useful information for student progress monitoring, using a powerful query language (namely SPARQL) allowing users to express their specific needs of data exploration and analysis. Since we do not expect end-users to master the raw syntax of SPARQL and to express directly complex queries in SPARQL, we have designed a set of parametrized queries that users can customize through a user-friendly interface.

Keywords: Learning Management System (LMS), Ontologies, RDF, rules,

☆This work has been partially supported by the LabEx PERSYVAL-Lab (11-LABX-0025-01) and the DUNE project SIDES 3.0 (ANR-16-DUNE -0002-02).

SPARQL, e-Learning, Medicine

1. Introduction

SIDES is a french national e-learning platform for medical training online since 2013. Currently, the platform is used by all the French medical schools (34 over the country) for assessments and graduations. SIDES provides a shared national training repository, filled by teachers who provide exam contents related to a national reference program organized by medical specialties and by items describing learning topics that are standardized across all universities through an official publication from the French Ministry of Higher Education in the so-called *Bulletin Officiel* [1]. In this document, each named and defined *item* can in turn contain *sub-items*. The identifiers of the items and sub-items were used for tagging the questions of exams in the platform. Table 1 shows 2 items (out of 368) of the ECN program, the national ranking examination for accessing the third cycle (residency or postgraduate training) of Medical studies in France.

Item ID	**About item**	**Required skill(s) for the item**
96	Myasthenia	- Diagnose a myasthenia - Know treatments and plan the follow-up of the patient - Know medicine dissuaded in the myasthenia - Identify emergency situations and plan their coverage
262	Urinary lithiasis	- Diagnose an urinary lithiasis - Argue the treatment and plan the follow-up of the patient

Table 1: Two items in the ECN program

SIDES is mainly used for *a)* academic ranking in medical schools and *b)* student self-assessment. In the last two years, more than 4 million self-assessment tests have been performed by students on SIDES. The adoption of computer-based exams allows the use of multimedia contents such as images, videos or sounds, and enables an automatic correction.

SIDES is based on an operational PostgreSQL relational database[1] and benefits from the well-established database technologies for multiple transactions and full accessibility and availability.

SIDES graphical user interface (GUI) is limited to predefined interactions that do not fulfill the needs of its three types of end-users (students, teachers, and institutional administrators) that may have different demands.

With the mediation of the national coordinator for e-learning in Medicine (a co-author of this article) who is both teaching (as a Professor of Medicine) and responsible for making SIDES evolve in terms of pedagogical content and services, we have identified the following *users requirements* that are not fulfilled by SIDES for data analytics (categorized by types of users):

- *Students requirements*: students need to customize the analysis of their own performance in a given (or each) topic or specialty of the program ; in a competitive setting as in Medical studies, each student is also willing to analyse how his/her results (in a given or for each topic or specialty) would be compared with the overall results on the same topic or specialty of groups of students at the same level of study over a given period.

- *Teachers requirements*: teachers needs to monitor groups of students by analysing overall or individual results in a given topic or specialty; they have also expressed the need for fine-grained and domain-specific analysis of the discriminative power of the questions that have been answered by a given population of students during a given test of the national ranking examination.

- *Institutional administrators requirements*: At the institutional level, an overall vision is necessary to identify for instance items without (enough) associated questions. This type of information provides teachers with instructions on the coverage of the program during exams. It also allows

[1] *https://www.postgresql.org*

to review the organization of the courses to improve the coverage of the program.

In this article, we describe OntoSIDES which is the ontology-based layer that we have developed on top of the SIDES database for facilitating the empowerment of SIDES end-users in data analytics, in particular to fulfill the above users requirements (i.e., students, teachers, and institutional administrators).

The remaining of the paper is organized as follows: in Section 2, we position our approach w.r.t. existing work ; in Section 3 we present our contributions ; finally, in Section 4 we conclude with some ongoing work and perspectives.

2. Related work

OntoSides is an implementation of the *Educational Semantic Web* vision initiated by several pioneering works [2, 3, 4] based on taking benefit of Semantic Web technologies for pushing forward Web-based educational environments. This vision has raised high expectations [5], and a lot has been done in this direction in the last decade (see [6] for a survey). However, a lot of things remain to be done for making this vision a reality [7]. Despite some work on ontology extraction from existing course materials (e.g., [8]), or on modeling ontologies about learners (e.g., [9, 10, 11, 12]), one of the main pending issues is the lack of fine-grained domain ontology covering the learning objects (courses, quiz) as well as the learning process (users, on-line actions) [13]. The Learning Analytics Initiative [14] promotes high-level educational standards such as xAPI [15] for describing learning resources and students activities. Existing educational standards, such as xAPI, but also IMS Question and Test Interoperability specification (QTI) [16] or SCORM [17], aim at re-usability and interoperability of learning content. In the existing learning analytics models, the assessment component is missing in all of them, except in the xAPI data model although insufficient [18, 19, 20].

However, these methods are not specialized enough for being used in practice by teachers in a specific domain such as Medicine. The concepts of a domain-specific e-learning ontology such as the one we developed in OntoSIDES can be declared as a specialization of high-level concepts of xAPI such as Activities, Responses or Interactions, or of IMS Tests and Questions model.

OntoSIDES follows the *Open Educational Resource movement* [21, 22], promoted by the Open University[2] that assigns to every course a dedicated URI acting both as an identifier for the course on the Web, and as a way to relate it with useful information such as its level, its title, the topic and description of the course, and with other relevant open educational resources. To the best of our knowledge, OntoSIDES is the first ontology-based open education resource in Medicine, in which the educational content, the traces of students' activities and the correction of exams are linked and related to items of a official reference program in a unified data model.

Instead, from a system and methodological viewpoint, OntoSIDES follows an *Ontology-based Data Access* approach. Ontology-based Data Access (OBDA) [23] is a novel paradigm at the crossroad of Artificial Intelligence and Databases that has recently received an increasing interest because it enables end-users to ask their own (possibly complex) queries using a domain vocabulary they are familiar with. In contrast with standard relational databases for which the focus is the technical optimization of query evaluation by formatting data according to a pre-established schema, linking data to a *domain ontology* makes the data more understandable for users not familiar with database technologies. OBDA systems have a three-level architecture, constituted by an ontology, (relational) data sources, and mappings between the two. In the standard OBDA approach (called the virtual approach), the mappings are exploited to reformulate users' queries into effective queries evaluated over the data sources. In contrast, in OntoSIDES we exploit the mappings for *data materialization*. As a result, OntoSIDES supports full SPARQL 1.1, thus allowing to handle complex aggregated and counting queries. This is a big difference with the OBDA virtual approach for which the state-of-the-art query rewriting algorithms supports (union of)

[2] *http://www.open.ac.uk*

conjunctive queries only.

Compared to *educational data mining* [24], the OBDA approach implemented in OntoSIDES makes learning analytics transparent and explainable to end-users (learners and teachers) which become actors of their data analytics.

3. Results and discussion

We have followed the Semantic Web [25] and Linked Data [26] standards for the semi-automatic and modular construction of OntoSIDES. OntoSIDES is a knowledge base comprised of a *domain ontology* that serves as a pivot high-level vocabulary of the query interface with users, and of a *dataset* made of factual statements relating individual entities to classes and properties of the ontology. Both the ontological statements and the factual statements are uniformly described in the RDF format [27], which makes possible to query OntoSIDES using the SPARQL query language [28]. RDF and SPARQL are standards recommended by W3C for the Semantic Web and Linked Data[3].

For the semi-automatic construction of OntoSIDES we have chosen an expert-based methodology for building manually a lightweight domain ontology enriched with rules, and a OBDA mapping-based materialization approach for the automatic population of the ontology.

The incremental construction of the ontology, rules and the mappings was guided by the needs to express queries that could capture the users requirements highlighted in Section 1.

Figure 1 shows the OBDA architecture of OntoSIDES. The *data source layer* corresponds to the SIDES relational database. The *mapping layer* is made of a set of mapping axioms (expressed using the Ontop syntax) that indicate how to transform relational data stored in SIDES into RDF data conform to the target RDFS schema specified in the ontology layer. The *ontology layer* is made of a set of RDFS statements and of a set of rules (expressed using the

[3] *http://linkeddata.org/*

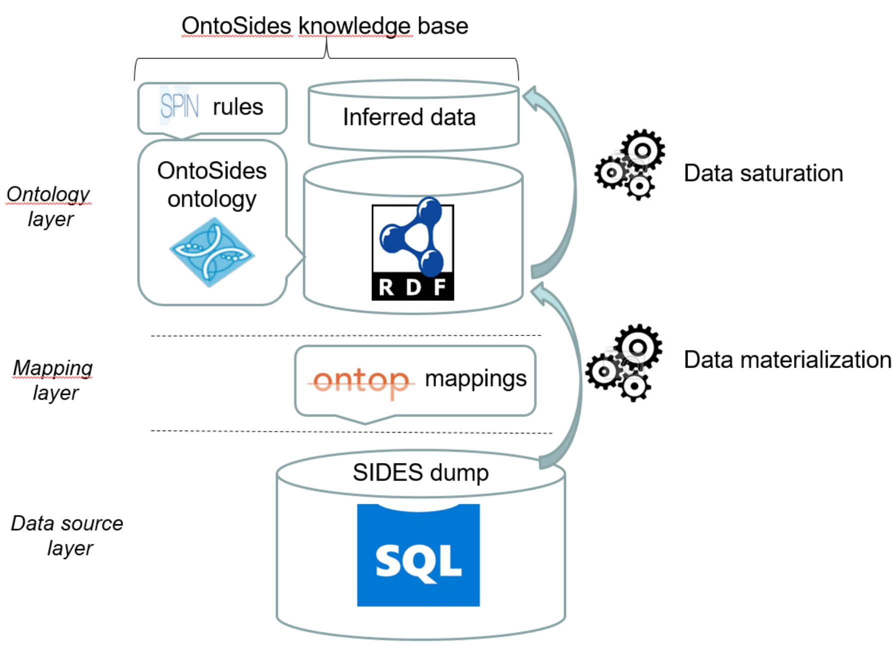


Figure 1: The OBDA architecture of OntoSIDES

SPIN syntax) for defining classes and properties that are meaningful for the target end-users. Within the OntoSIDES knowledge base, we distinguish the OntoSIDES ontology and rules from the OntoSIDES dataset that is obtained by two distinct processes: *data materialization*, which is the process of exploiting the mappings in order to generate the RDF data corresponding to the instances of the classes and the properties defined in the OntoSIDES ontology ; and *data saturation* which, given the materialized data, is the process of making explicit the RDF data that can be logically derived from the RDFS statements and the SPIN rules.

We now summarize the results that have been produced and we provide some elements for their validation.

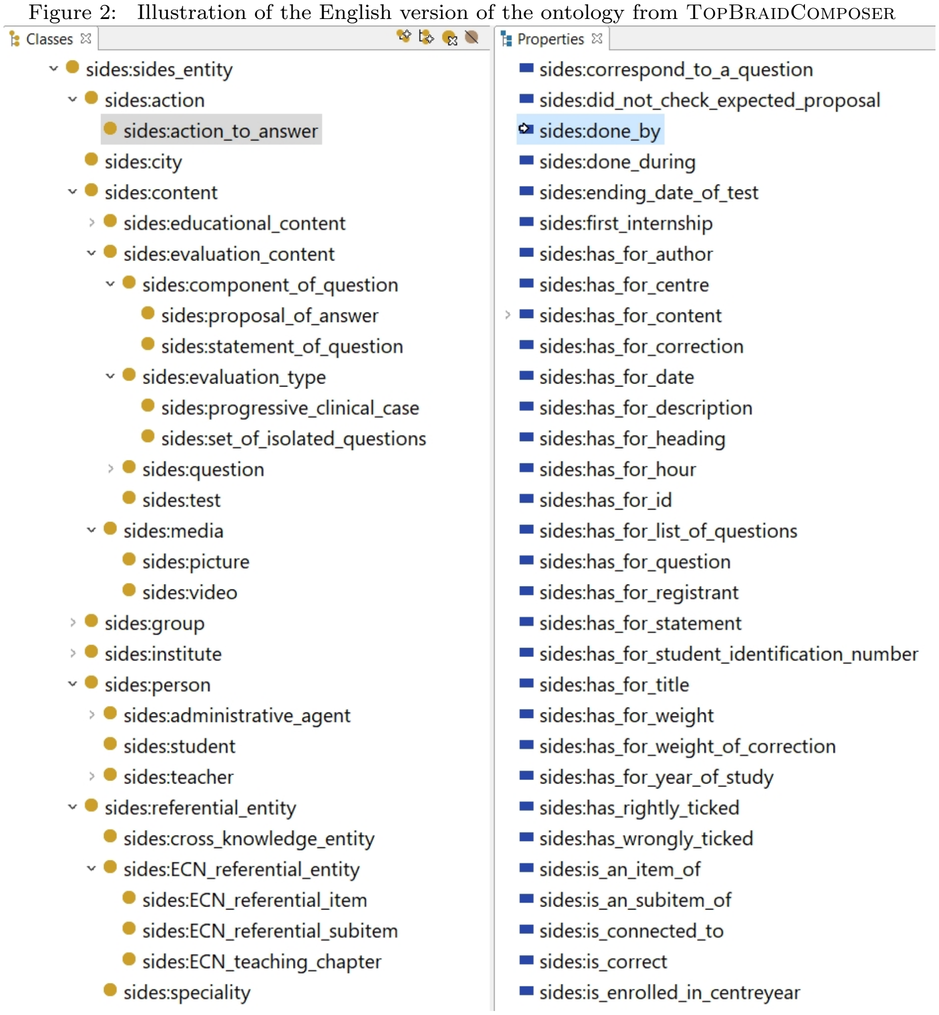
3.1. Ontology and rules

We have chosen the TopBraidComposer software suite [29] to support an incremental methodology for building the OntoSIDES ontology, as TopBraidComposer allows the definition of SPIN rules on top of RDF data and the application of these rules on RDF facts until saturation (i.e., until all the possible facts that can be inferred using the rules are obtained). RDF data equipped with rules capture most of OWL [30] constraints that are useful in practice, such as the transitivity or symmetry properties, as well as domain-specific rules with practical relevance for users in many domains of interest.

Using TopBraidComposer, the OntoSIDES ontology has been modelled by the national coordinator for e-learning in Medicine (a co-author of this article) in close cooperation with the other co-authors that are experienced in knowledge elicitation and ontology construction. Such an expert-based construction of the OntoSIDES ontology and rules has been facilitated by the unique position of the expert who has *(i)* a recognized and shared expertise on the organization and requirements of medical studies in France, *(ii)* a detailed technical knowledge on the SIDES database, and *(iii)* a previous experience in building ontologies in the related domain of anatomy [31, 32].

The resulting ontology consists of a taxonomy of 52 classes and a set of 50 properties among which 3 are defined by rules. Figure 2 shows an extract of the class taxonomy (left) and of properties (right) as displayed by the TopBraid-Composer editor for the english version of the ontology.

Figure 2: Illustration of the English version of the ontology from TopBraidComposer



The class taxonomy is built by successive refinements of *7 main classes* which respectively denote:

- the set of possible actions of students when using the SIDES pedagogical online resources (denoted by the class `sides:action` in Figure 2),

- the types of pedagogical resources (training or evaluation) available in the SIDES platform (denoted by the class `sides:content` in Figure 2),
- the set of reference items of the French educational program in Medicine (denoted by the class `sides:referential_entity` in Figure 2), published by the French Ministry of Higher Education in *Bulletin Officiel* [1] and also used in SIDES as metadata,
- the sets of French cities, universities and medical schools (denoted by the classes `sides:city` and `sides:institute` in Figure 2) for which there are as many local SIDES platforms that are indexed in the central SIDES database,
- the set of milestones (years, periods of practical internships) to register and validate by students to get their diploma (denoted by the class `sides:group` in Figure 2), that are encoded in the SIDES database by specific identifiers,
- the categories of persons (students, academic staff, administrative staff) involved in medical studies (denoted by the class `sides:person` in Figure 2), that correspond to specific roles of users in the SIDES database.

The declaration of properties and their signature (using the pre-defined *rdfs:domain* and *rdfs:range*) completes the ontology by establishing how the instances of different classes can be related or described. For instance, the following RDF statements

(done_by rdfs:domain Action) and *(done_by rdfs:range Student)*

express that the property *done_by* serves to relate (identifiers of) actions extracted from SIDES log traces with (the identifiers of) the students who performed these specific actions. Some temporal properties such as *starting_date_of_test* and *ending_date_of_test* serve to associate the starting / ending time and date to the training tests taken by students.

Some properties p are defined by rules of the form IF *Condition(?s)* IF *(?s p v)*, where *Condition(?s)* is a conjunction of triple patterns expressing conditions for instances of the variable ?s to have v as value for the property p . The different values of such properties are thus logically inferred by triggering the rules whose conditions can be matched with triples present in the dataset . For instance, the boolean property *is_validating_answer* is inferred with the value *true* for an action if it has been performed during an exam (validating test), with the value *false* if it has been realized during a training test (non validating test). This property is defined by the two following rules, in which the variables (preceded by a question mark) can be replaced by as many instances as facts in the knowledge base satisfying the rules conditions.

IF *(?a rdf:type Action)*, *(?e rdf:type Test)*, *(?a done_during ?e)*, *(?e is_validating true)* THEN *(?a is_validating_answer true)*.
IF *(?a rdf:type Action)*, *(?e rdf:type Test)*, *(?a done_during ?e)*, *(?e is_validating false)* THEN *(?a is_validating_answer false)*.

Similarly, the properties *nb_of_discordance* and *has_for_result* are defined by rules. For each answering action done by a student for a given question, they compute the number of discordances with the correct answers for the corresponding question, and the resulting score of the student to that question, respectively. For answer actions related to multiple choice questions, the number of discordances ranges from 0 to 5 since each question has 5 choices of answer. The computation of the different values of the property *nb_of_discordance* is done by 4 rules: one for handling the case of 0 discordance (if the student has ticked all the correct answers, without ticking wrong answers), one for counting the number of wrong ticks (if the student has ticked atleast one wrong answer), one for counting the number of missed right ticks (if the student has not ticked atleast one correct answer) and one for adding the latter two numbers. The property *has_for_result* is then defined in function of the property *nb_of_discordance* by the following rules that compute the score (ranging from 0 to 1) of an action depending on its number of discordances.

IF *(?a rdf:type Action)*, *(?a nb_of_discordance 0)*
THEN *(?a has_for_result 1)*.
IF *(?a rdf:type Action)*, *(?a nb_of_discordance 1)*
THEN *(?a has_for_result 0.5)*.
IF *(?a rdf:type Action)*, *(?a nb_of_discordance 2)*
THEN *(?a has_for_result 0.2)*.
IF *(?a rdf:type Action)*, *(?a nb_of_discordance ?n)*, $?n > 2$
THEN *(?a has_for_result 0)*.

It is important to note that such a rule-based definition of scores makes it easy for students to get explanations for their scores, and for teachers to change the calculation basis for scores just by modifying the rules. This is in contrast with the black-box approach currently implemented in the SIDES platform in which the scores are hard-coded and the number of discordances is not made explicit.

Methodology of validation

The validation of the ontology and of the rules was facilitated by their small size (52 classes, 50 properties, and 10 rules) and the fact that it has been designed by a recognized expert who has a global view of medical studies and of the users' expectations in terms of an e-learning platform in Medicine.

For the ontology, we just checked that all the properties were declared with a domain and a range and that the corresponding classes were correct types for the instances of the property.

For the rules, their fine granularity allowed us to check that all the cases were covered to compute correctly all the possible values of the corresponding properties.

For assessing the global completeness of the ontology enriched with rules, our baseline was to check whether the classes and properties defined in the ontology and the rules enable to specify queries that capture the users' requirements that had been identified (and reported in Section 1). The parametrized queries shown in Section 3.3 provide a positive answer to this validation question.

3.2. Data materialization and saturation

The OntoSIDES dataset is automatically constructed by a sequence of two processes that we explain below.

Mapping-based data materialization:

Using Ontop [33], we have defined as many mappings as classes and properties defined in the ontology. For each property or class of the ontology, a corresponding mapping specifies by an SQL query how to find, within the SIDES relational database, the elements corresponding to instances of this property or class, and how to transform them into RDF format.

For instance, the following mapping has been written to populate the property *correspond_to_a_question* that relates answering actions (performed by students) to questions. The source part is an SQL query constructed over different tables of the relational schema of the SIDES database. The target part indicates the corresponding RDF triples to add for each line of the result returned by the source SQL query.

```
mappingId urn:relation_answer_question
target sides:action{response_id} sides:correspond_to_a_question sides:q{question_id}.
source select concat(cast(response_id as text), cast(choice_id as text))
  as response_id_new, rt.id as response_id, rc.choice_id as choice_id,
  c.question_id as question_id
  from response.responses_choices rc
  inner join response.response_table rt on rt.id = rc.response_id
  inner join choice c on c.id = rc.choice_id
  inner join correction.choice_correction ccc on ccc.id = c.correction_id
```

Listing 1: OBDA mapping for the sides:correspond_to_a_question property

Its evaluation over the current SIDES dump, which corresponds to the activities over a period of 3 months (May, June and July 2015) of the students registered in Grenoble medical school, has resulted into 316,803 RDF triples that have been added to the OntoSIDES dataset as instances of the property *sides:correspond_to_a_question*. The result of the mapping-based data materialization from the current SIDES dump is an RDF graph containing 5,248,288 RDF triples.

It is worthwhile to emphasize that no personal data concerning students are extracted from SIDES. Instances of students in OntoSIDES are just represented by identifiers of the form `sides:etu12402` (as in Example 1). There are no associated properties giving their name or any other personal information stored in the OntoSIDES dataset.

Rule-based data saturation:

Our rule-based approach is implemented by setting up the TopBraidComposer inference mode to take into account RDFS and SPIN rules [34]. The inference mechanism is automatic and consists in applying the rules to the materialized dataset, in all the possible manners satisfying the conditions of the rules. For each possible instantiation of the variables (preceded by a question mark) appearing in a condition part of a given rule such that all its conditions are satisfied by explicit facts, the new facts corresponding to the (appropriately instantiated) conclusion of the rule are added. This saturation process is iterated as long as new facts can be generated. The termination is guaranteed by the form of the rules that are considered. They correspond to safe rules, also called Datalog rules [35], i.e., all the variables appearing in the conclusion of a rule also appears in the condition part.

This saturation process is done offline and the resulting set of inferred facts is added to the OntoSIDES dataset. The saturated version of the OntoSIDES dataset contains 7,962,247 triples. It has been processed in less than one hour on a standard notebook based on core i7 cpu, 16GB RAM and using Tobraid Composer 5 Rev 2.

Methodology of validation

The soundness and completeness of the ontoSIDES dataset, which is crucial for reliable data analytics, relies on the correctness of the mappings and of the rules that have generated it. Their verification is facilitated by the fact that each mapping and each rule targets a given class or property. Since the mappings and rules are specified in an explicit and declarative manner, their soundness is quite easy to check by a domain expert for whom the expressions in the source and target parts of the mappings (respectively, in the conditions and conclusions of the rules) are comprehensible. Understanding the expressions in the source part of the mappings requires in addition a good knowledge of the operational SIDES database since they are SQL queries expressed in the schema of the SIDES database. For helping the expert to validate and to complete the mappings if necessary, we use SPARQL queries to check whether properties known to be mandatory for a given class (e.g., the property *done_by* for the class *Action*) are valued for every instance of the class. If this is not the case, the corresponding mapping is incomplete and must be modified accordingly.

We follow a similar approach to check the completeness of a set of rules for defining a given property, like the 4 rules above for defining the property *has_for_result* that should assign a value to all the answering actions in the dataset. This can be seen as a cross-validation of the rules and the data.

3.3. Parametrized queries

The OntoSIDES knowledge base provides an integrated access to useful information extracted from different data sources, with a powerful query language (namely SPARQL) allowing users to express their specific needs of data exploration and analysis. Through SPARQL queries, a student can explore the set of reference items for which he/she has not trained yet, or launch a comparative analysis of his/her progress on a specific part of the program. Also a teacher may be interested in analyzing the strengths and weaknesses of his/her group of students compared to other groups at the same study level in other universities.

Example 1 provides the SPARQL syntax of a query to retrieve the list of ECN reference items for which a given student (`sides:etu12402`) has not answered any related question.

Example 1.

```
SELECT ?item WHERE {?item a sides:ECN_referential_item .
    FILTER NOT EXISTS {?q sides:is_linked_to ?item.
                       ?a sides:done_by sides:etu12402.
                       ?a sides:corresponds_to_question ?q}}
```

SPARQL can also be used as a declarative tool of tabular data preparation for advanced data analytics or visualization.

Full SPARQL (i.e., SPARQL 1.1) allows to specify on-demand calculation of simple statistical summaries of data blocks thanks to so-called *aggregation queries*. An aggregation query enables a user to declare data blocks of interest and numbers about these blocks that she wants to calculate based on aggregate functions. Such aggregation queries can be used to calculate distributions of numerical data that are computed on-the-fly by grouping facts in the knowledge base, such as grouping the questions by entity of the program to count them, or to compute the average marks obtained by (some groups of) students.

The distinguishing point with standard relational databases (or with Excel sheets) is that grouping can be defined on properties that are not explicitly instantiated but whose values must be inferred. In the setting of RDF knowledge bases, it is worth to outline that, without inference, the statistics obtained by evaluating aggregation queries may be incorrect even if they are correctly specified by users. For instance, Example 2 shows a simple SPARQL aggregation query that counts, for each entity of the program, the number of questions related to it.

Example 2.

```
SELECT ?entity  ((COUNT (DISTINCT ?q)) AS ?nbRelatedQuestions)
    WHERE { ?q rdf:type sides:question.
            ?q sides:is_linked_to  ?entity.
            ?entity rdf:type sides:referential_entity
    } GROUP BY ?entity
```

If it is evaluated against OntoSIDES before saturation, it will return an empty answer because the class `sides:question` has no explicit instance since each specific question appears in fact as an instance of one of the `sides:question`'s subclasses `sides:QMA` (denoting questions with multiple answers) or `sides:QUA` (denoting questions with unique answers). For the `sides:referential_entity` class, similarly, its (implicit) instances are in fact categorized as medical specialties (grouped as instances of the subclass `sides:speciality`), or pieces of transverse knowledge (instances of the `sides:cross_knowledge_entity` subclass), or items of the ECN program (denoted by the `sides:ECN_referential_entity` subclass). When evaluated *after the saturation* of OntoSIDES, it will provide a table with correct answers: a line per entity having related questions with the right number of related questions in each line.

We do not expect end-users to master the raw syntax of SPARQL and to express directly complex queries in SPARQL. In fact, we have designed a set of parametrized queries that users can customize through a user-friendly interface available at http://ontosides.univ-grenoble-alpes.fr/. To achieve realtime answering for multi-user access the Ontology layer uses the Virtuoso triplestore in version 7 on a server computer based on Intel Xeon 8 cpu E5-2699 @ 2.20Ghz with 16GB RAM.

We follow the categorization of the users' needs depicted in Section 1 to illustrate how these parametrized queries evaluated on top of the OntoSIDES knowledge base fulfill them.

Student-centered analyses

Figure 3 shows the interface for students who can choose to launch one of the six queries in the left column that will be parametrized by his/her identifier (`sides:etud12402` in the example). The chart in the right side of the Figure shows the result returned to the first query for that student, i.e., the average results obtained per specialty by the student (to all the questions he/she answered related to this specialty), compared to the overall average obtained per specialty by the whole group of active students.

Based on these results, it is easy to see that this student has an excellent grade profile with very high scores above average in all specialities.

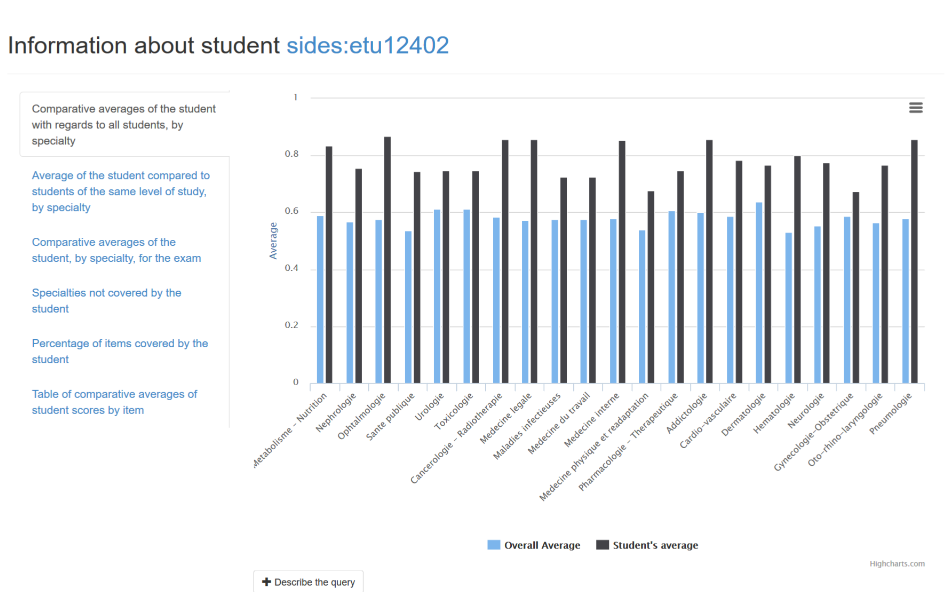


Figure 3: Parametrized queries for students

In contrast, Figure 4 exhibits the results of a student with a grade profile very close to that of the overall average. Personalized recommendations can however be given to him/her: he/she must focus on neurology, hepato-gastroenterology, gynecology and pediatrics if she wishes to improve her skills in these specialties.

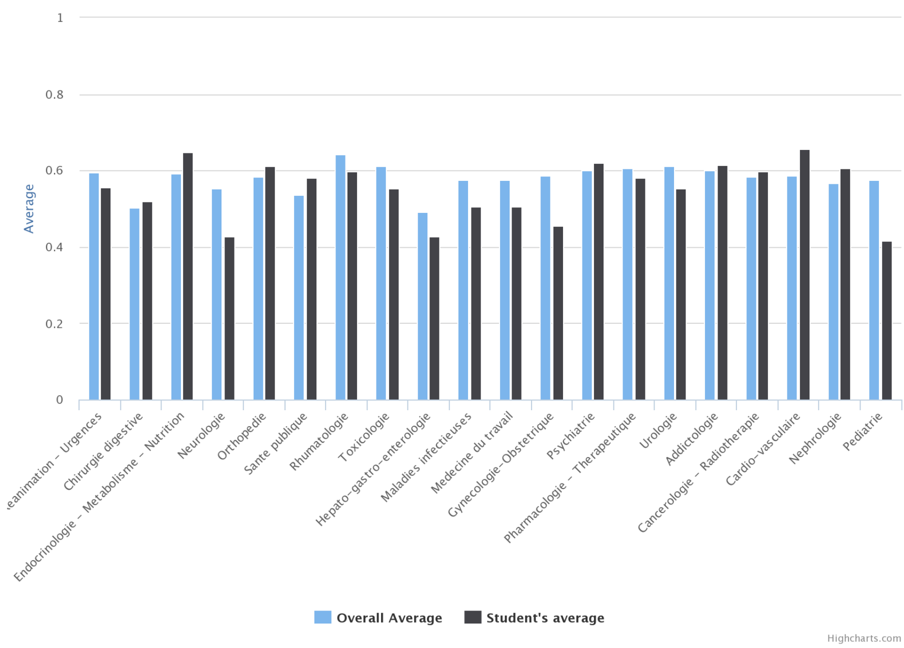


Figure 4: Results of an average student with small variations by specialties

Finally, Figure 5 exhibits the results of a student who is in trouble since he/she is below average in most of the specialties, except in cardiology, anesthesia and intensive care. A personalized procedure can be used to motivate him/her to work in the other specialties.

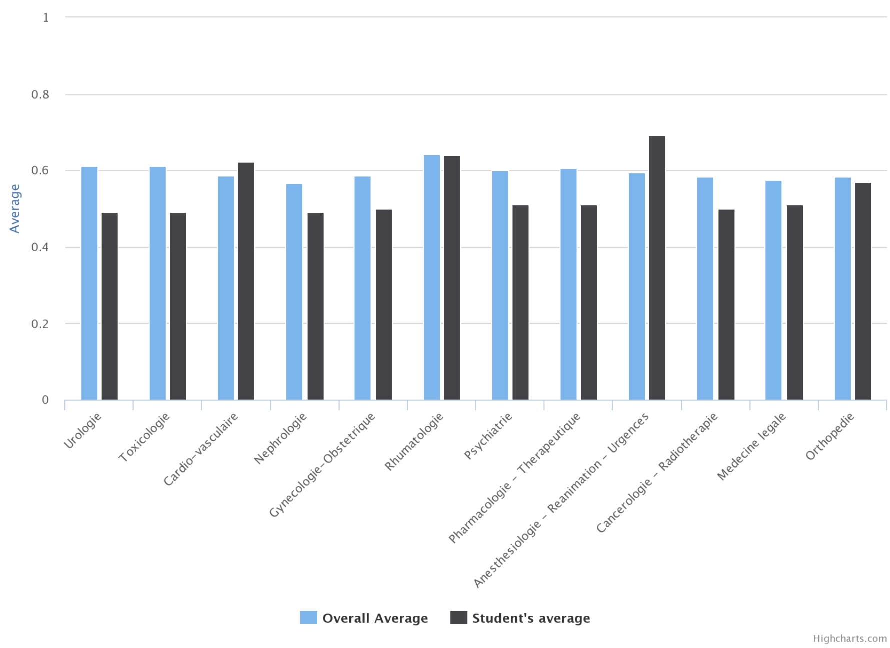


Figure 5: Results of a below average student in almost all the specialities

There exists plenty of other student-centered analyses. For instance, the query of Example 1 is an instantiation of a parametrized query that gets the list of items not covered by a given student. It is also possible to deepen and detail the analysis of average scores of one student per item to identify items to work specifically.

Teacher-centered analyses

Teachers may be interested by monitoring each student individually (as illustrated previously), or a group of students. They can also have more complex needs that require the evaluation of a chain of queries. This is the case for instance for finding which questions of a given test were able to discriminate between the students who took the test. This is a crucial criteria for adjusting the weight of questions for the final ranking of the annual national classification examination. For a given test, a question is judged discriminative by teachers if there is a significant difference between the average marks obtained for this question by the group of students who got the best global marks to the test compared to the group of students who got the worst global marks to the test. We were able to capture this definition by a *sequence of queries* parametrized by a given test, which can be chosen by users by clicking “Groups comparison” in the banner of the Web interface. First, 3 aggregated queries allow to construct 3 groups of students who took the test: the students are in Group 1 (respectively Group 2, Group 3) if they are in the first third (respectively second third, last third) of students ranked (in decreasing order) according to their global mark to this test. Then, a subsequent aggregation query computes, for each question of the test, the difference between the average mark obtained on this question by the students of Group1 and the average mark obtained on this question by the students of Group3. Figure 6 shows the top-4 most discriminative questions for a given test, according to the above specification.

For sides:eval52749 in sides:epreuve43969

+ Show the query

q	AverageQuestionGroup1	AverageQuestionGroup3	Difference
sides:q406589	0.8	0.181818181818182	0.618182
sides:q406588	0.961538461538462	0.491666666666667	0.469872
sides:q406586	0.707692307692308	0.284615384615385	0.423077
sides:q406582	0.884615384615385	0.469230769230769	0.415385

Figure 6: Comparison between the top and bottom student groups of their result for three questions of a given test

Administrator-centered analyses

At the institutional level, an overall vision is necessary to assess the global organization of the training activities. We have designed a set of parametrized queries (accessible by clicking “Management queries” in the banner of the Web interface) that provide some useful statistics for this. For instance, Figure 7 provides the list of items in the ECN program without associated questions.

item	intituleItem
sides:item_75	Addiction aux medicaments psychotropes (benzodiazepines et apparentes) (voir item 319)
sides:item_87	Alteration de la fonction auditive (voir item 127)
sides:item_79	Alteration de la fonction visuelle (voir item 127)
sides:item_133	Anesthesie locale, locoregionale et generale
sides:item_80	Anomalie de la vision d'apparition brutale

Figure 7: Items without associated question

By clicking the button “Describe the query”, users can see the SPARQL queries that compute the results, as shown in Figure 8 . Such functionality can help users to make sense of the computations that are done and to modify themselves the query to adjust the desired computation.

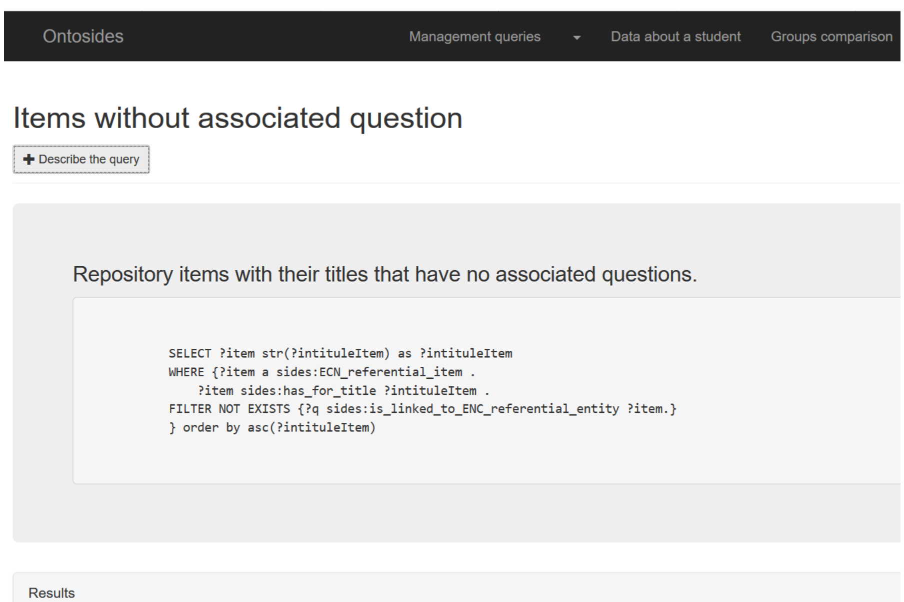


Figure 8: The SPARQL query that computes the list of items without associated question

4. Conclusion and Perspectives

OntoSIDES is a proof of concept for the feasability of an ontology-based learning management system in Medicine. In such an approach, a domain ontology provides a query interface for end-users to interact with data stored about their learning activities related to a given educational program. This enables end-users to receive personalized feedback and recommendations, and to become the major actors of their own progress monitoring and training. Our approach is centered on an ontology which serves as a pivot vocabulary relevant both for students, teachers and institutional administrators involved in education for a given discipline. The construction of the OntoSIDES ontology has been facilitated by the fact that Medicine is a discipline for which teaching objectives and evaluations are well codified standards shared by a whole community.

OntoSIDES has been a decisive step in setting up the ongoing SIDES 3.0 project that is conducted under the authority of UNESS (https://www.uness.fr/), Université Numérique en Santé et en Sport. The methodology and architecture set up for the version of OntoSIDES described in the current article have shown their robustness since the mapping-based data materialization of the OntoSIDES dataset has scaled up to 1,5 billions of triples just by applying it to a full dump of the SIDES database (corresponding to the activities of 65,000 students over several years). In SIDES 3.0, the goal is to remove the current operational SIDES database and to replace it by an OntoSIDES evolving knowledge graph.

The methodology that we developed is not specific to Medicine. It can be reused for enriching or building learning management systems in other disciplines. The effort is depending mainly on the existence of a standard referential for the educational program in the target discipline.

Our ongoing work aims at palliating the incompleteness of the SIDES database and thus of OntoSIDES in which explicit links are missing between questions and items of the programme or specialities for a significant fraction of the questions. We are investigating how to enrich OntoSIDES by automatic links discovery using text mining techniques.

References

[1] French Ministry for Higher Education and Research, Etudes Médicales, http://www.enseignementsup-recherche.gouv.fr/pid20536/bulletin-officiel.html?cid_bo=71544&cbo=1, bulletin Officiel n^o 20, May 16th (2013).

[2] V. Devedzic, Education and the Semantic Web, Int. J. Artif. Intell. Ed. 14 (2) (2004) 165–191.
URL http://dl.acm.org/citation.cfm?id=1434858.1434860

[3] T. Anderson, D. Whitelock, The Educational Semantic Web: Visioning and Practicing the Future of Education, Journal of Interactive Media in Education 2004 (1) (2004) Art. 1. doi:10.5334/2004-1.
URL http://jime.open.ac.uk/articles/10.5334/2004-1/

[4] L. Aroyo, D. Dicheva, The New Challenges for E-learning: The Educational Semantic Web, Educational Technology & Society 7 (2004) 59–69.

[5] P. Miranda, P. Isaias, C. J. Costa, E-Learning and web generations: Towards Web 3.0 and E-Learning 3.0, in: International Proceedings of Economics Development and Research, Vol. 81, IACSIT Press, 2014, p. 92.

[6] K. Krieger, D. Rosner, Linked data in e-learning: A survey, Semantic Web Journal 1.

[7] I. A. Bogoslov, Future Research Directions on Web-Based Educational Systems, in: R. Ortean, C. Ogrean, S. C. Mrginean (Eds.), Innovative Business DevelopmentA Global Perspective, Springer Proceedings in Business and Economics, Springer International Publishing, 2018, pp. 9–20.

[8] M. Hatala, D. Gaevi, M. Siadaty, J. Jovanovi, C. Torniai, Can Educators Develop Ontologies Using Ontology Extraction Tools: An End-User Study,

in: U. Cress, V. Dimitrova, M. Specht (Eds.), Learning in the Synergy of Multiple Disciplines, Springer Berlin Heidelberg, Berlin, Heidelberg, 2009, pp. 140–153.

[9] M. R. F. Sani, N. Mohammadian, M. Hoseini, Ontological Learner Modeling, Procedia - Social and Behavioral Sciences 46 (2012) 5238 – 5243. doi:https://doi.org/10.1016/j.sbspro.2012.06.416. URL http://www.sciencedirect.com/science/article/pii/S1877042812021520

[10] M. Rani, R. Nayak, O. P. Vyas, An ontology-based adaptive personalized e-learning system, assisted by software agents on cloud storage, Knowledge-Based Systems 90 (2015) 33–48. doi:10.1016/j.knosys.2015.10.002. URL http://www.sciencedirect.com/science/article/pii/S0950705115003779

[11] M. Rani, K. V. Srivastava, O. P. Vyas, An ontological learning management system, Computer Applications in Engineering Education 24 (5) (2016) 706–722. doi:10.1002/cae.21742. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/cae.21742

[12] M. Kultsova, A. Anikin, I. Zhukova, A. Dvoryankin, Ontology-based learning content management system in programming languages domain, in: Proceeding of the First Conference on Creativity in Intelligent Technologies and Data Science, Communications in Computer and Information Science, 2015, pp. 767–777. doi:10.1007/978-3-319-23766-4.

[13] R. D. Morris, Web 3.0: Implications for Online Learning I, TechTrends 55 (1). doi:https://doi.org/10.1007/s11528-011-0469-9. URL https://mafiadoc.com/web-30-implications-for-online-learning-i_5a262d341723dd4b03de09b5.html

[14] The Apereo Learning Analytics, https://www.apereo.org/communities/learning-analytics-initiative.

[15] The Experience API (xAPI) for learning technology, https://experienceapi.com/.

[16] The IMS Question and Test Interoperability specification, https://www.imsglobal.org/question/index.html.

[17] Sharable Content Object Reference Model (SCORM), https://scorm.com/.

[18] A. Nouira, L. Cheniti-Belcadhi, R. Braham, An enhanced xapi data model supporting assessment analytics, in: Knowledge-Based and Intelligent Information & Engineering Systems: Proceedings of the 22nd International Conference KES-2018, Belgrade, Serbia, 3-5 September 2018., 2018, pp. 566–575. doi:10.1016/j.procs.2018.07.291.
URL https://doi.org/10.1016/j.procs.2018.07.291

[19] A. Nouira, L. Cheniti-Belcadhi, R. Braham, A semantic web based architecture for assessment analytics, in: 29th IEEE International Conference on Tools with Artificial Intelligence, ICTAI 2017, Boston, MA, USA, November 6-8, 2017, 2017, pp. 1190–1197. doi:10.1109/ICTAI.2017.00181.
URL https://doi.org/10.1109/ICTAI.2017.00181

[20] A. Nouira, L. Cheniti-Belcadhi, R. Braham, An ontological model for assessment analytics, in: Proceedings of the 13th International Conference on Web Information Systems and Technologies, WEBIST 2017, Porto, Portugal, April 25-27, 2017, 2017, pp. 243–251. doi:10.5220/0006284302430251.
URL https://doi.org/10.5220/0006284302430251

[21] Open Educational Resource Handbook, http://wikieducator.org/OER_Handbook.

[22] D. Mouromtsev, Mathieud'Aquin (Eds.), Open Data for Education: Linked, Shared and Reusable Data for Teaching and Learning, Springer, 2016. doi:10.1007/2F978-3-319-30493-9.

[23] A. Poggi, D. Lembo, D. Calvanese, G. D. Giacomo, M. Lenzerini, R. Rosati, Linking data to ontologies, Journal f Data Semantics 10 (2008) 133–173.

[24] C. Romero, S. Ventura, M. Pecheniszkiy, R. Baker (Eds.), Handbook of Educational Data Mining, CRC Press, 2010.

[25] D. Allemang, J. Hendler, Semantic Web for the Working Ontologist: Modeling in RDF, RDFS and OWL, Morgan Kaufmann, 2011.

[26] T. Heath, C. Bizer, Linked Data : Evolving the Web into a Global Data Space, Morgan and Claypool, 2011.

[27] W3C, RDF Description Format, `http://www.w3.org/TR/rdf-concepts/` (2004).

[28] W3C, SPARQL 1.1 Overview, `https://www.w3.org/TR/sparql11-overview/` (2013).

[29] Top Braid Composer from TopQuadrant software, `http://www.topquadrant.com/`.

[30] W3C, Web Ontology Language (OWL), `http://www.w3.org/2004/OWL/` (2004).

[31] O. Palombi, F. Ulliana, V. Favier, J.-C. Leon, M.-C. Rousset, My corporis fabrica: an ontology-based tool for reasoning and querying on complex anatomical models, Journal of Biomedical Semantics (JOBS 2014) 5 (20).

[32] P.-Y. Rabattu, B. Masse, F. Ulliana, M.-C. Rousset, D. Rohmer, J.-C. Leon, O. Palombi, My corporis fabrica embryo: An ontology-based 3d spatio-temporal modeling of human embryo development, Journal of Biomedical Semantics (JOBS 2015) 6 (36).

[33] D. Calvanese, B. Cogrel, S. Komla-Ebri, R. Kontchakov, M. R. Davide Lanti, M. Rodriguez-Muro, G. Xiao, Ontop: Answering sparql queries over relational databases, Semantic Web Journal (SWJ 2017).

[34] W3C, SPIN Overview and Motivation, https://www.w3.org/Submission/spin-overview/ (2011).

[35] S. Abiteboul, R. Hull, V. Vianu, Foundations of Databases, Addison-Wesley, 1995.