Using the Semantic Web in Academic Websites

Loredana MOCEAN
Babes-Bolyai University, Cluj-Napoca, Romania
loredana_mocean@yahoo.com

One of the important activities in the academic environment is student orientation, professional counselling and work with students, as well as interaction with them. This collaboration becomes a feature of the professionalism of the management team as well as of the knowledge and abilities of its members. This activity is based on the information system on which it relies and so, becomes essential. A semantic Student Administration System can help university managers, teachers, students and administrative staff to integrate information into the higher education organization. This asset can be part of a strategy to improve students' management to meet the nowadays challenges and requirements for data integration. The paper motivates and presents an ontology adapted to the needs of educational data management. This new semantic web-based system helps students and members of the university management team. The presented paper describes the components and architecture of the system that underlie its implementation in a real platform and proposes an early stage ontology design for the area of application presented.
Keywords: Student Administration Systems, Semantic Web, Modelling Academic Websites, Local ontology, Knowledge

1 Introduction

A semantic Student Administration System (SAS) is a next-generation system for universities that combines the latest semantic technology with existing data management experience of education institutions (such as universities, institutes, high schools).

We propose to design an application that simplifies and reduces the system costs, provides students with an intuitive mobile experience, integrated in a knowledge graph model based on Semantic Web. The system must be sufficiently complete and robust to serve as an effective template for future research.

Starting from these considerations, the article has the following structure. In section 2 we analyze the problem of our research and the background, a literature review and key achievements for our research. Section 3 presents the workflow for the proposal, Section 4 describes modelling decisions and Section 5 presents the case study.

The paper ends with the Conclusions section.

2 Problem Statement, Background and State of the Art

Applications currently implemented in Romania's universities are mostly based on relational databases [1], isolated and lacking interoperability mechanisms with external systems or available open data sources.

That is why our proposal is to migrate the existing relational databases to semantic technology and to integrate all these applications according to the integration patterns of RDF Graphs and Linked Enterprise Data.

We want to build a new ontology and suggest a way to design a new model of such an ontology for our university, however the proposal should be generalizable for other institutions as well.

The actors affected are all those working in a university (professors, staff or students) but students are the main beneficiaries of the system. The university's employees can perceive a variety of benefits (stronger collaboration, automated operations, web of information, retrieval of connected data), which are now administered by different offices, both inside and outside the university.

In paper (Zhao et al.in 2004 [2]), the authors explore the use of Semantic Web technologies such as RDF and ontologies to support the

representation of information. Also they study existing initiatives such as Jena and LSID, to generate and store knowledge repositories.

The authors of paper [3], Lima, F. and Schwabe, D. argue the need for flexible models capable of representing semi-structured data processed in Web applications and also give the requirements for query languages that allow the specification of navigational aspects of Web applications.

The project presented in [4] describes the research methodological approach of a work in progress project. It is a Semantic Web related project that presents an architecture with several layers, one of them is Data Interchange based on RDF.

Gruber in 1993 [5] describes the role of ontologies in supporting knowledge sharing activities and then presents a set of criteria to guide the development of ontologies for these purposes. Paper [6] presents an ontology-based approach to e-learning system adapted to the needs of training human resource in a hospital. Also presents the components and system architecture underlying its implementation in a real platform, be a useful guide to how to build an ontology.

The study described in the paper of Saarsen [7] shows that models in which the ordering relations between tasks are captured in diagrammatic form, while the details of each task are captured in textual form, are more likely to be used on a sustained basis. Another paper that supports this thesis from a Knowledge Management perspective is [14].

In [8] the author state that "the design of a symbol influences whether it is easy to spot in a model and is correctly associated with the concept it represents".

By providing an ontology for representing risks and mitigation measures, the information can be represented through annotations in arbitrary types of enterprise models [9].

Studying the literature we can conclude that an infrastructure based on Semantic Web for SAS doesn't exist, there are only some synergetic structures.

We want to build a new architecture with semantic data services built upon existing sources – an earlier stage of this research was reported in [16], where we focused on semantic data links rather than the ontological aspect.

3 The Workflow of the Proposal
A large number of universities around the world design their individual database on a relational model for a variety of application domains.

Our proposal is to replace the existing applications - see the AcademicInfo Application existing at the level of the Romanian universities, implemented in relational database (https://academicinfo.ubbcluj.ro/Info/ CadreDidactice.aspx or https://www.amgd.ro/) with a D2RQ-lifted semantic database connected to the website of the university.

The construction of such an applicable model is complex especially when considering the entire spectrum of data and persons who interact with the platform.

Figure 1 presents the underlying recommended development process and the main steps in building the new system.

The steps are based on definition of terms, identification and motivation and end with communication, shared understanding, and agreement – most such systems project problems and errors are the result of failures of communication or a lack of shared understanding or agreement.

4 Modelling Decisions
Our work highlights the idea that "Artifacts must be proved upon existing solution to a problem or perhaps provide a first solution to an important problem" [17].

Among the methodologies studied, Method 101 helps ontology building the proposed system ([18] and [19]).

According to the study [20] there are a variety of tools that support ontology development process activities. In particular case study, domain knowledge graph construction system is presented in paper [23].

Protégé, an ontology editor based on Java (http://protege.stanford.edu/), is a standalone application, open source with an extensible architecture. The languages for the ontologies began to be created in the early 1990s, usually

as an evolution of existing knowledge representation languages. Among the most relevant languages include: Resource Description Framework and the Web Ontology Language (OWL), with the Extensible Markup Language (XML) as an early choice for serialization format.
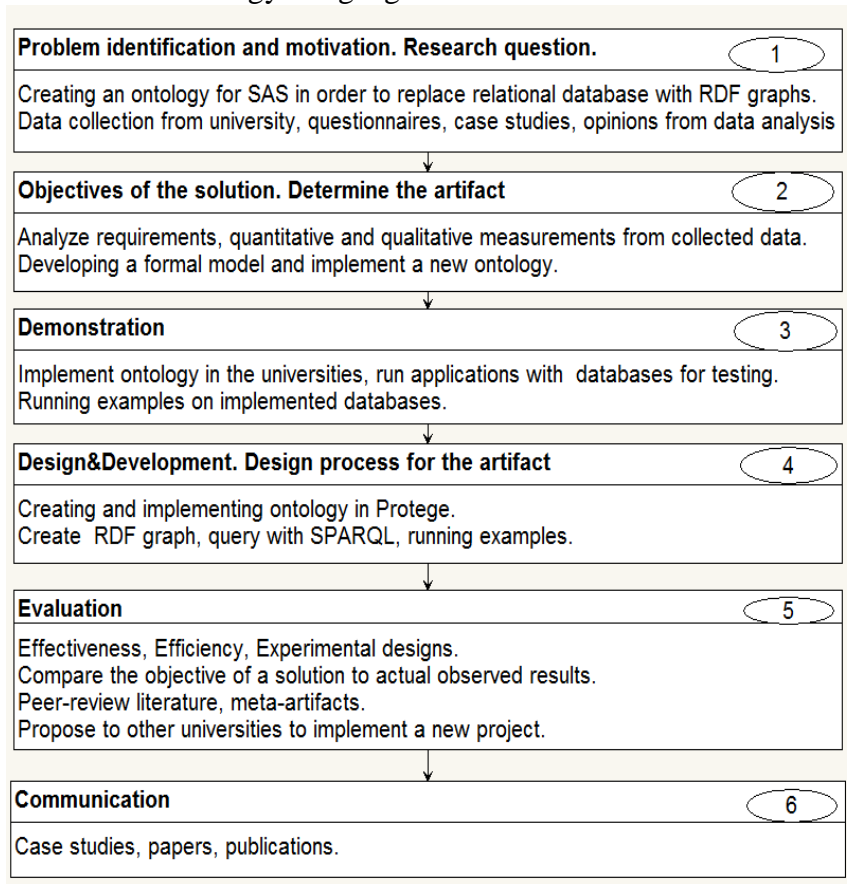
Problem identification and motivation. Research question.	1
Creating an ontology for SAS in order to replace relational database with RDF graphs. Data collection from university, questionnaires, case studies, opinions from data analysis	

Objectives of the solution. Determine the artifact	2
Analyze requirements, quantitative and qualitative measurements from collected data. Developing a formal model and implement a new ontology.	

Demonstration	3
Implement ontology in the universities, run applications with databases for testing. Running examples on implemented databases.	

Design&Development. Design process for the artifact	4
Creating and implementing ontology in Protege. Create RDF graph, query with SPARQL, running examples.	

Evaluation	5
Effectiveness, Efficiency, Experimental designs. Compare the objective of a solution to actual observed results. Peer-review literature, meta-artifacts. Propose to other universities to implement a new project.	

Communication	6
Case studies, papers, publications.	

Fig. 1. The steps in building a new system

OWL is designed to represent knowledge semantically enriched and complex representations about things, groups of things and relations between things.

Structural units for our proposal are represented in Figure 2. We start with the definition of system environment, we continue with the modelling of entities and we finalize with step 3, Knowledge Domain Modelling.
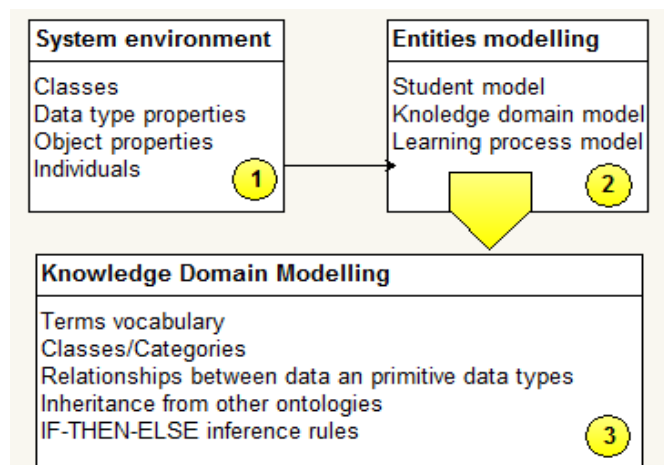
System environment

Classes
Data type properties
Object properties
Individuals   1

Entities modelling

Student model
Knoledge domain model
Learning process model   2

Knowledge Domain Modelling

Terms vocabulary
Classes/Categories
Relationships between data an primitive data types
Inheritance from other ontologies
IF-THEN-ELSE inference rules   3

Fig. 2. Structural units for modelling

The system meets the performance improvement, skills and capacity assessment for the students and staff of universities by proposing an automated information retrieval system.

For this purpose, ontologies are used in modelling the domain of interest and to build, organize and update specific learning resources (student profiles, learning pathways, learning objects). Pragmatic considerations (e.g., cost, effort of implementation, which may limit the number of artefacts) are also taken into account [14].

The design of this system is based on ontologies aimed to provide a comprehensive and systematic knowledge, skills and knowledge about the target group, providing training depending on the student profile and objectives. It proposes implementing the concept of personalization and demonstrate that personalization of learning requires new solutions to several issues such as identification profile, knowledge users, learning style, the aim of learning, educational attainment, purpose and context of student formalize knowledge, skills user evaluation level of training and feedback. We must define the domain and objectives of ontology, taking into consideration the activities to be supported, the main beneficiaries of ontology, those who will use and maintain it.

We define a list of classes using a mixed top-down and bottom-up approach. Then we define properties of classes, cardinality and domain and finally we create the instances of classes.

5 Case Study

The implementation of a semantic SAS is done starting with the *Student* concept, for determining the current level of knowledge and training objectives.

In Figure 3 we can see the model of Student repository and the graph representation.

The basic structure is built on four interconnected graphs:
- *Student,*
- *Hour,*
- *Mark*
- *Scholarship*.

The inter-table links are based on the primary-keys from relational database (see Figure 4) which are transformed by D2RQ into RDF object properties.

The model is built incrementally, using data source from the forms offered by the system, structured in two directions: determining the real level of knowledge of the student and the objectives of the student.

The knowledge domain is the management of student's data from the university. The concepts are organized in an ontology representing the domain to be managed.

In describing such a model we start by using **Mind Mapping**. This is a useful technique that helps us to structure and record information which will later be refined in the schema of the RDF knowledge graph. It also supports and enhances creative problem solving, in our case, and generates new ideas.

The mind mapping applications help us to build the relationships graphically, between concepts *Student, Scholarship, Hour* and *Marks.*

In this way our repository is organized into idea clusters that are centered on the main topic, in our case *Student* (see Figure 3). From Student Class are derived three important other classes:
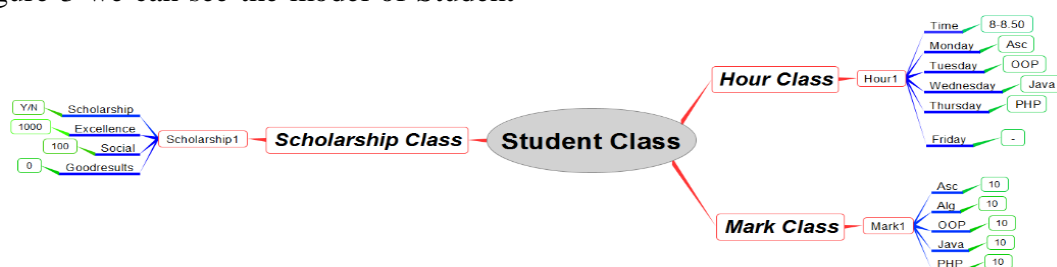- *Hour Class*
- *Scholarship Class*
- *Mark Class.*

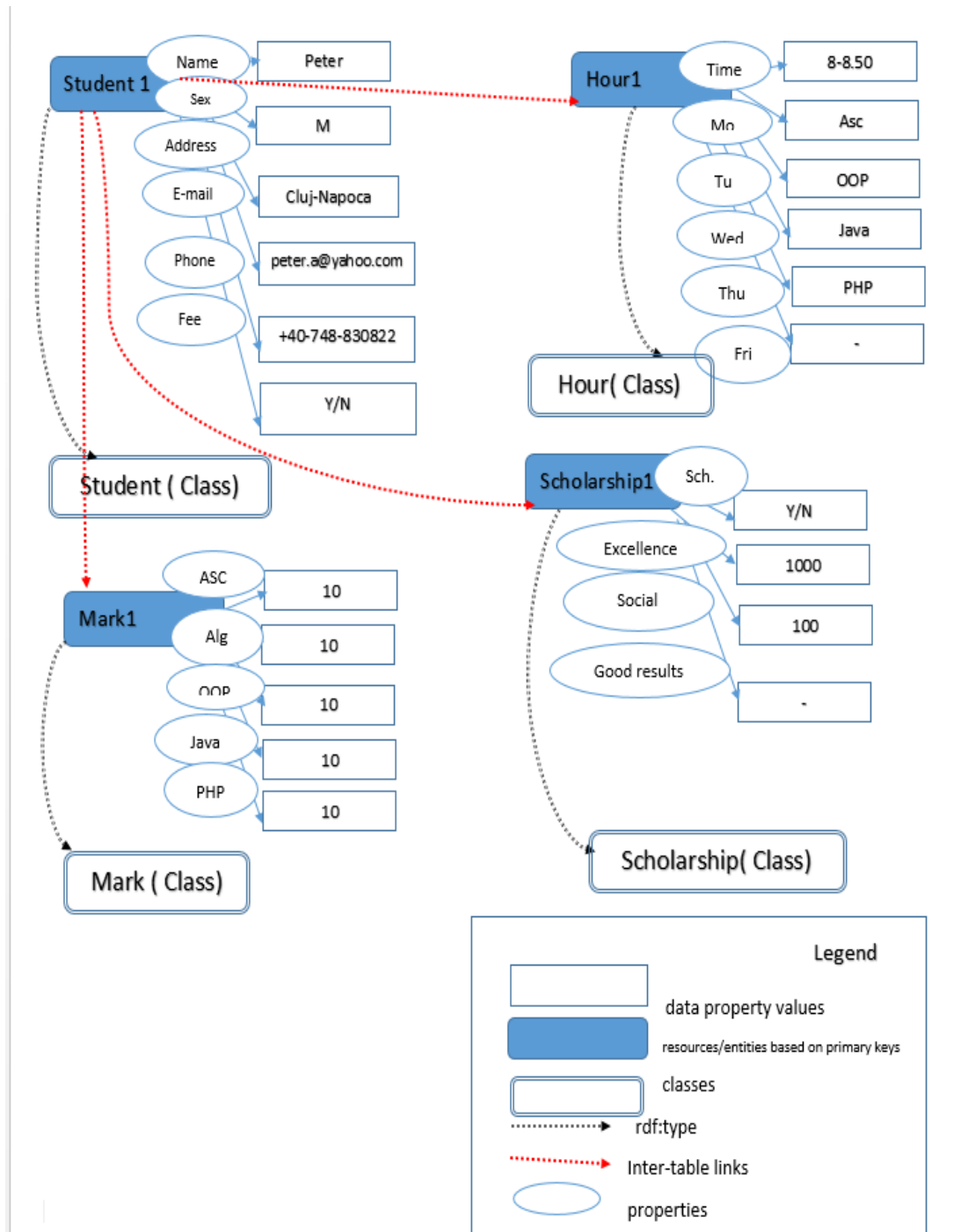**Fig. 3.** Mind-map diagram in a Student Class concept

Fig. 4. The graph model of application repository

The implementation of our ontology is performed with the help of Protégé (http://protege.stanford.edu/), as we see in Figure 5 and 6. We can find here some screen-shot from application.

Protégé is an ontology editor and a specialized platform for developing applications based on knowledge.
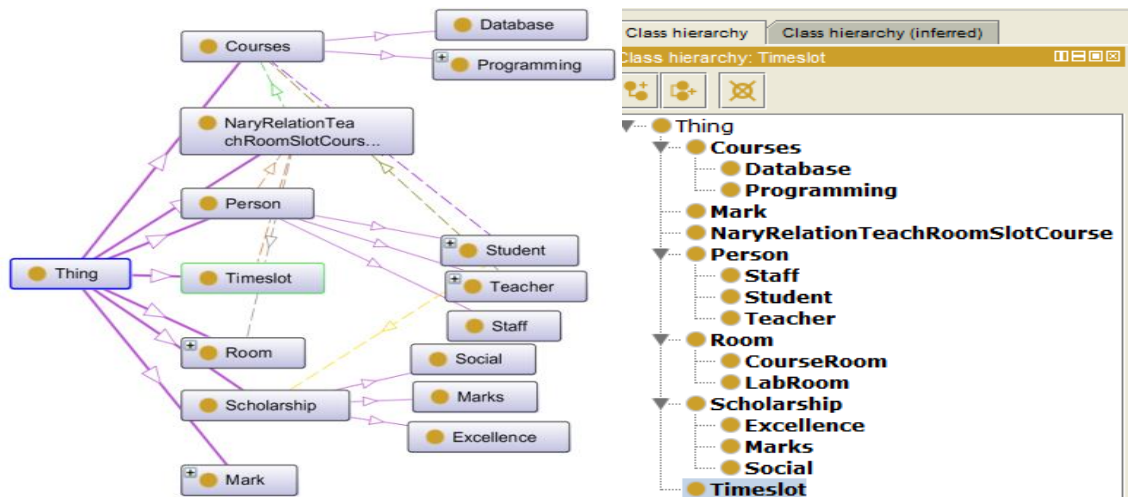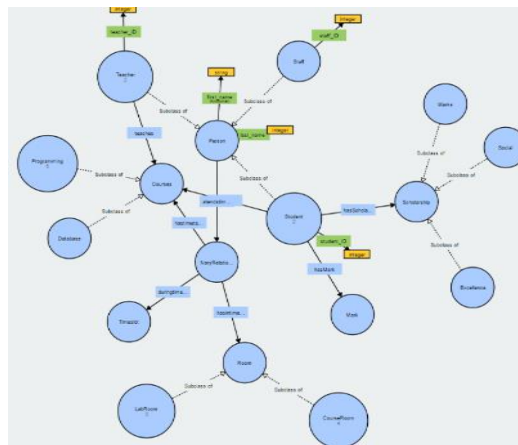
Fig. 5. The modelling of Student SAS Graph.

Fig. 6. The representation of Student SAS Graph represented in Protégé.

The class hierarchy is described within the entities:

Mark, Timeslot, NaryRelationTeachRoom-SlotCourse, Room, Courses, Scholarship and *Person*.

Some of classes have sub-classes and can be developed later.

TopObjectProperty subsumes the relationships used in graphs.
We implemented eight properties:

topObjectProperty → {attendstimetable, duringtimeslot, hasintimetableroom, hastimetablecourse, studies, teaches, hasMark, hasScholarship} (see Figure 7).
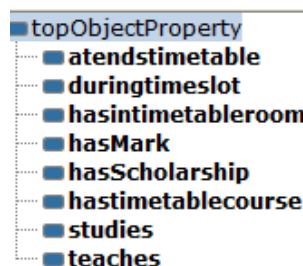
Fig. 7. Relationships between graphs represented by ObjectProperties.

The paper [24] proposes "Linked Open Models" as a possible additional step, whose aim is to enable users to externalize knowledge in the form of diagrammatic models – a type of

content that is human-readable, as well as linkable in the way promoted by the Linked Data paradigm (as authors state). Using such an example we can give an interesting trajectory for our research.

Domain modelling involves decomposing a body of knowledge in a set of elements called concepts that are basic pieces of knowledge or information.

Domain knowledge modelling allows the use of these structures to personalize learning by building, organizing and updating learning resources based on information from the student profile and the educational content provided by learning objects stored in the system.

The proposed ontological structure is a starting point that can potentially bridge data schema derived from legacy systems storing current data assets pertaining to educational activities.

In Figure 7 we give an example of domain knowledge for *Student1* derived from *Student* class.

During the phase of acquisition of knowledge, are gathered all the relevant information required for ontology conceptualization.

Our example contains a domain based on the following set:

$$Student1 \rightarrow \{Java, Alg, ASC, OOP, PHP\}$$

Sources for the knowledge database are the information from managers of universities and electronic documents with specific information requirements of the target group.

Ontology represented by concepts, relationships and properties is extensive with integrity and inference rules.

The rules allow inference implicit knowledge explanation about people participating in training. The knowledge base consists of information: student's profile, marks, skills and abilities of students, proposed posts and general concepts.
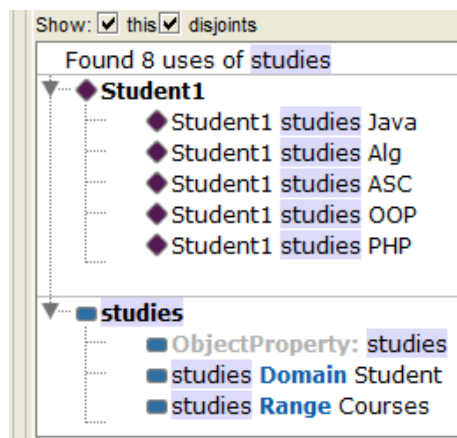

Fig. 8. Definition of concept Student

The representation of domain concepts in a structured and understandable manner can be done using classifiers from editor Protégé (representation of *Student* concept, see Figure 8). The modelling of domain that is specific in education and for students and the use of ontologies developed in the process of building customized system, will be implemented in an intelligent learning platform for Web.

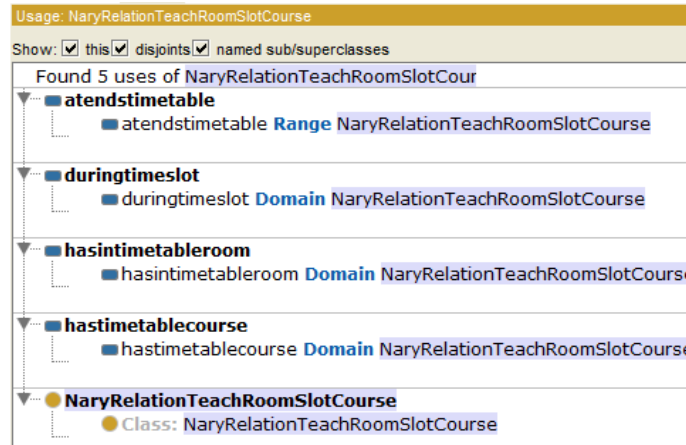A partial representation of conceptual structure of ontology is shown in Figure 9.

Fig. 9. A partial representation of conceptual structure

Data Properties for *Person Class* are the usual ones, as we describe in all databases when we implement a strategy related to persons (see Figure 9).

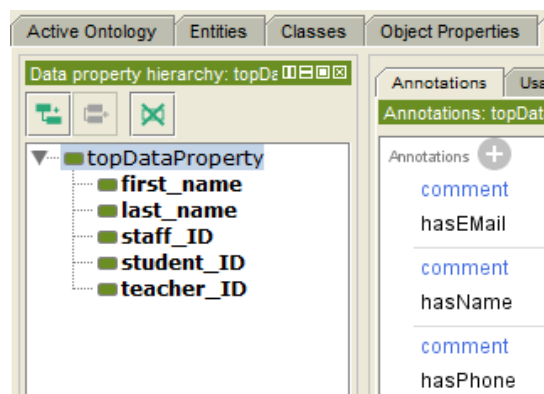$topDataProperty \rightarrow \{first\_name, last\_name, staff\_ID, student\_ID, teacher\_ID\}$

Fig. 10. Data properties for Person Class

During developing the new ontology we use all the facilities of queries from Protégé. So, we can see all the implemented classes with the following SQL statement:

```
SELECT DISTINCT * WHERE {
?Classes fromSAS a owl:Class}.
```

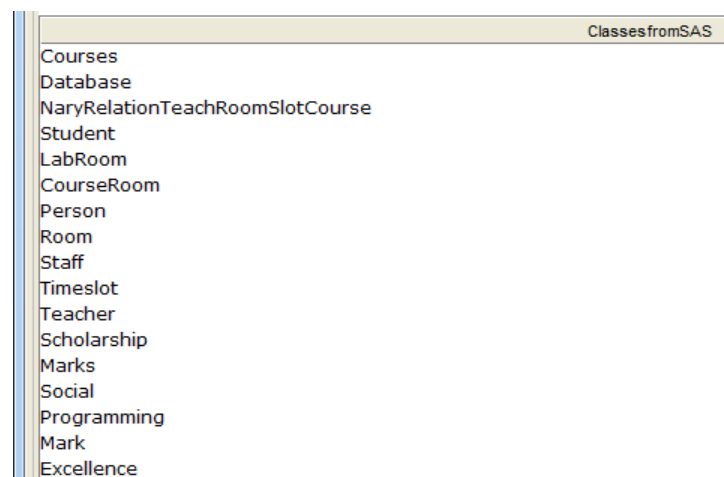The result is seen in figure 11.

Fig. 11. The query which determine all the classes from ontology.

All the predicate from the ontology we can see
with the group of statements:

```
SELECT ?predicate (COUNT(*)AS ?frequency)
WHERE {?subject ?predicate ?obDEject}
GROUP BY ?predicate
ORDER BY DESC(?frequency)
```

We can find all objects with the label writing
the statements below:

```
SELECT DISTINCT ?x(STR(?lab) AS ?label) WHERE {
?x rdf:type owl:Class.
OPTIONAL {?x rdfs:label ?lab}}
ORDER BY ?label
```

Considering that, we focus to the next stages: conceptualization in building the new architecture. The size and complexity of our architecture leads us to an architecture built in three phases.

Basic phase (*Level 1*) is the level where the system under study, applications scenarios and requirements are investigated. Here is built the RDF Graph and is made the conversion from relational database.

The second level (*middle*) defines domain ontology with relations, axioms and concepts. At the third level (*high level*) we find the communication with other platform and the outsourcing of data. A mind diagram which help us to understand and formalize the concepts is presented in figure 12.

These three levels that form the architecture are conducted to create and describe in detail, all the steps in building the SAS.
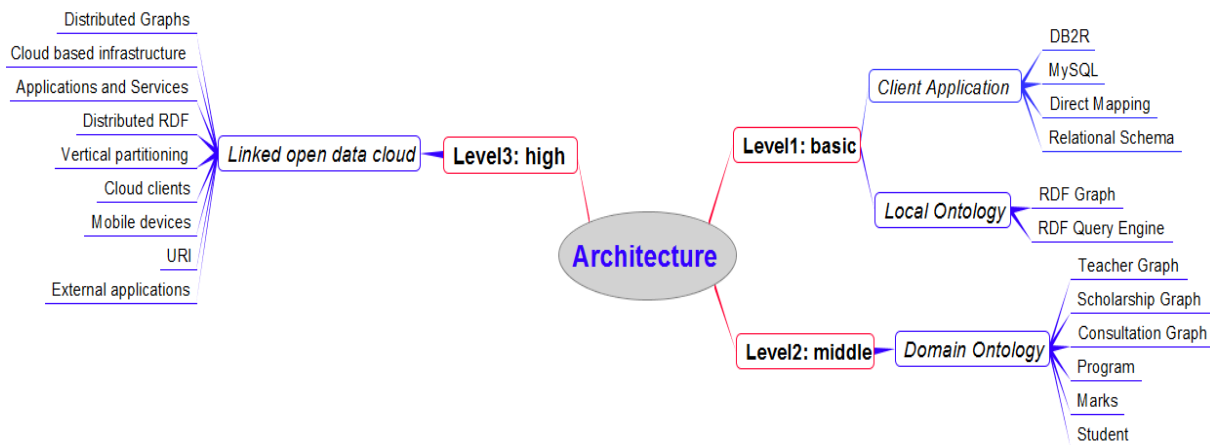

Fig. 12. Mind diagram designed for our architecture

After designing the mind diagram we can
build a conceptual architecture of the system.
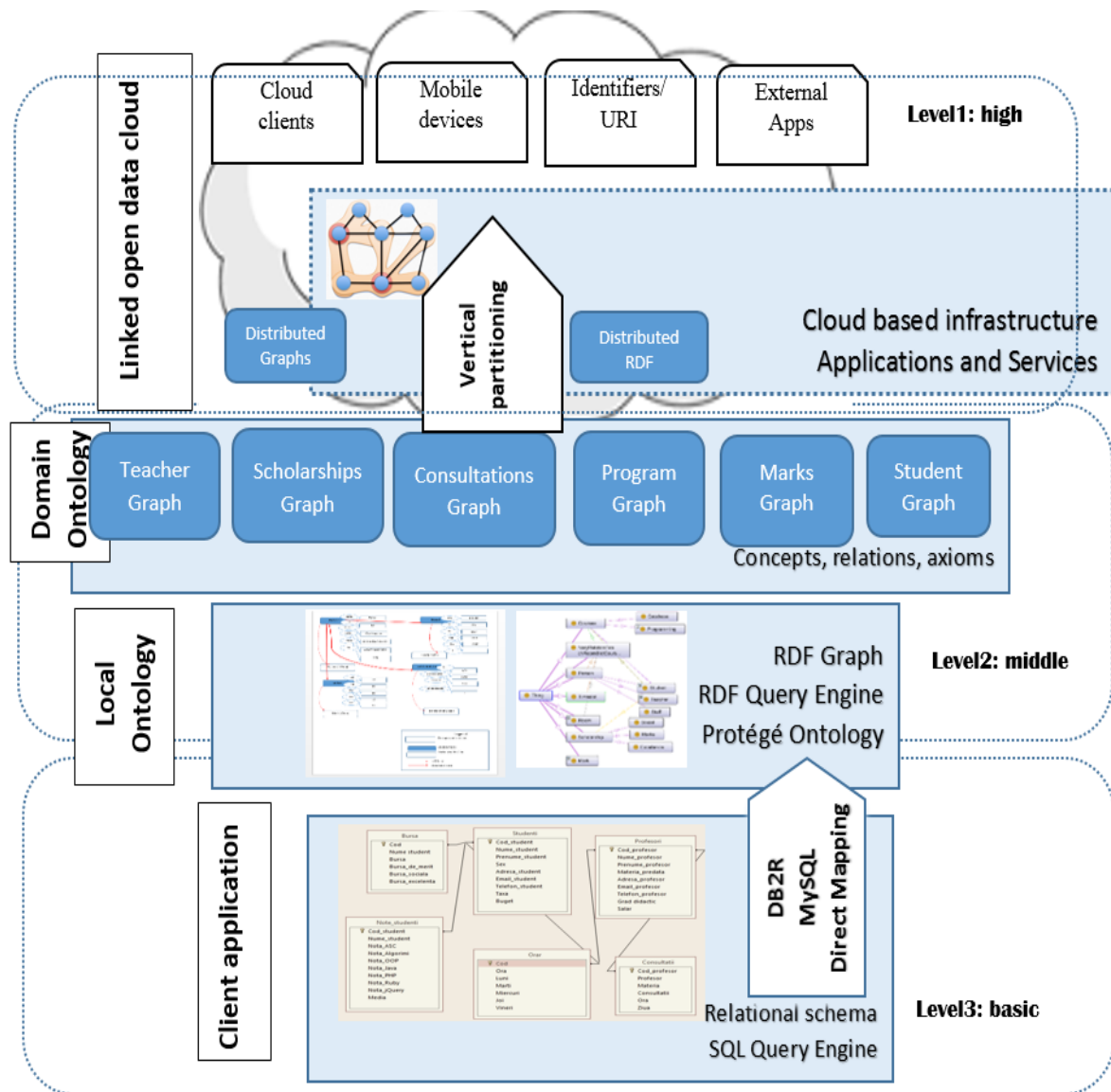The result is shown in Figure 13.

Fig. 13. The general architecture of the SAS system, based on ontologies and Semantic Web

The system allows students to access a set of domain concepts, each concept being associated with an explicit formal description.

The figure represents the reference architecture for our new SAS. The products of system and the definition of activities (system requirements, architecture and design) are inputs to system realization.

The logical view of the SAS architecture supports the logical operation of the system all along its life cycle, and include all the functional, behavioral, and temporal models. We find here collection of functions and dynamic structures that describe how the mission is performed. The physical view of the architecture is a set of system elements performing the functions of the system. Those system elements can be: devices, infrastructure, equipment made of hardware, software applications or cloud.

The decomposition of the SAS may include the decomposition of several layers of systems (intermediate levels of systems) until technological system elements (lowest level) are defined. Any layer of the decomposition may include systems and non-decomposable technological system elements.

The logical and physical representations of the system architecture are mapped onto each other. The interactions between system elements are defined by interfaces whose complexity strongly depends on the way the system architecture and design is defined.

6 Conclusions

The main contribution of our research is to provide a novel approach to Student Management Information Systems design. As we see in Figure 2, the three steps of the process (system environment, entities modelling and knowledge domain modelling) help researchers, reviewers, and editors, to find the template for a Student Information System based on Semantic Web. We hope that many researchers in IS will try to use the process and ontological model, thereby testing their viability. The case study we have provided with this paper demonstrates its use in technical details, but only within the scope of two research problems.

First, further use will tell us whether there are problem domains where it requires extension or where it just doesn't work well, and second, is that of the research entry point. We made a point of suggesting that there might be multiple possible entry points for such a system.

We hope that our proposal will inspire researchers to build student management ontologies based on existing relational databases, via their conversion to multigraphs (or distributed graphs) that can be processed with SPARQL semantic queries.

References

[1] T. Kuhn, "The structure of Scientific Revolutions", The University of Chicago (1962)

[2] I. Lakatos, "The methodology of scientific research programs", http://strangebeautiful.com/other-texts/lakatos-meth-sci-research-phil-papers-1.pdf (1978)

[3] J.E. Van Aken, Management research as a design science: Articulating the research products of mode 2 knowledge production in management, Br J Manage. 16(1): 19–36 (2005)

[4] J. Zhao, C. Wroe et al."Using Semantic Web Technologies for Representing E-science Provenance", Vol. 3298, pp. 92-106 (2004)

[5] A. Hevner, et al.,"Design Science in IS Research", MIS Quarterly Vol.28 No.1 (2004)

[6] T. Mclaren and P. Buijs, "A Design Science Approach for developing", Information systems research instruments, https://www.rug.nl/staff/p.buijs/design_science_approach_for_developing_isr_instruments.pdf (visited 10/2018)

[7] Peffers, K., Tuunanen,T. et al., "The design science research process: A model for producing and presenting information Systems research", http://www.wrsc.org/sites/default/files/documents/000designscresearch-proc_desrist_2006.pdf (visited 10/ 2018)

[8] F. Lima and D. Schwabe, "Exploring Semantic Web Modeling Approaches for Web Application Design", http://users.dsic.upv.es/~west/iwwost02/papers/lima.pdf (visited 10/2018)

[9] K. Peffers, T. Tuunanen, M. Rothenberger and S. Chatterjeea, "Design Science Research Methodology for Information Systems Research", http://wise.vub.ac.be/thesis_ info/Design_Science_Research_Methodology_2008.pdf (accessed 2017)

[10] M. Curado and A.A. Baptista, "A Design Science Research project: A method for the development of Dublin Core Application Profiles", ECRM (2013)

[11] T.R Gruber, "Toward Principles for the Design of Ontologies Used for Knowledge Sharing, International Journal Human-Computer Studies 43", p.907-928 (1993)

[12] L. Băjenaru and I. Smeureanu, "An ontology based approach for modeling e-learning in healthcare human resource management", Economic computation and economic cybernetics studies and research, vol. 49.Is.1 (2015)

[13] T. Saarsen and M. Dumas, "On the Effect of Mixing Text and Diagrams on Business Process Model Use", In Proceedings of the 25th European Conference on Information Systems (ECIS), pp. 552-565, https://aisel.aisnet.org/ecis2017_rp/36 (2017)

[14] F. Kathrin, "User Evaluation of Symbols for Core Business Process Modeling Concepts", in Proceedings of the 25th European Conference on Information Systems

(ECIS), pp. 581-594. https://aisel.aisnet.org/ecis2017_rp/38 (2017)

[15] B. Pittl et al.,"Enabling Risk-Aware Enterprise Modeling Using Semantic Annotations and Visual Rules", in Proceedings of the 25th European Conference on Information Systems (ECIS), https://aisel.aisnet.org/ecis2017_rp/22 (2017)

[16] Design Science Research in IS and technology, http://desrist.org/desrist/ (visited 2017)

[17] A. Hevner, S. Chatterjee, "DR in Information Systems. Theory and practice", vol.22, Springer (2010)

[18] Ostrowski, L. et al., "Ontology engineering step in design science research methodology: a technique to gather and reuse knowledge", Journal Behaviour & Inf. Tech., Vol.33 (2014)

[19] N. Noy and D.L. Mcguiness, "Ontology Development 101: A Guide to Creating Your First Ontology", National Snow and Ice Data Center. https://protege.stanford.edu/publications/ontology_development/ontology101.pdf (2015) (visited 11/2017)

[20] S. Youn, A. Arora, P. Chandrasekhar, P. Jayanty , A. Mestry and S. Sethi, "Survey about Ontology Development Tools for Ontology-based Knowledge Management", http://wwwscf.usc.edu/~csci586/projects/ (2005)

[21] N. Noy ,F. Fergerson, R. Mark and M.A. Musen, "The knowledge model of Protégé-2000: combining interoperability and flexibility", Proceedings of the 12th European Workshop on Knowledge Acquisition, Modelling and Management (2000)

[22] R. Lukyanenko, J. Parsons, J. et al., "Presentation of Toward Artifact Sampling in IS Design Research", https://www.researchgate.net/profile/Roman_Lukyanenko/publication/311734379 _Presentation_of_Toward_Artifact_Sampling_in_IS_Design_Research/links /58582b7e08ae3852d25440ac.pdf (2016)

[23] Y. Fan, C. Wang, G. Zhou et al. "An Architecture for Building a Domain Knowledge Graph from Scratch", Proceedings of DASFAA 2017

[24] D. Karagiannis and R. Buchmann, "Linked Open Models: Extending Linked Open Data with Conceptual Model Information, Information Systems", Vol. 56, pp.174-197, Elsevier, 2016

Loredana MOCEAN has graduated Babes-Bolyai University of Cluj-Napoca, the Faculty of Computer Science, she holds a PhD diploma in Economics and she had gone through didactic position of assistant, lecturer and as professor, since 2000 when she joined the staff of the Babes- Bolyai University of Cluj-Napoca, Faculty of Economics and Business Administration. Also, she graduated Faculty of Economics and Business Administration. She is the author of more than 20 books and over 35 journal articles in the field of Databases, Data mining, Web Services, Web Ontology, ERP Systems and much more. She is member or director in more than 20 grants and research projects, national and international.