A General Semantic Web Approach for Data Analysis on Graduates Statistics

Antonella Carbonaro
University of Bologna,
Bologna, Italy
antonella.carbnaro@unibo.it

Luca Santandrea
Consorzio Interuniversitario AlmaLaurea
Bologna, Italy
luca.santandrea@almalaurea.it

***Abstract*—Currently, several datasets released in a Linked Open Data format are available at a national and international level, but the lack of shared strategies concerning the definition of concepts related to the statistical publishing community makes difficult a comparison among given facts starting from different data sources. In order to guarantee a shared representation framework for what concerns the dissemination of statistical concepts about graduates, we developed SW4AL, an ontology-based system for graduate's surveys domain. The developed system transforms low-level data into an enriched information model and is based on the AlmaLaurea surveys covering more than 90% of Italian graduates. SW4AL: i) semantically describes the different peculiarities of the graduates; ii) promotes the structured definition of the AlmaLaurea data and the following publication in the Linked Open Data context; iii) provides their reuse in the open data scope; iv) enables logical reasoning about knowledge representation. SW4AL establishes a common semantic for addressing the concept of graduate's surveys domain by proposing the creation of a SPARQL endpoint and a Web based interface for the query and the visualization of the structured data.**

I. Introduction

Nowadays, several studies are exploring how the growing Web of Data will change our world. Research organizations, governments and private firms are involved in this transformation. Topics such as ontology building, usage of semantics technologies and open data on the Web, and future applications that will be supported by these technologies are becoming important research areas in their own right [1].

Whilst government organizations have begun to enhance transparency, communicate, and interact with citizens via the Web, the development of appropriate Web of Data strategies has demanded a better understanding of user requirements for tailoring solutions. Many Italian public institutions have provided their Linked Open Data (LOD) and the government has proposed a centralized management (http: //www.datiopen.it/). In particular, the Italian government took several actions through the Digital Italian agency. Among others, the publication of national guidelines for the valorisation of the public sector information, the definition of a license named Italian Open Data License (IODL) and the creation of a centralized catalogue of the open data of the public administration (http://dati.gov.it). Despite the centralization, in some domains there are not standards to facilitate the interoperability and data integration yet. For example, international and national governments lack of shared strategies concerning the definition of concepts related to the statistical publishing community, in particular to the statistics on graduates [2].

The described scenario has led to the development of a formal model for what concerns the dissemination of statistical information about graduates. In particular, the reference is given by the surveys effectuated by the AlmaLaurea consortium. The analysis of these surveys represented the beginning for the creation of Semantic Web for AlmaLaurea (SW4AL), an ontology-based system for graduate's surveys domain which transforms low-level data into an enriched information model encoded using RDF describing the different peculiarities of the graduates. SW4AL uses standard Semantic Web (SW) technologies promoted by W3C. This effort is useful also for the expression of the AlmaLaurea data into meaningful information, providing their possible reuse in the open data scope via the publication of the resulting Resource Description Framework (RDF) datasets. Indeed, ontology-based SW technologies can support interoperability among data and can be used for semantic annotation, information sharing, resource discovery and knowledge transformation [3] [4] [5]. The idea of giving a semantic structure to the AlmaLaurea data information has the purpose to overcome the limits of the application dependency for their retrieving, letting the final user capable to freely obtain information in an interoperable way without the constraint imposed by a software. So doing, the data becomes usable in different contexts, for instance as basis of mashup applications, gaining also the possibility to enhance their value by combining them with external sources, getting the rid of the information silos problem. For instance, the integration with international open datasets like the ones exposed by the European Union open data portal, can help to compare the Italian graduates' performances with those from others countries. Developed SW4AL ontology actually consists of 74 classes, 121 object properties, 107 data properties, and 1628 axioms. Dataset contains 3.161.153 triples.

The aim of this paper is to describe SW4AL. SW4AL: i) semantically describes the different peculiarities of the graduates; ii) promotes a standard definition of the AlmaLaurea data and the following publication in the Linked Open Data context; iii) provides their reuse in the open data scope; iv) enables logical reasoning about knowledge representation. SW4AL is able to: i) establish a common semantic for addressing the concept of graduate's surveys domain; ii) create an infrastructure for networked concepts; iii) support more general knowledge management and sharing; iv) provide semantic reasoning. Moreover, we developed a SPARQL endpoint and a Web based interface for the query and the visualization of the structured data.

ISSN 2305-7254

The novelty of the proposed approach lies in exploiting SW technologies to explicitly describe the meaning of graduate's surveys domain and to facilitate interoperability and data integration in order to construct a unified interlinked data model and enable semantic reasoning capabilities over it. In compliance with the open paradigm, the structured data proposed are freely available and unconstrained in proprietary applications. The additional wizard interface for the SPARQL query creation and the consequent possibility of query and result download assure a full availability.

The remainder of this paper is organised as follows: Section 2 presents a survey of both the work done by the AlmaLaurea interuniversity consortium and of input data. Section 3 describes the proposed SW4AL architecture, the knowledge representation decisions and the incremental work followed to obtain the final ontologies for the domain representation. Section 4 describes the creation of a SPARQL endpoint and a Web based interface for the querying and the visualization of the SW4AL data. Section 5 reports some evaluations about the effected. Finally, some considerations and conclusions close the paper.

II. THE ALMALAUREA SURVEYS

A significant contribute to the graduates statistics in the national sphere is given by the work done by the AlmaLaurea interuniversity consortium. Founded in 1994 after a first project started by the Statistical Observatory of the University of Bologna, the consortium, supported by the Italian Ministry of Education, University and Research (MIUR), has as main mission in the production of statistical surveys about the situation of the Italian graduates.

The surveys done have a wide representativeness due to the high number of member universities (75 at the first months of 2018), which guarantees a coverage of more than 90% of Italian graduates. This diffusion made the AlmaLaurea surveys a reference point for the academic community and for the economical and political world. The analysed aspects cover two distinct surveys, published annually:

- Survey on the profile of graduates: delineates characteristics and performances of the graduates by providing a picture of the situation considering criteria such as study condition, satisfaction on study careers and university success (in terms of final mark and regularity of studies). Data derive from questionnaires distributed to students at the end of their course of study and are integrated with administrative documentation coming from the Universities.
- Survey on employment condition of graduates: monitors the insertion of the graduates in the business world by collecting data deriving from interviews conducted at one, three and five years from the achievement of the degree. Through it, it is possible to obtain information about the typology of work done, the average satisfaction, the average retribution and the inherence with the studies.

The data derived from the interviews, effectuated both in telephonic and in Web modality (CATI and CAWI) and characterized by more than hundred variables, are publicly available on the AlmaLaurea website (https://www.almalaurea.it/) and can be consulted in a single modality at different granularity levels. Moreover, it is possible to perform comparison among different collectives, basing on different variable like gender, degree class or degree course. AlmaLaurea also highlights observations on specific themes and employment patterns resulting from the interpretation of the data. The AlmaLaurea surveys are presented every year during a dedicated convention. The high number of effectuated questionnaires (more than 200.000 every year) allows obtaining a significant dataset [6].

Basing on the data of the AlmaLaurea statistics, it is possible to generate reports with information on transparency requirements for each course of study for which data of at least one graduate exist within the database. These data concur to the creation of the single annual reports (SUA) for each course of study of each University of the consortium. The SUA is a management tool useful for planning, realization, self-evaluation and redesign of the course of study. The SUA reports are published by the National Agency of the Evaluation of The University system and of the Research (ANVUR, http://www.anvur.org/) and accessible on the platform UniversItaly (https://www.universitaly.it/). The generation of the reports occurs by selecting a reduced set of indicators starting from the profile and employment condition surveys. Moreover, comparisons and aggregated visualizations with equivalent pre-reform courses are possible for the extended surveys.

A. Linked Open Data for statistics on graduates

Within the SW context, ontologies play a central role in resource representation, since they explicitly define concepts and relationships related to a particular domain, in a structured and formal way (i.e. ontologies are machine-processable) [7], [8], [9], [10]. The state of the art of open data availability in the education domain is currently quite fragmented [11]. Taking as reference the portal of the LOD of the Italian public administration, a research on the topic "degree" returns only a set of datasets of a few specific territorial realities, and therefore does not capture the majority of aspects at a national level. In a similar way, the same non- comprehensiveness problem happens in the European Data Portal (https://www.europeandataportal.eu/), which collects data from the single national sources and merges data in a bottom-up modality guaranteeing a standardization thanks to the respect of the principles of the open data paradigm. Despite the difficulties of attainment of a complete picture form the holistic point of view, the LOD phenomenon concerning the educational theme is growing and contributes to the creation of a global knowledge which is very important for the future generations [12]. Particularly interesting is the LOIUS project (Linking Italian University Statistics), which proposes the definition of an ontology for the representation of university statistics published by MIUR, by effectuating their exposition in RDFa format with the goal of providing their Web-based representation [13].

The integration of the AlmaLaurea statistics in the LOD scope can give an important contribution to the available information on the status of graduates in Italy. The availability of exhaustive information about the graduate's employment condition supports others deriving from Italian National Institute of Statistics (ISTAT) sample surveys, and enriches

them with more specific data thanks to the numerous variables present in the questionnaires. In a similar way, the survey on the profile of the graduates further improves the picture, giving exhaustive and reliable information about the quality of the study experience of the graduates, though data collected at the end which also provide a vision of subjective aspects like the personal satisfaction, and that can complete those deriving from the EUROSTAT surveys.

III. SW4AL PROJECT

In this Section, we describe the SW4AL architecture and the effectuated knowledge representation choices taken according to the analysis of the input data.

A. Architecture

The SW4AL main architecture adopted for the addition of semantic metadata to the SUA information follows a workflow, as reported in Fig. 1. Starting from the reified cube of each fact table connected to a specific survey, the data are extracted and transformed in RDF format following a RDF/XML syntax, according to the rules of a defined ontology. After that, the generated triple store file is uploaded on a triple store server, which has the task of interpret the data and to provide an endpoint for the query, performed via SPARQL (Query Language for RDF). The structured data can also be used to release a graphical user interface able to better explain certain queried facts.

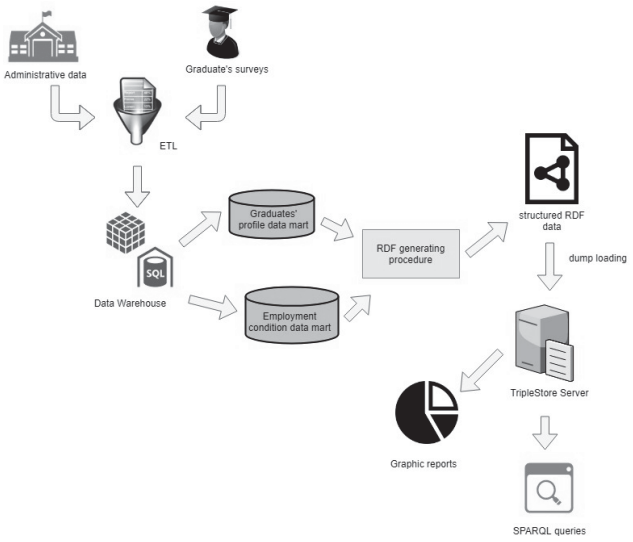


Fig. 1. SW4AL architecture representing data enriching process. Data are semantically annotated according to the developed ontology producing information represented using RDF graphs. Querying and reasoning over the ontologies and information generates knowledge

B. Knowledge representation and data modelling

Domain ontologies express in a formal and unambiguous manner all the aspects that have to be stated into the RDF [14], [15], [16], [17], [18]. As not all the traits reported in the SUA sheet have a covering reference ontology, an ad hoc ontology was developed, including the import of external validated third-part ontologies. The AlmaLaurea ontology is represented using Web Ontology Language (OWL) [19].

IV. SW4AL DATA VISUALIZATION

The definition of the AlmaLaurea ontology and the following generation of the triple store have been accompanied by the development of a reporting tool which permits the visualization of the data. In addition, a modular interface for the creation of queries and their subsequent launch has been created.

A. Graphical reports

Inspired by the existing SUA reports available at the AlmaLaurea's university staff website, the main goal is to create a graphical report, providing the same level of knowledge expressivity, starting from the data available in the newly defined open format. Furthermore, to accomplish the openness paradigm, the creation of these tools has been done with the help of open source and freely available software instruments. The full process has also the aim to point out an important advantage of the exploitation of the structured open data: not only a powerful tool for automatic semantic reasoning, but also a way to share data among humans in a standard defined style, so that to promote the distribution and the reuse of the information. The visualization of the open data demonstrates an immediate possible way of their reutilization.

1) Technological stack

The graphical reports development has been done with the purpose of the creation of a Web-based software platform, dynamically populated. The main technological solutions used for the realization of the SW4AL Web interface are HTML5, CSS and JavaScript. For the visualization of the data, the JavaScript charting library proposed by Google (Google charts) have been used, while its instances have been fed thanks to the data AJAX retrieving, which have been done with the support of the jQuery library.

The data retrieving has been made exploiting a feature of the Fuseki server: thanks to its support of SPARQL over HTTP (SOH) commands, the queries have been resolved with a set of HTTP requests. In particular, Fuseki exposes a SPARQL endpoint, queryable in a RESTFul way. The possibility of the attainment of the results in Javascript Object Notation (JSON) format ensured the possibility of use from the defined Web application.

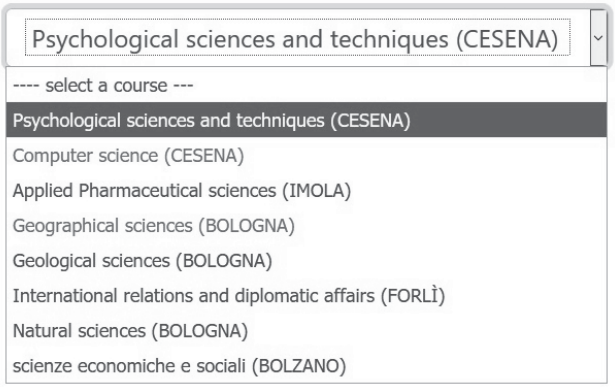
The developed reporting tool followed several considerations regarding the knowledge pattern to show, starting from the retrieved RDF data. The main idea in this case is to explain the meaning of the data in a more human-friendly mode. The report about the graduates' profile is divided into ten different questions, each of them having different possible values for the reply. Thus, for each question, a different pie chart has been created. The total on which the profile survey is calculated is the number of graduates interviewed in the last three years. This values equals the sum of all possible values (including the "not responding" option) of each question. Using the pie chart is then possible to see the full distribution of all possible values over a given question.

2) Dynamic retrieval of course list

The starting point of the visualization report is the degree course. In order to create a more dynamical Web application, it has been performed an opposite query aiming at retrieving all

the possible courses for which a correct survey value exists. This query constructs the graph until a degree course granularity level, returning distinct degree codes values. Particular conditions to be satisfied concerns the presence or absence of the variables regarding the questions wanted to be shown into the report. In fact, if for a given course the cardinality of the graduates is under threshold, the information about the instance of the survey questions for that course is not present. This is due to the privacy duties (for cohort with very low cardinality it is necessary to envisage a different visualization).

The course retrieving SPARQL query is then based on the previously mentioned course graph together with the presence



of at least one value for a predicate regarding a question of the questionnaire. The results of both the previous queries creates a full list of courses, either available or not for visualization. This has been used for the population of a dropdown for the selection of the course data to visualize. Fig. 2 reports the result of this process in the Web application, containing available courses and disabled ones.

Fig. 2. Example of implementation of the course dropdown within the application

3) Update after ontology clarification

The visualization tool has been developed in according to AlmaLaurea ontologies. The system allows the visualization of all the three different employment surveys results, reporting the values of each single question in paired columns in the chart. Fig. 3 shows the visual outcome for a question.

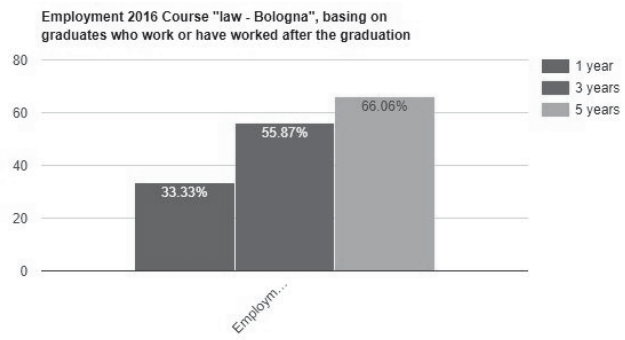


Fig. 3. Example of multi-year employment survey question visualization

4) Cohort comparison

Moreover, the system offers the possibility to compare the data of a given course with the information coming from the related university or degree class collective. In a similar way to the existing SUA reports, each chart of a survey variable for a given course is placed side by side to the correspondent one of the chosen collective, allowing the comparison of the results.

5) System interfaces

Pursuing the simplification of the instrument and its usage improvement, the created system consists in two reports, one for each survey type. These reports allow a global data visualization experience through rapid switches on the collectives to discover. The proposed report contains a parametric form that consents the choice of the degree course on which to perform the data visualization. Notably, it has been defined a set of conditional dropdowns which filter the full list of courses. The dropdowns regard the following criteria: i) year of execution of the survey, ii) university and iii) type of degree (level).

Similarly to the construction of the dropdown of the courses, also these ones have been populated starting from the RDF triplestore, by executing a query via the HTTP restful endpoint. The form is then completed with the presence of other input controls which allow the change of the cohort for the comparison and the inclusion of aggregated values. Fig. 4 reports the aspect of the visualization form.

To maintain a better truthfulness of the information shown, several warnings about possible data inconsistencies have been implemented. The particular cases managed regard:

- Inter-class courses: when showing a comparison over the degree class for these types of degrees a warning is shown, because the degree class data refers to the aggregation of all the possible single degree class values.
- Previous / next version of a course: implemented as fallback in the aggregated data mode of the employment report, this information is used to retrieve data of previous version of a course when there is an empty result of the current course.

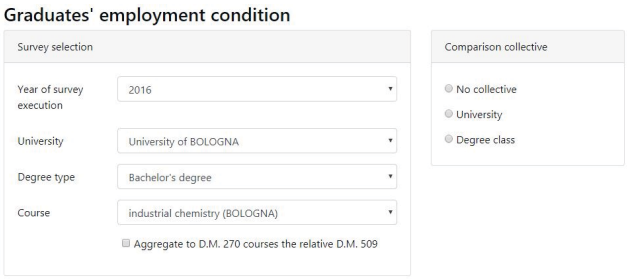


Fig. 4. Parametric form for the visualization of data about employment condition's survey

B. A wizard for query building

In order to guarantee the full utilization of the open knowledge base generated, we developed a wizard interface for the incremental creation of SPARQL queries, to be launched on the available Fuseki endpoint. The idea of the tool has been inspired by other well known examples, like the ISTAT open data query construction platform or the European Data Portal

Linked data query wizard. The reasons of the work consist in letting the non-technical users able to perform specific data retrieving requests, exploiting the structure defined by the created ontologies.

C. Interface design

The system aims at the growing granular construction of SPARQL queries, starting from the university level until the degree course level, focusing on the institution didactic hierarchy. The proposed interface, built with the same technologies used for the data visualization platform, is combined by three sections, each containing different aspects for the query construction:

- Geographical map: used to refine the dataset basing on the establishment region of the university.
- Survey / degree choice: same set of filters present in the data visualization consists on performing year of survey, degree choice and aggregated visualization option.
- Variables to be extracted: a series of checkboxes is listed, corresponding to the existing variables of the survey. The choice of a variable includes its possible values into the SPARQL query.

In the employment condition survey, also a fourth box is showed, regarding the choice of the kind of employment survey by year, for filtering the results in one of the three available employment surveys. Basing on the configuration of the form, several scenarios of non consistency can happen. Therefore, the implemented software considers also the validity of the chosen combination of parameters, pledging the creation of valid queries. As an example, within the employment survey, the form disables the selection of the question "graduates currently enrolled to a master degree course" if the selected degree type is not a bachelor level. This because that question is present only for the first-level degree surveys, and the inclusion of the graph pattern in other degree types would lead to an empty result. The described filter boxes are followed by a parameters recap box and by a text area where the generated query is present, ready to be launched. On Fig. 5 is reported the resulting form.

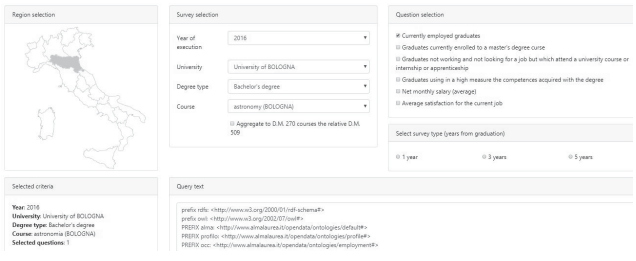


Fig. 5. Wizard form for the employment condition queries

D. Query construction logic

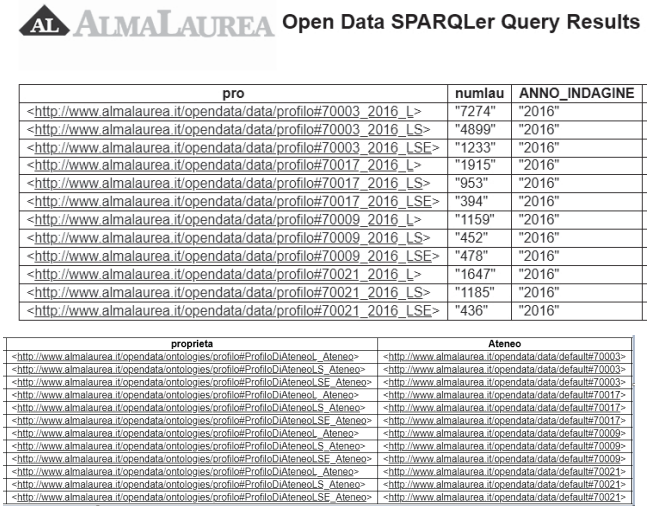
The designed form allows the interrogation over the dataset at different granularity levels. According to the ontology structure, the different levels correspond to different objects instances of subclasses of the main classes representing the survey values. Thus, to maintain the correctness of the result retrieval, different possible paths have been individuated, produced by the possible combinations of filters in the form. The possibility to act on the hierarchy of concepts is given by the definitions present into the ontologies. In particular, the queries have been constructed including the checking of properties searched using the rdf:subPropertyOf property. For example, some filtering selections for the query building could regard choice of survey year, region, university, type of degree.

E. Results

After the parameterization of the form, the final step is the launch on the query on the Fuseki server. This has been done exploiting the same HTTP endpoint used for the AJAX calls performed in the data visualization. In this case, the chosen format is XML. Thank to the application of layout rules defined in extensible stylesheet language (XSL), the result appear in a styled tabular design. The result of a query launch is visible in Fig. 6.

V. Conclusions

Our research demonstrates the advantages of explicitly encoding specific domain concepts by formally representing the semantics of the collected data, the domain and their relationships. In particular, in this paper we proposed SW4AL, an ontology-based system for graduate's surveys domain which transforms low-level data into an enriched information model encoded using RDF describing the different peculiarities of the graduates. The proposed system is also able to map the information into a specialized domain model by providing support for logical reasoning. SW4AL proposes the creation of a SPARQL endpoint and a Web based interface for the querying and the visualization of the structured data. Regarding the data visualization tools, most of the responsibilities for the performances are prerogative of Apache Jena Fuseki server and Javascript optimized code, while the first difficulty scales with the growth of the number of the triples over a single scheme, for the second case an important improvement have been noticed thanks to the usage of functional constructs of JavaScript. A satisfying benchmark is anyway guaranteed with the generation of all the survey data for a 3-year period.


ALMALAUREA Open Data SPARQLer Query Results

pro	numlau	ANNO_INDAGINE
<http://www.almalaurea.it/opendata/data/profilo#70003_2016_L>	"7274"	"2016"
<http://www.almalaurea.it/opendata/data/profilo#70003_2016_LS>	"4899"	"2016"
<http://www.almalaurea.it/opendata/data/profilo#70003_2016_LSE>	"1233"	"2016"
<http://www.almalaurea.it/opendata/data/profilo#70017_2016_L>	"1915"	"2016"
<http://www.almalaurea.it/opendata/data/profilo#70017_2016_LS>	"953"	"2016"
<http://www.almalaurea.it/opendata/data/profilo#70017_2016_LSE>	"394"	"2016"
<http://www.almalaurea.it/opendata/data/profilo#70009_2016_L>	"1159"	"2016"
<http://www.almalaurea.it/opendata/data/profilo#70009_2016_LS>	"452"	"2016"
<http://www.almalaurea.it/opendata/data/profilo#70009_2016_LSE>	"478"	"2016"
<http://www.almalaurea.it/opendata/data/profilo#70021_2016_L>	"1647"	"2016"
<http://www.almalaurea.it/opendata/data/profilo#70021_2016_LS>	"1185"	"2016"
<http://www.almalaurea.it/opendata/data/profilo#70021_2016_LSE>	"436"	"2016"

proprieta	Ateneo
<http://www.almalaurea.it/opendata/ontologies/profilo#ProfiloDiAteneoL_Ateneo>	<http://www.almalaurea.it/opendata/data/default#70003>
<http://www.almalaurea.it/opendata/ontologies/profilo#ProfiloDiAteneoLS_Ateneo>	<http://www.almalaurea.it/opendata/data/default#70003>
<http://www.almalaurea.it/opendata/ontologies/profilo#ProfiloDiAteneoLSE_Ateneo>	<http://www.almalaurea.it/opendata/data/default#70003>
<http://www.almalaurea.it/opendata/ontologies/profilo#ProfiloDiAteneoL_Ateneo>	<http://www.almalaurea.it/opendata/data/default#70017>
<http://www.almalaurea.it/opendata/ontologies/profilo#ProfiloDiAteneoLS_Ateneo>	<http://www.almalaurea.it/opendata/data/default#70017>
<http://www.almalaurea.it/opendata/ontologies/profilo#ProfiloDiAteneoLSE_Ateneo>	<http://www.almalaurea.it/opendata/data/default#70017>
<http://www.almalaurea.it/opendata/ontologies/profilo#ProfiloDiAteneoL_Ateneo>	<http://www.almalaurea.it/opendata/data/default#70009>
<http://www.almalaurea.it/opendata/ontologies/profilo#ProfiloDiAteneoLS_Ateneo>	<http://www.almalaurea.it/opendata/data/default#70009>
<http://www.almalaurea.it/opendata/ontologies/profilo#ProfiloDiAteneoLSE_Ateneo>	<http://www.almalaurea.it/opendata/data/default#70009>
<http://www.almalaurea.it/opendata/ontologies/profilo#ProfiloDiAteneoL_Ateneo>	<http://www.almalaurea.it/opendata/data/default#70021>
<http://www.almalaurea.it/opendata/ontologies/profilo#ProfiloDiAteneoLS_Ateneo>	<http://www.almalaurea.it/opendata/data/default#70021>
<http://www.almalaurea.it/opendata/ontologies/profilo#ProfiloDiAteneoLSE_Ateneo>	<http://www.almalaurea.it/opendata/data/default#70021>

Fig. 6. Results of the query launch

The novelty of our work lies on the application of SW technologies to provide knowledge representation of semi

structured data sources and external domain datasets in order to construct a unified interlinked data model and enable semantic reasoning capabilities over it. The presented approach can significantly facilitate sharing, exploitation and creative reuse of existing data sources thus fully exploiting their intrinsic potential. Our innovative semantic system can be used wherever there is the need to create knowledge from information and information from data using semantic representation, reasoning technologies and incorporating domain knowledge into the computation.

REFERENCES

[1] Heath Tom and Christian Bizer (2011) Linked Data: Evolving the Web into a Global Data Space. Synthesis Lectures on the Semantic Web: Theory and Technology, 1:1, (2011) 1-136. Morgan & Claypool

[2] Keegan McBride, Ricardo Matheus, Maarja Toots, Tarmo Kalvet, and Robert Krimmer (2018) The Role of Linked Open Statistical Data in Public Service Co-Creation. In Proceedings of the 11th International Conference on Theory and Practice of Electronic Governance (ICEGOV '18), Atreyi Kankanhalli, Adegboyega Ojo, and Delfina Soares (Eds.). ACM, New York, NY, USA, 679-681.

[3] Bischof Stefan, Andreas Harth, Benedikt Kämpgen, Axel Polleres, Patrik Schneider (2018) Enriching integrated statistical open city data by combining equational knowledge and missing value imputation, *Journal of Web Semantics*, Volume 48, 22-47.

[4] Carbonaro A. (2009) Collaborative and semantic information retrieval for technology-enhanced learning. In Proceedings of the 3rd International Workshop on Social Information Retrieval for Technology-Enhanced Learning, Aachen, Germany.

[5] Carbonaro A. (2010) WordNet-based summarization to enhance learning interaction tutoring. Journal of e-Learning and Knowledge Society 6, 2, 67–74.

[6] AlmaLaurea, Indagini e ricerche, 2018, http://www.almalaurea.it/universita/statistiche. Accessed 2018-03-12

[7] Reda R., F. Piccinini, and A. Carbonaro (2018) Towards consistent data representation in the IoT healthcare landscape. In ACM DH'18: 2018 International Digital Health Conference, April 23–26, Lyon, France.

[8] Carbonaro A. and R. Ferrini (2007) Ontology-based video annotation in multimedia entertainment. In Consumer Communications and Networking Conference. CCNC 2007. 4th IEEE. Citeseer, 1087–1091.

[9] V.C. Storey, I.Y. Song (2017) Big data technologies and management:What conceptual modeling can do, Data & Knowledge Engineering.

[10] Riccucci S., A. Carbonaro, and G. Casadei (2007) Knowledge acquisition in intelligent tutoring system: A data mining approach. In Mexican International Conference on Artificial Intelligence. Springer, 1195–1205.

[11] Aslam, Muhammad Ahtisham and Naif Radi Aljohani. (2018) "SPedia: A Central Hub for the Linked Open Data of Scientific Publications." IJSWIS 13.1: 128-147. Web.

[12] European Data Portal, Education: Open Data in Schools (2018) https://www.europeandataportal.eu/highlights/open-data-schools. Accessed 2018-03-12.

[13] Pirrotta G. (2010) Linking Italian university statistics. In Proceedings of the 6th International Conference on Semantic Systems (I-SEMANTICS '10), Adrian Paschke, Nicola Henze, and Tassilo Pellegrini (Eds.). ACM, New York, NY, USA.

[14] Carbonaro A. and R. Ferrini (2007) Personalized Information Retrieval in a Semantic-based Learning Environment. In Dion Hoe-Lian Goh and Schubert Foo, editors, Social Information Retrieval Systems, pages 270–288. Idea Group Reference.

[15] Carbonaro A. (2010) Improving Web Search and Navigation Using Summarization Process. In Knowledge Management, Information Systems, E-Learning, and Sustainability Research, pages 131–138. Springer.

[16] D. Dou, H. Wang and H. Liu (2015) Semantic data mining: A survey of ontology-based approaches, Proceedings of the 2015 IEEE 9th International Conference on Semantic Computing (IEEE ICSC 2015), Anaheim, CA, pp. 244-251.

[17] D. Dermeval, J. Vilela, I. I. Bittencourt, J. Castro, S. Isotani, P. Brito and A. Silva (2016) Applications of ontologies in requirements engineering: a systematic review of the literature, Requirements Engineering, vol. 21, no. 4, pp. 405–437.

[18] Carbonaro A. (2010) Towards an automatic forum summarization to support tutoring. In Lytras, M., Ordonez De Pablos, P., Avison, D., Sipior, J., Jin, Q., Leal, W., Uden, L., Thomas, M., Cervai, S., and Horner, D., editors, Technology Enhanced Learning. Quality of Teaching and Educational Reform, volume 73 of Communications in Computer and Information Science, pages 141–147. Springer Berlin Heidelberg.

[19] Horrocks Ian, Peter F Patel-Schneider, and Frank Van Harmelen. (2003). From SHIQ and RDF to OWL: The making of a Web ontology language. Web semantics: science, services and agents on the World Wide Web 1, 1 7–26.