Semantic Web Applications: Current Trends in Datasets, Tools and Technologies' Development for Linked Open Data

Sabina-Cristiana NECULA
Alexandru Ioan Cuza University of Iasi, Iasi, Romania
sabina.necula@uaic.ro

We report a survey on the actual state of the art about semantic web and its applications. Semantic Web plays a major role in integrating data, especially open data, publicly available on the Internet. We reviewed scientifically research papers, study cases, web sites and specialty books in order to discuss the main applicative areas, especially in the field of governmental use, the main technologies and the architectures involved. Both quantitative and qualitative analyses were carried out on the data, which related to 1460 ontologies belonging to Linked Open Data Cloud. The second analysis was on the content of scientific articles belonging to Clarivate Analytics, Scopus and Google Scholar databases. We identified and analyzed 84941 articles written on the subject of ontology, from which computer science is represented by 36264 articles. The results of our research proved that semantic web technologies are an important tool for describing and integrating data and an important component in the data layer of any intelligent application. This study contributed to the mainstream of the research literature by presenting the applicative areas of semantic web and semantic web applications' development tools, architectures, and methodology.
Keywords: semantic web, Linked Open Data, Resource Description Framework, SPARQL Protocol and RDF Query Language

1 Introduction

Our live is surrounded by modern technologies such as cloud, Artificial Intelligence (AI) and Internet of Things (IoT). We talk about the existence of different technologies, but, very often, these technologies process the same data or, at least, same information coming from heterogeneous sources of data. Technical and semantic interoperability are subjects of study for both researchers and practitioners. Because freely offering integrated data is subject of public interest, Europe established a strategy for data that aims creating a single market for data that will ensure Europe's global competitiveness and data sovereignty. [1]. The EU Open Data Portal is the point of access to public data published by the EU institutions, agencies, and other bodies. It uses Semantic Web and offers a SPARQL Protocol and RDF Query Language endpoint to all its datasets. Linked Open Data (Linked+Open data) has been proposed as the basis for open government and for solving many of the data integration issues.

Although integrating and offering data is a clear application of Semantic Web, it does not mean that this is without obstacles or challenges. A key obstacle for the Semantic Web is the lack of availability of usable tools [2]. It is not very clear what Semantic web role is alongside cloud, Artificial Intelligence, and Internet of Things technologies, but definitely applying AI for discovering relationships between data or integrated processed data coming from IoT devices are the next challenges.

This research studies the opportunities and challenges of semantic web technology by presenting the main applications of the semantic web, the current state and directions of research, the technologies and architectures involved. After identifying the main application areas based on Linked Open Data (LOD) cloud, we discussed the main topics debated in the scientific literature and by practitioners. One of the used technologies for sharing common representations are ontologies. An ontology is a formal explicit description of concepts in a domain of discourse (classes),

properties of each concept describing various features and attributes of the concept, and restrictions on slots. [3]

The term of Semantic Web was coined in 2001 by Tim Berners Lee. In 2006, the research belonging to ontology engineering field emerged into Linked Data, as it was named. Diverse groups of researchers developed tools (i.e. Protégé, Neon), created publicly available ontologies (all of them are available online at Linked Open Data cloud website) and provided SPARQL Protocol and RDF Query Language (SPARQL) endpoints in order to offer access to data described with ontologies. In this moment, there are 1460 datasets on Linked Open Data (as of November 2020) website belonging to diverse domains grouped into:

- *Government;*
- *Geography;*
- *Life Sciences;*
- *Linguistics;*
- *Media;*
- *Publications;*
- *Social Networking;*
- *User Generated.*

Google Trends revealed a constant interest on the subject of "semantic web application" in the last 5 years. Figure 1 presents the results shown by Google trends about searches on "semantic web" string.
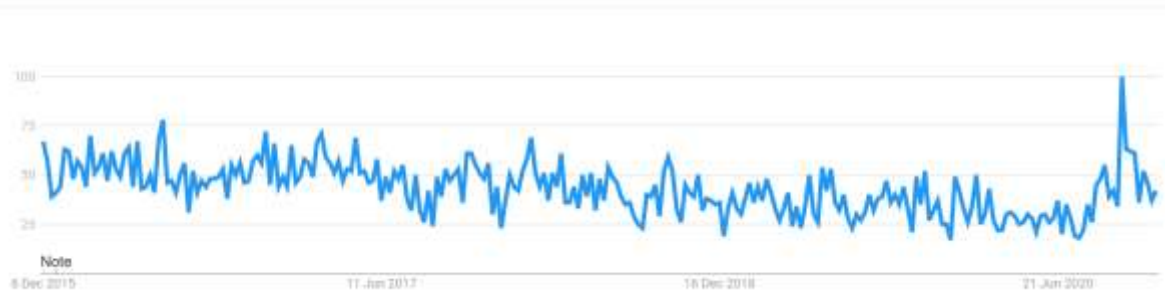


Fig. 1 Google Trends results on "semantic web"

The first research question of this study (RQ1) was: what are the main Semantic Web applications area from Linked Open Data Cloud.

However, we asked what exactly can a developer do with these ontologies? Normally, the answer is to develop semantic web applications. In this regard, what would be Semantic Web applications' architecture? Bizer, et. al., [4] presented the general architecture of a semantic web application (Figure 2). The architecture of a semantic web application when Internet of Things or Artificial Intelligence is involved generally implies the presence of the IoT in the publication layer (as a source of data) and Artificial Intelligence in the Application layer (in order to apply Artificial Intelligence or machine learning on integrated queried data).

As we easily notice when we analyze Figure 2, we can say that, if we are interested to find out what exactly are the research subjects to study in order to actually develop a semantic web application, we must study about (RQ2):

- *How to implement a Linked data Wrapper.*
- *How to develop a web data access module.*
- *How to develop a vocabulary-mapping module.*
- *How to implement an identity resolution module.*
- *How to implement a quality evaluation module.*
- *Besides actually developing the application layer.*

Therefore, we expect to find studies on these subjects in the scientific literature. This is our second research focus.
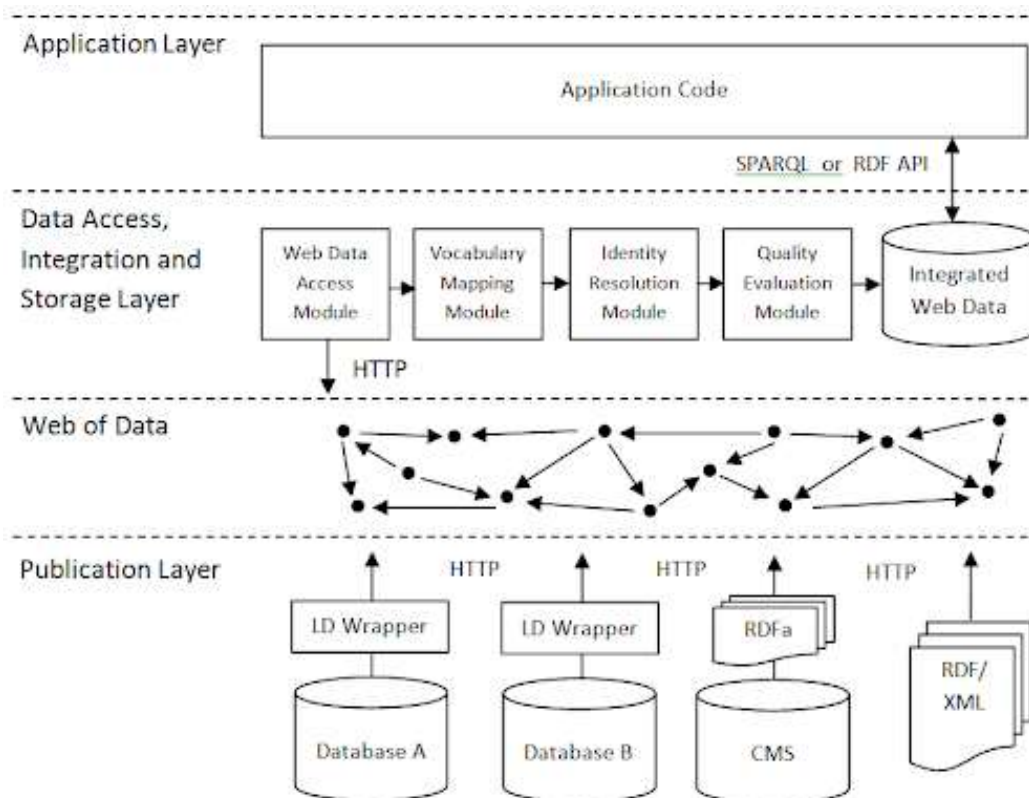
Fig. 2 The architecture of a semantic Web application [4]

In addition, we observed that there are already established technologies to describe data with terms belonging to ontologies. Resource Description Framework in Attributes (RDFa) is a technology to embed data in HTML. Resource Description Framework (RDF) is the standard to describe data in triples of subject-predicate-object. Each subject has a unique Uniform Resource Identifier (URI) or it can be a blank node, each predicate is a property defined by a publicly available ontology and represented with an Uniform Resource Identifier (URI), and each object has an Uniform Resource Identifier (URI) or it can be a literal.

Another question to ask, if not the most important one, is why exactly use semantic web? What is the necessity? The literature and practitioners thought us that Semantic Web is useful to integrate heterogeneous data in order to gather knowledge, insights or supplementary relationships between data. Are there already developed tools that could help us to analyze data and get insights from data described with ontologies? (RQ3) This is another question that this study is answering.

2. Materials and Methods
This study explores articles obtained from the following literature databases: Clarivate, Scopus, and Google Scholar. We took into consideration the title, abstract and keywords, but also information about authors' address and total number of citations. In addition, the study explored Linked Data Open Cloud website and took into consideration ontologies belonging to each activity. Besides investigating articles and ontologies, we investigated also the W3C website (https://www.w3.org), as W3C established Resource Description Framework, Ontology Web Language, SPARQL Protocol and RDF Query Language (SPARQL) and methodology and/or best practices to develop Semantic Web applications. The entire dataset contains 1461 datasets. We analyzed each dataset by categories as described in the Methods section. The data contains an identifier, an eventual SPARQL endpoint and a description about the total number of triples. We studied the availability

of the links towards the SPARQL endpoints and the purpose on which the dataset was build.

We searched information about using ontologies also in/by research scientific papers. On the Clarivate Analytics we found 84941 articles written on the subject of ontology, from which computer science is represented by 36264 articles. We analyzed the data by considering authors' country, the title of the paper, the abstract, the keywords, the type of the paper, and the publication year.

We established the methods to study this subject by anchoring on methods used in other research articles. Our methodology consists in five steps:

1. Defining the research questions
2. Establishing the search keywords (in the mainstream of publications or on official websites known as standards' recommender or on the Linked Open Data Cloud website, as it was our case)
3. Establishing criteria for facts selection
4. Extracting data
5. Data analysis
6. After establishing the purpose of our research, we identified the main three research questions and enlisted them in the Introduction section.
7. What are the main Semantic Web applications area from Linked Open Data Cloud.
8. What exactly are the research subjects to study in order to actually develop a semantic web application
9. Are there already developed tools that could help us to analyze data and get insights from data described with ontologies?

For each question, to establish research terms, we identified sub-questions and their purpose. We identified 8 research sub-questions. Table 1 presents the main research questions.

Table 1. The main research questions

RQ	RQ	Research question	Goal
RQ1	RQ11	What are the main activities from the Linked Open Data cloud	Identify applicative areas
	RQ12	What is the current state of using those ontologies	Assessing the use
	RQ13	What are the methodologies to develop semantic web apps	Gather insights on developing semantic web applications
	RQ14	What are the technologies involved	Gather insights on developing semantic web applications
RQ2	RQ21	What are the technologies used to develop LD wrappers, data access module, vocabulary-mapping module or identity resolution module?	Gather insights on developing semantic web applications
	RQ22	What is their scalability?	Assessing the use
RQ3	RQ31	What are the most encountered tools to gather knowledge from semantic web data	Identify applicative areas, assessing the use
	RQ32	Is there any study that proves an impact of using semantic web technology when analyzing data?	Identify applicative areas, assessing the use

Table 2 presents the search terms that we used to retrieve facts from databases.

Table 2. Search terms and source data

RQ	Search terms	Source data
RQ11	Government, Geography, Life sciences, Linguistics, Media, Publications	Linked Open Data cloud
RQ12	Ontology- only articles that have an associated dataset and only articles that were highly cited	Clarivate Analytics, Scopus and Google Scholar
RQ13	W3C recommendations	W3C website
RQ14	W3C standards and tools	W3C website
RQ21	Linked data Wrapper, web data access module, vocabulary mapping module	Clarivate Analytics, Scopus and Google Scholar
RQ22	Linked data Wrapper, web data access module, vocabulary mapping module	Clarivate Analytics, Scopus and Google Scholar
RQ31	Mining Semantic Web data	Clarivate Analytics, Scopus and Google Scholar
RQ31	Mining Semantic Web data	Clarivate Analytics, Scopus and Google Scholar

Both quantitative and qualitative analyses were carried out on the data. We present the main results in the next Section.

3. Results
The main results are presented on each sub-question.
RQ11 What are the main activities from the Linked Open Data cloud

If we search at the Linked Open Data Cloud website, we notice that it is not easy to query the cloud about ontologies from a specific applicative area. In order to count an approximatively number of datasets, we searched after each applicative area in the Browse page. The results are those presented in Table 3.

Table 3 The number of datasets from LOD cloud on each applicative area

Applicative areas	Number of datasets from the Linked Open Data cloud diagram
Government	18
Geography	6
Life sciences	2
Linguistics	5
Media	21
Publications	32
Social Networking	21
User generated	35

Each main activity has an important number of datasets, which connects with other datasets from the Linked Open Data cloud. As ontology engineering is a field emerged from knowledge representation, we observe that the main applicative areas are life sciences (back, in 1970, there were an important number of knowledge based systems developed for life sciences), government and geography, as open data invites to make publicly available

governmental data; linguistic, media, and publications as the concept of semantic web itself is related to semantics, relationships between words, synonyms, antonyms, and/or homonyms.
RQ12 What is the current state of using ontologies
In order to answer to this question, we analyzed diverse data sources from the Linked Open Data cloud. Table 4 presents their most

important characteristics.

<p style="text-align:center">Table 4 Characteristics of some datasets publicly available on
Linked Open Data Cloud website</p>

Dataset	Description	Number of triples	Access
EU: fintrans.pub-licdata.eu Namespace:	The names of the beneficiaries of ca. 10 billion of grants and other forms of support, procurement contracts	1,897,840 triples	SPARQL endpoint does not function. Only the metadata is available
Brown Corpus in RDF/NIF	newspapers texts on diverse topics, non-fiction and fiction books as well as government documents	14,335,131 triples	No SPARQL endpoint available, no data to download
data.gov.uk Time Intervals	This is an infinite set of linked data. It includes government years and properly handles the transition to the Gregorian calendar within the UK	1,000,000,000 triples	Download options in RDF
Diavgeia	Greece governmental data	48,051,108 triples	SPARQL endpoint does not function, the official website mentions the current miss functionalities
Entrez Gene	Public domain information on the National Library of Medicine	1,327,212 triples	No SPARQL endpoint, download does not function
EPA-TRI USA	toxic chemical releases and pollution prevention activities reported by industrial and federal facilities	137,423,327 triples	EPA Linked Data Services (Platform for accessing all EPA Linked Data sets using query tools, services, applications, URI resolution, mapping and more.)
GovWILD - Government Web Integration for Linked Data	Data about politicians, parties, government agencies, funds, companies, and industrial leaders	42,773,056 triples	SPARQL endpoint or download options do not function
IATI as Linked Data	International Aid Transparency Initiative data	36,629,045 triples	Available SARQL endpoint

Dataset	Description	Number of triples	Access
Instance Hub Fiscal Years	This dataset lists the variety of fiscal years for different countries and different domains (corporate, government, personal).		Nothing available
The London Gazette - Gazettes Data	The London Gazette, A wide range of notices, the majority of which are mandatory by law appear in the following broad categories (bold = very high numbers of notices): State, Parliament, Ecclesiastical, Public Finance, Transport, Planning, Health, Environment, Water, Agriculture & Fisheries, Energy, Post & Telecom., Other Notices, Competition, Corporate Insolvency, Personal Insolvency, Companies & Financial Regulation, Partnerships, Societies Regulation, Personal Legal.	12,000,000 triples	No SPARQL endpoint available, download does not function
N-Lex as Linked Data	national legislation databases (Germany)		Nothing available
Organigram and staff pay data for Thurrock Thames Gateway Development Corporation	A list of most Senior Civil Service posts	310 triples	Nothing available
statistics.data.gov.uk	Linked data about administrative areas used within UK government official statistics	343,733 triples	SPARQL endpoint and download data available
Street level crime reports for England and Wales	a linked data representation of the street-level crime reports first released for England and Wales in 2011	7,780,271	SARQL endpoint and download

We can enumerate several semantic web applications that we found after searching the public ontology on their original website: Aid Transparency Tracker, AidView, Akvo Openaid, Development Tracker, IATI Explorer, International Aid Transparency Initiative Data, Crime finder, Governmental data, Health data. After analyzing data, we observed that the most encountered applicative problems are: public spending, public data about contracts, funds, governmental actors and/or legislative acts.

After a brief analysis of the other dataset, we can say the same facts. We noticed though that most of the SPARQL Protocol and RDF Query Language (SPARQL) endpoints do not function, but the spreading is massive, the intention is, generally, to make available, publicly, data in the same format, namely: Resource Description Framework. This unavailability of SPARQL Protocol and RDF Query Language (SPARQL) endpoints could lead us to draw the conclusion that these datasets are not used, but it is early to derive this conclusion.

The results obtained after searching over the

Clarivate Analytics database were of 2402 articles with associated datasets, which we consider to be important. The average citations per each article belonging to the most cited articles is around 390. The main WebOf-Science categories are those depicted in Figure 3. The publication years are presented in Figure 4.

Fig. 3 WebOfScience categories of the articles written on the subject of ontology that have associated datasets
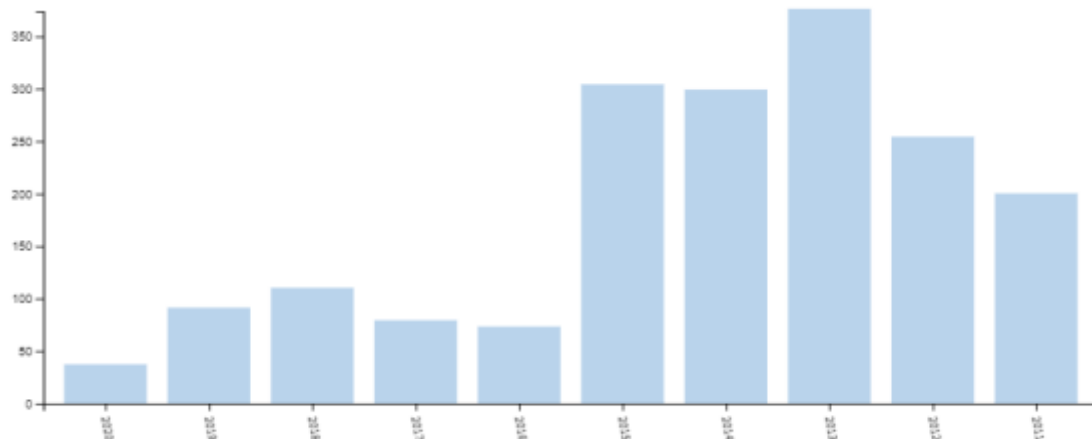
Fig. 4 The publication years of the articles written on the subject of ontology that have associated datasets

We were interested to find about the countries in which this subject presents interest. Figure 5 presents the main countries from which the authors writing on the subject of ontology belong.

Fig. 5 The countries from which the authors writing on the subject of ontology belong

We can say that using ontologies is a debated subject, it is prospected by many researchers belonging to different fields of science. They develop tools, datasets, frameworks, and

applications.

Q13 What are the methodologies to develop semantic web apps

To answer to this question, we investigated the (World Wide Web Consortium) W3C website, besides analyzing the mainstream of research. W3C recommends [5] to use:

- Predicate Uniform Resource Identifiers should be HTTP Uniform Resource Locators (URLs)
- Use and include the predicate rdf:type to represent the concept of type in Linked Data Platforms (LDPRs)
- Avoid dot-segments in Uniform Resource Identifiers of POSTed content or use with caution
- Represent container membership with hierarchical Uniform Resource Identifiers
- Re-use established linked data vocabularies instead of (re-)inventing duplicates
- Representing relationships between resources
- Open source tools and libraries for the semantic Web include:
- Apache TinkerPop: https://tinkerpop.apache.org
- RDFLib: https://github.com/RDFLib/rdflib
- Apache Jena: http://jena.apache.org/
- Protégé: https://protege.stanford.edu
- Sesame: http://www.openrdf.org/

- Linked Media Framework: https://code.google.com/p/lmf/
- Open Semantic Framework: http://opensemanticframework.org/
- D2R: http://d2rq.org/d2r-server
- Paget: http://code.google.com/p/paget/
- Semantic Media: http://semantic-mediawiki.org/wiki/Semantic_MediaWiki

To develop semantic web applications, there are a several number of software libraries for each important programming language. For example, the standard Python library ElementTree44 is used to process the XML files, which are then converted to RDF using the RDFLib45 library. For Java programming language, there is Apache Jena library, Apache Jena TDB and Fuseki server. Cliopatria triple store was used by some authors, for example [6].

Another important subject of research is how to visualize the semantic web data. For example, EPA United States Environmental Protection Agency developed a platform (Linked Data as a Service) on which users could query and visualize data.

Q21 What are the technologies used to develop LD wrappers, identity resolution and so on?

To answer this question, we searched over Clarivate analytics website. The terms used in our queries are those from Table 5.

Table 5. Search term used to identify articles written on the subject concerning RQ21

Search string	Results
Linked Data Wrapper	3 articles: Java, Python
vocabulary mapping	26 articles
identity resolution	83 articles, after analyzing the title and the abstract, we identified subjects as: Identity Resolution in Ontology Based Data Access to Structured Data Sources [7], Distant supervision of relation extraction in sparse data [8], CODA: Computer-aided ontology development architecture [9], Integrating Product Data from Websites Offering Microdata Markup [10], A Framework for Identity Resolution and Merging for Multi-source Information Extraction [11]

Different programming language are used to implement Linked Data wrappers and all the other modules. Among them, we list Java and Python. As of scalability, authors [12] writes

that LDWs are fragile upon upgrades on the underlying APIs, compromising LDW stability. The authors proposed SYQL, a Linked data web (LDW) platform on top of Yahoo's YQL.

Q31 What are the most encountered tools to gather knowledge from semantic web data and

Q32 Is there any study that proves an impact of using semantic web technology when analyzing data?

Mining Semantic Web ontologies provides a great possibility to get better results [13]-[14], discovers new and valuable insights data from the semantic annotations [15], solves problems that deals with complex and heterogeneous data [13], [16] and improves in easy, and effective ways the results of the web mining [17], [18]. We encountered SemWeb (working within RapidMiner). It transforms RDF triples to an example set and then any kind of learning can be applied on this data [19], [20].

4. Discussions
This study conducts a review of the main applicative areas of semantic web technology, of the tools needed to develop SW applications and the actual trends of study. Our research prospected the current state of the mainstream research, focused on semantic web technologies and ontologies with specific search keywords. The results provided a clear view on each research question that our study established.

Semantic web integrates with all the other ICT technologies. The following ideas are revealed by our research:

- The applicative areas of semantic web technologies are diverse, with a raising trend in making publicly available datasets and prove their use with potential impact on improving AI, IoT or other integrated systems;
- Semantic web technology is useful in government and support the development of smart governmental business intelligence.
- Semantic web technology is a technology used in the data layer of the application architecture. It helps describing data with

terms belonging to ontologies. There is a growing interest in developing quality assessment frameworks for Linked Data Open datasets [22], [23], [24] which improves the trust in represented data. The trust improves the use and using data, semantically enriched, improves the quality of the output data.

- Semantic web belongs to AI research, it is therefore naturally integrated with machine learning, natural language processing and sensing from data.
- Semantic web technology is used for improving semantics, therefore the future developments of AI, analytics, business intelligence is better prospected.

The results of our research proved that semantic web technologies are an important tool for describing data and an important component in the data layer of any intelligent application. The level of acceptance is determined by the trust gained at the users and companies' level. Therefore, it becomes important to involve the final users in semantic web applications' development. They can, for example, describe datasets or define the terms belonging to different vocabularies. The users are diverse, but usually engaged in analyzing publicly available open data, either it is about governmental data, educational data, medical data or linguistic data. For each applicative area, we presented a selection of applications to identify potential users.

Research directions
Semantic web applicability did not reach its full potential yet due to many reasons discussed before by specialty studies (link discovery, standards too technical to use, appropriate tools to develop semantic web applications). Based on the finding in Results section, this section provides some future research directions:

- On answering the question "what are the main activities in which semantic web technologies can be applied" we noticed that the specialty studies are lacking concrete results. The existent review studies focus on development tools or ontologies. Most of the studies conducted so far focus

on the use of semantic web technologies and tools applied to contexts such as biology, social sciences, libraries, research, and education [21]. We suggest applying AI techniques on LOD cloud datasets to gather knowledge about specific activities for which semantic web applications are developed.

- About the current use of ontologies belonging to LOD cloud we can state that they are used for describing data that different research projects intended to. There are European research funded projects which developed, maintained, and publicly offered ontologies and datasets for different purposes. Among them we cite: Europeana (libraries, museums, art gallery data https://www.europeana.eu/ro), Inception (cultural heritage), ALIGN (software quality and data engineering) and others. On the Cordis website, there are listed 312 research projects that uses semantic web technologies, for example. What is important to notice is the fact that ontologies and semantic web are financed, prospected yet as research activities, and that they span from FP7 to HORIZON 2020 in the effort to integrate heterogeneous data in order to implement an information layer of bigger applications, useful to society and citizens and which integrate various emerging technologies like IoT, Blockchain or AI. In this respect, we suggest ontologies discovery by analyzing research projects that developed them.

- Concerning the methodologies used to develop semantic web applications we consider that they are very well represented, but not standardized as methodologies to develop applications. Each author is defining its own methodology and, in general, the recommendations provided by W3C are the only one considered when publishing and consuming Linked Open Data. Therefore, we suggest as future research direction establishing methodologies for consuming linked open data. As for publishing open data we consider that developing methodologies to discover links between different datasets is a promising

research direction.

5. Conclusions

This study contributed to the research literature by presenting the applicative areas of semantic web and semantic web applications' development tools, architectures, and methodology. This study intended to investigate how semantic web applications contribute to the improvement of our lives. Our lives are surrounded by unstructured data, heterogeneous sources of data and, generally, by a lot of data. The semantic web applications' development must integrate into the web of data. This depends on the standards' development and integration. It is expected that improved meaning of information will be achieved. The general findings of our research findings are: (1) the potential of semantic web technology to provide positive changes in developing web applications is huge if multiple users use them; (2) the development tools continually evolve. The specific findings are: (1) the applicative areas of semantic web technologies are diverse; (2) the methodology and the tools to develop semantic web applications exist and improve on continual bases; (3) there is a big potential in gather insights on data when combining machine learning and semantic web.

This article has some limitations. One limitation is the big number of scientific articles written on the subject and the big number of datasets. To address this, we searched with specific keywords that we established by studying the practitioner's results: the available ontologies on the Linked Open Data cloud.

Successful implementation of these technologies involves integration and acceptance. As our study revealed there are premises to solve these future engagements.

References

[1] A European Strategy for Data | Shaping Europe's digital future, https://ec.europa.eu/digital-single-market/en/european-strategy-data

[2] Hogan, A. "The Semantic Web: Two Decades On." Semantic Web Preprint 2020: 1-17, doi: 10.3233/SW-190387

[3] Protégé, https://protege.stanford.edu/

[4] Bizer, C.; Heath, T.; Idehen, K.; Berners-Lee, T. Linked data: evolving the web into a global data space. Morgan & Claypool Publishers 2011.

[5] World Wide Web Consortium https://www.w3.org/

[6] Brandt, K.; de Boer, V. Linked Data for IATI (Doctoral dissertation, MSc Thesis), Vrije Universiteit Amsterdam) 2013

[7] Toman, D.; Weddell, G. Identity Resolution in Ontology Based Data Access to Structured Data Sources. In: Nayak A., Sharma A. (eds) PRICAI 2019: Trends in Artificial Intelligence. PRICAI 2019. Lecture Notes in Computer Science, vol 11670. Springer, Cham. https://doi.org/10.1007/978-3-030-29908-8_38

[8] Ranjbar-Sahraei, B.; Rahmani, H., Weiss, G.; Tuyls, K. Distant supervision of relation extraction in sparse data. Intelligent Data Analysis, 2019, 23(5), 1145-1166, https://doi.org/ 10.3233/IDA-184238

[9] Fiorelli, M.; Pazienza, M. T.; Stellato, A.; Turbati, A. CODA: Computer-aided ontology development architecture. IBM Journal of Research and Development, 58(2/3), 2014, 14-1, https://doi.org/ 10.1147/JRD.2014.2307518

[10] Petrovski, P.; Bryl, V.; Bizer, C. Integrating product data from websites offering microdata markup. In C.W. Chung (Ed.), WWW '14 Companion: Proceedings of the 23rd International Conference on World Wide Web (pp. 1299–1304), 2014, https://doi.org/10.1145/2567948.2579704

[11] Yankova, M.; Saggion, H.; Cunningham, H. A Framework for Identity Resolution and Merging for Multi-source Information Extraction. In Proceedings of the International Conference on Language Resources and Evaluation, 2008.

[12] Azpeitia, I.; Iturrioz, J.; Díaz, O. Volunteering for Linked Data Wrapper maintenance: A platform perspective. Information Systems, 2020, 89, 101468, https://doi.org/10.1016/j.is.2019.101468

[13] Nebot, V.; Berlanga, R. Finding association rules in semantic web data. Knowledge-Based Systems, 2012, 25(1), 51-62, https://doi.org/10.1016/j.knosys.2011.05.009

[14] Segura, A.; Vidal-Castro, C.; Menéndez-Domínguez, V.; Campos, P. G.; Prieto, M. Using data mining techniques for exploring learning object repositories. The Electronic Library, 2011, https://doi.org/10.1108/02640471111125140

[15] Abedjan, Z.; Naumann, F. Context and target configurations for mining RDF data. In Proceedings of the 1st international workshop on Search and mining entity-relationship data (pp. 23-24), 2011.

[16] Nebot V.; Berlanga R. Mining Association Rules from Semantic Web Data. In: García-Pedrajas N., Herrera F., Fyfe C., Benítez J.M., Ali M. (eds) Trends in Applied Intelligent Systems. IEA/AIE 2010. Lecture Notes in Computer Science, vol 6097. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-13025-0_52, 2010

[17] W. Yong-gui; J. Zhen, Research on semantic Web mining, 2010 International Conference On Computer Design and Applications, Qinhuangdao, 2010, IEEE pp. V1-67-V1-70, doi: 10.1109/ICCDA.2010.5541057.

[18] Velásquez, J. D.; Dujovne, L. E.; L'Huillier, G. Extracting significant website key objects: A semantic web mining approach. Engineering Applications of Artificial Intelligence, 24(8), 1532-1541, https://doi.org/10.1016/j.engappai.2011.02.001, 2011

[19] Ristoski, P.; Bizer, C.; Paulheim, H. Mining the web of linked data with rapidminer. Journal of Web Semantics, 35, 142-151, https://doi.org/10.1016/j.websem.2015.06.004, 2015

[20] Potoniec, J.; Lawrynowicz, A. Rmonto: ontological extension to rapidminer. In 10th International Semantic Web Conference (Vol. 111), 2011.

[21]   Avila-Garzon, C. Applications, Methodologies, and Technologies for Linked Open Data: A Systematic Literature Review. International Journal on Semantic Web and Information Systems (IJSWIS) 16.3 (2020): 53-69. doi: 10.4018/IJSWIS.2020070104

[22]   Debattista, J., C. Sören Lange, and A. Sören. "A.: LUZZU–A Framework for Linked Data Quality Assessment', CoRR abs/1412.3750." (2014).

[23]   Zaveri, Amrapali, et al. "Quality assessment for linked data: A survey." Semantic Web 7.1 (2016): 63-93.

[24]   Mendes, Pablo N., Hannes Mühleisen, and Christian Bizer. "Sieve: linked data quality assessment and fusion." Proceedings of the 2012 Joint EDBT/ICDT Workshops. 2012.

Sabina-Cristiana NECULA (b. July 27, 1979) received her PhD in Accounting (Business In-formation Systems) (2007). She is currently a Scientific Researcher at Alexandru Ioan Cuza University of Iasi, Faculty of Economics and Business Administration. Her current research interests include Semantic Web standards and technologies, Decision Support Systems, Business Information Systems. She is the author of more than 30 scientific papers.