Linked Data: perspectives for IT professionals

Ricardo Colomo-Palacios

ricardo.colomo@uc3m.es
Computer Science Department
Universidad Carlos III de Madrid
Av. Universidad 30, Leganés, 28911, Madrid, SPAIN
Phone: +34 91 624 5958
Fax: +34 91 624 9129

José Luis Sánchez-Cervantes

joseluis.s.cervantes@alumnos.uc3m.es
Computer Science Department
Universidad Carlos III de Madrid
Av. Universidad 30, Leganés, 28911, Madrid, SPAIN
Phone: +34 91 624 5936
Fax: +34 91 624 9129

Giner Alor-Hernandez

galor@itorizaba.edu.mx
Division of Research and Postgraduate Studies
Instituto Tecnológico de Orizaba
Avenida Oriente 9 No. 852 Col. Emiliano Zapata
Orizaba, 94320, Veracruz, México

Alejandro Rodríguez-González

alejandro.rodriguez@uc3m.es
Computer Science Department
Universidad Carlos III de Madrid
Av. Universidad 30, Leganés, 28911, Madrid, SPAIN
Phone: +34 91 624 6257
Fax: +34 91 624 9129

ABSTRACT

The Semantic Web provides a common framework that allows data to be shared and reused across application, enterprise, and community boundaries. To make the Semantic Web or Web of Data a reality, it is necessary to have a large volume of data available on the Web in a standard, reachable and manageable format. In addition the relationships among data also need to be made available. This collection of interrelated data on the Web can also be referred to as Linked Data. Linked Data lies at the heart of the Semantic Web: large scale integration of, and reasoning on, data on the Web. Supporting the adoption of semantic Web technologies, currently, there exists a great of tools oriented to creation, publication and management of data, and a big subset for Linked Data. However, an important weakness in this area of Web engineering is that it has not been established completely a formal reference that integrates the necessary infrastructure in terms of components, and the order that this infrastructure must be implanted. This lack implies a slower technological adoption, covering both public sector and private sector. This paper explores the emergence of the Semantic Web and in particular of Linked Data, and their potential impact on IT industry. The main advantages of using Linked Data are discussed from an IT professional perspective where the capability of having standard technologies and techniques to access and manipulate the information is one of the most important achievements in the application of Linked Data.

Keywords:

Information Technology Professionals, Linked Data, Ontologies, Semantic Web.

1. Introduction

Today, the successful exploitation of Information Technology (IT) for corporations dependent upon the availability of IT professionals to design and integrate IT infrastructure and applications (Agarwal & Ferratt, 2002). As a result of this, IT workers represent a strategic resource for firms which have the ability to bestow competitive advantages (Bharadwaj, 2000; Wade & Hulland, 2004). Given that a company's human resources can be a source of competitive advantage that is difficult for competitors to imitate (Kuean, Kaur& Wong, 2010), IT professionals are in the eye of the hurricane of firms assets. In this context, IT practitioners are gaining importance in today's changing and more and more competitive environment (López-Fernández, Martín-Alcázar & Romero-Fernández, 2010).

Beard, Schwieger and Surendran (2010) indicated that the skills and knowledge of IT professionals should prove invaluable in seeking and implementing innovations. Given that according to Sullivan and Dooley (2010), organizations must improve their competitive advantage and respond faster to changing markets by increasing productivity (O'Sullivan & Dooley, 2010), organizations must face a iterative process of revision and development of IT workers' competences. That is the rationale behind Casado-Lumbreras et al. (2009) words: IT workers professional practice must be continually revised and improved in order to adapt workers' competencies to technical innovations.

However, the turbulent nature of the IT industry may complicate long-term IT skill forecasting (Schwarzkopf et al., 2004). In this sense, according to Joseph et al. (2007), IT professionals experience lower job satisfaction, possibly because of having to cope with the changing IT skills set required by the profession. Moreover, changing technology has been pointed out as one of the main stressors for IT professionals (Love et al., 2007).

In any case, advances in technology and developments in the business environment have driven changes in desired skill sets of IT professionals (Goles, Hawk & Kaiser, 2008). As a consequence of this, skill development programs for IT roles are well established in the industry.

One of the main trends in the IT scenario in the last years is Linked Data. In summary, Linked Data is simply about using the Web to create typed links between data from different sources (Bizer, Heath & Berners-Lee, 2009). Unlike Web 2.0 mashups that work against a fixed set of data sources, Linked Data applications can discover new data sources at runtime by following data-level links, and can thus deliver more complete answers as new data sources appear on the Web (Bizer, 2009). In this scenario, Linked Data, based on Semantic Technologies seem to be a new deal for IT professionals. The aim of this paper is to explore the implications of this set of technologies for IT practitioners.

This paper is structured as follows. Section 2 outlines semantic technologies. Section 3 describes Linked Data concepts. Section 4 proposes the main implications of Linked Data technologies for IT professionals. Finally, section 5 offers several conclusions and provides future lines of research.

2. Semantic Technologies

The Web has changed the way users get information, but its success and exponential growth have made it increasingly difficult to find, access, present, and maintain such information for a wide variety of users (Lassila, et al. 2000). However, during years the World Wide Web has evolved allowing the transformation of the traditional Web to Semantic Web called also "the Web of Data" (Shadbolt, et al. 2006). The Semantic Web is described by Berners-Lee, Hendler and Lassila (2001) as an extension of the current Web in which information is given a well-defined meaning, better enabling computers and people to work in cooperation in a better way.

The transformation of traditional Web in Semantic Web consists in the increase of Web Pages with data targeted at computers and by adding documents solely for computers therefore, the Semantic Web is not a separate Web but an extension of the current one (Allemang & Hendler, 2011). For its success, the Semantic Web has a set of technologies among which the Ontology is the main one (Castellanos, et al. 2011). The Ontologies define common, shareable and reusable views of a domain, and they give meaning to information structures that are exchanged by information systems (Brewster & Hara, 2004). In the literature, Ontology has various definitions, one of the most cited is proposed by (Gruber, 1993) where an Ontology is a specification of a conceptualization. Figure 1 shows the elements of the Semantic Web stack.

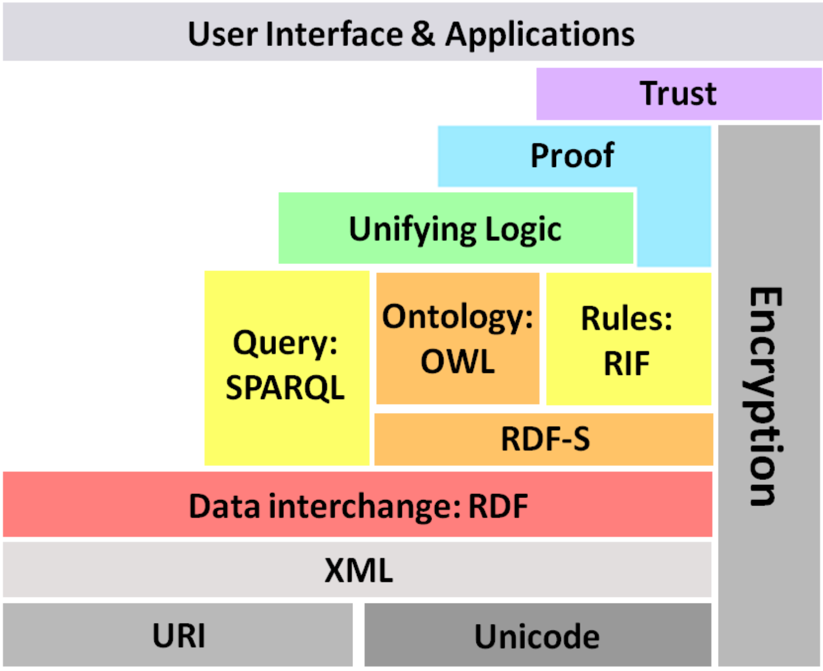

Figure 1. Elements of the Semantic Web Stack

The Semantic Web technologies are used in several domains that include the data integration and the improvement of the information search. Some examples of its application are briefly described below:

Semantic Annotations: It is a process that allows add semantic metadata to Web resources and consists in assigning unambiguous meaning to such resources in order to enable a more efficient discovery mechanism (Kiryakov, et al. 2004).

Semantic Search Methods: Its aim is augment and improve traditional search results by using not just words, but concepts and logical relationships (Fazzingaa, et al. 2011).

Semantic Content Discovery: It makes use of taxonomies and Ontologies for describing different types of content and for dynamic linking between content items (Sheth, et al. 2002).

Social Networks Technologies: These are used to help people track, discover, and share content around topics they are interested in, thus forming social networks (Dietrich & Jones, 2007).

Natural Language Interfaces: It consists in the application of methods and technologies useful to provide natural, human like interaction with the computer (Kaufmann & Bernstein, 2010).

Service integration: It adds semantics to Web services with the aims to augment service integration, i.e. discovery, composition, ranking, selection and mediation of services (Janev & Vranes, 2010).

With the Semantic Web Technologies, the traditional Web has evolved from a global information space of linked documents to one where both documents and data are linked. The result of this evolution is a set of best practices for publishing and connecting structured data on the Web known as Linked Data (Segaran, et al. 2009). Figure 2 shows some Web technologies and their relationship with Semantic Web technologies.

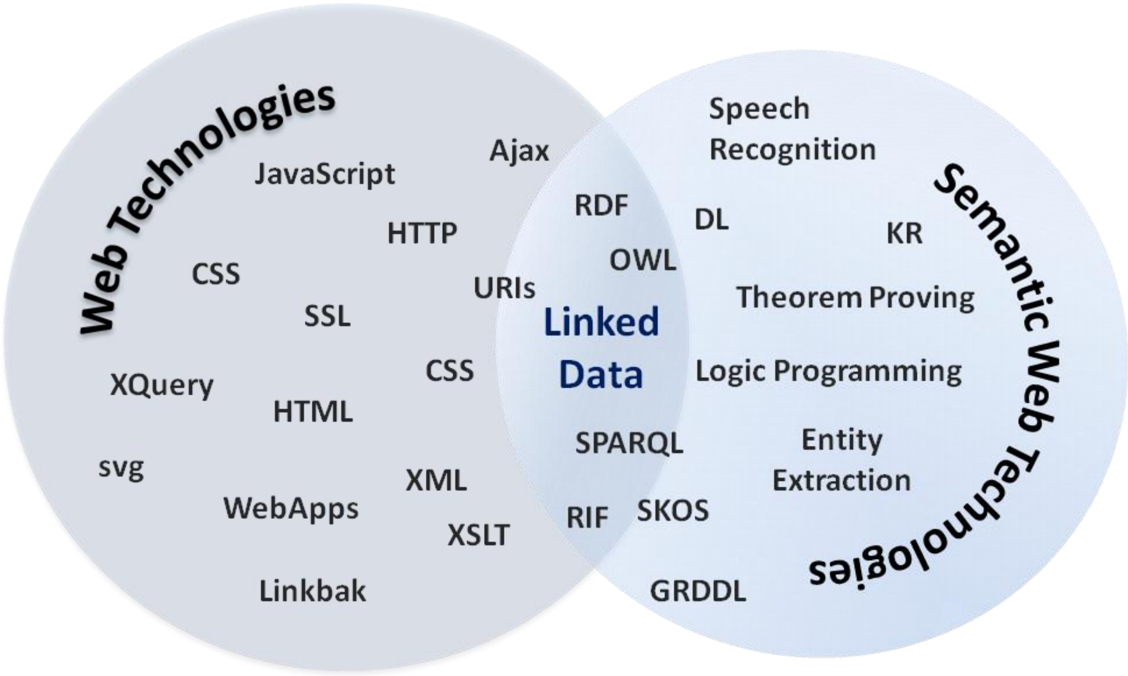

Figure 2. Web technologies and Semantic Web technologies

3. Linked Data

Bizer, Heath and Berners-Lee, (2009) mentioned that Linked Data is simply about using the Web to create typed links among data from different content and data sources. These may be as diverse as databases maintained by two organizations in different geographical locations, or simply heterogeneous systems within one organization that, historically, have not easily interoperated at the data level. Technically, Linked Data refers to data published on the Web in such a way that it is machine-readable, its meaning is explicitly defined, it is linked to other external data sets, and can in turn be linked to from external data sets.

The Hypertext Web is conformed of HTML pages (Hypertext Markup Language) connected by untyped hyperlinks. Unlike these, Linked Data relies on documents containing data in RDF (Resource Description Framework) format (Klyne & Carroll, 2004).

However, rather than simply connecting these documents, Linked Data uses RDF to make typed statements that link arbitrary things in the world. The result, which we will refer to as the Web of Data, may more accurately be described as a Web of things in the world, described by data on the Web. For the publication of data on the Web outlined a set of "rules" called "Linked Data principles" that provides a basic recipe for publishing and connecting data using the infrastructure of the Web in a way that all published data becomes part of a single global data space (Berners-Lee, 2006):

1. Use URIs as names for things

2. Use HTTP URIs so that people can look up those names
3. When someone looks up a URI, provide useful information, using the standards (RDF, SPARQL)
4. Include links to other URIs, so that they can discover more things

The application of Semantic Web technologies such as RDF, SPARQL, URIs, among others, and the correct use of its architecture and standards provides an environment that allows perform various operations among which are included the query of data and obtain inferences using vocabularies.

Uniform Resource Identifier (URI)

The URI syntax has been designed with global transcription as one of its main considerations. A URI is a sequence of characters from a very limited set: the letters of the basic Latin alphabet, digits, and a few special characters and it is composed by a hierarchical sequence of components (Berners-Lee, Fielding & Masinter, 2005). The URI's generic syntax is shown in Figure 3:

```
URI = scheme ":" authority "/" path ["?" query]["#" frag]
authority = [userinfo "@"] host [":" port]
host = IP-literal/IPv4Address/reg-name
IP-literal = "[" (IPv6Address/IPvFuture)  "]"
IPvFuture = "v" 1*HEXDIGIT "." 1*(unreserv/sub-delims/":")
reg-name = *(unreserv/pct-encoded/sub-delims)
port = *DIGIT
```

Figure 3. URI's generic syntax

In the URL/URI of a client-server application, the scheme identifies the application protocol to be employed (e.g. http), the authority part defines the location of the peer entity (e.g. www.example.com:80), while the optional path, query and fragment sections usually refer to the piece of information to be handled by the protocol (e.g. index.html#top) (Urueña & Larrabeiti, 2005).

Resource Description Framework (RDF)

RDF is a standard model for data interchange on the Web. RDF has features that facilitate data merging even if the underlying schemas differ, and it specifically supports the evolution of schemas over time without requiring all the data consumers to be changed. RDF extends the linking structure of the Web to use URIs to name the relationship among things as well as the two ends of the link (this is usually referred to as a “triple”). Using this simple model, it allows structured and semi-structured data to be mixed, exposed, and shared across different applications (Klyne & Carroll, 2004).
A “triple” consists of a subject, predicate and an object. It can be viewed as a resource identifier, an attribute or property name, and an attribute or property value (DuCharme, 2011). This linking structure forms a directed, labeled graph, where the edges represent the

named link between two resources, represented by the graph nodes. This graph view is the easiest possible mental model for RDF and it is often used in easy-to-understand visual explanations (Klyne & Carroll, 2004).

Protocol and RDF Query Language (SPARQL)

SPARQL is a query language for data that follows a particular model, but the Semantic Web is not about the query language or about the mode it is about the data. The booming amount of data becoming available on the Semantic Web is making great new kinds of applications possible, and as a well-implemented, mature standard designed with the Semantic Web in mind, SPARQL is the best way to get that data and put it to work in Web applications. SPARQL is based in a set of W3C standards for querying and updating data conforming to the RDF model (DuCharme, 2011).

Linked Data relies on two technologies that are fundamental to the Web: 1) Uniform Resource Identifiers (URIs) and 2) Hyper Text Transfer Protocol (HTTP). While Uniform Resource Locators (URLs) have become familiar as addresses for documents and other entities that can be located on the Web, Uniform Resource Identifiers provide a more generic means to identify any entity that exists in the world (Berners-Lee, Fielding & Masinter, 2005).

An example of Linked Data application

Linked Data provides direct access to the data sets entities within (or behind), Web pages enabling the construction of powerful data-mashups across heterogeneous data source collections without any programming. Also, Linked Data provides bridges among areas of interest and knowledge revealing relationships where information might be seen as unrelated and through its principles and practices have been adopted by an increasing number of data providers. A basic example of functionality of Linked Data is the direct navigation among Web pages using a SPARQL-based query to obtain information. Suppose the following scenario: we need access the NASA Web site and we require know the URL with the Apollo 7 information. Given Virtuoso SPARQL Query Editor available in: http://demo.openlinksw.com/sparql and using the FOAF vocabulary for names and homepages which is: http://xmlns.com/foaf/0.1/.

The SPARQL Query corresponding to get the required information is the following:

```
PREFIX foaf: <http://xmlns.com/foaf/0.1/>
SELECT ?homepage ?craft
{
	?craft foaf:homepage ?homepage .
	?craft foaf:name "Apollo 7"
}
```

Figure 4 shows the results obtained, once the SPARQL-based query was executed.

homepage	craft
http://nssdc.gsfc.nasa.gov/database/MasterCatalog?sc=1968-089A	http://nasa.dataincubator.org/spacecraft/1968-089A

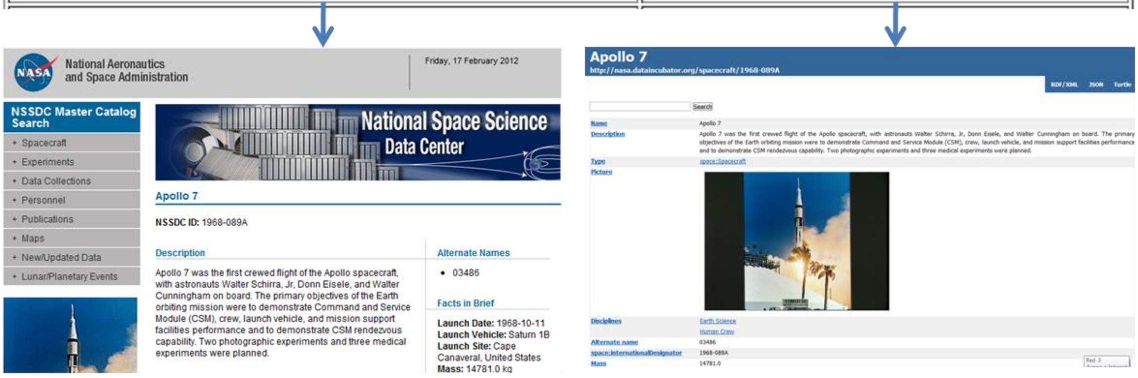

Figure 4. Results obtained of a SPARQL-based Query

There are many benefits obtained by the implementation of Linked Data practices on Web applications, where we can mention: 1) acquiring semantic information from RDF triples describing billions of data resources, 2) obtaining information more accurate through the links to navigate among Web pages by using data mining techniques, artificial intelligence, among others; in order to obtain and generate new knowledge.

4. Linked Data: implications for IT professionals

The term Linked Data was born in 2006 (Berners-Lee, 2006), as a new form of publishing and linking data from the Web. This new concept, which is behind of the well-known Semantic Technologies has emerged in the last years in several and different areas such as Economics and Finances (O'Riain, Harth & Curry, 2012), Bioinformatics (Zhao et al., 2009), Pharmaceutical research (Samwald et al., 2011), Enterprises (Wood, 2010), E-Government (Alvarez-Rodríguez et al., 2012) books (Bizer & Cyganiak & Gauss, 2007), scientific publications (Van de Sompel et al., 2009), films (Hassanzadeh & Consens, 2009), music, television and radio programs (Kobilarov et al, 2009), genes, proteins, drugs and clinical trials (Belleau et al., 2008, Jentzsch et al., 2009), online communities, statistical data, census results, and reviews (Heath & Motta, 2008) to mention a few.

Linked Data techniques and related technologies have become interesting to organizations and institutions from several domains. Nowadays, European projects such as Linked Open Data 2 (LOD2), Linked Open Data Around-the-Clock (LATC) and PlanetData are examples of this new situation.

However, an open question which has not been addressed yet is in which way Linked Data will affect to the perspectives of IT professionals. One of the main advantages of Linked Data concept is that from a broader perspective seems that researchers and scientists has finally reached an agreement in standardization. The technologies behind Linked Data allow knowledge sharing and formalization, and the wide adoption that these technologies are an important factor of the agreement which exists between different communities. From an IT professional perspective, this agreement represents an optimal scenario, because the capability of having standard technologies and techniques to access and manipulate the information is one of the most important achievements in the application of information technology techniques.

In this paper, five aspects are mentioned as the main features in which Linked Data takes part in IT professional's perspectives: multidisciplinary, rigor, availability, big data and training.

Multidisciplinary

IT professionals have to deal with several disciplines in their day by day, which reflects the high degree of multidisciplinary of these professionals. This multidisciplinary is one of the main points which allow supporting the introduction and use of Linked Data from IT professional's perspective. Linked Data allows to link information from several areas, representing each of these areas a discipline. The ability of link information from several disciplines and being able to navigate from one discipline to another will allow IT professionals to perform fast and accurate searches over the Internet by using Linked Data.

A good representation of this multidisciplinary from is the isolation of Linked Open Data Cloud Datasets as a knowledge owner. This information which is contained in the datasets can contain information from several areas or domains (i.e. DBpedia). However, most of the information which contains each dataset is referred to a unique domain or discipline.

For hence, the multidisciplinary of Linked Data allows to IT professionals to consume data from different sources and from different domains, allowing them to navigate through the Linked Open Data Cloud to search information from several domains. However, the consumption is not the only approach that should be remarked about Linked Data. This consumption is clearly linked with the concept of generation. The idea behind this approach is that IT professionals not only can consume information; they can create new data and publish it into the Linked Open Data Cloud, making it accessible to the community.

Trust, Quality, Relevance and Rigor:

As was exposed by Bizer, Heath and Berners-Lee (2009), an important challenge of Linked Data applications is the capability of ensuring that the most relevant or appropriate data is available to the user's needs. The quality of the data which will be available to the user depends on several aspects which only can be tackled by the users which publish the data (context, amount of information, tagging, area, vocabulary or linking with appropriate concepts among others). An overview of the different content, context, and rating-based techniques that can be used to assess the relevance, quality and trustworthiness of the data was provided by Bizer & Cyganiak (2009) or Heath (2008).

However, other important concept is the data modeling and vocabularies which have been used in the modeling process. As Bizer, Heath and Berners-Lee (2009) stated "different communities have specific preferences on the vocabularies, they prefer to use for publishing data on the Web". This is an important issue regarding the quality or rigor of the data, since the Web has capacity to use different vocabularies in parallel. It is considered a good practice to reuse concepts from well-known RDF vocabularies such as FOAF, SKOS, vCard, SIOC, Dublin Core or GoodRelations among others. However, it is also necessary the modeling of new bases or vocabularies to support the new information. In this context, when current vocabularies do not provide the necessary concepts, the necessity to create these vocabularies emerges. Hence, it is important to adopt quality schemas when these new vocabularies are defined. The aim is to guarantee the quality of the data which will be published under these vocabularies.

Availability

The availability of the data is other of the challenges for every Web application. However, availability presents two main meanings: 1) Availability as the term used to describe products and services that ensure that data continues to be available at a required level of performance in situation ranging from normal to disastrous; or 2) Availability as a term to describe products and services that ensure that data, in a normal situation, is always available, in this case, through Internet.

Both cases are interesting and can be explored in depth; however, the most pertinent definition of availability in the context of IT professionals is related with the second case. IT professionals have, in several contexts, the necessity of consuming large amounts of data, requiring that this data is always available. In this context, Linked Data offers a good set of opportunities for IT professionals. Each dataset of information is normally stored in one or several servers which contain the RDF triples that can be queries through Internet, offering a high degree of availability. However, even in the cases were the access to the service is unavailable, preventing the access to the data of a concrete dataset, Linked Data allows through the use of, for example, *sameAs* links, access to the information in alternative datasets as a solution.

Big Data

In the past few years, "big data" is a buzzword; even though different people mean different things when they use this term, the consensus is that data volumes are exploding, and that new tools are needed to handle these volume (Madden & van Steen, 2012). Big Data makes reference to the collection of datasets that grow so large that they become awkward to work using on-hand database management tools. This is not a simple problem, since big data is now part of every sector and function of the global economy (Manyika et al., 2011). The value of big data to an organization falls into two categories: analytical use, and enabling new products (Dumbill, 2012). The main difficulties to deal with this kind of data include capture, storage, search, share, analytics and visualization. A recent review of Big Data challenges can be found in the work of Bizer et al. (2012). In this context, Linked Data is one approach to cope with Big Data. The main differences between both approaches is the nature of the data and how it is stored (in Linked Data, we store the information as RDF triples whereas in Big Data not always it is possible to have a defined structure which allows processing the data in an easy way).

Nowadays, several times, IT professionals have to deal with large amounts of unstructured data. The nature of this data makes very difficult to these professionals the realization of several tasks such as search, analyze or store the information. Linked Data features can help with those problems allowing having a structured way to store the information, which permits an improved way to search and analyze the information which is behind the data. However, the main problem behind this is the process to convert this large amount of unstructured data in Linked Data triples.

Training

Keeping practitioner competencies current and avoiding professional obsolescence is crucial for both professionals and industry. Technical obsolescence is an intrinsic characteristic of IT profession (e.g. Fu, 2011; Thong & Yap, 2000). According to Ferdinand (1966), professional obsolescence refers to professionals whose scope of technical competence does not encompass the reaches of knowledge and technique that inform their discipline. In the IT field, practitioners are facing constant technical skill depletion (Tsai, Compeau & Haggerty, 2007). IT professionals need to update their knowledge and necessary skills to maintain effective performance (Fu, 2010). Unlike other professionals where basic knowledge remains enduring, the half-life of knowledge and skills in the IT profession is estimated at less than two years (Dubin, 1990). In this scenario, keeping these technical competences up-to-date requires a defined long life learning program (LLP). In the case of Linked Data, it is necessary to perform two different actions. On the one hand, the inclusion of Linked Data and Semantic Technologies knowledge areas in both graduate and undergraduate levels and on the other hand the generalization of LLPs devoted to the topic.

5. Conclusions

In a few years, Linked Data will revolutionize the world of data access. Linked Data takes a much more important in the context of reuse of public data. Linked Data obtains its best results if the information is provided in a standard way to encourage reuse. In this paper, we have discussed the importance of reusing public and private data trough Linked Data and the main risks associated with its use and non-use by public and private organizations and individuals. Furthermore, we have discussed the implications of a set of related technologies to Linked Data creating the potential for changing the way information research can be carried out in future.

Acknowledgements

This work is supported by the Spanish Ministry of Science and Innovation under the project "FLORA: Financial Linked Open Data-based Reasoning and Management for Web Science".

References

Agarwal, R., & Ferratt, T. W. (2002). Enduring practices for managing IT professionals. Communications of the ACM, 45(9), 73-79.

Allemang, D., & Hendler. J. (2011). Semantic Web for the Working Ontologist, Second Edition: Effective Modeling in RDFS and OWL. Elsevier.

Álvarez-Rodríguez, J.M., Labra-Gayo, J.E., Cifuentes, F., Alor-Hérnandez, G., Sánchez, C., & Guzmán-Luna, J.A. (2012). Towards a Pan-European E-Procurement platform to Aggregate, Publish and Search Public Procurement Notices powered by Linked Open Data: The MOLDEAS Approach. International Journal of Software Engineering and Knowledge Engineering, In press.

Beard, D., Schwieger, D., & Surendran, K (2010). A Value Chain Approach for Attracting, Educating, and Transitioning Students to the IT Profession. Information Systems Education

Journal, 8(7), 1-12.

Belleau, F., Nolin, M., Tourigny, N., Rigault, P., & Morissette, J. (2008). Bio2RDF: Towards a mashup to build bioinformatics knowledge systems. Journal of Biomedical Informatics, 41(5), 706-16.

Berners-Lee, T. (2006). Linked Data - Design Issues. W3C. Available online at: http://www.w3.org/DesignIssues/LinkedData.html

Berners-Lee, T., Hendler, J., & Lassila, O. (2001). The semantic web. Scientific American, 284(5), 34-43.

Berners-Lee., T. Fielding, R., & Masinter, L. (2005). Uniform Resource Identifier (URI): Generic Syntax. IETF, STD 66, RFC 3986.

Bharadwaj, A.S. (2000). A Resource-Based Perspective on Information Technology Capability and Firm Performance: An Empirical Investigation. MIS Quarterly, 24(1), 169-196.

Bizer, C. (2009). The emerging web of linked data. IEEE Intelligent Systems, 24(5), 87-92.

Bizer, C., & Cyganiak, R. (2006). D2R Server - Publishing Relational Databases on the Semantic Web. Poster at the 5th International Semantic Web Conference (ISWC2006).

Bizer, C., Boncz, P., Brodie, M.L., & Erling, O. (2012). The meaningful use of big data: four perspectives--four challenges. ACM SIGMOD Record, 40(4), 56—60.

Bizer, C., Cyganiak, R., & Gauss, T. (2007). The RDF Book Mashup: From Web APIs to a Web of Data. In Proceedings of the 3rd Workshop on Scripting for the Semantic Web (SFSW2007).

Bizer, C., Heath, T., & Berners-Lee, T. (2009). Linked Data - The Story So Far. International Journal on Semantic Web and Information Systems, 5(3), 1-22.

Brewster, C., & O'Hara, K. (2007). Knowledge representation with Ontologies: present challenges-future possibilities. International Journal of Human-Computer Studies, 65(7), 653-568.

Casado-Lumbreras, C., Colomo-Palacios, R., Gómez-Berbís, J.M., & García-Crespo, Á. (2009). Mentoring programmes: a study of the Spanish software industry. International Journal of Learning and Intellectual Capital, 6(3), 293-302.

Castellanos-Nieves, D., Fernández-Breis, J.T., Valencia-García, R., Martínez-Béjar, R., & Iniesta-Moreno, M. (2011). Semantic Web Technologies for supporting learning assessment. Information Sciences, 181(9), 1517-1537.

Dietrich, J., & Jones, N. (2007). Using Social Networking and Semantic Web Technology in Software Engineering--Use Cases, Patterns, and a Case Study. In Proceedings of the 18th Australian Software Engineering Conference, pp. 129-136.

DuCharme, B. (2011). Learning SPARQL, O'REILLY Media.

Dumbill, E. (2012). What is big data? An introduction to the big data landscape. O'Reilly Radar. Available online at http://radar.oreilly.com/2012/01/what-is-big-data.html

Fazzingaa, B., Gianformeb, G., Gottlobc, G., & Lukasiewiczc, T. (2011). Semantic Web search based on ontological conjunctive queries. Services and Agents on the World Wide Web, 9(4), 453-473.

Ferdinand, T. (1966). On the obsolescence of scientists and engineers. American Scientist, 54 (1), 46-56.

Fu, J.R. (2010). Is Information Technology Career Unique? Exploring Differences in Career Commitment and its Determinants among IT and Non-IT Employees. International Journal of Electronic Business, 8(4), 272—281.

Fu, J.R. (2011). Understanding career commitment of IT professionals: Perspectives of push–pull–mooring framework and investment model. International Journal of Information Management, 31 (3), 279–293.

Goles, T., Hawk, S., & Kaiser, K.M. (2008). Information technology workforce skills: The software and IT services provider perspective, Information Systems Frontiers, 10(2), 179—194.

Gruber, T. (1993). A translation approach to portable ontology specifications. Knowledge Acquisition. 5(2), 199–220.

Hassanzadeh, O., & Consens, M. (2009). Linked Movie Data Base. In Proceedings of the 2nd Workshop on Linked Data on the Web (LDOW2009).

Heath, T. (2008). Information-seeking on the Web with Trusted Social Networks – from Theory to Systems. PhD Thesis, The Open University.

Heath, T., & Motta, E. (2008). Revyu: Linking reviews and ratings into the Web of Data. Journal of Web Semantics, 6(4), 266-273.

Janev, V., & Vranes, S. (2010). Applicability assessment of Semantic Web technologies. Information Processing and Management, 47(4), 507-517.

Jentzsch, A., Hassanzadeh, O., Bizer, C., Andersson, B., & Stephens, S. (2009). Enabling Tailored Therapeutics with Linked Data. In Proceedings of the 2nd Workshop on Linked Data on the Web (LDOW2009).

Joseph, D., Ng, K., Koh, C., & Ang, S. (2007). Turnover of information technology professionals: a narrative review, meta-analytic structural equation modeling, and model development. Management Information Systems Quarterly, 31(3), 547–577.

Kaufmann, E., & Bernstein, A. (2010). Evaluating the usability of natural language query languages and interfaces to Semantic Web knowledge bases. Web Semantics: Science, Services and Agents on the World Wide Web, 8(4), 377-393.

Kiryakov, A., Popov, B., Terziev, I., Manov, D., & Ognyanoff, D. (2004). Semantic annotation, indexing, and retrieval. Web Semantics: Science, Services and Agents on the World Wide Web, 2(1), 49-79.

Klyne, G., & Carroll, J. (2004). Resource Description Framework (RDF): Concepts and Abstract Syntax. W3C Recommendation, Available at: http://www.w3.org/TR/rdf-concepts/

Kobilarov, G., Scott, T., Raimond, Y., Oliver, S., Sizemor, C., Smethurst, M., Bizer, C., & Lee, R. (2009). Media Meets Semantic Web - How the BBC Uses DBpedia and Linked Data to Make Conections. In Proceedings of the 6th European Semantic Web Conference (ESWC2009).

Kuean, W.L., Kaur, S., & Wong, E.S.K. (2010). The relationship between organizational commitment and intention to quit: the Malaysian companies perspective. Journal of Applied Sciences, 10(19), 2251 2260.

Lassila, O., Harmelen-Van, F., Horrocks, I., Hendler, J., & McGuinness, D.L. (2000). The semantic Web and its languages. IEEE Intelligent Systems, 15 (6), 67-73.

López-Fernández, M., Martín-Alcázar, F., & Romero-Fernández, P. M. (2010). Human resource management on social capital. International Journal of Human Capital and Information Technology Professionals, 1(2), 36–48.

Love, P.E.D., Irani, Z., Standing, C., & Themistocleous, M. (2007). Influence of job demands, job control and social support on information systems professionals' psychological well-being. International Journal of Manpower, 28 (6), 513-528.

Madden, S., & van Steen, M. (2012). Guest Editors' Introduction: Internet-Scale Data Management. IEEE Internet Computing, 16(1), 10-12.

Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., & Hung Byers, A. (2011). Big data: The next frontier for innovation, competition, and productivity. McKinsey Global Institute. Available online at: http://www.mckinsey.com/Insights/MGI/Research/Technology_and_Innovation/Big_data_The_next_frontier_for_innovation

O'Riain, S., Harth, A., & Curry, E. (2012). Linked Data Driven Information Systems as an Enabler for Integrating Financial Data. In A.Y. Yap (Eds.), Information Systems for Global Financial Markets: Emerging Developments and Effects. Palo Alto, USA: IGI-Global.

Samwald, M., Jentzsch, A., Bouton, C., Kallesøe, C.S., Willighagen, E., Hajagos, J., Marshall, M.S., Prud'hommeaux, E., Hassanzadeh, O., Pichler, E., & Stephens, S. (2011). Linked open drug data for pharmaceutical research and development. Journal of Cheminformatics, 3(1), 19.

Schwarzkopf, A.B., Mejias, R.J., Jasperson, J., Saunders, C.S., & Gruenwald, H. (2004). Effective practices for IT skills staffing. Communications of the ACM, 47(1), 83—88.

Segaran, T., Evans, C., & Taylor, J. (2009). Programming the Semantic Web, O'REILLY Media.

Shadbolt, N., Hall, W., & Berners-Lee, T. (2006). The Semantic Web Revisited. IEEE Intelligent Systems, 21 (3), 96-101.

Sheth, A., Bertram, C., Avant, D., Hammond, B., Kochut, K., & Warke, Y. (2002). Managing

Semantic Content for the Web. IEEE Internet Computing, 6(4), 80-87.

Thong, J. Y. L. & Yap, C. S. (2000). Information systems and occupational stress: A theoretical framework. Omega, 28(6), 681-692.

Tsai, H., Compeau, D., & Haggerty, N. (2007). Of races to run and battles to be won: technical skill updating, stress, and coping of IT professionals. Human Resource Management, 46(3), 395-409.

Uruena, M., & Larrabeiti, D. (2005). Nested uniform resource identifiers. In Proceedings of the 31st EUROMICRO Conference on Software Engineering and Advanced Applications, pp. 380-385.

Van de Sompel, H., Lagoze, C., Nelson, M., Warner, S., Sanderson, R., & Johnston, P. (2009). Adding eScience Assets to the Data Web. In Proceedings of the 2nd Workshop on Linked Data on the Web (LDOW2009).

Wood, D. (2010). Linking Enterprise Data. New-York, USA: Springer-Verlag.

Zhao, J., Miles, A., Klyne, G. & Shotton, D. (2009). Linked data and provenance in biological data webs. Briefings in Bioinformatics, 10(2), 139-152.

Authors Bio

Ricardo Colomo-Palacios is an Associate Professor at the Computer Science Department of the Universidad Carlos III de Madrid. His research interests include applied research in Information Systems, software project management, people in software projects and social and Semantic Web. He received his PhD in Computer Science from the Universidad Politécnica of Madrid (2005). He also holds a MBA from the Instituto de Empresa (2002). He has been working as Software Engineer, Project Manager and Software Engineering Consultant in several companies including Spanish IT leader INDRA. He is also an Editorial Board Member and Associate Editor for several international journals and conferences and Editor in Chief of International Journal of Human Capital and Information Technology Professionals.

José Luis Sánchez-Cervantes is PhD student in Computer Science in the Universidad Carlos III de Madrid, he obtained his master degree in Computer Systems with Software Engineering specialty and he is received as an Engineer in Computer Systems in the Instituto Tecnológico de Orizaba. His research interests include Semantic Web, Linked Data, Cloud Computing, Collective Intelligence and distributed systems.

Giner Alor-Hernandez is a full-time researcher of the Division of Research and Postgraduate Studies of the Instituto Tecnologico de Orizaba. He received a MSc and a PhD in Computer Science of the Center for Research and Advanced Studies of the National Polytechnic Institute (CINVESTAV), Mexico. He is involved in several Mexican research projects and he has been a committee program member of several international conferences in computer science. His research interests include Web services, e-commerce, Semantic Web, Web 2.0, service-oriented and event-driven architectures and enterprise application integration. He heads the Research Group on Computing and Systems of the Insitituto Tecnologico de Orizaba. He is an IEEE and ACM Member.

Alejandro Rodríguez-González, PhD, is a Researcher of the Computer Science Department at Universidad Carlos III de Madrid. His main research interests are the Semantic Web, Artificial Intelligence and

Biomedical informatics field, with a particular interest on the creation of Medical Diagnosis Systems. He has a degree in Computer Science, a M.Sc in Computer Science and technology in the specialty of Artificial Intelligence and a M.Sc in Engineering Decision systems. He has published several papers in international journals and edited various special issues in high quality journals.