Computers in Industry 61 (2010) 729–741

Contents lists available at ScienceDirect

Computers in Industry

journal homepage: www.elsevier.com/locate/compind

Semantic Web computing in industry

John G. Breslin *, David O'Sullivan, Alexandre Passant, Laurentiu Vasiliu

School of Engineering and Informatics, National University of Ireland, Galway, Ireland

ARTICLE INFO

Article history:
Available online 22 June 2010

Keywords:
Semantic Web
Semantic Web Services
Knowledge Management
E-Commerce
Enterprise Application Integration

ABSTRACT

The Semantic Web has attracted significant attention during the last decade. On the one hand, many research groups have changed their focus towards Semantic Web research and research funding agencies particularly in Europe have explicitly mentioned Semantic Web in their calls for proposals. On the other hand, industry has also begun to watch developments with interest and a number of large companies have started to experiment with Semantic Web technologies to ascertain if these new technologies can be leveraged to add more value for their customers or internally within the company, while there are already several offers of vendors of Semantic Web solutions on the market. The essence of the Semantic Web is to structure Web-based information to make it more interoperable, machine-readable and thereafter to provide a means to relate various information concepts more easily and in a reusable way. The Semantic Web acts as an additional layer on the top of the Web, and is built around explicit representations of information concepts and their relationships such as ontologies and taxonomies. Furthermore, Semantic Web technologies are not only valuable on an open environment like the Web, but also in closed systems such as in industrial settings. Hence, these technologies can be efficiently deployed for domains including Web Services, Enterprise Application Integration, Knowledge Management and E-Commerce, fulfilling existing gaps in current applications. This paper focuses on this synthesis between Semantic Web technologies and systems problems within industrial applications. There will be a short review of Semantic Web standards, languages and technologies followed by a more detailed review of applications of Semantic Web computing in industry. The paper covers theoretical considerations as well as use cases and experience reports on the topic, and we also present some current challenges and opportunities in the domain.

1. Introduction

The Semantic Web is now becoming a well-established branch of computer science and software engineering with its own standards, languages, technologies and applications. It is also a foundation for what is termed 'Web Science', where the Web itself is the object of a dedicated science of its own when it is deployed in a wide range of domains.[1] There are a number of research institutes now feeding new knowledge into the associated research community, and a large number of new and existing industries are deploying Semantic Web techniques to provide goods and services to customers. While the World Wide Web and its associated technologies and applications have become a 'disruptive technology' over a relatively short period of time, it remains to be seen whether the Semantic Web with its related new technologies and applications will do the same. There are nevertheless some encouraging indications. The number of new business start-ups that now deploy Semantic Web technologies has become noticeable. Web 2.0 companies such as Freebase, Faviki and Zemanta have embraced Semantic Web technologies. The New York Times also identified commercial industries around the world that are using Semantic Web technologies as part of their core business offerings to customers [1]. Giants such as Oracle, Vodafone, Amazon, Adobe, Microsoft, Yahoo and Google are now experimenting with Semantic Web technologies to provide new value to customers [2], with some recent efforts including the Yahoo! SearchMonkey search engine [3] and Google's support in indexing structured RDF information from the Web.[2]

The core of the Semantic Web contains a number of fundamental formal models, languages and technologies for interoperability and reuse of information, including RDF, RDFS, the OWL family of languages, the WSML family of languages and SPARQL. Semantic Web Services build on the Semantic Web and previous work regarding Web Services to power semi-automated or automated interoperable applications. In this paper, we will

* Corresponding author.
E-mail address: John.Breslin@nuigalway.ie (J.G. Breslin).

[1] http://webscience.org.

[2] http://googlewebmastercentral.blogspot.com/2009/05/introducing-rich-snippets.html.

0166-3615/$ – see front matter
doi:10.1016/j.compind.2010.05.002

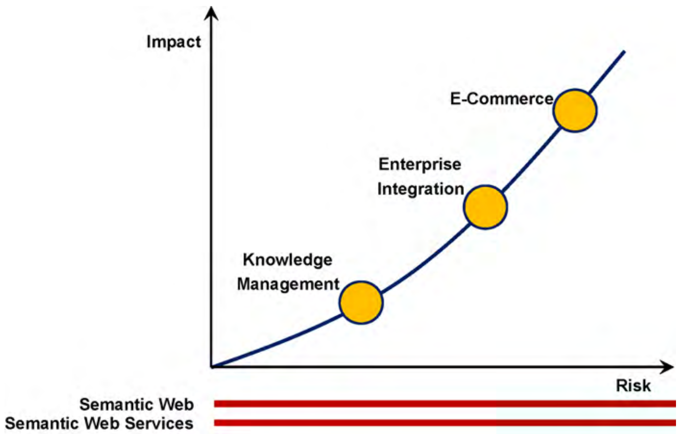

Fig. 1. Industrial applications for the Semantic Web and Semantic Web Services.

describe the Semantic Web and Semantic Web Services technologies which act as foundational layers for a variety of semantic industrial applications. We will detail three key areas for applications: Knowledge Management, Enterprise Application Integration and E-Commerce. These three application areas are shown in Fig. 1, where the *Y*-axis represents the reward or potential impact for semantics within industry, and the *X*-axis represents the risk involved in implementing change to existing technologies. For example, deploying semantics in the field of Knowledge Management may not yield the same rewards as applications in e-commerce, but the risks involved in successful e-commerce deployment are much higher. Each of these three major application domains is built upon two major areas of research—the Semantic Web and Semantic Web Services. We begin our review by looking at these two areas and then looking at each of the major application domains in turn.

In the following section, we will introduce the main directions of development within the Semantic Web along with their main technologies, tools and achievements to date, as well as describing various limitations and possible future developments.

2. The Semantic Web

The 'Semantic Web' can be thought of as the next generation of the Web where computers that can aid humans with their daily web-related tasks as more meaningful structured information is added to the Web (manually and automatically) [4]. For example, using a combination of facts like "John works_at NUI Galway", "Mary knows John", "a Person works_at an Organisation", and "a Person knows a Person", you can allow computers to answer relatively straightforward questions like "Find me all the people who know others who work at NUI Galway" which at this moment is quite difficult to do without significant manual processing of the information returned from search results. The Semantic Web represents these facts through the use of metadata that is associated with Web resources, and behind this metadata there are specific vocabularies or 'ontologies' [5] that describe what are the semantics (or meaning) of this metadata and how it is all related to each other.

Metadata can be thought of as 'data about data'. Similar to how librarians traditionally put information about books into catalogues or library cards, metadata on the Web commonly refers to descriptive information about Web resources that can support a wide range of operations [6] ranging from retrieving to re-contextualising content. Metadata elements are used to give structure to the description of a resource. For example in an educational course, metadata elements will include title, description, keywords, author, educational level, version, location, language, date created, and so on. RDF (Resource Description Framework) is used to express metadata about resources [7] while these resources are defined using URIs (Uniform Resource Identifier) such that they are provided with unique and non-ambiguous identifiers at Web-scale, enabling interoperability between various applications. Led by the W3C consortium, RDF is supported by a wide range of stakeholders ranging from digital librarians to B2B industries and has achieved significant industrial momentum.

RDF consists of two aspects: a graph-based abstract model for the data (made up of multiple statements, or triples) and the RDF syntax (with a variety of serialisations to represent these triples in a computer-readable form such as N3, Turtle, RDF/XML or RDFa which allows RDF annotations to be directly embedded within XHTML pages). For example, to say that Alice knows Bob, we could use the Notation3 (N3) syntax for the corresponding RDF triple: "http://example.org#Alice http://xmlns.com/foaf/0.1/knows http://example.org/#Bob". All triples are in the form of a directed graph, from subject via a directed arc (the predicate) to an object. In the previous example, Alice would be the subject, the 'knows' relationship is the predicate and Bob is the object. URIs are normally used to give identifiers to the subject, predicate and object, but the object may sometimes be a literal or text string if an attribute is to be assigned to a subject, e.g. "http://example.org#Alice http://xmlns.com/foaf/0.1/name 'Alice Cooper'". A sample RDF graph model is shown in Fig. 2.

Further structure is provided by a metadata schema or ontology (e.g. as shown in the bottom layer of Fig. 2). For example, if there is metadata about a soccer team, an underlying ontology will say that a soccer team always has a goalkeeper and always has one and only one manager, so each metadata entry for a soccer team should have that information. Ontologies are formal and consensual specifications of conceptualisations that provide a shared and common understanding of a domain [5]. In order to deploy ontologies on the Web, two languages have been put forward as standard proposals by the W3C, namely RDFS and OWL. RDF schema (RDFS) is commonly used for the definition of RDF ontologies (and written in RDF) on the Semantic Web [8]. Some of the more popular Semantic Web lightweight ontologies include FOAF (Friend-of-a-Friend, for social networks) [9], Dublin Core (for resources online or in libraries), SIOC (Semantically-Interlinked Online Communities, for online communities and content) [8], and the Geo vocabulary[3] (for geographic locations). Recently, Bizer et al. [10] provided a list of popular and core vocabularies that people should use when publishing data on the Semantic Web as well as some best practices for publishing RDF data on the Web.

While popular, RDFS is somewhat limited in various regards. In order to overcome some of the limits of RDFS, ontology developers can use OWL (the Web Ontology Language) [11] (currently being revised towards OWL2)[4] to define more precise axioms within their ontologies, for example, transitivity of some properties (e.g. in an "ancestor" property), symmetry (e.g. "sibling") or cardinality constraints (such as the "has one and only one manager" in the previous example). In addition, ontologies also act as a support for reasoning systems, both to derive new facts or to check the consistency of the model. OWL provides three increasingly expressive sublanguages designed for use by specific communities of implementers and users: OWL Lite, OWL DL and OWL Full. OWL Lite supports those users primarily needing a classification hierarchy with simple constraints. OWL DL supports those users who want the maximum expressiveness while retaining computational completeness, while OWL Full is meant for users who want

[3] http://www.w3.org/2003/01/geo/.

[4] http://www.w3.org/2007/OWL/wiki/OWL_Working_Group.