Semantic Web in Action
Abstract

The Semantic Web is the extension of the World Wide Web that enables people to share content beyond the boundaries of applications and websites [10]. Although it is a relatively new concept, the first time the term was coined and printed in a fairly accessible place was in the year 2001 [20]. There are already many successful industrial applications that use Semantic Web technologies. This work introduces some basic concepts relative to semantic web and presents interesting examples of the advantages of incorporating these technologies in different areas such as industry and commerce.

1 Introduction

The web content of today has been designed for humans to read and understand. Computer programs cannot interpret and manipulate web content meaningfully [20]. Computers can parse pages and find titles, links, figures, tables, etc. but there is no reliable way to understand semantics. The web is then a collection of documents for people to consume rather than a set of data that computer software can process automatically. The objective of Semantic Web technologies is then to help in the construction of a Web that can be processed by both, humans and machines [23, 31, 24].

The goal of this article is to briefly introduce some ideas about Semantic Web as well as to present some use cases. Section 2 introduces two technologies used by the Semantic Web. Then, Section 3 defines the concept of ontology, one of the pillars of the Semantic Web. After that, Section 4 presents Semantic Web resources defined for E-Commerce. Some use cases related to E-Commerce are presented in Section 5. Afterwards, in Section 6, use cases related to other fields are described. Finally, conclusions are stated in Section 7.
There are two main technologies that help in the development of the Semantic Web: eXtensible Markup Language (XML) and the Resource Description Framework (RDF).

People can create their own tags using XML. Then, scripts or programs can use these tags in many different ways. However, the programmer (the writer of the script or program) has to know the exact meaning of these tags. In short, XML allows users to create arbitrary structures in documents but says nothing about their meaning.

Meaning is expressed by RDF, which encodes it in sets of triples. Each triple is composed of a subject, a predicate and an object, corresponding to subject, verb and object of an elementary sentence. These triples can be written using different syntaxes, one of them is RDF/XML, which uses XML tags. A typical graphical representation of a triple is shown in Figure 1. In RDF, a document makes assertions that certain things (people, Web pages, etc.) have some properties (such as “is the father of,” “is the author of”) with certain values (another person, another Web page). With this structure it is possible to describe most of the data that can be processed by computer programs. Each part of the triples, subject, property and object, must be identified by a Universal Resource Identifier (URI), just like a link on a Web page (URLs, Uniform Resource Locators, are the most common type of URI.) This allows people to create new concepts just by defining a new URI for it somewhere on the Web.

It is very frequent for humans to use a same term to reference different things. This is very rarely a problem since it is easy for people to understand the intended meaning. Computer programs are different and cannot understand these differences. This task, that is natural for a human, is difficult or sometimes even impossible for a machine, unless it is hard-coded. Using different URIs for different concepts solves the problem.

RDF triples form a Web of content. This content is made available from different systems. Equivalent concepts in this Web of content are identified thanks to the use of URIs.

Other important components of the Semantic Web are ontologies. In 1993, Gruber originally defined the notion of an ontology as an “explicit specification of a conceptualization” [26]. The most typical kind of ontology for the Web has a taxonomy and a set of inference rules.

The taxonomy defines classes of objects and relations among them. Classes, subclasses and relations
among entities are a very powerful tool for Web use. We can express a large number of relations among entities by assigning properties to classes and allowing subclasses to inherit such properties.

Inference rules in ontologies supply further power. Reasoning support is important because it allows to check the consistency of ontologies and the knowledge to check for unintended relationships between classes and automatically classify instances in classes. Automated reasoning support allows one to check more cases than could be checked manually. Checks like the preceding ones are valuable for designing large ontologies, where multiple authors are involved, and for integrating and sharing ontologies from various sources [17]. The computer doesn’t truly “understand” any of this information, but it can now manipulate the terms much more effectively in ways that are useful and meaningful to the human user.

With ontology pages on the Web, solutions to terminology (and other) problems begin to emerge. The meaning of terms or XML codes used on a Web page can be defined by pointers from the page to an ontology. Of course, the same problems as before can arise if one person points to an ontology that defines addresses as containing, for example, a zip code and another person points to one that uses a postal code instead. This kind of confusion can be resolved if ontologies (or other Web services) provide equivalent relations: one or both of the ontologies used may contain information that the zip code is equivalent to the postal code.

Ontologies can enhance the functioning of the Web in many ways, they can be used in a simple fashion to improve the accuracy of Web searches. The search program can look for only those pages that refer to a precise concept instead of all the ones using ambiguous keywords. More advanced applications will use ontologies to relate the information on a page to the associated knowledge structures and inference rules.

Ontologies enable knowledge sharing and reuse. There are many different ontologies already defined out there in the Web, depending on the application it can be useful to use an already existent ontology. Sometimes, however, it is necessary to modify one of those available ontologies or to create a new one, more adapted to the concerned domain. For example, in Section 6.1 we shall see that the NASA and FAA have defined 200 ontologies to model the different data they work with. Another example can be seen in the next section, where an ontology specially defined for E-Commerce is mentioned.

4 E-Commerce Resources

If you have a Web site where you offer certain products and services, only 5% of all potential visitors will actually have access to these offers in their original beauty. 95% will never get beyond a reduced preview of your products and services as provided by a Web search engine [27]. The Semantic Web aims to help you to provide more visibility to your company. With this aim in mind, the GoodRelations ontology and the Facebook Open Graph Protocol have been defined.
4.1 GoodRelations Ontology

GoodRelations is an ontology for product, price, store, and company data that can be embedded into existing static and dynamic Web pages in a way accessible to search engines, mobile applications, and browser extensions. By adding some extra code to your Web content, you make sure that potential customers can reach your products and services because their computers can extract and present this information with ease. This increases the visibility of your products and services in the latest generation of search engines, recommender systems, and other novel applications [28].

Figure 2 shows an extract of GoodRelations Ontology related to offers made by companies. In that ontology an offer is related to a company and it has its name and description. Besides, a price is also related to the offer. This price have a value and a currency.

![Figure 2: Extract from GoodRelations Ontology](image)

There are several tools that can help add GoodRelations snippets into Web pages [29]. *GoodRelations Snippet Generator* creates simple snippets of additional HTML/RDFa\(^1\) markup for copy-and-paste. There are some extensions for Shop Software that add GoodRelations to shop applications with a few mouse-clicks: *Magento, Joomla/Virtuemart, Prestashop, xtCommerce*. Besides, very powerful HTML templates, such as *Django-/Jinja-style HTML templates*, are used for adding GoodRelations data to Web applications. There is also a module that adds Good Relations markup on *Drupal Commerce* product displays, allowing search engines to display Rich Snippets for products [11].

GoodRelations is being used by 10.000 small and large shops world-wide. Some prominent users are [28]:

**Google** officially recommends GoodRelations for sending structured information for Google Rich Snippets to Google (since 11/2010).

---

\(^1\)A RDF syntax used for embedding RDF subject-predicate-object expressions within XHTML documents.
Yahoo officially recommends GoodRelations for sending structured information for their SearchMonkey feature (since 10/2008).

Best Buy is using GoodRelations as a fundamental part of their digital marketing strategy and publishes full catalog, store, and special offers with GoodRelations on their production Web sites.

O’Reilly is using GoodRelations for Semantic SEO\(^2\) of all of their book titles.

Volkswagen is using GoodRelations for exposing car feature and car component information at massive scale, in its United Kingdom site.

Renault is using GoodRelations for Semantic SEO for their merchandise shop, also in the United Kingdom.

OpenLink Software is using GoodRelations as the fundamental vocabulary for E-Commerce technology based on Virtuoso and other products.

Peek & Cloppenburg is using GoodRelations for publishing information on all European stores plus the brands available in each one of them.

CSN Stores is using GoodRelations for Semantic SEO of all of their 2,000,000 item pages and substores.

Arzneimittel.de (one of Germany’s leading mail order pharmacies) is using GoodRelations in RDFa\(^3\) in all of their catalogs of 250,000 item pages.

Rakuten.de (a leading German online mall) aggregates approx. 6500 merchants with more than 16 million item pages.

4.2 Facebook Open Graph Protocol

Open Graph is a protocol \([2]\) provided by the Facebook platform which allows web pages to integrate with Facebook social graph in such a way that the pages are considered graph objects. In particular, Facebook provides developers with a SDK (Software Development Kit) for iOS (iPhone OS), Android, and the web allowing Open Graph to be put into practice. Thus, by means of integrating meta-information in an app (e.g., web sites, mobile applications) and establishing the relation between this one and Facebook, visitors of a web site can publish, in the social network, pages and actions carried out from the site. That is to say, once a relation between a page, Facebook and a page visitor is established, it is possible that the visitor browsing in the page carries out actions through Open Graph according to her or his interests, such as “Like” (a product or page), “Read” or “Wants to Read”

\(^2\)Search Engine Optimization

\(^3\)A RDF syntax used for embedding RDF subject-predicate-object expressions within XHTML documents.
Figure 3: Open Graph action buttons

(a book), etc. These actions will have direct effect on Facebook when publishing in the social network the page and the action result.

Open Graph is based on RDFa which allows to include additional tags in a page and then to add a button (“Like”, “Rate”, etc) with a determined action which will be available to the user. In particular, these actions can be associated to a page, a particular product, an article, etc. When the button (see Figure 3) is selected by the visitor, the app asks her or his authentication in Facebook in order to establish the connection; and thus, publish the result on the social network.

Many companies use Open Graph on their websites in order to propose a more social space and increase the visits. Some examples of these companies are described below. The interested reader can find further examples in [1].

TripAdvisor was one of the first to adopt Open Graph by integrating its site with Facebook in order to increase social sharing capabilities and its portfolio of travel-related apps. This strategy led to the following results [22]. More than double of new users were acquired in 2012, by implementing Open Graph, increasing the number of marketing members to 44 million year to year. From Open Graph it was possible to detect that Facebook users have 27% higher participation on TripAdvisor than those not connected to Facebook and that over 1/3 of new TripAdvisor reviews are written by users of this social network. Consequently, this highlights that Open Graph is a medium which allows to capture users of Facebook in order to increase visits and activities in a site.

Skyscanner, a global travel search site, has also incorporated Open Graph to its site. Inquiries accomplished by this company [6] revealed that 52% of respondents were inspired by Facebook friends’ holidays pictures to select the same holiday destinations. Besides, 42% of respondents said that this social network encourages them to visit their friends abroad and 46% of people stated that they either organized or were invited to a trip through Facebook. Thus, by incorporating Open Graph in its site, Skyscanner allows visitors to recommend or share searching results or bookings, promoting then the activity on its site.

The Guardian, the well-known newspaper, aimed to reach a new audience on Facebook, to increase participation on its site and ultimately to grow its base of readers through Open Graph [7]. Thus, the newspaper was integrated with Facebook through Open Graph in 2011. By incorporating Open Graph actions such as “agree” and “disagree” [3], The Guardian has grown to over 3.9 million monthly active users, over half of which are under the age of 25, an audience
traditionally hard to reach for this newspaper. In this way, Open Graph has demonstrated to be a strategic resource for the mentioned newspaper.

5 E-Commerce Use Cases

In this section, some real use cases related to E-Commerce are introduced.

5.1 Best Buy

The first idea to include Semantic Web technologies into the Best Buy site was to promote machine discovery of information without sacrificing the human readability component. Best Buy started adding RDFa to its web pages in 2008. The objective was to get more visibility for its stores on the Web. The solution chosen by Best Buy consisted of giving each store a blog. Employees then started sharing information through these blogs on a daily basis, using online forms that output RDFa. This made input from store employees more visible on the Web.

As a result, search traffic to these pages increased by 30% [14]. Best Buy hadn’t expected to see a SEO benefit, but it has resulted in an advantage to the company since the company relies on search engines for product discovery and store locations. Myers, who works for Best Buy, said [15], “We saw a pretty big SEO jump when we started using semantic descriptions, way more than we could get using clever keyword strategies.”

To improve machine discovery of product and store information, Best Buy started using GoodRelations. In this way, Best Buy opened up its product catalog to humans and machines. “I’m personally very much into making data more contextual and more available - and if it raises us in search engine results or ranking on pages, so be it,” said Myers. He’s hearing a lot from outside developers interested in parsing the data and seeing what they can do with it. “The first step was exposing the data and I’m excited and enthused to see what will come out of it,” he said.

Best Buy has also used other ontologies. For example, to encode its music product information with RDFa Best Buy used not only GoodRelations but also the Music Ontology, Dublin Core and Google Rich Snippet Breadcrumbs vocabulary, in addition to the Facebook Open Graph Protocol.

All this work aimed to help machines get valuable product data directly through the browser. Best Buy plans to extend the work done over music product information to other type of products, using relevant ontologies for the type of products concerned.

Best Buy has also implemented a semantically-driven “Like for Like” API. This API helps customers find the most similar products to a product chosen by him/her. The results are ordered according to the number of product attributes that match.

The following example has been taken from [13]. Let’s say, for instance, that the customer wants to identify other laptops similar to the fine Toshiba C855D-S5100 with a 15.6” display, 4GB RAM and
320GB hard drive (and 41 other important product attributes). Querying the like for like endpoint (using content negotiation) at http://metis.bbyopen.com/product/lf1/sku/7334096 the customer would get the following products (smaller text sample provided for brevity):

- Toshiba - Satellite 15.6” Laptop - 4GB Memory - 320GB Hard Drive - Satin Black, SKU 7704082, price $279.99, 44 matching attributes to the control
- Toshiba - Satellite 15.6” Laptop - 4GB Memory - 500GB Hard Drive - Satin Black, SKU 7337066, price $369.99, 36 matching attributes to the control
- Toshiba - Satellite 15.6” Laptop - 4GB Memory - 320GB Hard Drive - Satin Black, SKU 6927628, price $376.98, 35 matching attributes to the control
- HP - 15.6” Laptop - 4GB Memory - 320GB Hard Drive - Black Licorice, SKU 7594164, price $349.99, 32 matching attributes to the control

In conclusion, Like for Like API helps customers to find the most similar products using available semantic data.

5.2 Volkswagen

Volkswagen also decided to introduce Semantic Web technologies in its United Kingdom site [16]. In this case, the idea behind this decision was to provide a new site search and browse engine. This enables people to get the information they want, quicker. Evidently, getting the information and content quicker helps people make decisions more quickly and increases the rate of conversion.

Another advantage of the Semantic Web is the ability to promote and re-purpose content in different contexts as well as the facility to suggest and find relations between things that were not directly questioned.

Volkswagen has decided to add RDFa, using relevant ontologies, for enriching its HTML pages. The already existent vocabularies that were found useful for this use case are: GoodRelations, Dublin Core, Vehicle Sales Ontology and Ontology for Innovation. Besides, new vocabularies were created where terms did not exist. These new vocabularies were procured from Prof. Dr. Martin Hepp, author of the popular GoodRelations vocabulary and Vehicle Sales Ontology. The vocabularies created consisted of a generic, industry wide Car Options Ontology to describe models, trims, derivatives, components and component compatibility along with the Volkswagen Vehicles Ontology to describe Volkswagen specific concepts. The former extends both the Good Relations vocabulary and the Vehicle Sales Ontology, whilst the latter extends the former.

One of the key advantages provided by the system is the degree of expressivity open to application developers. Whereas previous searches were syntactic, based on keywords and phrases, across unstructured
and meaningless content (in the eyes of a search engine), it has now moved to a model of semantics where meaning and aggregation can be derived and applied. By combining the expressivity and richness provided by semantics and structured data, Volkswagen can start to apply semantics to unstructured content and data, enriching its knowledge base.

It is possible to show the different abilities of the system through some examples [16]. For example, we can ask the system: “Find me all the derivatives priced between x and y that have an engine power greater than z and come in red.”

Additionally by using ontology and knowledge representation it is possible to search for data across a range of contexts, combining different sets of content in different contexts, for example: “Find me all the derivatives priced between x and y that have an engine power greater than z and come in red and have a user review greater that 4 stars.”

Finally by harnessing the features of SPARQL 1.1\(^4\) federated extensions it is easy to cross domains, including data from other 3rd party data sources in a single expression, example: “Find me all the derivatives priced between x and y that have an engine power greater than z and come in red and have a user review greater that 4 stars. Include results not only from VW, but also other car manufacturers, external car review sites and second hand car dealers.”

SPARQL provides a standardized web interface for content, so it is not necessary for the developer to write a new publishing code every time the organization wants to share data with 3rd parties. Furthermore, system administrators don’t have to maintain multiple architectures or systems lending to a separation of concerns.

In summary, the usage of Semantic Web technology provided Volkswagen a standardized interface to data and content, it helped on the separation of concerns between information and application, both logically and physically, and increased value, reusability and accessibility of data.

5.3 Trust You

TrustYou is an online reputation management company which provides services and tools to collect reviews across the Internet and compile sentiments and trends in order to understand what is being said about certain hotels, restaurants, car dealers, etc. Thus, TrustYou provides reputation management tools which allow to analyze reviews, tweets and posts across the social web and make sense of big social data. These tools are based on semantic technology and software as a service (SaaS). Sentiments and opinions are captured and analyzed from millions of comments coming from different sources such as TripAdvisor, Facebook, Google+, Twitter and Yelp, and considering 23 different languages (English, German, French, Spanish, Chinese, Russian, Portuguese, etc.).

---

\(^4\)SPARQL (pronounced “sparkle”, a recursive acronym for SPARQL Protocol and RDF Query Language) is an RDF query language, that is, a query language for databases, able to retrieve and manipulate data stored in Resource Description Framework format.
TrustYou\textsuperscript{5} provides the followings products (see Figure 4).

- **Reputation Surveys**: by means of free online survey tools and Facebook review applications, customers’ insights and feedbacks are taken into account and listened to.

- **Reputation Monitoring**: from different sources (TripAdvisor, Facebook, Google+, etc), positive and negative trends are identified and analyzed.

- **Reputation Marketing**: business reputation is evaluated with the TrustScore, a consumer trusted score which represents customers’ online opinions, in order to improve business strategies and convince customers with trusted reviews.

According to [8],

- “Social media directly influences more than 83\% of all online bookings, and 49\% of guests won’t even consider booking a hotel that doesn’t have reviews.”

- “Guests want to know that they booked the right hotel, so they are willing to pay more for hotels with higher scores.”

- “Sites with trusted reviews and scores integrated on their website are crawled up to 200\% more frequently.”

- “92\% of all users trust reviews of their friends […]. Having reviews on the company’s own site encourages customers to book/purchase there rather than on a competing site.”

\textsuperscript{5}http://www.trustyou.com
TrustYou already has more than 10,000 clients in 60 countries and monitors over 400,000 hotels. The company’s clients include important brands such as Accor, Marriott, Starwood, Hard Rock Cafes, Best Western Hotels, Fairmont and Trump Hotel Collection [9]. Results of a case study are presented below. The interested reader can find other case studies at http://www.trustyou.com/case-studies.

Ridges Hotels & Resorts
Ridges is a family of 47 hotels and resorts in Australia, New Zealand and London. This company was an early adopter of TrustYour technology. Using TrustYou automatic surveys, the company gained over 4,000 additional surveys in 4 months, survey scores rose nearly 9% in 30 days and the company’s overall review scores increased over 2% at the same time. In particular, the average number of reviews on TripAdvisor increased over 20% per month.

6 Use Cases in Other Fields
In this section, some other industrial use cases are presented. Even though they are not directly related to E-Commerce, the benefits obtained can be considered relevant for any organization or company.

6.1 NASA and FAA

NASA (National Aeronautics and Space Administration) and FAA (Federal Aviation Administration) have resorted to semantic web concepts and technologies in order to develop practical application systems which allow to manage information coming from different sources in a centralized way [18, 19].

Occasionally, controllers in an airport are required to suspend landings and take-offs during a period of time, for example due to bad weather, which entails re-routing or rescheduling of flights. Consequently, this causes flight delays, flight connections missed, etc. Ideally, this kind of situation should be efficiently addressed so that flight delays and missed connections are minimized and the additional financial costs are minimal. In order to better solve problems like this, NASA in collaboration with FAA have implemented systems based on semantic web technology so as to improve activities carried out in different missions (Exploration, Space, Science and Aeronautics). Some of these systems are presented below.

SWIM - System Wide Information Management
Scientists and engineers need to understand data coming from different sources which are created through several NASA technologies and projects (see Figure 5(a)). SWIM aims to integrate information in the NAS (National Airspace System) domain where different kinds of information (airports location, weather, air-traffic, security, etc.) come from a wide variety of sources, different kinds of systems provide information

---

6http://www.hospitalitynet.org/organization/17016016.html accessed on 22/08/2013
and access to this data (security systems, enroute systems, traffic flow management systems, etc.) and support tools also require the mentioned information.

Figure 5: Airspace information integration

This project is based on using ontologies and marking up aviation related data by using languages based on XML such as AIXM (Aeronautical Information Exchange Model) and the NIXL (NAS Information Exchange Language).

**SemanticOrganizer**

SemanticOrnagizer (SO) is a collaborative knowledge management system that supports distributed NASA teams, which consists of a customizable and semantically structured repository and provides a common access point for information. SO users can upload, store and query information in different formats into the repository. Besides SO maintains data items, relations between them and relations among items, people and groups. SO is based on shared ontologies and uses over half a million of RDF triplets. In particular, SO being a NASA semantic web application, it has over 500 users and over 45,000 information nodes that are connected by over 150,000 links.

**NASA Taxonomy**

The NASA taxonomy aims to make access to information easier by helping NASA scientists and engineers find desired data. This taxonomy reflects NASA organisations and divisions, by means of categories such as engineering and scientific disciplines, NASA locations, etc., and is used for intelligent browsing and navigation. In particular, Dublin Core metadata specifications and XML Scheme are used in the mentioned

---

7Taken from the slide presentation by Naveen Ashish. *NASA and Semantic Web*. Research Institute for Advanced Computer Science. NASA Ames Research Center
SWEET - Semantic Web for Earth and Environmental Technologies

SWEET [30] establishes a set of 200 ontologies written in the OWL ontology language defining 6000 concepts for earth science information sources, earth science data and earth science subject domains of interest in order to provide information access in a more effective and intuitive way.

6.2 SIRI

Another interesting use case is Siri, the Virtual Personal Assistant acquired by Apple in 2010. Siri has its roots in the DARPA CALO project (“Cognitive Agent that Learns and Organizes”) which was led by SRI. The goal of CALO was to develop AI technologies (dialog and natural language understanding, machine learning, evidential and probabilistic reasoning, ontology and knowledge representation, planning, reasoning, service delegation) all integrated into a virtual assistant that helps people do things [12].

It pushed the limits on machine learning and speech, and also showed the technical feasibility of a task-focused virtual assistant that uses knowledge of user context and multiple sources to help solve problems.

Siri is integrating, commercializing, scaling, and applying these technologies to a consumer-focused virtual assistant. Siri was under development for several years during and after the CALO project at SRI. It was designed as an independent architecture, tightly integrating the best ideas from CALO but free from the constraints of a national distributed research project.

Siri wasn’t built using Semantic Web open-standards (like RDF, etc.) However, it connects to partners on the web using structured APIs, some of which do use the Semantic Web standards. A site that exposes RDF usually has an API that is easy to deal with. For instance, Siri uses geonames.org as one of its geospatial information sources. It is a full-on Semantic Web endpoint.

Siri’s knowledge is represented in a unified modeling system that combines ontologies, inference networks, pattern matching agents, dictionaries, and dialog models. As much as possible things are represented declaratively (i.e., as data in models, not lines of code). This is a tried and true best practice for complex AI systems. This makes the whole system more robust and scalable, and the development process more agile. It also helps with reasoning and learning, since Siri can look at what it knows and think about similarities and generalizations at a semantic level.

Siri isn’t a source of data, so it doesn’t expose data using Semantic Web standards. In the Semantic

---

9The Web Ontology Language is a family of knowledge representation languages for authoring ontologies. The languages are characterized by formal semantics and RDF/XML-based serializations for the Semantic Web.
10The Defense Advanced Research Projects Agency (DARPA) is an agency of the United States Department of Defense responsible for the development of new technologies for use by the military.
11http://www.sri.com/
Web ecosystem, it is doing something like the vision of a semantic desktop - an intelligent interface that knows about user needs and sources of information to meet those needs, and intermediates.

Siri does, however, benefit greatly from standards for talking about space and time, identity (of people, places, and things), and authentication.

All players in the E-Commerce value chain (customer, producer, distributor, advertiser, etc.) profit when the information can be unambiguously identified, compared, and combined. Legitimate service providers on the supply end of the value chain also benefit, because structured data is harder to scan than text. So if some service provider offers a multi-criteria decision making service, say, to help make a product purchase in some domain, it is much easier to do fraud detection when the product instances, features, prices, and transaction availability information are all structured data.

6.3 BBC - British Broadcasting Corporation

BBC is a public service broadcaster which publishes large amounts of content online (text, audio and video). Historically, the website has focused on supporting broadcast brands and a series of domain-specific sites by providing separate and standalone HTML sites which were not related to each other; and consequently, the lack of data integration represented a difficult navigation.

Thereby, BBC incorporated semantic web technology in order to interpret information and integrate data of different domains and sources. In particular, in line with this goal, the following BBC ontologies\textsuperscript{12} were defined.

- **Programmes Ontology** provides a vocabulary and describes how brands, series, episodes, particular versions of episodes and broadcasts interact with each other.

- **Wildlife Ontology** is a vocabulary for describing biological species and related taxa (taxonomic groups or entities). Thus this ontology defines terms for describing the names and ranking of taxa, as well as providing support for describing their habitats, conservation status, and behavioral characteristics, etc. In particular, it is used in the BBC Wildlife Finder application\textsuperscript{13}.

- **Sport Ontology** is an ontology for publishing data about competitive sports events. In particular, it was used to deliver the BBC London 2012 website and is used in the BBC Sport online section.

- **The News Storyline Ontology** is a generic model for describing and organising news. The ontology is intended to be flexible so as to support any given news or media publisher’s approach to handling news stories.

\textsuperscript{12}http://www.bbc.co.uk/ontologies/
\textsuperscript{13}http://www.bbc.co.uk/nature/wildlife
• **The Corenews Ontology** describes the basic concepts that journalists can use. These concepts can be tangible entities (People, Places, Organizations), intangible entities (Themes), or compound entities (Events).

• **Curriculum Ontology** provides a data model and vocabularies (e.g., topic, field of study and programmes of study, etc.) for describing the National Curricula within the UK.

These ontologies are used on the BBC website in order to provide semantic support to BBC sections such as BBC Programmes, BBC Sport, BBC Wildlife Finder, etc.

**BBC World Cup 2010 and Olympics 2012**

One of the first BBC forays into semantic web was the BBC World Cup 2010 website. BBC utilized metadata and linked data in order to organize and manage dynamically the content published on the BBC World Cup 2010 site [4, 5].

The goal was to design a website with a page for each team, group and player as well as results, fixtures, etc. Considering that the World Cup 2010 consisted of 32 teams, 8 groups and 736 players, it was necessary to create over 750 pages of dynamic content. However, to manage this amount of pages maintaining updated information would have required hundreds of journalists. Thereby, a high-performance dynamic semantic publishing (DSP) framework was used in order to facilitate the publication of automated metadata-driven web pages, requiring minimal journalistic management. This framework published metadata describing the world cup content. By consulting this metadata, dynamic page aggregations for groups, teams, players, etc. can be created. These dynamic aggregations are based on an ontology domain model where an ontology describes groups and relations between the concepts. This ontology also allows the description of stories, blogs, profiles, images, etc. established by a journalist and the association of them to concepts within the domain model. Journalists use a web tool called Graffiti for selective/manual association of concepts to content (tagging). Besides manual tagging, journalist content is automatically analyzed against the mentioned ontology by means of a natural language and ontological determiner process, which automatically extracts World Cup concepts embedded within a textual representation of a story. In this way, a journalist’s story results in published metadata which is captured and made persistent in a RDF representation for querying through SPARQL.

Thus, the World Cup 2010 site consisted of over 800 dynamically generated pages, serving an average of 2 million page requests per day and an average of 1 million SPARQL queries per day, accepting hundreds of repository updates/inserts per minute.

The success of the the World Cup 2010 site led to apply the same approach to BBC Olympics 2012 website which consisted of over 10,000 athlete pages, over 200 country pages, 500 event and discipline pages and dozens of venue pages.

From these good experiences, BBC continued incorporating semantic web technology to its websites.
7 Conclusions

The Semantic web, a recent research area, arose about a decade ago and has demonstrated to be a rich study field for industry and commerce.

In this paper, we introduced basic notions relative to the Semantic Web along with resources provided by this approach which can be used for the benefit of industry and commerce. Besides, we presented concrete use cases applied to E-Commerce and other areas, which highlight some important advantages that can be obtained from the Semantic Web. In particular, the obtained benefits include standardized interfaces to data and content, separation of concerns between information and applications, manipulation of large numbers of data and/or systems with ease, between others. Consequently, this improves users searching experience, the visibility of the enterprise and its products on the Web and it could, eventually, implicate better rankings from search engines.

This article proves that, even though Semantic Web technologies are relatively new, there are already many enterprises that are investing in them, positioning thus the Semantic Web as one of the main strongholds of the new technologies.

References


