



# Journal of Applied Sciences

ISSN 1812-5654

**science**  
alert

**ANSI***net*  
an open access publisher  
<http://ansinet.com>

## Semantic Aggregator of Public Professional Events

Lilac Al-Safadi, Nour Alkhatib, Rawan Babaier and Lama Assum  
Department of Information Technology, College of Computer and Information Sciences,  
King Saud University, Saudi Arabia

**Abstract:** The data on the web has been created by a large number of people and is hosted in a large number of independent resources across the globe. Therefore, there is a need to develop data aggregators to help the process of knowledge-based integration. The data without central coordination resulted in heterogeneity. The integration of data on the Internet continues to be a challenge for searching and discovery efforts. Semantic aggregators are agents for data integration using semantic web techniques. Semantic web technologies provide the capability to more easily aggregate data and thus can be utilized to improve the efficiency of information discovery. This study describes an implementation of a semantic web infrastructure that collects and integrates data in Extensible Markup Language (XML) format and utilizes the Resource Description Framework (RDF) Data model as the repository. One of the main challenges addressed in this paper is the heterogeneity of existing XML schemata and semantics of websites. The proposed approach is designed to be implemented and tested on publicly available professional events.

**Key words:** Semantic aggregation, data integration, data heterogeneous, semantic web, XML, RDF

### INTRODUCTION

The Internet succeeded in making huge amount of data available on the web. However, these data were developed and hosted in a large number of independent resources, which resulted in a data heterogeneous environment that makes it challenging for computers to automatically exploit and integrate the available information on the web using existing tools.

Search engines are known ways to look for information on the web. However, the result is a list of suggested links to pages. In most search engines, these results are based on mapping keywords content and do not take into consideration the meanings. The main issues of keyword-based query model lie in the difficulty of query formulation and the inherent word ambiguity in natural language (Fiaidhi *et al.*, 2003). Therefore, the results require human interpretation and filtration. Semantic search engines provide searching and retrieving resources conceptually related to the user informational needs. The concept-based retrieval is classified by the Web Intelligence Consortium (WIC) as one of the nine topics in the area of web intelligence (Curran *et al.*, 2004). It performs content-based search of documents on the web focusing on the concepts structure of the content rather than the lexical content. Usually what makes semantic search engines effective is that it focuses on specific domain. An aggregator is a website or software

program that gathers (aggregates) and combines web content from multiple websites to a single location after eliminating redundant content. The semantic web is the idea of having data on the web defined and linked in a way that it can be used by machines not just for display purposes, but for automation, integration and reuse of data across various applications (Cruz *et al.*, 2002). Hence, the semantic aggregator helps the users in gathering resources on the web that are relevant to their needs rather than just by keyword. It uses the different databases available on the Internet and provides semantic-based data integration. Aggregators often associated with gathering items from different sources more than data integration. This work describes semantic aggregators as an approach for data integration using a semantic web approach. It considers data integration a major advantage of aggregators over search engines.

Current websites does not use a uniform way to name entities. For example, an event is named as "Advanced AdWords Seminars for Success Seattle 2011" in Eventful website ([www.eventful.com](http://www.eventful.com)) and named as "The Seattle AdWords Seminar in September" in Certified Knowledge website ([www.certifiedknowledge.org](http://www.certifiedknowledge.org)). As a result, the same entity is identified in various sources with different names. Currently it is very difficult to connect these entities seamlessly unless they are transformed into a common format with IDs connecting each of them. This heterogeneity of data on the Internet eliminates data

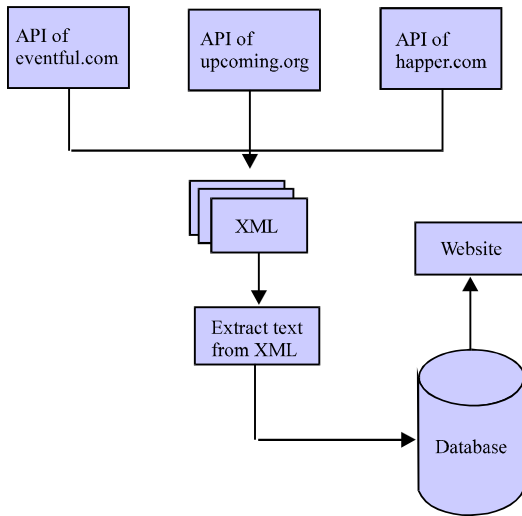


Fig. 1: Abstract model for eventology

interoperability; easily finding, sharing and integrating information. Data interoperability can be achieved by semantic web, which aims at moving from a web of documents, only understandable by humans, to a web that can be read and used by software agents, thus permitting them to find, share and integrate information more easily (Berners-Lee *et al.*, 2001).

The knowledge integration is a complicated task because it requires creating a common data model, finding semantic correspondence between two entities, satisfying the merge requirements and generating the duplicate free entities (Lu and Feng, 2010). A number of works have been conducted to highlight the challenges of data integration in specific domains and illustrate the difficulty involved in integrating the many publicly available data sources with in-house ones (Wang *et al.*, 2005). The challenges lie in the syntactic and semantic heterogeneity of data sources. The semantic web community, led by the W3C (www.w3.org), addressed the data integration challenge by proposing a series of standards; the Resource Description Framework (RDF) format for document (Manola and Miller, 2004) and the Web Ontology Language (WOL) language for ontology specification (McGuinness and van Harmelen, 2003). They believed that semantic web technologies would reduce the trouble of data integration and sharing. The benefits promised by the semantic web include aggregation of heterogeneous data using explicit semantics, simplified annotation and sharing of findings, the expression of rich and well-defined models for data aggregation and search, easier reuse of data in unanticipated ways and the application of logic to infer additional insights (Neumann *et al.*, 2004). Yet, most of data currently

available in the web are in Extensible Markup Language (XML) format. Therefore, the purpose of the present paper is to describe the semantic aggregation and integration approach of XML documents.

Eventology is an aggregator application, following the definition of semantic aggregators given above. It aggregates professional events on the web by integrating publicly available data from some of the most popular databases in professional events over the web such as eventful.com, upcoming.org and eventbrite.com. It is described as a data aggregator using a semantic web approach for data integration. The goal is to build a vertical portal allowing researchers to query and visualize, in a coherent presentation, various information automatically mined from public sources.

Figure 1 illustrates how Eventology works. Since most web services or Application programming interfaces (APIs) use XML format that follow the Simple Object Access Protocol (SOAP) standard, the aggregator first authenticate itself in the desired websites in order to get their API. When aggregator needs data from a website it sends a SOAP message to their web service with the parameters needed for retrieval. The site would then return an XML-formatted document with the resulting data (XML feed). The aggregator then reads the XML feed using feed reader tools or web scripting language, performs syntactic and semantic integration of the results and store them in a centralized repository available for queries. We made this choice to store aggregated data since the key goal in building the prototypes is to design a system that will search millions of professional events.

**Data integration:** An important aspect of aggregation is to eliminate duplication. This requires mapping and integrating equivalent entities collected from the various sources. In Eventology, semantic web is used to assist in data search, integration and retrieval. A pre-requisite for the semantic web is a shared understanding of the domain in question (semantic mapping) and the schema of the websites (schema mapping). XML data has no fixed architecture of format, so it is named semi-structured data (Zhao *et al.*, 2010). Therefore, Eventology data integration uses two level of mapping approaches; schema-level and semantic-level.

**Schema-level mapping:** Schema-level mapping considers the schema of the information, not instance data (Al-Safadi, 2009). The available information of the XML schema elements are the schema name and schema structure. Existing XML documents are defined with different element names and structures; therefore the schema-level integration requires the analysis of the XML

structure of documents (Masood, 2004) presents a schema analysis approach based on the semantic of the component database schema. The work neglects an important aspect of matching which is the structure of the schema.

Two levels of schema-level mapping are supported in our proposed work; element mapping performed for individual schema elements and structured mapping performed for combinations of elements. Element mapping is an atomic level match. Structure mapping, on the other hand, refers to mapping combinations of elements that appear together in a structure. In schema-level mapping, mapping between collected events usually falls in one of two categories; equivalent or not equivalent. The match operation takes as input two XML schema and determines a value between 0 and 1 for each schema element, where 0 is dissimilar between elements (i.e., not equivalent) and 1 is a strong similarity between elements (i.e., equivalent).

**Element mapping:** Element mapping matches elements in the atomic level of the two schema. It uses a linguistic-based approach, which is based on names of schema elements that results in equal, similar or dissimilar names. Equality of names indicates textually equivalent such as “venue” → “Venue”. Similarity of names involves an ontology of professional events to find synonyms, abbreviations or pre-processed word (removing punctuations, prefix, postfix, etc.), for example, it can match “venue” with “location”. Eventology ontology is explained later.

**Structured mapping:** Rajesh and Srivatsa (2009) propose an algorithm to match one to one correspondences between schema elements. There are other types of correspondences neglected in their work such as one to many, many to one and many to many. In our proposed structured mapping, the overall mapping process may match one or more elements of one schema to one or more elements of the other, resulting into four cases: 1:1, 1:n, n:1, n:m. An element of the schema can participate in zero, one or many matched elements of the mapping operation result between the two input schema. Moreover, within an individual mapping operation, one or more elements of the first schema can be matched to one or more elements of the second schema. e.g.,:

Cardinality	S1 element(s)	S2 element(s)
1:1	Title	Title
n:1, 1:n	City, zip code	Address

As for n:m cases, it is treated as 1:1 relationship for every two mapped elements.

**Data set analysis:** Deciding on the schema and elements used for mapping data in the professional events domain involved the analysis of a data set publicly available. Ten publicly available data sets were identified. These collectively contained a wide range of data relevant to professional events on the Internet. This is an arbitrary choice intended to illustrate the variety of data available and the way these data are processed. Available data are expressed in XML. Each data set was manually examined to identify important properties associated with events descriptions. The goal was to identify the representation of equivalent events in the different events data sources. A sample of the data sets used is provided in Table 1.

The previous analysis shows that:

- The 100% of the events occur in different websites
- The 60% of the events are repeated several times with different dates, times, country and/or city
- The 96% of the events are identified by the title, start date, start time, country name, city name and venue

As a conclusion, predicates used for the schema-level mapping of professional events are identified as (title, start time, start date, country, city and venue). When two or more resources identify the same object, we unify them in the repository. In the cases where the equivalences between resources are not specified in an existing mapping file, the identification of naming variants for a same resource was manually performed.

**Semantic-level mapping:** Semantic integration requires mapping and unification of the semantic descriptions extracted from the various data sources. The unification allows multiples specifications, defined with different ontologies to be queried in a unified way. Being able to recognize homonyms, synonyms and related terms is critical in data integration. Therefore, semantic mapping requires the use of domain-specific ontologies. By linking the website content to the existing Eventology ontology and instances, the equivalences between the classes and properties defined in the sources can be described. Each representation of the concepts in the sources is declared as a subclass or an equivalent class of the unified concept defined in Eventology ontology, for example, “application” is mapped with “software”. The same principle is applied for properties by specifying that several equivalent properties are sub properties of a unified one, for example, event title “Google Talk” is mapped with “Gtalk”. A detailed formal solution for ontology matching is presented in (Kolli and Boufaida, 2011).

Table 1: Sample of the data set analyzed to identify properties associated with professional events

Title	Start date	Start time	Country	City	Venue	Sources
Google small business training	20 May 2011	--	United States	Fort Worth	Norris Conference Center	(www.eventful.com), (www.eventbrite.com)
Computer club: google calendar	18 May 2011	10:00:00	United States	Florissant	Jamestown Bluffs Branch Library	(www.eventful.com), (events.fox2now.com)
Making maps the google way	8 June 2011	9:00:00	United States	Vermont	Hampton Inn	(upcoming.yahoo.com), (www.vcgi.org)
The evangelical network 2011 conference	7 July 2011	8:00:00	United States	Texas	Omni Austin Hotel Downtown	(upcoming.yahoo.com)
Google analytics training course	11 July 2011	10:00:00	England	Brighton	Silicon Beach Training	(upcoming.yahoo.com), (www.webcredible.co.uk)
Advanced google analytics training course	12 July 2011	10:00:00	United Kingdom	Brighton	Silicon Beach Training	(www.eventful.com), (upcoming.yahoo.com)
Google AdWords-generate web leads at higher ROI course/class	27 Aug 2011	9:00:59	Australia	Sydney	Centre for Continuing Education (CCE), The University of Sydney	(upcoming.yahoo.com), (cce.usyd.edu.au)
Google and online marketing seminar	14 June 2011	18:00:00	United States	Plainview	1600 Round Swamp Road	(upcoming.yahoo.com, (long-island-website-design.com)
SHRUG computer club meeting	14 May 2011	10:30:00	United States	Portsmouth	Portsmouth Public Library Churchland Branch	(events.wtkr.com), (informationtechnologyedge.com)
Google apps for education workshop	6 Aug 2011	9:00:00	United Kingdom	London	Lambeth City Learning Centre	(www.zvents.com)

Another important aspect in semantic-level mapping is to determine the relevance of a concept in a website to the topic. Only related concepts are used in the semantic-level mapping process. Relevance is determined by the density of a concept in a website. Concept density measures the volume of a concept in a document.

**Ontology design:** In order to solve the problem of polysemy in the semantic description of XML and RDF, we need to introduce ontology (Liu *et al.*, 2011). Ontology is employed to tackle not only structure but also semantic interoperability in information integration (Lu and Feng, 2010). An ontology is an explicit formal specifications of the terms in the domain and relations among them (Gruber, 1993). Most ontologies describe individuals (instances), classes (concepts), attributes and relations. OWL is the conceptual language chosen by the semantic web community for ontology representation. In semantic web, many ontologies we developed. Dupre (1993) stated that here is no unique ontology. But there are multiple ontologies which each models a specific domain. In Eventology, the ontology is used to hold the vocabulary used to generate the RDF descriptions. It is a unified ontology describing the equivalences between the classes and properties related to professional events defined in the different sources. Classes such as “event”, “software” and “certificate” and properties such as “name” and “date”, are defined in the ontology. Figure 2 shows a snapshot of the class hierarchy of the ontology developed for Eventology.

**Data sets analysis:** Ten publicly available data sets were used to determine the density of a concept in a website.

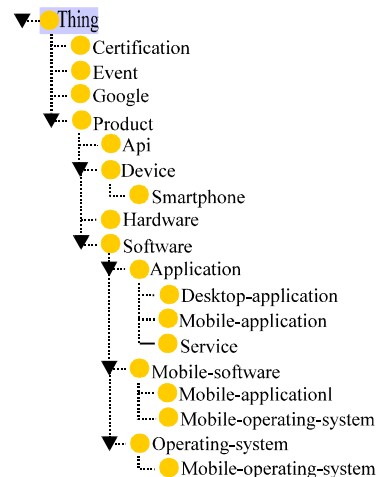


Fig. 2: A snapshot of the ontology class hierarchy

GATE, discussed later, used to extract concepts represented in the ontology from the website content. These sets collectively contained a wide range of data relevant to professional events. Each data set was manually examined to identify existing salient concepts and explore their representations in documents. Salient concepts are notable thematic relevant concepts in the website. The objective of the analysis is to study the relevance the concept and the ratio of its occurrence compared with the number of concepts in the website and the relevance of the concept and the location (title or description) where it occurred. The goal was to create a concept map that linked all of the contained concepts to one another and to explore the interoperability among them in the different events data

Table 2: Sample of the data sets analyzed to determine concept density

Website (#)	Title			Description			
	Salient concepts	Total No. of concepts in title	Occurrence ratio	Salient concepts	Total No. of concepts in description	No. of occurrences	Occurrence ratio
1	AdWords seminar	3	33.3	AdWords	137	4	2.9
			33.3	seminar		2	1.4
				event		1	0.7
2	Analytics event	6	16.6	Analytics	166	5	3
			16.6	Event		5	3
				Search engine		3	1.8
3	Analytics workshop	8	12.3	Analytics	214	9	4.2
			12.3	seminar		1	0.4
				training		1	0.4
				AdWords		1	0.4
4	app Engine	6	16.6	app engine	163	5	3
				events		2	1.5
				workshop		1	0.6
5	google apps workshop	3	33.3	google apps	88	1	1.1
			33.3	workshop		1	1.1
				google docs		1	1.1
				google sites		1	1.1
				AdWords		1	1.1
6	google maps workshop	3	33.3	workshop	318	5	1.6
			33.3	google earth		7	2.2
				conference		1	0.3
				event		1	0.3
				gmail		2	0.58
7	wave android workshop	4	25	wave	198	3	1.5
			25	google buzz		2	1
			25	workshop		3	1.5
				android		4	2

sources. Further details regarding the data sets used are provided in Table 2.

From the previous analysis, we noticed that:

- Not all the extracted concepts reflect the meaning of the event
- The concepts that reflect the meaning are found more than once in the description
- The concepts that occur in the event’s title reflect the meaning of the event (i.e., salient concepts)
- The relevant concepts have an occurrence ratio of 0.65 and above

As a conclusion, two important factors were analyzed to identify the important of a concept in a document referred to as concept density; the ratio of occurrence and location of occurrence. All data mappings were manually examined to ensure correctness to write a program that could automatically perform the mappings between the data sets depicted in the below algorithm.

```

If a concept occurs in the title add it to the list of concepts
If the concepts occurs in the description:
Return equivalent classes from ontology O
Calculate the number of occurrences for the concept in description
Calculate the ratio = (number of occurrence for the certain concept/number of
concepts in description) *100
Compare the resulting ratio with thrush-hold value:
If ratio>thrushhold value then add the concept to the list
    
```

**Architecture:** The two main semantic web standards are RDF and OWL. RDF represents data using subject-predicate-object triples. This triple representation connects data in a directed labeled graph form. Triplestores are database systems able to handle large datasets. Sesame is a known open source triplestore server which provides storage and querying capabilities of triples ([www.openrdf.org](http://www.openrdf.org)).

A peer-to-peer (P2P) approach is used in the data integration. It is a recent paradigm and used in the data integration (Ougouti *et al.*, 2010). The integration system follows a decentralized approach for integration of autonomous and distributed peers containing data which can be shared. The principle objective of such a system is to provide a semantic interoperability between several sources with the absence of global schema.

Three main ideas have directed the development of our software: The integration of XMLs, the conversion of XML into RDF and the use of existing semantic web software to query and visualize the data. The Eventology architecture is composed of two main phases; data aggregation and data retrieval. This section focuses on the data aggregation phase shown in Fig. 3.

The data aggregation (or integration) phase is composed of the following processes:

- Data gathering; collects external websites in XML format that are related to professional events. Data

from these websites is stored in a MySQL database. The process then extracts the schema from the XML file and performs schema-level mapping of websites. Results are stored in the metadata. Users are involved in the mapping process to accept or reject possible mapping candidates

- Content integration; determines equivalent elements on the semantic level. It is the task of extracting the XML content, performing semantic-level mapping and assigning meaning to extracted content. It creates a one unified representation of the event and stores it in a tabular form. In this stage, words and structures found in events' titles and descriptions are mapped to existing ontology to determine equivalence. Equivalent concepts are determined by calculating their concept density. Equivalent concept are unified and integrated in the database
- Data conversion; transforms the returned event into RDF and makes it available to the application. Tabular data are converted in RDF with a simple procedure. Each column which had to be converted in RDF was associated with a namespace that was used to construct the Uniform Resource Identifier (URI) (Berners-Lee *et al.*, 1998) identifying the values of the column. The relationship between the content of two columns was expressed in RDF by a triple having the content of the first column as subject

A number of knowledge stores and data repositories are used. Technically, we choose a centralized data repository which all the data are aggregated and stored in.

- Metadata; stores websites and XML descriptions
- Ontology; represent the knowledge base of professional events. It defines the vocabulary used in generating RDF descriptions
- RDF store; stores the semantic content of the received websites. This represents the global view used to match with the imposed queries

**IMPLEMENTATION AND RESULTS**

The initial web page interface shown in Fig. 4 was designed to assist the user in identifying the data that they would be interested in retrieving. Subsequent web pages were designed to help guide the user to relevant information of interest. The interface was designed to enable users to either retrieve data about a single event from a single source, single event from different sources, or to retrieve data about number of events from different sources.

When the user submits a search query, graph representation of the query is created. For example, using the interface shown in Fig. 5, when the user searches for

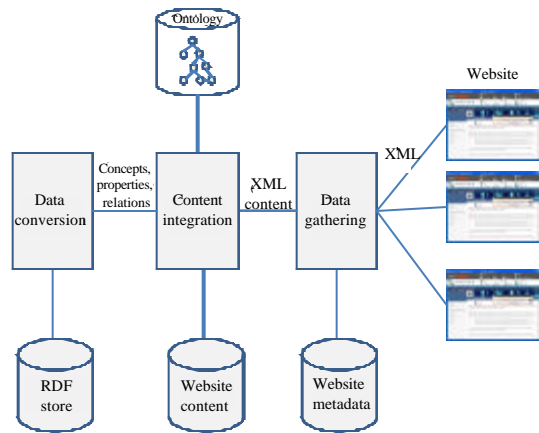


Fig. 3: The architecture of eventology

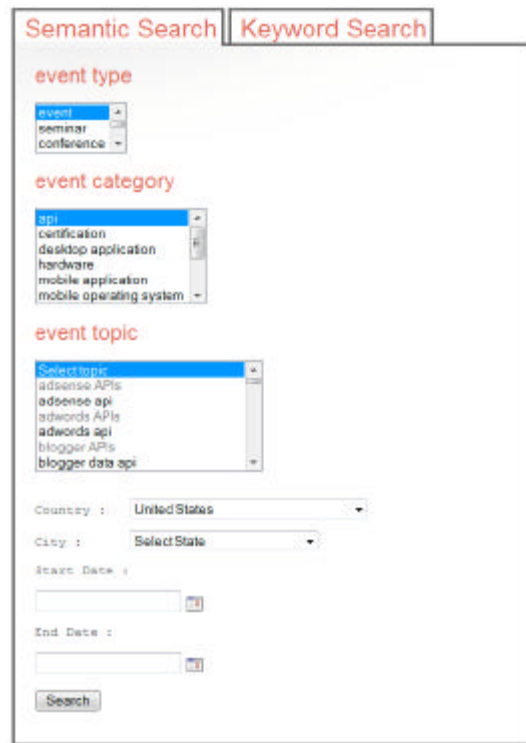


Fig. 4: Snapshot of eventology query screen

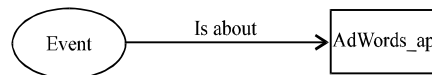


Fig. 5: Graph representation of ‘AdWords API event’

“event” type of event, “API” category of an event and “AdWords API” topic of event. Figure 7 shows the graph representation created for the query. Figure 6 shows the equivalent graphs returned from the ontology.

To design the ontology of Eventology, we used the Protégé open source framework (protégé.stanford.edu)

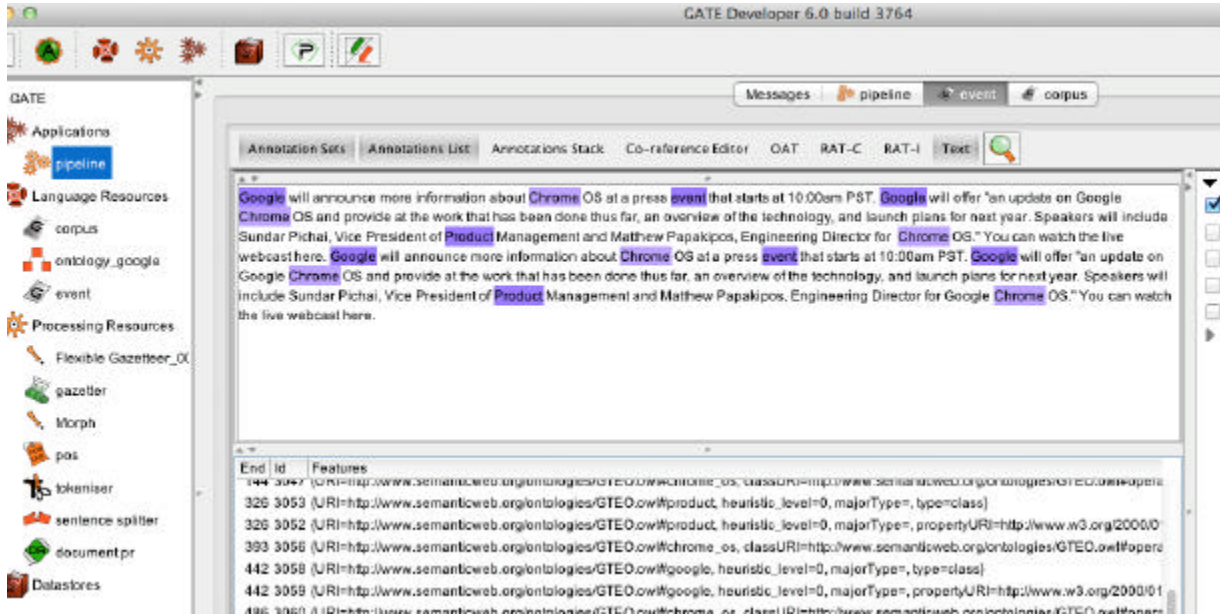


Fig. 6: Snapshot of GATE

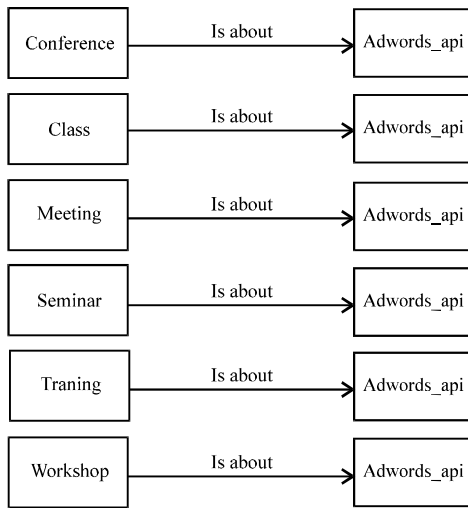


Fig. 7: Equivalent graphs generated using eventology ontology

and its OWL editor Protégé-OWL. After the ontology was created, the second step consisted of writing the necessary programs in JSP in order to map the data elements of the original document with the predicates in the RDF version. RDF is processed and stored using the Jena2 Semantic Web toolkit (McBride, 2000) and is stored in the MySQL database (www.mysql.com).

GATE (GATE, 2010) is an open-source, world-leading platform for language technology, which provides support for multiple languages and formats, based on established standards, such as Unicode and XML. GATE is currently being extended to use semantic web standards to store, index and retrieve language resources in RDF repositories, such as Sesame. In Eventology, GATE was used to extract the concepts found in web sites' titles and descriptions based on ontology information as illustrated in Fig. 6. It is also interoperable with well-established tools such as the Protégé ontology editor, which is also integrated within GATE's visual development tools.

**CONCLUSION**

As data volumes continue to grow over the web, it becomes increasingly important to have effective mechanisms to be able to retrieve data of interest. Search engines helped in overcoming some of the difficulties related to information retrieval. Yet, they still retrieve a huge number of irrelevant and redundant results. Search engines require users to know in advance what keywords stored in the data repository and how it is classified. Aggregating data on the Web has proven to be a challenge due to data heterogeneity on the web. Semantic web technology had promised the ability to more easily aggregate such data. Semantic web standards have



matured to a point where commercial software solutions are available to address real-world problems. Semantic aggregators are expected to overcome the limitations of search engines.

The infrastructure highlighted in the study takes advantage of XML data format, which is widely used in the Internet, with the goal of showing how this format can be used to successfully aggregate data. Furthermore, the approach transforms integrated XML into RDF to have a good understanding of the semantics of the data and how they are interrelated for a semantic search of website content. However, one of the main challenges of this approach is the heterogeneity of existing XML schemata and semantics of websites.

This study presents the design and the implementation of Eventology, an application that collects and semantically aggregates professional events on the Web encoded in XML. The resulting data is cleansed, transformed into RDF to be available for semantic query by software agents via remote query servers.

#### ACKNOWLEDGMENTS

Our thanks to Lamia Al-Manea and Rehab Aljaloud, who have contributed towards development of the prototype.

This research project was supported by a grant from the "Research Center of the Center for Female Scientific and Medical Colleges". Deanship of Scientific Research, King Saud University.

#### REFERENCES

- Al-Safadi, L.A.E., 2009. Electronic medical records ontology mapper. *Int. J. Adv. Comput. Technol.*, 1: 85-97.
- Berners-Lee, T., J. Hendler and O. Lassila, 2001. The semantic web. *Sci. Am.*, 284: 34-43.
- Berners-Lee, T., R. Fielding and L. Masinter, 1998. Uniform resource identifier (URI): Generic syntax. Internet Engineering Task Force. <http://www.mendeley.com/research/uniform-resource-identifiers-uri-generic-syntax/>
- Cruz, I., S. Decker, J. Euzenat and D. McGuinness, 2002. The Emerging Semantic Web, Selected Papers From the First Semantic Web Working Symposium. IOS Press, Amsterdam, The Netherlands.
- Curran, K., C. Murphy and S. Annesley, 2004. Web intelligence in information retrieval. *Inform. Technol. J.*, 3: 196-201.
- Dupre, J., 1993. *The Disorder of Things: Metaphysical Foundations of the Disunity of Science*. Harvard University Press, USA.
- Fiaidhi, J., S. Mohammed, J. Jaam and A. Hasnah, 2003. A standard framework for personalization via ontology-based query expansion. *Inform. Technol. J.*, 2: 96-103.
- GATE, 2010. General Architecture for Text Engineering. The University of Sheffield, South Yorkshire, England.
- Gruber, T.R., 1993. A translation approach to portable ontology specification. *Knowledge Acquisit.*, 5: 199-220.
- Kolli, M. and Z. Boufaida, 2011. Composing semantic relations among ontologies with a description logics. *Inform. Technol. J.*, 10: 1106-1112.
- Liu, C., D. Yang, Y. Wang and Q. Pan, 2011. Web intelligence analysis in the semantic web based on domain ontology. *Inform. Technol. J.*, 10: 2343-2349.
- Lu, H. and B. Feng, 2010. Distributed knowledge integration based on intelligent topic map. *Inform. Technol. J.*, 9: 132-138.
- Manola, F. and E. Miller, 2004. RDF primer. W3C Recommendation. <http://www.w3.org/TR/rdf-primer/>
- Masood, N., 2004. A schema comparison approach to determine semantic similarity among schema elements. *Inform. Technol. J.*, 3: 57-68.
- McBride, B., 2000. Jena: Implementing the RDF model and syntax specification. Technical Report, HP Labs at Bristol, UK.
- McGuinness, D.L. and F. van Harmelen, 2003. OWL web ontology language overview, W3C proposed recommendation. <http://www.w3.org/TR/2003/PR-owl-features-20031215/>
- Neumann, E., E. Miller and J. Wilbanks, 2004. What the semantic web could do for the life sciences. *Drug Discovery Today Biosilico*, 2: 228-236.
- Ougouti, N.S., H. Belbachir, Y. Amghar and N.A. Benharkat, 2010. Integration of heterogeneous data sources. *J. Applied Sci.*, 10: 2923-2928.
- Rajesh, A. and S.K. Srivatsa, 2009. A hybrid path matching algorithm for XML schemas. *Inform. Technol. J.*, 8: 378-382.
- Wang, X., R. Gorlitsky and J. Almeida, 2005. From XML to RDF: How semantic web technologies will change the design of 'omic' standards. *Nat. Biotechnol.*, 23: 1099-1103.
- Zhao, Y., T. Ma and F. Liu, 2010. Research on index technology for group-by aggregation query in XML cube. *Inform. Technol. J.*, 9: 116-123.