The BOEMIE Semantic Browser: An Application Exploiting Rich Semantic Metadata

Sofia Espinosa Peraldi, Atila Kaya, Ralf Möller

Abstract: Recent advances in the standardization of knowledge representation languages to realize the Semantic Web as well as advances in natural language processing techniques have resulted in increased commercial efforts to create so called annotation services which aim to annotate web content with metadata. Description Logics-based metadata are symbolic descriptions of some content, e.g., web content, with semantics defined through so called ontologies. Ontologies play a central role for the definition of semantics in the context of the Semantic Web. Currently, annotation services can be exploited to build content-driven applications. However for such applications to offer more valuable functionality, current annotation services should provide for richer metadata. Rich metadata describe a deeper and more abstract level of information through complex relational structures. In this paper we describe use cases that require richer metadata than currently provided by annotation services and describe a content-driven application that implements them. The use cases focus on geography-aware multimedia navigation, content activation and retrieval of semantically related multimedia. Furthermore a declarative process to extract richer metadata is introduced.

1 Introduction

The Semantic Web vision promises to pave the way for the development of content-driven applications and services. To this end the semantics of content in the Semantic Web should be available. Nowadays the majority of information reside in web pages. Typically a web page has rich media content and includes visual and/or audio information in addition to textual information. In the last decade, information retrieval from the Web has become the underlaying basis of daily information access. At the same time, natural language processing techniques have reached a high level of maturity such that text analysis tools with good performance become widely-used in practical applications. Analysis tools such as *Ellogon*[1] [FPT+08] provide a comprehensive infrastructure for natural language processing (NLP) including support for principal tasks such as information extraction (IE) and machine learning (ML). Facilitated by these advances, commercial interest arose in the creation of annotation services. Companies such as *ClearForest*[2] and *Reuters*[3] provide annotation services e.g., OpenCalais, about various entities such as persons, organizations, locations as well as facts, events and generic relations from textual content. Currently some information portals already contain web pages that exploit metadata for highlighting particular words in the text or presenting advertisements. These can be considered

[1] www.ellogon.org
[2] www.clearforest.com
[3] www.reuters.com

as forerunners of upcoming more 'intelligent' content-driven applications that can exploit rich metadata to offer more valuable services. The sucess of current annotation services represent a positive indicator of progress towards the objectives of the Semantic Web. But the metadata obtained by current annotation services is not enough for content-driven applications to provide for more valuable functionality. For this, richer metadata is required. Rich metadata means symbolic descriptions that describe a deeper level of information i.e. more abstract entities and relations among those. Rich metadata is also multimodal, thus describes the content of multimedia documents. This paper has two main contributions: First we describe use cases for 'intelligent' content-driven applications that exploit richer metadata automatically extracted from multimedia documents. Furthermore, we describe the BOEMIE Semantic Browser (BSB), a content-driven application that implements the use cases as a proof of concept for the practical exploitation of rich semantic metadata. Second, to better understand what is meant with rich semantic metadata, we describe a framework for the extraction of rich semantic metadata. The rest of this paper is organized as follows: In Section 2, the approach followed for the generation of rich semantic metadata is introduced. In Section 3, use cases to demonstrate the advantages of exploiting rich semantic metadata are discussed together with examples and their implementation in BSB. The architecture underlying the BSB is the topic of Section 4. Finally, we conclude this work in Section 5.

2 An Ontology-based Framework for the Extraction of Rich Semantic Metadata

The framework for the extraction of rich semantic metadata that is described here is called the BOEMIE framework. The acronym BOEMIE means *Bootstrapping Ontology Evolution with Multimedia Information Extraction*. The framework integrates a set of modality-specific analysis tools and a modality-independent interpretation and fusion engine. In this framework, a multimedia document is processed along three phases of metadata extraction. Consider that extraction techniques are not the central part of this paper therefore references are give where necessary.

First phase: Analysis
In the first phase, modality specific analysis takes place. Here the structure of the multimedia document is analyzed to identify the different modalities of content that a document contains such that the document is divided into modality specific content items, e.g., a web page containing text and two images is divided into three content items, one text item and two image items. Each content item is analysed with the corresponding modality-specific analysis tool. The analysis tools generate symbolic descriptions as metadata, whose semantics are defined through ontologies. The framework integrates three different OWL-DL ontologies, two which are domain specific, these are the *Athletics Event Ontology* (AEO)[4] [DEDT07] and the *Geographic Information Ontology* (GIO)[5] [DEDT07] and one

[4]http://repository.boemie.org/ontology_repository_tbox/aeo-1.owl
[5]http://repository.boemie.org/ontology_repository_tbox/gio-1.owl

more to address the structural aspects of a multimedia document, the *Multimedia Content Ontology* (MCO)[6] [DEDT07]. The resulting metadata describes *surface-level information*. Surface-level information consists of *observable* entities, e.g., objects in visual content or words in non-visual content and *observable* relations between the entities, e.g., spatial relations between objects in an image or domain-specific relations between words in text. For a better illustration consider the captioned image in Figure 1, its analysis results consist of two sets of symbolic descriptions one from the image content item and one from the textual content item. The results can be observed in the upper part of the graphic in Figure 2.

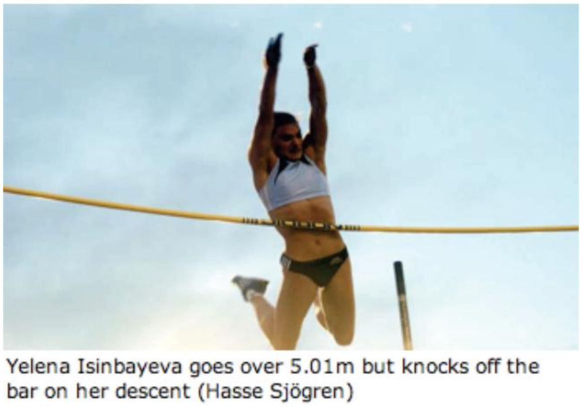

Figure 1: A captioned image from a web page about athletics news

Notice that the term *observable* refers to the extraction capabilities of the analysis tools, for instance, in this example image analysis can extract a person's face and a person's body, but more sofisticated image analysis techniques with face recognition could extract the name of the person. In [PTK$^+$06] the analysis tools are discussed in more detail.

Second phase: Interpretation

In the second phase, modality specific interpretation takes place. The interpretation process which is described in [PKM$^+$07] with more detail, is a generic and declarative process that can be used for various domains of interest. The domain of interest is described through ontologies and rules which constitute the background knowledge of the interpetation process. The process uses the Description Logics (DLs) reasoner RacerPro[7] to execute abductive and deductive reasoning over the results of the previous phase, such that metadata describing more abstract information is obtained. Abstract information is what we call deep-level information. The metadata is expressed through symbolic descriptions that we call *aggregates*. Aggregates have been defined by Neumann and Möller [NM06] in the context of DLs as representational units for object configurations consisting of a set of parts tied together to form a concept and satifying certain constraints

[6]http://repository.boemie.org/ontology_repository_tbox/mco-1.owl
[7]www.racer-systems.com

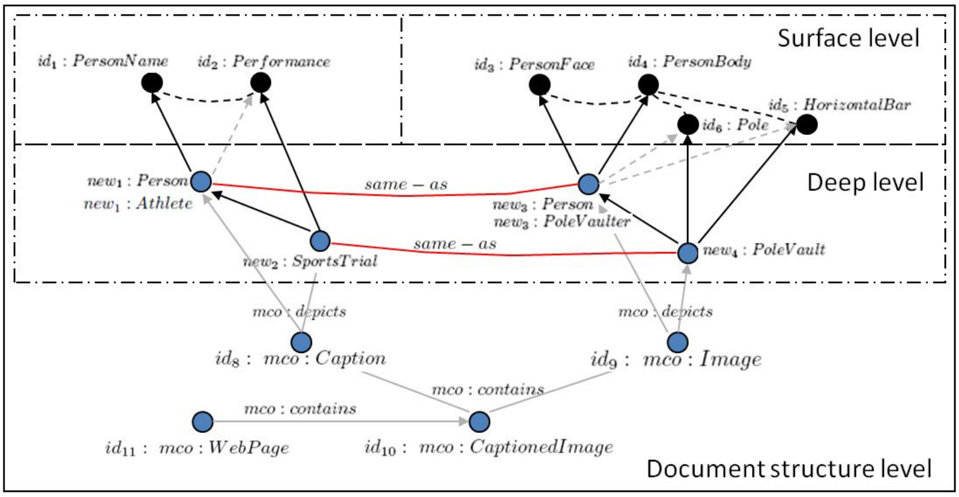

Figure 2: A graphic representation of rich semantic metadata resulting from analysis, interpretation and fusion extraction phases of Figure 1

w.r.t. a background knowledge. In our athletics example (see the central-right part of the graphic in Figure 2), from image interpretation, two aggregates are extracted a *PoleVault* trial and a *Person*. The aggregate *PoleVault* has three parts, the aggregate *Person* and two sports equipment, a *HorizontalBar* and a *Pole*. The aggregate *Person* has two parts, a *PersonFace* and a *PersonBody*. Thus the results of interpretation is a set of relational structures that represent aggregates that have as parts other aggregates. These configuration of relational structures has a specific advantage, namely that it allows information to flow from aggregate to aggregate according to the semantics described in the background knowledge. Such that information flow contributes to increase the precision of the extracted metadata. For example, consider the following formula from the AEO ontology, $PoleVault \sqsubseteq SportsTrial \sqcap \forall_{<1} hasParticipant.PoleVaulter \sqcap \exists_{<1} hasPart.Pole \sqcap \exists_{<1} hasPart.HorizontalBar$, the universal restriction ($\forall$) over the role *hasParticipant* restricts the range of the role to *PoleVaulters*, for this reason in Figure 2 the instance of *Person* (new_3) is also an instance of *PoleVaulter* given that new_3 is in a relation *hasParticipant* with an instance of type *PoleVault*. In [EKM$^+$07] the interpretation of image segments, and in [EKM08] the interpretation of text segments have been discussed in detail.

Third phase: Multimedia fusion

Since the ultimate goal of a multimedia document is to illustrate the domain of interest in a richer way than a single modality would possibly do, it is expected that some of the information extracted from one modality is related to information from another modality such that they are complementary. In this phase, the purpose is to identify the entities from each modality that refer to the same real-worlds entities and make this explicit through *same-as* relational structures. We argue that multimedia fusion is only possible over deep-level information that results from the interpretation phase. Thus, most of the time different modalities contain surface-level information of a different nature. For example, facts can be obtained from textual content such as person names, dates, age, etc. and it is unlikely that these facts can be obtained from visual content (with the exception of OCR (Optical Character Recognition)). In Figure 2 it can be observed that same real world entities are found in the deep-level information layer, since interpretation provides the necessary abstract information that place observations of a different nature into the same context.

Notice that the resulting rich semantic metadata (see Figure 2) of a multimedia document should be consistent with respect to the background knowledge and contains not only domain specific information but also information about the compositional structure of the multimedia document. The structural metadata contains information about the type of content item, e.g., text, image, audio, etc. and the segments within the content item where objects, words or utherances can be localized. Segments can be of different types according to the modality of the content item, for example space segment for images and text, time segments for audio and space-time segments for video. The symbolic descriptions in Figure 2 is what we call rich semantic metadata that is obtained through the three phases of extraction in the BOEMIE framework. A more detailed description of the framework can be found in [CEF$^+$49].

3 Use Cases For Rich Semantic Metadata

Rich semantic metadata obtained from multimedia analysis, interpretation and fusion can be exploited in practice for various purposes beneficial for 'intelligent' content-driven applications. In this section, three use cases are described that are inspired by the current commercial interest of producing applications that aim to exploit semantically related information. They are addressed in a web application called *BOEMIE Semantic Browser* (BSB)[8]. A generic description of how they have been implemented is included, as well as a comparison with current application that can be found in the market is found in related work.

Use case 1: Geography-aware Information Navigation
Digital maps have nowadays become rich geographic databases that provide a common platform for various types of information with geographical references, e.g., museums, shops, venues, offices, facilities, etc. Moreover, digital media objects, e.g., audio, video, text documents, images, etc. are constantly being produced and are even publicly available on the web. As a result, various applications have been developed that use those geographic databases together with publicly available media objects to provide for geography-aware information navigation. *Google Earth*, *Yahoo Maps*, *OpenStreetMaps* are examples of such applications. The wide acceptance of this kind of applications proves how appealing it is for end-users to use geographical references as anchors for browsing multimedia content. The problem is that this kind of functionality requires some kind of geographic

[8]BOEMIE Semantic Browser: http://boemie.sts.tu-harburg.de

annotation. In this paper a *geographic annotation* refers to the process of attaching metadata to a content item that describes geographic information. Geographic information can be a geocode (longitude and latitude coordinates) or anything that can be characterized with a geocode, for example, the name of a place, a point of interest, a street name, etc. A geographic annotation can be done manually or automatically. Clearly manual annotation is a tedious and expensive activity. Therefore, until now, such applications have obtained manual annotation from *hobby end-users* on the web. This is useful for scenarios such as on-line communities or for personal use. But it is not sufficient for professional scenarios. Here it is called *professional scenario*, to such scenarios where information providers can not allow any end-user to annotate their content and therefore are forced to pay for manual annotation work. Moreover the information providers are the owners of huge amounts of information, want to avoid manual annotation costs and want to keep their rights over their information. Current NLP techniques have facilitated the automatic geographic annotation of textual content by identifying geographic name entities (NE) such as geopolitical names. We argue that rich semantic metadata obtained after analysis, interpretation and fusion can contribute in the following:

- For the geographic annotation of visual content by the use of fusion with metadata of non-visual content that contains geographic references.
- To support the geolocalization of content items, where the the results of interpretation and fusion provide a context to improve the precision.

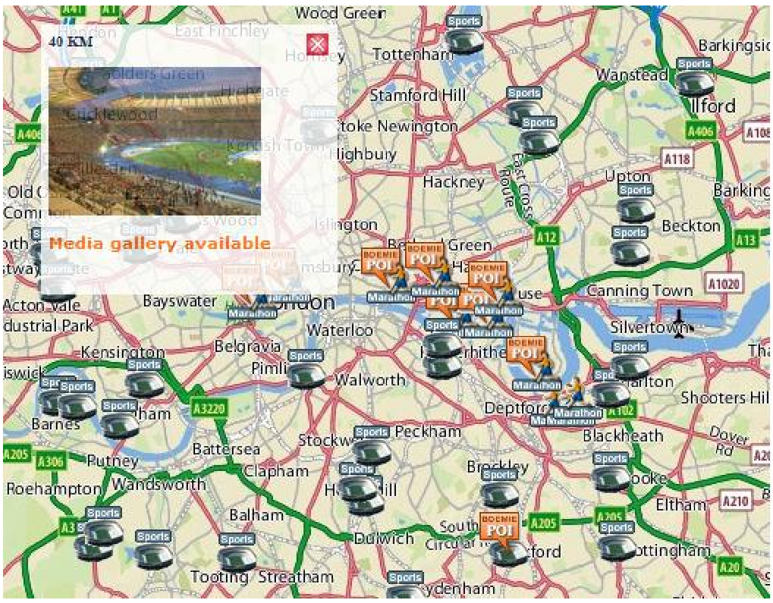

Figure 3: Geography-aware Information Navigation in the BSB.

Geolocalization, refers to the association of a content item with a geocode to support geography-aware information navigation as done by the BSB (see Figure 3). The precision of geolocalization requires more than a geographic NE recongized in the content. It requires a context around that geographic NE that helps in determining how relevant that NE is inside the content, in order to be considered as basis for geolocalization. We consider that the relevance of a geographic reference is determined by the domain of interest. In our example the domain of interest is athletics events. In a web page about athletics news[9], typically various geographic references appear, such as country names to provide the nationality of athletes, city names to provide the location where an athletics event takes place, as well as athletics POIs such as stadiums can be found, etc. Thus, the results of analysis, interpretation and fusion of such web pages contain a set of aggregates having geographic references as parts. From the various aggregates in the domain, the athletics event aggregate is the most relevant one given that is the parent of all the other aggregates. According to the AEO ontology, an athletics event has competitions as parts, competitions have trials, and trials have athletes as participants. In this context, the geographic reference that is part of an athletics event is the most relevant one in the domain. The most relevant geographic reference is considered as basis to associate the content item with a geocode in a map. More over, if there exist a second geographic reference that is in a *contains* qualitative-spatial relationship with the most relevant reference, e.g., a city contains a stadium, then the second reference is prefered since it provides for a geocode that increases the precision of the geolocalization.

Use case 2: Interactive content

Interacting with maps for information navigation as shown in the previous use case is practical and well accepted. In a similar way the multimedia content can also be made interactive. In this respect the structural part of rich semantic metadata helps in activating the multimedia content for various purposes:

- To provide for information about specific segments in visual and non-visual content e.g. objects in images or words in text (see Figure 4).
- To use the active content to support informationa retrieval (IR), where the deep-level metadata of the active content act as a context to define a specific query.

Suppose that the image in Figure 4 has been accessed through the image gallery of a sports news portal and its content is activated such that it can be used for interaction purposes. The interaction could facilitate the access to information about specific objects that have been recognized in the image, such that the user can select among different commands provided in a context menu. Such commands can involve:

- Specific queries, e.g. Further images of the same person in high jump trials.
- Advertisements, e.g. showing the price, model and producer of the sport shoes used by the athlete.

[9] www.iaaf.org

The manual work that is necessary to obtain the structural aspects of the content such as polygon coordinates in an image or character position in a text is demanding. Moreover, the association of structural aspects of a segment with domain semantics is also necessary. For this reason the rich semantic metadata extracted by the BOEMIE framework is helpful since in the process of analysis the necessary structural information is obtained. Interactive media content is now starting to find acceptance in the market, mainly for advertisement purposes on top of text, so called *in-text advertisement.* Clearly highlighting the relevant words in a text and placing the relevant advertisement to the active content requires annotation work, for which commercial services are already available.

Figure 4: Active content given text and image analysis

BSB demonstrates similarly the use of analysis (surface-level information extraction) to automatically highlight relevant content of a specific domain on top of text or images to prepare the interface for further interaction possibilities. For example Figure 4 contains an excerpt of a web page about athletics news. The words that are relevant for the athletics domain are highlighted and related semantics, e.g. 'PersonName' are automatically annotated. Figure 4 contains an image in which an object has been recognized by image analysis and related semantics are also annotated, e.g. 'Horizontal Bar'. Similar to the geographic annotations, the AEO ontology is used as infrastructure for annotating the multimedia content, which is used to obtain the required labels for the domain specific annotations which are useful as relational structures to support IR.

Use case 3: Dynamic triggering of commands

Besides the standard GUI elements such as drop-down lists, check boxes, etc. that a web application can provide for interactivity purposes, the activation of content allows for new interaction possibilities that can be determined by the semantics of the content. The objective of this use case is to demonstrate that by exploiting reasoning services over interpretation and fusion results is helpful to dynamically determine the commands of a context menu that are applicable with respect to the content. In the BSB the commands of a context menu are used to support IR and to place advertisements.

For a better illustration of this use case an example is useful. Consider the image in Figure 4 and corresponding metadata in Figure 5. It can be observed that metadata coming from surface-level information was extracted e.g., a horizontal bar, a person's body and face

surface-level			deep-level		
pb_1	:	$PersonsBody$	ind_1	:	$Person$
pf_1	:	$PersonsFace$	(ind_1, pb_1)	:	$hasPart$
hb_1	:	$HorizontalBar$	(ind_1, pf_1)	:	$hasPart$
(pb_1, pf_1)	:	$adjacent$	ind_2	:	$HighJump$
(pb_1, hb_1)	:	$adjacent$	(ind_2, ind_1)	:	$hasParticipant$
			(ind_2, hb_1)	:	$hasPart$

Figure 5: Surface-level and deep-level information extracted from the image in Figure 4.

$HorizontalBar$	$\sqsubseteq$	$SportsEquipment$
$Athlete$	$\sqsubseteq$	$Person$
$HighJumper$	$\sqsubseteq$	$Athlete$
$SportsTrial$	$\sqsubseteq$	$\exists hasParticipant.Athlete$
		$\sqcap \exists hasPerformance.Performance$
		$\sqcap \exists hasRanking.Ranking$
$JumpingTrial$	$\sqsubseteq$	$SportsTrial$
$HighJump$	$\sqsubseteq$	$Jumping \sqcap \forall hasParticipant.HighJumper$
		$\sqcap \exists hasPart.HorizontalBar$
		$\sqcap \neg PoleVault$
$PoleVault$	$\sqsubseteq$	$Jumping \sqcap \forall hasParticipant.PoleVaulter$
		$\sqcap \exists hasPart.Pole$
		$\sqcap \exists hasPart.HorizontalBar$

Figure 6: Relevant axioms of the athletics ontology.

and relations between them. These extractions are used to relate domain semantics to the structural part of the content, e.g., corresponding polygons in the image that should be activated given a domain specific object that has been extracted. Moreover deep-level information was also extracted, namely that the image depicts a high jump trial which has a participant and a horizontal bar as parts (see Figure 5). If a user interacts with the active content, for example with the horizontal bar, the following commands are applicable:

- Buy athletics equipment at sports-warehouse.
- More images of high jump trials.
- More images of jumping trials.
- View athlete's biography.

To support this kind of functionality it is necessary to previously define a set of so called Dynamite Commands. In this example four commands are required. Currently BSB supports commands for IR and links to other web portals. Due to space restrictions we include here an example for the first two commands:

- **Command Id:** 1
  Menu-name: Buy athletics equipment at sports-warehouse
  Arguments: x : SportsEquipment

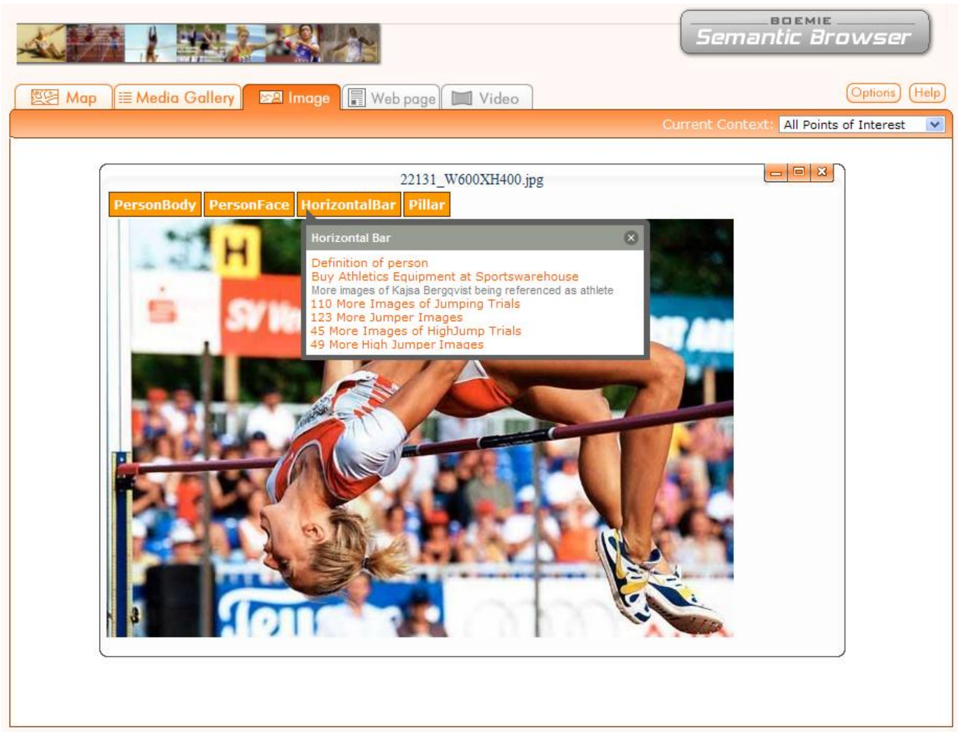

Figure 7: Context menus for active content in the BOEMIE Semantic Browser.

Type: WebNavigation
URL: http://www.sportswarehouse.co.uk/acatalog/Athletics

- **Command Id:** 2
  Menu-name: More images of high jump trials
  Arguments: x : HighJump
  Type: RespositoryNavigation
  Query:
  SELECT DISTINCT ?u WHERE {
  ?w rdf:type mco:Image .
  ?w mco:hasURL ?u .
  ?w mco:depicts ?y .
  ?y rdf:type aeo:HighJump .}

The applicability of a command is determined by the arguments it requires. The arguments have a type that can be found in the signature of an ontology, in this case the AEO ontology. From the extractions in Figure 5, $hb_1 : HorizontalBar$ triggers the first command since it requires an argument of type *SportsEquipment*, thus the first command is activated by reasoning on surface-level information since according to the AEO ontology (see Figure 6) *HorizontalBar* is a special type of *SportsEquipment*. The second command requires an argument of type *HighJump*, therefore this command is activated by deep-level information, namely by $ind_2 : HighJump$. The third command requires an argument of type *Jumping*, thus it is activated by reasoning on deep-level information, since *HighJump* is a specific type of a jumping trial. Finally, the forth command requires an argument of type *Athlete*. Therefore it is activated by reasoning on deep-level information, namely on $ind_1 : Person$ since, according to the ontology, a person participating in a jumping event is also an athlete.

Currently the BSB triggers the commands of context menus which are accessed through a click on the active objects of an image (see Figure 7) or words in the text of a web page. Observe that the relevance of the use case relays on the dynamic triggering of commands and not on the interface widgets, thus the same principle can be used to support other types of interface widgets like side bars, drop-down lists, etc.

Related work

Currently various tools provide for geography-aware information navigation. Leading tools like *Yahoo Maps*[10] and *Google Earth*[11] offer a wide variety of services ranging from informal collaboration scenarios for hobby end-users on the internet to professional enterprise solutions. Their strength lays on their rich variety of services and rich information content both in geographic and multimedia content such as 3D content, images and text documents, etc. More over annotation services such as OpenCalais offer the annotation of geographic references with corresponding geocode. We argue that the type of rich semantic metadata presented here is a valuable complement to current commercial technologies to support on the one side, the automatic geographic annotation of multimedia content beyond text and on the other side to increase the precision of multimedia geolocalization. And in this way support also what we call professional scenarios. We gave an example of how the extraction of geographic NE from text is not sufficient, but also interpretation and fusion processes are necessary to improve the precision of geolocalization.

Various companies exist which provide services of active textual content for advertisement such as *VibrantMedia*[12], *Kontera*[13] to name a few. With the BSB we have proposed a new way of using active content that goes beyond advertisement. We have proposed the exploitation of interpretation and fusion results for browsing purposes. Commercial providers also exist for the automatic recognition of surface-level semantics from text such as *Reuters* and *ClearForest*. Such services are helpfull for the implementation of semantic-based content activation, such as the browser plug-in called *Gnosis*[14]. With *Gnosis* words in a web page are activated with semantics. For interaction with the active content, context menus are used. The commands of such context menus help to transfer the semantics to other web portals, such as news agencies, wikipedia and search engines to find further references to the active word in other web portals. This is what they call in-depth browsing. We argue that in-depth browsing is achieved if rich semantic metadata resulting from interpretation and fusion is used. Thus, interpretation results help in increasing semantic precision. This can be appreciated through the specific queries that are behind the

[10] http://maps.yahoo.com
[11] http://earth.google.com/intl/en
[12] http://www.vibrantmedia.com
[13] http://www.kontera.com
[14] https://addons.mozilla.org/en-US/firefox/addon/3999

commands of the BSB context menus as appreciated in Figure 7.

4 Architecture and Implementation of BSB

The BOEMIE Semantic Browser has a client-server architecture, as depicted in Figure 8. On the client side a web application was developed using AJAX technologies to support the communication with the server for the retrieval of multimedia and map information. To provide for geography-aware information navigation, three elements of this architecture are relevant, namely the *TeleAtlas*[15] Web Map Research Platform (WMRP), the BOEMIE Ontologies repository and the BOEMIE Multimedia Objects repository. From the WMRP, the Geocoder is used to obtain coordinate information for a specific location such as street names, cities, countries, etc. With this information a bounding box is created such that coordinates for lower-left corner and upper-right corner of a geographic area can be specified, as required by the Web Feature Service (WFS). The WFS provides a list of coordinates for all points of interest in a city such as marathon routes, sports points of interest, etc. within the specified bounding box. The WMS provides geographic maps for specific layers of information in a specific area described by a bounding box. Examples of information layers are street name layers, transport means layers, sports POI layers, etc.

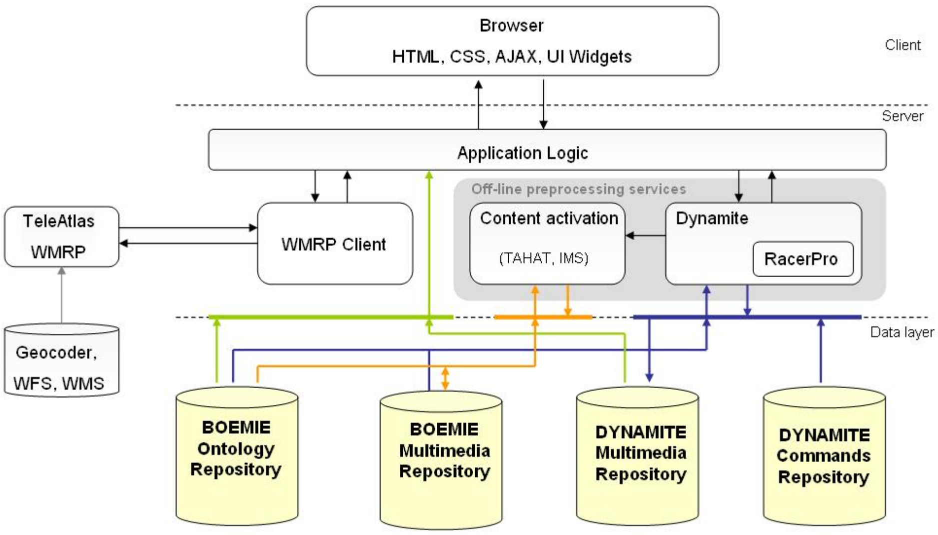

Figure 8: Overall architecture of the BOEMIE Semantic Browser

The BOEMIE Ontologies repository is used to query for all extracted points of interest (POIs) that are located in a specific city. If there exist a match between the extracted POIs and the POIs delivered by the WFS, then media objects corresponding to the extraction

[15] http://www.teleatlas.com

results are associated with corresponding coordinates on the map and a BOEMIE POI (see Figure 3) is created. The BOEMIE Ontology repository comprehends a set of OWL files containing assertional information, also called ABoxes, where each ABox contains the rich semantic metadata of a specific multimedia content item from the BOEMIE Multimedia Objects repository. The ABoxes are stored in an Allegro Graph RDFStore. Moreover the athletics ontology (AEO), the geographic ontology (GIO) and the multimedia content ontology (MCO) are also found in the BOEMIE Ontology repository. To provide for content activation and corresponding context menus, two services are used off-line to preprocess the multimedia content. Currently content activation is done over text in web pages and images only. The Text Analysis HTML Annotation Tool (TAHAT) is used for the semantic annotation and activation of web pages with surface-level metadata from text analysis. The Image Map Service (IMS) is used for the semantic annotation of images with surface-level metadata from image analysis and HTML image maps are created to support content activation. The resulting annotated media objects are stored in the Multimedia Objects repository. From this repository, the DYNAMITE (DYNAMIc InTEractive web pages) module retrieves annotated objects and extends the annotation with command identifiers that are used by JavaScript on the client side in order to render the context menus when required. The extended annotated objects are stored in the Dynamite Multimedia Objects repository. Finally the server delivers to the web clients only Dynamite Multimedia Objects.

4.1 Evaluation

The evaluation of the BOEMIE Semantic Browser gave positive feedback given different evaluation strategies that were applied consisting of testing sessions and meetings with companies such as new agencies[16]. Both evaluation strategies focused on qualitative measures, namely a generic view of the end-users' perspective with respect to the quality of extraction results and the usefulness of extraction results on graphical user interfaces. The testing sessions were organized such that the evaluators could interact themselves with the BSB. These sessions had a structured organization divided intro three parts. The first part containing an introduction to the BSB explaining the use cases to test. The second part focused on testing the interface along the use cases (or exercises). Finally, the third part consisted in the completion of an online questionnaire. Meetings with IAAF the International Association of Athletics Federation (IAAF)[17] and DeltaTre[18] were organized. IAAF was a key evaluator for the BSB due to their special interest in enhancing their web portal with semantics, thus they represent the final user of the information extraction results for the construction of semantic web applications. DeltaTre, as providers of multimedia solutions for sport events they have expressed their interest in the usage of ontologies as background knowledge to manage multimedia information. The evaluators noticed the advantage of using rich semantic metadata since it can be easily perceived the amount of

[16] BOEMIE showcase short video: http://www.boemie.org/video_gallery

[17] www.iaaf.org

[18] www.deltatre.com

manual annotation work that would require the activation of content and relation of domain metadata to a corpus of multimedia content. Geography-aware information navigation is interesting and currently more used due to the well known application of Google-maps. The context-based activation of information through active content was found interesting, but can cause disturbances when too many options are offered. More over, this scenario demands precise analysis results, since the quality of the extraction becomes obvious to the end user. Our conclusion is that context-based functionality is well accepted and this acceptance can be increased when using other interface widgets, e.g., banners, drop-down lists, buttons, etc.

5 Conclusion

Rich semantic metadata is necessary if more (than possible today) valuable functionality is desired for 'intelligent' content-driven applications. In this paper we have described how content-driven applications can exploit such rich metadata by focusing on three specific use cases which are popular on the web, namely for geography-aware information navigation, content activation and retrieval of semantically related multimedia. For the use case of geography-aware information navigation we contribute in two ways. First, the rich semantic metadata obtained by a process of analysis, interpretation and fusion of multimedia helps in increasing the precision required for the automatic association of media objects to a geocode on a map. Thus, the resutls of interpretation and fusion give a context to determine the relevance of a geographic reference in a content of a specific domain. In this way geography-aware information navigation can be exploited in professional environments and not only in informal collaboration scenarios as currently done through hobby end-users on the web. Second, the results of interpretation provide the required level of metadata abstraction that place observations of a different nature (given the type of multimedia) into the same context to be able to fuse information of different modalities. Such that with the exploitation of fusion results, the automatic geographic annotation of visual content is facilitated. For content activation, we have proved with examples and with the BOEMIE Semantic Browser that, on the one side structural information can be used for content activation. And on the other side interpretation results increase the precision of surface level semantics, by exploiting the implicit information that can be obtained through reasoning over the rich semantic metadata. We have proposed a new way of using active content that goes beyond advertisement. We have proposed the exploitation of interpretation and fusion results as context for the definition of more complex queries that allow for multimedia IR. Such that the process of multimedia IR is done on a richer semantic basis. Different to current annotation services such as OpenCalais which produce annotation in RDF, our approach uses ontologies in OWL-DL, allowing the use of reasoning services which is the basis for the semantic web. The sucess of current annotation services represent a positive indicator of progress towards the objectives of the semantic web, but to fully exploit its possibilities, we encourage annotation services to adopt a more expressive language that could allow the use of reasoning services.

Acknowledgments

This work has been partially supported by the EU-funded project BOEMIE (Bootstrapping Ontology Evolution with Multimedia Information Extraction) under the 6th Framework Programme (IST-FP6-027538).

References

[CEF+49] S. Castano, S. Espinosa, A. Ferrara, V. Karkaletsis, A. Kaya, R. Möller, S. Montanelli, G. Petasis, and M. Wessel. Multimedia Interpretation for Dynamic Ontology Evolution. *Journal of Logic and Computation*, Advance Access published on September 30, 2008. doi:10.1093/logcom/exn049.

[DEDT07] K. Dalakleidi, C. Evangelou, S. Dasiopoulou, and V. Tzouvaras. Domain Ontologies - version 2, 2007. Internet http://www.boemie.org/system/files/D3.5_v2.1.pdf.

[EKM+07] S. Espinosa, A. Kaya, S. Melzer, R. Möller, and M. Wessel. Towards a Media Interpretation Framework for the Semantic Web. In *Proceedings of the IEEE/WIC/ACM International Conference on Web Intelligence (WI'07)*, pages 374–380, Washington, DC, USA, November 2007. IEEE Computer Society.

[EKM08] S. Espinosa, A. Kaya, and R. Möller. On Ontology Based Abduction for Text Interpretation. In *Proceedings of 9th International Conference on Intelligent Text Processing and Computational Linguistics (CICLing-2008)*, number 4919 in Lecture Notes in Computer Science, pages 194–2005, Haifa, Israel, February 2008. Springer.

[FPT+08] P. Fragkou, G. Petasis, A. Theodorakos, V. Karkaletsis, and C. D. Spyropoulos. BOEMIE ontology-based text annotation tool. In *In Proceedings of the 6th Language Resources and Evaluation Conference (LREC 2008)*, 2008.

[NM06] B. Neumann and R. Möller. On Scene Interpretation with Description Logics. In H.I. Christensen and H.-H. Nagel, editors, *Cognitive Vision Systems: Sampling the Spectrum of Approaches*, number 3948 in LNCS, pages 247–278. Springer, 2006.

[PKM+07] S. Espinosa Peraldi, A. Kaya, S. Melzer, R. Möller, and M. Wessel. Multimedia Interpretation as Abduction. In *Proc. DL-2007: International Workshop on Description Logics*, 2007.

[PTK+06] S. Petridis, N. Tsapatsoulis, D. Kosmopoulos, Y. Pratikakis, V. Gatos, S. Perantonis, G. Petasis, P. Fragou, V. Karkaletsis, K. Biatov, C. Seibert, S. Espinosa, S. Melzer, A. Kaya, and R. Moeller. D2.1 Methodology for Semantics Extraction from Multimedia Content, 2006. Internet http://www.boemie.org/system/files/.