



**HAL**  
open science

## End-to-End Adaptive Framework for Multimedia Information Retrieval

Maria Sokhn, Elena Mugellini, Omar Abou Khaled, Ahmed Serhrouchni

► **To cite this version:**

Maria Sokhn, Elena Mugellini, Omar Abou Khaled, Ahmed Serhrouchni. End-to-End Adaptive Framework for Multimedia Information Retrieval. 9th Wired/Wireless Internet Communications (WWIC), Jun 2011, Vilanova i la Geltrú, Spain. pp.197-206, 10.1007/978-3-642-21560-5\_17. hal-01583629

**HAL Id: hal-01583629**

**<https://inria.hal.science/hal-01583629>**

Submitted on 7 Sep 2017

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

# End-to-end adaptive framework for multimedia information retrieval

Maria Sokhn, Elena Mugellini, Omar Abou Khaled, and Ahmed Serhrouchni\*

University of applied sciences of western Switzerland

\*Telecom ParisTech France

{`maria.sokhn,elena.mugellini,omar.aboukhaled`}@hefr.ch

\*{`ahmed.serhrouchni`}@telecom-paristech.fr

**Abstract.** The evolution of the web in the last decades has created the need for new requirements towards intelligent information retrieval capabilities and advanced user-oriented services. The current web integrates heterogeneous and distributed data such as XML database, relational database, P2P networks etc. leading to the coexistence of different data models and consequently different query languages. In this context, effective retrieval and usage of multimedia resources have to deal with the issues of creating efficient content based indexes, developing retrieval tools and improving user-oriented visualization interfaces. To that end we put forward an end-to-end adaptive framework based on an ontological model. This framework aims at enhancing the management, retrieval and visualization of multimedia information resources based on semantic techniques.

**Keywords:** knowledge management, querying, visualization, ontology, P2P network

## 1 Introduction

We witnessed in the last decades a massive growth of the available multimedia information on the web. This information presents several challenges in the information retrieval process. This is caused by the dynamic nature of the medium and the lack of advanced and precise methods that handle non-textual features [1]. This multimedia information lacks semantic content annotation, giving place to the so-called semantic gap. This gap was defined by Smeulders [2] and other papers such as [3] as “the lack of coincidence between the information that one can extract from the visual data and the interpretation that the same data has for a user in a given situation”. In recent years, semantic techniques such as ontologies have been adopted to enhance content based annotation services and, accordingly, retrieval and visualization of multimedia resources [1, 4, 5]. Indeed, as argued in Hare [6], the use of ontologies improves both automatic annotation and retrieval process. Using ontologies to describe multimedia resources provides methods to define well structured and related concepts facilitating the burden of annotation and retrieval [7]. Ontologies are also very useful for visual representation of resources [8, 9], and may considerably enhance browsing, navigation

and data accessibility. In our work we aim at handling the scientific conference multimedia information. We propose an end-to-end scalable solution that meets the needs of massively distributed and heterogeneous multimedia information produced by different services along a scientific conference.

This paper presents an end-to-end framework for multimedia retrieval based on semantic techniques. The remaining parts of the paper are structured as follows: In section 2, we present some of the related works. In section 3, we describe the framework CALIMERA. In section 4 and 5, we describe the use of the CALIMERA framework. Finally, in the section 6 we conclude the paper and present future works.

## 2 Related work

The semantic approach to existing web resources is one of the major challenges for building the semantic web. Few tools among those we found use systems that exploit the power of millions of users every day seeking information on the web. Today, many applications of file sharing peer-to-peer exist. These applications do not have a major characteristic: the semantic search. Although most applications of peer-to-peer sharing offer a keyword search based on indexing data, few are actually accessing the data content and use powerful annotation systems that provide detailed information on data. This situation is due to lack of use of metadata information generated by annotations and ontology models that describe the data content rather than a general description of the file.

Many publications and projects have been designed to manage multimedia information retrieval [7, 10–14]. Some of them, such as VIKEF [10], were developed in order to provide support for advanced semantic information, content production and knowledge acquisition, processing, annotation and sharing by empowering information and knowledge environments for scientific and commercial communities. MATTERHORN [14] produces recordings of lectures, manages existing videos, distribution channels and provides user interfaces for students with educational videos. These projects did not use the possibilities of a distributed system such as the P2P architecture, whereas other projects such as PHAROS[11] and SAPIR [13] use the power of P2P. However these projects do not make use of semantic techniques. In our work we propose a novel end-to-end adaptive framework named CALIMERA that takes advantage of semantic techniques and exploit the power of the peer to peer network.

## 3 CALIMERA Framework

CALIMERA stands for Conference Advanced Level Information ManagemEnt and Retrieval. This framework based on a ontological conference model and aims at facilitating the management and retrieval of the multimedia information generated within the scientific conferences [15]. The figure 1 outlines the global view of the framework architecture.

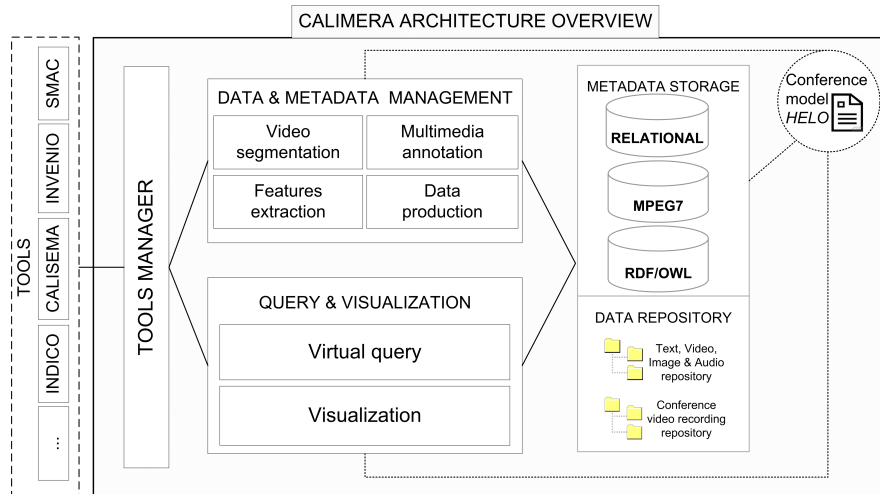


Fig. 1. CALIMERA architecture overview

### 3.1 Architecture overview

(1) **Tools manager:** CALIMERA is a tool independent framework. The tool manager allows any user to integrate a tool that may be used for data and metadata management, query and visualization or both. For a proof of concept we integrated four principal tools. 1.INDICO a tool developed by the CERN [22] that manages the administrative part of a conference (such as the conference planning, logistics, etc.). 2.SMAC is a tool we developed, to record conference talk and automatically segment the recording of this talk based on slide change detections. 3.CALISEMA is a tool for semi-manual and manual video annotation based on semantic techniques. 4.INVENIO is a set of modules for automatic features extraction, information indexing and retrieval.

(2) **Data and metadata management module:** It consists on handling the conference high-level information, such as recording talks, segmenting video recordings, annotating video segments, managing the context information of these talks, etc.

(3) **Data and Metadata storage:** It integrates existing data and metadata formats such as MPEG-7 [23] which is one of the most widely used standard for multimedia description, and RDF [24] and OWL [25], which are a more semantically oriented standards.

(4) **Query and visualization module:** It queries the data and metadata storage in order to return the video or the set of video sequences of the recorded talks that the users are seeking for.

(5) **Conference model, HELO:** It is an ontological model that describes and structures the information conveyed within a conference life cycle. the model has been designed to enhance the multimedia information retrieval and more precisely the scientific conference recordings. It is based on the effort made in

ESWC2006, ISWC2006 and SWC [26] and a study we conducted on user' needs and requirements when searching for scientific recorded talks. The work is detailed in the paper [16].

### 3.2 Layered model overview

In this section we introduce the proposed layered model (Figure 2) that maps the architecture of the framework presented in the previous section. The layered model describes the adaptive approach offered by the framework aiming to guarantee an end-to-end service. The model is composed of four main layers, the **Network layer**, the **Integrator layer**, the **Knowledge Management layer** and the **End User layer**. Each layer uses performed actions of the un-

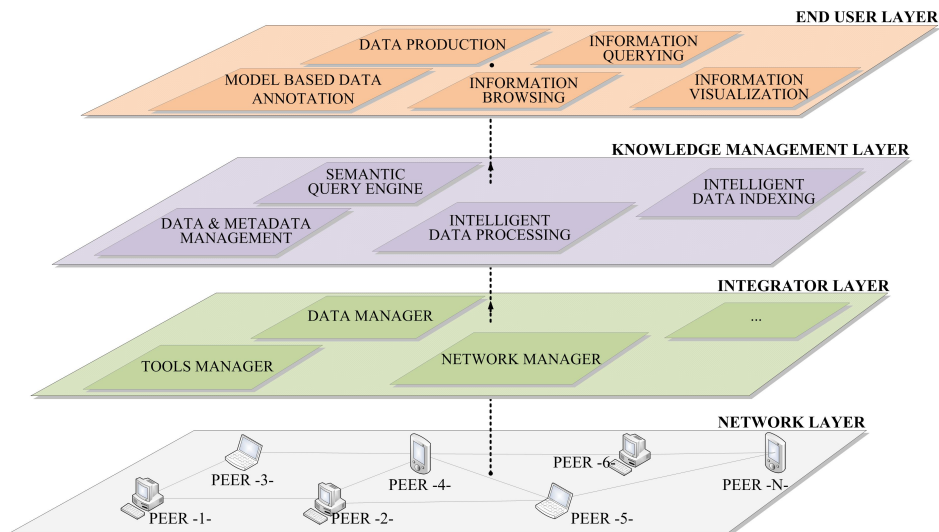


Fig. 2. CALIMERA layered model overview

derneath layer. The Network layer deals with the P2P protocols and network. The Integrator layer manages the peer integration within the network on one hand and the communication between the peers and the knowledge database on the other hand. The Knowledge Management layer handles the data and metadata management (processing, indexing, querying, extracting, etc.). This layer is a key point for the End User top layer which focuses on the users needs and performs modules such as data semantic annotation and information browsing and visualization.

## 4 Heterogeneous information retrieval

Multimedia information retrieval, more precisely scientific conference data retrieval, is enhanced by using a conference ontology model throughout the three main activities of an information retrieval process: the modal based annotation (section 4.1), the semantic querying process (section 4.2) and the multimedia results rendering/visualization (section 4.3).

### 4.1 Model based annotation

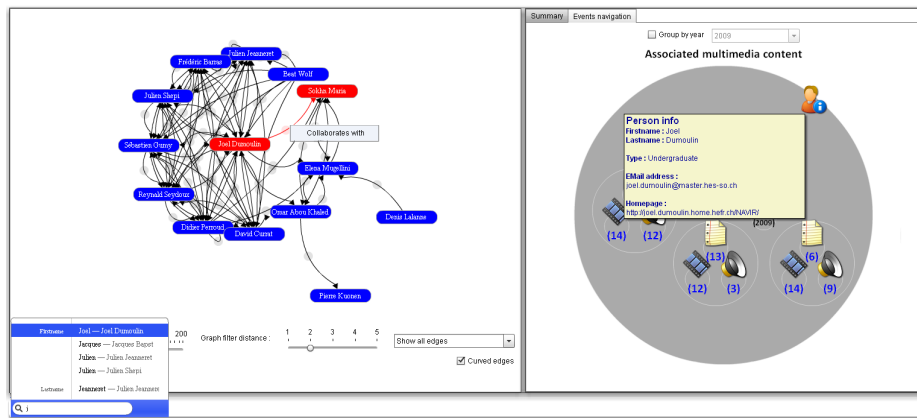
The Semantic web initiative seeks to add semantics to web resources in order to enhance their management, retrieval and access. In order to add semantics, web resources should move from *machine-readable* to *machine-understandable*. For that, effort should be provided to create content based description for these resources. This description is called annotation. Annotations are an important key-success in the emergence of the semantic web. Since manual annotation is a laborious task, it is important and required to develop annotation tools that facilitate this task. In our work we developed CALISEMA, a tool for video annotation: manually with a model based interface and automatically through a speech to text plug-in module. What characterizes it from other existing tools is that it is fully integrated and adapted to conferences. In fact CALISEMA is based on the conference model HELO and is integrated in the information management process of the framework CALIMERA. This enables the annotator to -automatically- collect information (e.g. name of the speaker, duration of the video, etc.) from other tools used in our framework as well as peers annotation. Moreover, it integrates several video segmentation algorithms (such as slide change detection algorithm) allowing users to choose the one that most meets their requirements. The annotation is exported in MPEG-7 format, OWL format or both.

### 4.2 Semantic query engine

CALIMERA allows the integration of heterogeneous data sources such as the output of the manual or automatic annotation, etc. These resources (may) have different data storage models. This leads to the coexistence of different models and scheme and therefore different query languages. The structural and semantic heterogeneity of data makes the development of custom solutions for querying these data time-consuming and complex. This issue can be divided into three main parts: dispatching important search information depending on the data sources, creating database-specific queries and merging results from several sources. To handle this issue we designed a query system named Virtual-Q. Virtual-Q aims to provide users transparent and easy access to retrieve data from heterogeneous sources. This system is a novel approach based on a virtual-query engine. This engine integrates concepts such as query analysis or sub-queries formulation, which facilitate transparent access to heterogeneous data. More details about this approach can be found in the paper [17].

### 4.3 Model based browsing & visualization

As mentioned earlier, using ontology is a powerful key in multimedia information retrieval. Still, to be effective the rendering of the query result has to be easy to read and simple to navigate through. To address this issue and further enhance the retrieval of scientific recordings of the talks, we designed a set of graphical renderings of information based on the semantic model HELO. Using this model for visualization, provides the users with an interface where the descriptions are grouped based on their use (the multimedia information can be searched by person, by date, by topic, etc.) , offering them the ability to explore the information content in an interactive way. These features become essential in multimedia retrieval which has a complex content hardly searched by “traditional” keyword-based searching. We implement a proof-of-concept prototype named NAVIR which focuses on information related to the scientific community member relationships and their associated multimedia data such as scientific publications, recorded talks, etc. (Figure 3). Once the media is retrieved, users



**Fig. 3.** NAVIR - Scientific community graph (left), Associated multimedia resources (right)

have the possibility either to play the video within the search interface (Youtube-like visualization) or to get redirected to a new window where the video of the recorded talk is played within a conference based visualization interface using our prototype SMAC. This visualization interface presents three distinct blocks (Figure 4): the recorded talk block on the left, the slides set block on the right and the information block on the bottom. The recorded talk is synchronized with the slides extracted from the presentation slides set. The navigation banner at the bottom of the slide set block allows the user to navigate through the slides. When the user chooses a slide the corresponding talk video sequence is played. Using SMAC to replay the recorded talks provides users with simple

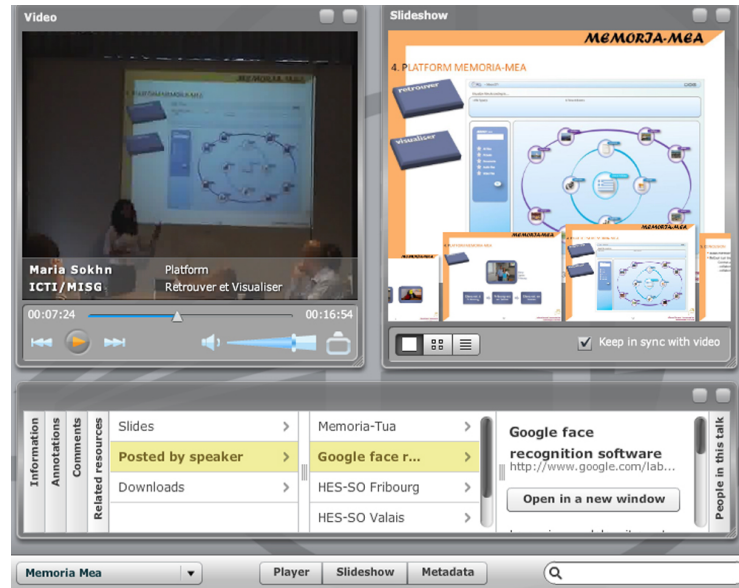


Fig. 4. SMAC visualization - Recorded talk replay interface

navigation through the different sequences of the talk, in addition to the display of information provided by the framework such as the annotations coming from the annotations of the peers or the output of the automatic metadata extraction modules.

## 5 Distributed information retrieval

The advances in wireless communication technology and the increasing number of mobile users networks brought a shift from the traditional client-server computing model to the P2P computing model. In this type of model, each peer takes both roles, client and server. In our work we addressed the issue of the distributed information retrieval. Every peer of the Network layer (Figure 2) may host a database to be shared with its peers in the network. Through the Network Manager of the Integrator layer every peer can access the P2P network. Each peer can make use of the Integrator layer to push information (through the Data Manager) into the Knowledge database of the framework and can integrates knowledge management tools into the framework. The peers make use of the End User layer to perform tasks such as the information browsing and visualization.

### 5.1 P2P network

We based our work on the JXTA framework [27], into which we introduced the RDF technology as a descriptor for available multimedia information. The



JXTA platform is a series of classes and methods for managing and transmitting application and control data between JXTA compatible peer platforms. Using our platform the user can annotate the data to be shared, based on the designed conference model HELO and can initiate traditional keyword search: The query is performed over the neighbors that search in turns for information related to the concerned keyword. The application is composed of the following modules: **Graphic User Interface (GUI)**: It allows the user to interact with the application by sharing and managing the files and performing search queries. **Controller**: It collaborates with the user interface to send the query to the query manager and manage the ontology using Jena, a Java framework for building Semantic Web applications [28]. **Query Manager**: The intermediate module between the application modules (GUI, Controller, Jena) and the network modules (JXTA manager, Download manager). **JXTA Manager**: It is responsible of the connections and the discovery of new peers in the network. **Jena**: The programming toolkit that manages tasks related to the semantic annotation. **Download Manager**: It is responsible of uploading and downloading the shared files.

## 5.2 Process over the P2P

We used in our work a decentralized network architecture. When the application is launched, the peer connects to the JXTA net peer group then sends advertisements to the peers on the network to exchange some information about their respective sockets. The classes and functions of the JXTA framework provide these functionalities. The class JXTAManager is responsible for connecting the peer to the network. It uses the NetworkManager class provided by JXTA framework and creates a multicast JXTA socket on each peer, allowing it to send and receive requests and responses. The application makes use of the Discovery Service offered by JXTA in order to periodically detect the new peers in the network and connect to them.

**Data annotation**: When the users want to share/annotate a file, they choose the type of file they want to share in the annotation tab of the interface (Figure 5). According to the choice, the fields to annotate are shown dynamically in the tab. The user may also be redirect to a specific annotation tools such as CALISEMA. Once the user submits the annotation, the Knowledge Management layer takes the lead to make the necessary updates.

**Information retrieval**: Our application allows any peer to join the network. In our scenario, we have two *types* of peers: the *normal* peers that represent the users that might want to share/annotate files and search for information and the *INVENIO servers* peer that acts like a super peer integrating the Metadata Storage and Data Repository of CALIMERA (ref. Figure 1). The search is performed over all the connected peers through the semantic query engine. The address of the results is mention in the search interface: Resources coming from

the super peer are designated by "invenio:IPadress", resources coming from a *normal* peer are designated by the "IPadress" only (Figure 5).

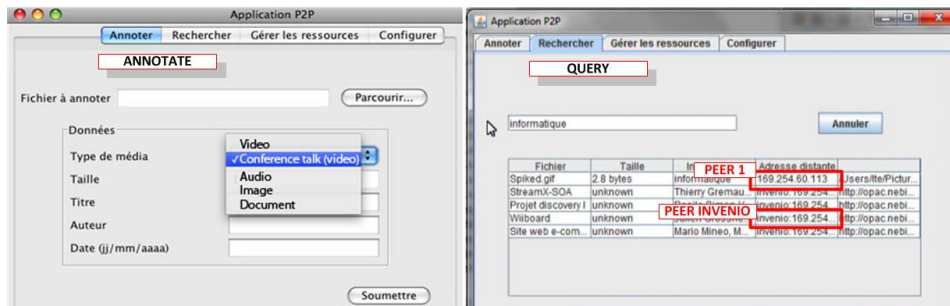


Fig. 5. Distributed data annotating and querying

## 6 Conclusion and future work

In this paper we presented an end-to-end adaptive framework that meets the needs of massively distributed multimedia information produced by different services along a conference life cycle. This end-to-end solution is supported by the framework CALIMERA. A framework for multimedia information retrieval based on the ontological model HELO and a layered approach model. HELO was built on the base of a study that analyses the user's needs and requirements when searching for videos of recorded talks. The model is used over the layered model to perform and build efficient content based annotations and to enhance browsing, navigation and data accessibility. As next steps, we intend to evaluate the proposed frameworks with its different modules. This evaluation focuses mainly on four axes: 1- The ontological model HELO: This evaluation should validate the adaptability of the scopes and their coherence from different users point of view. It should also measure the facility of learning and using this model. 2- The semantic search interface: This interface should be evaluated in a set of experiments to verify both its expressiveness and its effectiveness. 3- The multimedia rendering and browsing interfaces: This evaluation should compare the proposed interfaces versus the existing ones in different retrieval situations. 4- The information retrieval over the P2P network: This evaluation should analyze the results efficiency and pertinence and the added value compared to the use of the framework without the P2P network.

## References

1. Andrew D. et al., Semantic annotation and retrieval of video events using multimedia ontologies. In ICSC, p. 713-720, (2007)

2. Smeulders A. et al., Content-based image retrieval at the end of the early years, IEEE Transactions on Pattern Analysis and Machine Intelligence, (2000)
3. Liu Y. et al., A survey of content-based image retrieval with high-level semantics, (2007)
4. L. Hollink et al., Building a visual ontology for video retrieval. 13th annual ACM international conference on Multimedia, Singapore (2005)
5. Knud, M. et al., The ESWC and ISWC Metadata Projects. In ISWC2007, pp.802–815, Korea (2007)
6. Hare J. S. et al., Bridging the semantic gap in multimedia information retrieval - top-down and bottom-up approaches. In ESWC2006, Budva (Montenegro), (2006)
7. Bertini, M. et al. Soccer Video Annotation Using Ontologies Extended with Visual Prototypes. In CBMI07, Bordeaux (2007)
8. Vladimir G. et al. Visualizing the Semantic Web: XML-based Internet and information visualization. (2006)
9. Heiko P. et al., Ontology-based information visualization in integrated UIs, IUI2011, ACM, (2011)
10. An advanced software framework for enabling the integrated development of semantic-based information, content, and knowledge management systems. [http://cordis.europa.eu/ist/kct/vikef\\_synopsis.htm](http://cordis.europa.eu/ist/kct/vikef_synopsis.htm)
11. Laurier C. et al., Pharos: an audiovisual search platform using music information retrieval techniques (2009)
12. Huang W. et al., Conkml: A contextual knowledge management framework to support multimedia e-learning, (2006)
13. SAPIR - Search In Audio Visual Content Using Peer-to-peer IR, <http://www.sapir.eu/>
14. MATTERHORN - Platform to support the management of educational audio and video content, <http://www.opencastproject.org/project/matterhorn>
15. Sokhn M. et al., Knowledge management framework for conference video-recording retrieval. SEKE2009, Boston, July (2009)
16. Sokhn M. et al., Conference Knowledge modelling for conference-video-recording querying and visualization, Journal of Multimedia Processing and Technologies, Vol.1 No.2 June (2010)
17. Sokhn M. et al., Virtual-Q: Querying over heterogeneous and distributed data sources, AWIC2011. Advances in Soft Computing series, Springer. January (2011)
18. AKT, <http://www.aktors.org/publications/ontology/>
19. Conference ontology by Jen Golbeck <http://www.mindswap.org/~golbeck/web/www04photo.owl/>
20. eBiquity Conference Ontology, <http://ebiquity.umbc.edu/ontology/conference.owl/>
21. Semantic Web Technologies ESWC2006, <http://www.eswc2006.org/technologies/ontology-content/2006-09-21.html>
22. CERN, Centre Europeen de Recherche Nucleaire, <http://www.cern.ch>
23. General introduction to MPEG-7, <http://mpeg.chiariglione.org/standards/mpeg-7/mpeg-7.htm>
24. RDF, Resource Description Framework, <http://www.w3.org/RDF/>
25. OWL, Web Ontology Language, <http://www.w3.org/TR/owl-features/>
26. Semantic Web Conference Ontology, [http://data.semanticweb.org/ns/swc/swc\\_2009-02-27.html](http://data.semanticweb.org/ns/swc/swc_2009-02-27.html)
27. The Language and Platform Independent Protocol for P2P Networking, Semantic Web Conference Ontology, <http://jxta.kenai.com/>
28. Jena A Semantic Web Framework for Java, <http://jena.sourceforge.net/index.html>