A schema-based P2P network to enable publish-subscribe for multimedia content in open hypermedia systems

Thanassis Tiropanis*

Athens Information Technology (AIT)
19.5km Markopoulo Ave.
GR-19002, Peania, Athens, Greece
E-mail: ttir@ait.edu.gr
*Corresponding author

Dimitris Kanellopoulos

Department of Mathematics
University of Patras
GR-26500, Rion, Patras, Greece
E-mail: dkanellop@teipat.gr

Abstract: Open Hypermedia Systems (OHS) aim to provide efficient dissemination, adaptation and integration of hyperlinked multimedia resources. Content available in Peer-to-Peer (P2P) networks could add significant value to OHS provided that challenges for efficient discovery and prompt delivery of rich and up-to-date content are successfully addressed. This paper proposes an architecture that enables the operation of OHS over a P2P overlay network of OHS servers based on semantic annotation of (a) peer OHS servers and of (b) multimedia resources that can be obtained through the link services of the OHS. The architecture provides efficient resource discovery. Semantic query-based subscriptions over this P2P network can enable access to up-to-date content, while caching at certain peers enables prompt delivery of multimedia content. Advanced query resolution techniques are employed to match different parts of subscription queries (subqueries). These subscriptions can be shared among different interested peers, thus increasing the efficiency of multimedia content dissemination.

Keywords: Open Hypermedia Systems; OHS; semantic annotation; publish-subscribe P2P; query languages; query.

Reference to this paper should be made as follows: Tiropanis, T. and Kanellopoulos, D. (2008) 'A schema-based P2P network to enable publish-subscribe for multimedia content in open hypermedia systems', *Int. J. Web Engineering and Technology*, Vol. 4, No. 1, pp.21–43.

Biographical notes: Thanassis Tiropanis is an Assistant Professor of Software Systems and Web Applications at the Athens Information Technology where he is Co-founder and Co-head of the Software Engineering and Web Applications Research Group. He holds a DipIng in Computer Engineering and Informatics from the University of Patras and a PhD in Computer Science from the University of London. He teaches courses on web technologies and e-commerce. His research interests are in the areas of web technologies,

semantic web, grid computing, P2P systems and applications in e-commerce and e-learning. His work has been presented and published in international conferences and journals.

Dimitris Kanellopoulos graduated from the Department of Electrical and Computer Engineering of the University of Patras, Greece, in 1990. He holds a PhD degree in Multimedia Communications from the same department. Currently, he is with the Department of Tourism Management at the Technological Educational Institute of Patras. He is also a member of the Educational Software Development Laboratory in the Department of Mathematics at the University of Patras. His research interests include multimedia communications, web engineering, P2P systems and applications, intelligent tourism information systems and knowledge representation. He has authored numerous papers in international journals and conferences in these areas. He serves as an editor in several academic journals.

1 Introduction

A Peer-to-Peer (P2P) infrastructure provides a suitable architecture for distributed content management as it offers user-centred, data-centred and computing-centred models. Recent research in P2P systems focuses on providing techniques for evolving from basic P2P networks, supporting only file exchanges that use simple filenames as metadata, to more complex systems such as schema-based P2P networks, capable of supporting explicit schemas to describe knowledge, usually using RDF and thematic ontologies as metadata (Nejdl *et al*., 2003). Schema-based P2P systems are the next step in distributed P2P networks with advanced data management and storage features. RDF-based P2P networks allow complex and extendable descriptions of resources instead of fixed and limited ones, and they provide query facilities against these metadata instead of simple keyword-based searches.

The work by Benatallah *et al*. (2006), as well as by Zhou *et al*. (2003), demonstrates that efficient discovery of hypermedia resources in a P2P network is feasible in the context of Open Hypermedia Systems (OHS). Thus, we can efficiently obtain links to hypermedia resources by employing semantic descriptions of topics and content.

Nevertheless, there are additional issues in OHS over P2P that need to be addressed (Zhou *et al*., 2007). These issues are:

- Availability of hypermedia content on P2P networks cannot be guaranteed given the *ad hoc* nature of a P2P network where peers can join and leave the network at any time.
- Just-in-time delivery of bandwidth-demanding hypermedia content is often a challenge in P2P networks, when a peer needs to obtain content that is not widely available and is located at remote peers.
- Availability of up-to-date hypermedia at a given time cannot be guaranteed in P2P networks since peers hosting the most recent content may be offline.

To address the above challenges, we propose a publish-subscribe approach combined with caching for efficient OHS over P2P networks. With caching, the availability of content will be ensured even after the peer from whom the content originates goes offline. By employing subscription mechanisms, content that is relevant to a number of peers can be cached at other peers near them thus enabling just-in-time delivery of multimedia content. The availability of up-to-date content can be facilitated given the proposed publish-subscribe and caching mechanisms.

Further, the use of ontologies will enable the precise description of available hypermedia resources and resource discovery as well as the precise specification of the capabilities of peers. Ontology-based, semantic queries can be decomposed to enable subscription aggregation for more efficient use of network resources, while the use of conceptual models and integration models can provide for synthesising content subscribed to in specific contexts.

In this paper we propose an unstructured P2P architecture that employs semantic annotation of multimedia resources and peers, and publish-subscribe mechanisms for efficient discovery and delivery of rich and up-to-date hypermedia content. This is to provide an approach for OHS over publish-subscribe P2P networks. The difference from existing approaches on OHS over P2P lies (a) on the use of publish-subscribe mechanisms and caching and (b) on aiming for a comprehensive framework to enable OHS over P2P.

Our architecture was inspired by Benatallah *et al*. (2006), Zhou *et al*. (2007), and Lewis *et al*. (2004), whose works helped us formulate a publish-subscribe approach to content discovery and delivery for OHS over a P2P infrastructure. In addition, our architecture proposes the consideration of a data integration framework in the above context.

The remainder of the paper is organised as follows: Section 2 presents the state of the art on publish-subscribe and OHS over P2P and discusses it in the context of the concerns addressed in this paper. Section 3 presents the proposed architecture and its elements, namely, peer and super-peer OHS servers. The interfaces each peer/super-peer will have to expose and use are identified. In addition, Section 3 proposes an ontological structure for the peers/super-peers and outlines mechanisms for semantic query-based subscriptions on the proposed P2P network, as well as mechanisms for the P2P network configuration. Section 4 provides an example of employing the proposed architecture in the domain of the tourism industry. Complex queries are employed in this example to demonstrate the potential of the proposed architecture for efficient delivery of up-to-date information based on semantic query publish-subscribe. Section 5 discusses challenges and open issues, and Section 6 concludes the paper.

2 Related work

2.1 Querying and searching in P2P-based hypermedia systems

Benatallah *et al*. (2006) have provided query-matching algorithms that can be employed in the context of peer relationships among (but not only) e-catalogue communities. Complex queries formulated in the context of one community can be fully or partially matched to queries in different contexts by other communities; this may require reformulating the initial query. Thus, it is possible to determine the e-catalogue communities that match a query (or part of it) and it is the responsibility of the peer from which the query originated to choose the combination of potentially reformulated queries, choose the peers to query and synthesise the obtained results.

Zhou *et al*. (2004) focused on searching in P2P-based hypermedia systems and Zhou *et al*. (2007) identified the advantages of P2P architectures for OHS. Considering the potential of P2P systems for supporting activities among groups of people, Zhou *et al*. (2007) propose a web-based Distributed Dynamic Link Service (DDLS) that addresses issues of "routing queries to located resources, organising and manipulating resources to facilitate discovery and presentation and enhancing resource discovery and performance". In this direction, a search algorithm has already been proposed (Zhou *et al*., 2003) that is based on a notion of semantic equivalent data to rearrange the network topology in order to cluster semantic-related entities. This is achieved by letting each peer publish a topic list. Other peers use these lists to maintain information of neighbours that hold semantically equivalent topics. Zhou *et al*. (2007) also recognise that in such *ad hoc* environments "resources available at any particular time is unknown and unpredictable, and both services and resources from a particular user are more probabilistic than deterministic".

2.2 Publish-subscribe systems

There have been proposals on how publish-subscribe infrastructures can be implemented over P2P networks using semantic query-based routing (Lewis *et al*., 2004). These proposals envisage an infrastructure consisting of Information and Collaboration Spaces (ICS), which are set up and operated by user communities. A network of Semantic Query-Based Routing servers facilitates subscriptions to information resources on behalf of the communities to populate their corresponding ICSs with relevant content. Additionally, caching is employed to make sure that the community members of each ICS will efficiently access bandwidth-demanding multimedia content.

More than a few distributed RDF metadata management systems provide publish-subscribe mechanisms. Chirita *et al*. (2004) proposed a P2P RDF publish-subscribe system on a super-peer-based RDF P2P network. Keidl *et al*. (2002) proposed a distributed RDF metadata management system that supports caching and replication. Their system is called MDV and is based on a 3-tier architecture. In the MDV framework, a filter algorithm was implemented based on relational database technology that powerfully computes all subscribers for created, updated and deleted RDF data. There are also several other publish-subscribe systems that support topic-based or type-based subscriptions (Castro *et al*., 2002; OMG, 2004; Graham *et al*., 2006). In addition, recent progress in distributed hash tables enables a new category of publish-subscribe systems (Tam *et al*., 2004) that neither relies on any centralised server nor subscription broadcasting.

2.3 OHS architectures

The strength of hypermedia systems lies in their ability to combine content in different formats with hyperlinks, thus enabling the creation of non-linear media. OHS try to establish a consensus on the way in which hypermedia systems are deployed to ensure that they are open. Openness requires that no specific data model format (coupling structure and content) is imposed on an OHS; only the use of a specific structure format is necessary (Wiil, 1997). OHS requires discussion and agreement on communication protocols to be employed among hypermedia systems, hypermedia system architectures, hypermedia services, hypermedia linking structures, retrieval structures or scenarios for hypermedia use. The Open Hypermedia Systems Working Group (OHSWG) is coordinating such effort. Work on the design of an open hypermedia protocol has also been undertaken (Reich *et al.*, 1999). A number of research areas for open hypermedia have been identified (Nürnberg *et al.*, 1998) addressing the human perception and use of hypermedia, hypermedia client-servers, structure servers and hypermedia systems backend. An approach for offering OHS to the WWW has also been proposed by Karousos *et al.* (2003).

OHS systems have been developed in different domains. Chimera is an OHS that employs hypermedia to support software development environments (Anderson *et al.*, 2000) and provides extensibility mechanisms (Anderson, 2001).

2.4 P2P and OHS

P2P systems can be categorised into:

- centralised ones such as Napster
- unstructured ones such as Gnutella
- structured ones such as Chord (Zhou *et al.*, 2007).

The placement of data in unstructured P2P networks does not require knowledge of the network topology and thus resembles real-life situations better (Zhou *et al.*, 2007). Centralised P2P systems present challenges when it comes to scalability. In structured P2P, the peers on which data is placed are controlled and although such networks are efficient for locating resources, they do not present efficiency when it comes to serving complex queries. Therefore, unstructured P2P systems could provide better support for semantic-based complex queries, if the issue of discovery and prompt delivery of such resources is sufficiently addressed.

In an envisaged unstructured P2P network with OHS servers as peers, the peers are independently operating nodes. The information stored by each peer and the links to it are accompanied by semantic descriptions. In addition, the capabilities of the nodes are also semantically described. Therefore, we can consider an unstructured P2P knowledge network among OHS servers that provides publish-subscribe services.

In this network, nodes can join or disconnect at any moment. A node acts as a peer in the network and provides access to knowledge sources (resources for which semantic descriptions are available) in the P2P network and may need to interact with other peers to achieve its goals. The resources to which a peer can offer access can be multimedia ones placing requirements in terms of adaptation. Thus, an OHS server must be designed to meet the following requirements:

- supporting the distribution of information within the network
- supporting query answering and routing
- providing multiple views on available information
- enabling access to multiple sources of information
- supporting adaptation of information view according to the user's personalisation requirements.

The semantic descriptions of peers and resources will enable improved resource discovery and multimedia resource delivery. In a schema-based P2P network, any metadata model needs to reflect the following requirements:

- *Integration* – Each piece of knowledge requires metadata about its origin. When retrieving external information, additional information about where each piece of information was obtained needs to be captured. This information will allow identifying a peer and locating resources in its repositories.
- *Information heterogeneity* – As information is added from a variety of peers, inconsistencies may occur in a local repository. For example, if two peers have different information about an element, inconsistencies may occur in the local repository of a peer. Therefore, each piece of information needs to be assigned a confidence rating in order for the system to be able to handle heterogeneity and provide useful information. In the same way, a level of trust can be assigned to peers to model their reliability. The level of trust (that the trusting peer has in the trusted peer at a given time slot with a given type of association relationship) is depicted by the trustworthiness that is computed in Hussain *et al*. (2004) and Dillon *et al*. (2004). Nejdl and Olmedilla (2004) considered the problem of passing credentials in P2P systems. More specifically, they recommended sharing credentials with third parties if allowed by trust negotiation strategies (Yu *et al*., 2003). In the PeerTrust project (Basney *et al*., 2004), trust negotiation in semantic web and P2P environments is investigated. Digital credentials can be signed XML or RDF statements that express peer properties, and policies are expressed as logic programs that tie resource access to required credentials. Besides, the semantic interoperability is addressed through ontology mapping. This is a process whereby two ontologies are semantically related at conceptual level, and the source ontology instances are transformed into the target ontology entities according to those semantic relations. An interesting approach to ontology mapping has been taken in the GLUE system (Doan *et al*., 2003). Furthermore, as each peer uses its own local ontology, mapping may be required to avoid problems of heterogeneous labelling for certain objects.
- *Security* – Some information may be of private nature and should not be visible to others peers. Other information may be restricted to a specific set of peers. A metadata model needs to provide the means to express such security policies.
- *Caching* – The availability of other peers is not always guaranteed within P2P systems. In particular, some peers may have better connectivity to the rest of the network in terms of bandwidth than other peers. Therefore, to improve network efficiency, caching of information is needed. The caching mechanisms need to be transparent to the user, but must be captured by the proposed metadata model.
- *Adaptability* – Information concepts can also be used as a basis for implementing the adaptive behaviour of an OHS server node. Consequently, context-specific configuration of information elements and their adaptation according to the user's personalisation requirements can be enabled as semantic metadata are attached to elements. To achieve this, the ontologies used must be aligned to the ontologies defining the context and user profile.

In OHS the following features affect collaboration, and need to be addressed by appropriate techniques (Castano *et al*., 2005):

- dynamism of the system, considering the fact that the peers are allowed to join and leave the network at any moment
- autonomy of peer, in that each peer is responsible for its own information resource management and representation
- absence of *a priori* agreement about ontology specification vocabulary and language to be used for knowledge specification
- equality of responsibilities, no centralised entities with coordinating tasks are recognised and each peer enforces interaction facilities with other peer for resource sharing and collaboration.

Peers are not usually created equal but have different characteristics with respect to their capabilities (*e.g.*, bandwidth, storage space or processing power). Exploiting the different capabilities in a P2P network can lead to an efficient network architecture, where a small subset of peers, called super-peers, takes over specific responsibilities for peer aggregation, query routing and possibly mediation (Yang and Garcia-Molina, 2002).

2.5 Data integration in the context of OHS

The main issue in Web Information Systems (WIS) design is the specification of the hypermedia generation process including the specification of which heterogeneous resources to use and how to map their data (integration) into the system. Vdovjak *et al.* (2003) presented their model-driven approach called Hera and introduced the transformation software that builds the heart of the hypermedia presentation process. Hera distinguishes three parts in the design of a WIS: integration, data retrieval and presentation generation.

The integration part manages the gathering of data from different sources on the basis of source ontologies and the mapping between those source ontologies and the Conceptual Model (CM) of the WIS. The data retrieval part handles the user queries and produces the data that represents the query results. In the presentation generation part, this query result is transformed into a web presentation and that presentation is adapted to the user platform (*e.g.*, HTML, WML or SMIL).

The CM of a WIS serves as an interface between data retrieval and presentation generation. It provides a uniform semantic view over multiple data sources and it is composed of concepts and concept properties that together define the domain ontology. During the ‘integration and data retrieval phase’, several autonomous sources are connected to a CM by creating channels through which the data will populate the concepts from a CM on request. There are two types of concept properties: (1) concept attributes, which associate media items to concepts and (2) concept relationships, which define associations between concepts. A CM is expressed using two descriptions: CM properties and system media types. CM properties define the cardinality and inverse of concept relationships, while system media types define the multimedia ontology. The root of this ontology is the `Media` class, which is further refined in the `Text` and the `Image` classes. `Text` is characterised by the length attribute (expressed in number of characters) and `Image` has width and height attributes (expressed in pixels) plus a string containing a URL of the actual picture. `Text` has two subclasses: `Integer` and `String`.

Their Integration Model (IM) addresses the problem of relating concepts from the source ontologies to those from the CM. This problem can also be seen as the problem of merging or aligning ontologies. To automate the solution to this problem, the Integration Model Ontology (IMO) is proposed, which is a meta-ontology describing integration primitives that are used both for ranking the sources within a cluster and for specifying links between them and the CM. In their IMO, the main concepts are decoration and articulation. Decorations serve as a means to label 'appropriateness' of different sources (and their concepts) grouped within one semantically close cluster, while articulations describe actual links between the CM and the source ontologies and clarify also the notion of the concept's uniqueness, which is necessary to combine several sources.

Adopting the Hera approach in our schema-based P2P network, several CMs can be revealed/formed on request. In addition, complex user queries can generate query results, which are transformed to web presentations. Every time, the multimedia content of these presentations can be fully adapted to different user platforms. From another perspective, many semantic web applications, such as Shared-HiKE, need a distributed RDF infrastructure in order to let their users to create, organise and share RDF data. For example, in Shared-Hierarchical Knowledge Editor (HiKE) (Cai *et al.*, 2004), a collaborative HiKE lets users create, organise and share their RDF data. Each participant has his/her local hierarchical knowledge and also shares the external knowledge from other participants.

The combination of the IM model (Vdovjak *et al.*, 2003) with schema-based P2P publish-subscribe networks is proposed in our work. By combining Adaptive Hypermedia (AH), semantic web and P2P technologies in the information management context, novel value-added services can be provided enabling efficient use of network resources for multimedia content exchange and delivery.

A combined semantic web and P2P solution can provide proper peer selection service in order to receive the right answers without flooding the network with queries. For this purpose, ontology-based peer selection mechanisms can exploit similarity of available ontologies. As different peers may use different, though overlapping, ontologies alignment, mapping and visualisation tools need to be applied. This is achieved by using methods of the 'emergent semantics' field (Madche *et al.*, 2002). Emergent semantics deal with assigning semantic categories and relations to objects by observing user interactions or patterns of user behaviour, taking into account the user's current context. For example, a file categorised to different concepts indicates alignment. Emergent semantics are closely related to adaptive hypermedia for the provision of personalised services. In addition, the emergent semantics field is related to ontology learning where the aim is to reduce the overhead of building up and maintaining ontologies. Ontologies have to update themselves to cope with ontological drift. In fact, users will often not know what is in the ontologies on their machine. Finally, an inference engine is necessary to ask and answer queries to peers in a robust, scalable, often locally contained manner.

2.6 Web services and OHS

The description of services available in the context of OHS as web services has been under consideration (Pandis *et al.*, 2005). There are proposals for a semantic description of Web Services (WS) such as Web Service Modelling Ontology (WSMO) (Roman *et al.*, 2005) and OWL-S (Dean and Schreiber, 2004).

The WSMO (Roman *et al*., 2005) describes all relevant aspects related to WS with the ultimate goal of enabling the automation of tasks such as discovery, selection, composition, mediation, execution, monitoring, *etc*. Novel techniques have been proposed to facilitate semantic discovery and interoperability of WSs that manage and deliver web media content (Sakkopoulos *et al*., 2006). WSMO has its conceptual basis in the Web Service Modelling Framework (WSMF) (Fensel and Bussler, 2002) refining and extending this framework and developing a formal ontology and set of languages.

OWL-S is an ontology for service description based on the Web Ontology Language (OWL) (Dean and Schreiber, 2004). It can facilitate the design of semantic WS and can be considered as "a language for describing services, reflecting the fact that it provides a standard vocabulary that can be used together with the other aspects of the OWL description language to create service descriptions". The OWL-S ontology consists of the following parts:

- a service profile for advertising and discovering services
- a process model that describes the operation of a service in detail
- the grounding that provides details on how to interoperate with a service, via messages.

The vocabulary defined by OWL-S may be used to provide semantic annotations of web services, and automatic agents may process this information (Hendler, 2001). The adoption of OWL-S for providing self-descriptive hypermedia services has already been proposed (Pandis *et al*., 2005).

3 The architecture

The proposed architecture is an overlay P2P network with OHS servers as peers. This network is unstructured and enables efficient exchange of multimedia content among peers by employing publish-subscribe techniques (Lewis *et al*., 2004) and semantic query resolution (Benatallah *et al*., 2006). Communication among peers is facilitated via web service ports exposed by the peers, which provides for interoperability. The CM and IM (Vdovjak *et al*., 2003) are used for the integration of multimedia content received at the OHS side according to an IMO.

In addition, the proposed P2P architecture considers peers and super-peers (Yang and Garcia-Molina, 2002) that collaborate in the context of the P2P architecture by exposing and using web service ports.

Such an approach provides for scalable, efficient, relevant, up-to-date and content-rich OHS for the following reasons:

- By employing semantic annotation of resources, it is possible to match subscription queries (or parts of them) to the content of certain OHS servers.
- By employing advanced query resolution, subscription queries (or parts of queries) can be reformulated to match potentially different ontologies, which are used to annotate relevant multimedia resources.

- By adopting a publish-subscribe model for the multimedia content exchange among OHS servers, it is possible to have OHSs that will be content-rich, up-to-date and relevant. New content is integrated as soon as it becomes available. If certain peers go offline it is likely that the content they provided when connected will have been cached by the network.
- By adopting an unstructured P2P architecture, a scalable network of OHS servers can be formed. As new OHS servers join the network, their introduced content in the network will be automatically matched to existing subscription queries, thus making the new resources immediately available to interested peers.
- By using a CM and an IM, content received through subscriptions to various OHS servers can be contextualised. As the addition of OHS servers enriches the set of query results, the CM and IM can potentially be updated by the OHS server administrators.
- By employing super-peers along with peers, the P2P publish-subscribe network can become more efficient and scalable. Certain responsibilities of super-peers such as ontology matching or subscription aggregation can add such value to the P2P network.
- By employing web services for the interaction among peers and super-peers, the ease of introducing new peers to the network and its scalability are further ensured.

The two kinds of peers in the proposed architecture are:

1 OHS servers (peers)

2 Advanced OHS servers (super-peers).

The following sections describe these peers and present different scenarios with regard to the operation of the P2P publish-subscribe network.

3.1 OHS servers (peers)

The ordinary peers in the proposed architecture provide access to semantically annotated information via specific web service ports. These web service ports provide access to WS, which are interoperable software components that can be used in application integration and component-based application development.

In particular, the OHS servers provide the following types of ports:

- *Peer status and ontology management ports*; for providing information about the peer, its identification or the Quality of Service (QoS) levels supported. In addition, these ports are used for obtaining the ontologies used for the CM (as per Vdovjak *et al*., 2003) that the OHS server employs or for enabling mapping among the ontologies used in the OHS and those of other domains. These ports can support a number of different operations such as:

  `isSetOfOntologiesSupported`, or
  `returnOntologiesUsed`, or
  `addOntologyMapping`, *etc*.

- *Query resolution ports*; for the resolution of RDF queries (other query languages may also be supported).
- *Subscription management ports*; for obtaining the set of queries for which the particular OHS server provides with subscriptions other peers or super-peers and for managing those subscriptions (add new ones, delete or modify old ones, *etc*.).

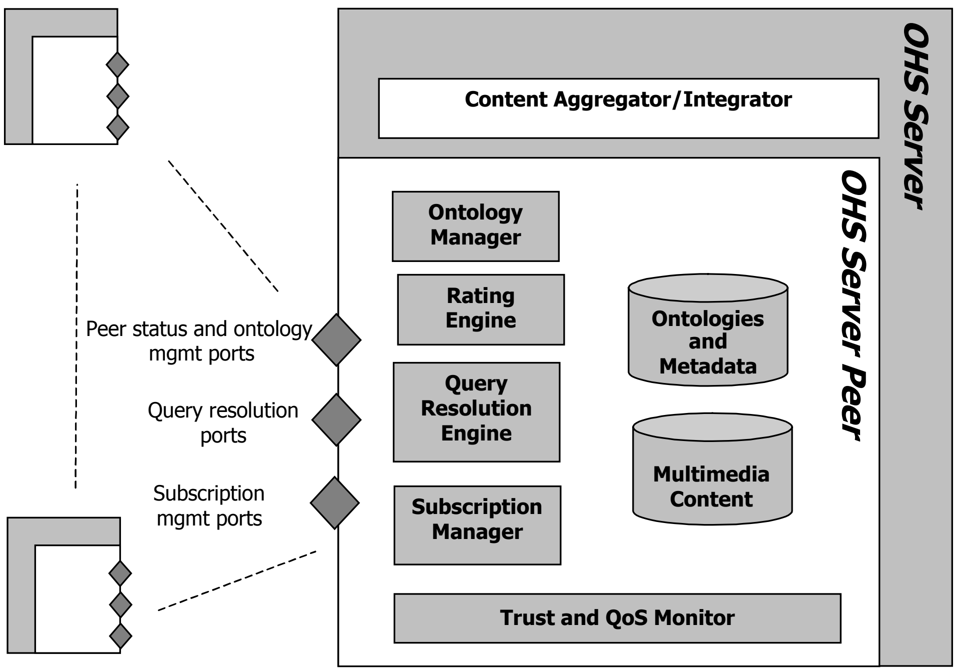
The proposed WS can be ontologically described using the WSMO (Roman *et al*., 2005) or the OWL-S (OWL-S Services Coalition, 2004).

Using the above WS ports, a peer can respond to four types of queries:

1. *Peer status queries* (peer status and ontology management ports) for establishing the status of the peer, the supported QoS level, *etc*.
2. *Scoping queries* (peer status and ontology management ports) that enable the determination of the degree of relevance of a peer to a specific domain/context of another peer.
3. *Query resolution queries* (query resolution ports) for determining whether a subscription query (or part of it) can be matched by the resources available by the particular peer.
4. *Subscription management queries* (subscription management ports) for setting up and managing subscriptions to specific queries on this peer.

Figure 1 illustrates the proposed architecture of an OHS peer. As shown on this diagram, each peer exposes ports to other peers to implement the proposed P2P publish-subscribe mechanisms.

Figure 1 The OHS server peer architecture



Each peer stores multimedia content, links, ontologies and metadata, which are processed by a content aggregator or integrator to populate the OHS server (peer) with rich and relevant content. Each OHS peer implements logic in terms of Ontology Manager, Rating Engine (to rate different subscriptions for a given query based on QoS and Trust metrics), Query Resolution and Subscription Manager. At the same time, each peer implements a Trust and QoS monitor for given subscriptions or connections to other peers.

3.2 Advanced OHS servers (super-peers)

In the proposed architecture the super-peers expose the same WS ports as the ordinary peers. However, super-peers can have additional capabilities such as:

- They can provide ontology mappings for a larger number of ontologies than the standard community-operated peers. Therefore, they can respond more frequently to scoping queries.
- They can provide more advanced and efficient query resolution than ordinary peers.
- They can establish and maintain a larger number of subscriptions to other peers than the standard peers of the network.
- They can provide subscriptions to complex queries (if necessary) by making separate subscriptions to match different parts of a complex query of a requesting peer, while offering a single subscription to the requesting peer.
- They can monitor their connection to other super-peers or peers and record the experienced QoS during multimedia content exchange.
- They can frequently monitor the P2P network and ask other peers or super-peers to modify their subscription portfolios to ensure optimised flow of content; this can be achieved by asking other peers or super-peers to perform subscription aggregation taking into account the recorded QoS levels.
- They can frequently monitor the P2P network using peer status and scoping queries, in order to discover whether newly introduced peers offer resources that match existing subscription queries or parts of them.

3.3 Ontological structure for the peers/super-peers

The ontological structure for peers includes the following metadata:

- *Peer* – Each exchanged resource originates at a peer. The local repository of each peer stores information about the peer itself and other peers in the network with which resources are exchanged.
- *peerID* – Each peer has a unique ID. This is the first attribute stored within the peer object, and it can be a JXTA UID. JXTA is a set of XML-based protocols to cover typical P2P functionality (Gong, 2001). It is a platform-independent set of open protocols to facilitate P2P communication and application development. JXTA provides a Java binding offering a layered approach for creating P2P applications. It provides P2P protocols and services, including peer discovery and peer groups.

- *peerLabel* – It is the peer label, which is a human readable description of the peer.
- *peerTrust* – It is a measure of peer trust. Some peers might be more reliable than others. The value of peerTrust is defined to be between 0 and 1, with 0 meaning that the peer is not reliable at all and 1 meaning that the peer is very trustworthy.
- *Cache* – Caching of information is necessary to increase the network efficiency. Information about the cache includes its location, its capacity, its content and its update policies.

For each resource available at a peer, the following information is maintained:

- *URI* – To keep track of the origin of a resource its primary Universal Resource Identifier (URI) is explicitly saved among the metadata. Across the network a resource object can be identified by means of its URI.
- *Location* – To access a multimedia resource, we need not only the URI of the resource in the domain ontology but also links to it. When a resource is cached at a certain peer, its cached location will be added to the set of links to the resource.
- *additionDate* – This attribute keeps track of the date a resource was added to the local cache. This could be used to determine confidence; old information might become less reliable. This also determines when a resource will be updated subject to the cache update policies of each peer.
- *Security* – Access control is required to ensure proper usage of the information and further mechanisms to ensure its integrity.
- *Visibility* – Some resources may need to be hidden, instead of being completely removed. This can be determined in combination with the cache update policies of each peer.

This ontological structure is enhanced with further concepts in the context of the proposed P2P framework. The main elements of the enhanced ontology structure are listed below:

- *The Peer*, which includes concepts such as detailed above:
  - a Peer Identification
  - b Peer Status (*i.e.*, peer, super-peer)
  - c Stated QoS (quantitative or qualitative)
  - d Other concepts (location, additionDate, *etc*.)
- *The Neighbouring Peers*, which include concepts such as:
  - a Neighbouring peer identification
  - b Peer status (*i.e.*, peer, super-peer)
  - c Online status (*i.e.*, on–off)
  - d Stated QoS (quantitative or qualitative as stated by neighbouring peer)

  e Observed QoS (quantitative or qualitative as measured by peer)
  f Peer trust/confidence as discussed above

- *Ontologies*, which include concepts such as:
  a Ontologies employed by the peer for content annotation
  b Ontology mappings supported by the peer
- *Subscriptions obtained*, which include, for each subscription obtained, concepts such as:
  a Subscription Query ID, the original Query ID requesting the subscription
  b Subscription Query Expression(s), as formulated in the original query expression and as potentially reformulated to query resources annotated using different ontologies
  c Subscription configuration, meaning
    1 Subscription query (sub)expression
    2 Peer ID of the peer providing the subscription
    3 Last update (the sequence number of last update received)
  d Alternative subscription configurations
- *Subscriptions offered*, which include per subscription offered concepts such as:
  a Subscription ID, globally unique, managed by the peer
  b Subscription Query ID List, the list of the original Query IDs making use of the particular subscription
  c Subscription Query Expression(s) List
  d Subscription configuration
  e Alternative subscription configurations

This information is used or updated by the components illustrated in the peer architecture of Figure 1.

3.4 Performing semantic query-based subscription

To illustrate the envisaged operation of this P2P network, we outline a scenario where a peer performs a query-based subscription for access to resources in the P2P network. This scenario follows four steps:

Step 1 The peer sends a query resolution request to the P2P network for the given subscription query. Each query resolution request has a unique identifier potentially based on the domain of the peer and/or on the unique peerID.

Step 2 Peers who are the original providers of content that matches the query as well as peers who already have subscriptions for the subscription query (or part of it) will forward their replies through the P2P network.

Step 3 Super-peers will be able to selectively forward query resolution requests to other peers/super-peers and combine their responses to provide a smaller set of candidate subscriptions. If necessary, a super-peer will establish the necessary subscriptions for different parts of a query itself, and provide the requesting peer with a single subscription. In order to do that, super-peers will perform peer status and scoping queries on the requesting peer or other peers/super-peers of the network.

Step 4 Through the P2P network, the requesting peer will receive back a set of candidate subscription configurations for the original query. In an ideal case, the requesting peer will be offered a single subscription potentially to a super-peer and will be relieved of the task of combining results of different subqueries to synthesise the results of the original query (these tasks will be performed by the super-peer). In addition, it will be easier for a peer to resume its subscriptions when getting back online after having stayed offline for a period of time.

3.5 The P2P network configuration

Super-peers can take action to improve the performance of the P2P network. Action can be triggered by new subscription requests in the P2P network or by certain events monitored by the super-peers in the network. Such events can involve:

- *Subscription request.* During this scenario, a super-peer will be forwarded a uniquely identified subscription query. The super-peer will then take the following steps:

  1 Check whether it already has a subscription to this query; if this is the case, the super-peer will respond positively.

  2 Check if part of the query can be matched by itself or other peers/super-peers within a specified range of hops and respond positively.

  Following the query resolution phase, each super-peer will rate the potential subscription configurations in a way that is commonly agreed within the P2P network taking into account QoS, number of subscriptions required, level of trust of peers/super-peers involved, *etc*. Subsequently, it will only return to the requestor the top one or top two highest-rated subscription configurations for the uniquely identified subscription query.

  While a query is being resolved by the P2P network, a peer or super-peer may be asked to forward responses (proposed subscription configurations) to subscription queries to other peers or super-peers. In this case the behaviour of a peer can be described as follows:

  a If the peer/super-peer has already forwarded a higher-rated response for the particular query, it will discard the response received.

  b If the peer/super-peer has already forwarded a different response with the same rating, it will forward the response received.

  c If the peer/super-peer has not forwarded a response with a rating as high as the one of the response received, it will forward it.

The peer from whom the subscription query originated may receive a number of top-rated subscription configurations.

As a second step, the peer can choose to subscribe to one of the proposed configurations; for this purpose it will send subscription management queries to the peers/super-peers identified in the proposed subscription configuration quoting the (uniquely identified) original subscription query.

- *Continuous rating of subscription configurations*. The QoS levels observed in connections to other peers/super-peers, the trust levels and alternative subscription configurations for each subscription query are continuously monitored by the super-peers of the P2P network. For each existing subscription (that corresponds to a uniquely identified query) the rating of the chosen subscription configuration can be reevaluated following a request by a super-peer. In this case, by employing the mechanisms used when a new subscription query is received, an alternative subscription configuration can be adopted for a particular subscription query.

  a *Peers or super-peers newly introduced to the P2P network*. When peers/super-peers are newly introduced to the network, a rerating of existing subscription configurations can be triggered.

  b *Peers/Super-peers becoming unavailable*. The availability of each uniquely identified peer/super-peer is recorded in the P2P network by super-peers and is taken into account when subscription query rerating takes place.

When subscription reevaluation and rerating takes place, the P2P network can establish whether other subscriptions will be affected by rating-related subscription queries. The P2P network needs to establish a number of policies to determine how potential conflicts can be resolved.

The scenarios discussed above highlight the importance of super-peers in the operation of the P2P network. In a real-life scenario, the super-peers will be operated in combination with certain OHS servers for which there will be an interest in gaining more efficient access to multimedia content of a potentially larger number of other OHS servers. This is expected to be the motive for the operation of certain OHS servers as super-peers.

3.6 Relevant network configuration approaches

The following approaches can also be considered for the configuration and operation of the proposed P2P network and the efficiency of the mechanisms outlined in the previous sections can be compared to the ones below:

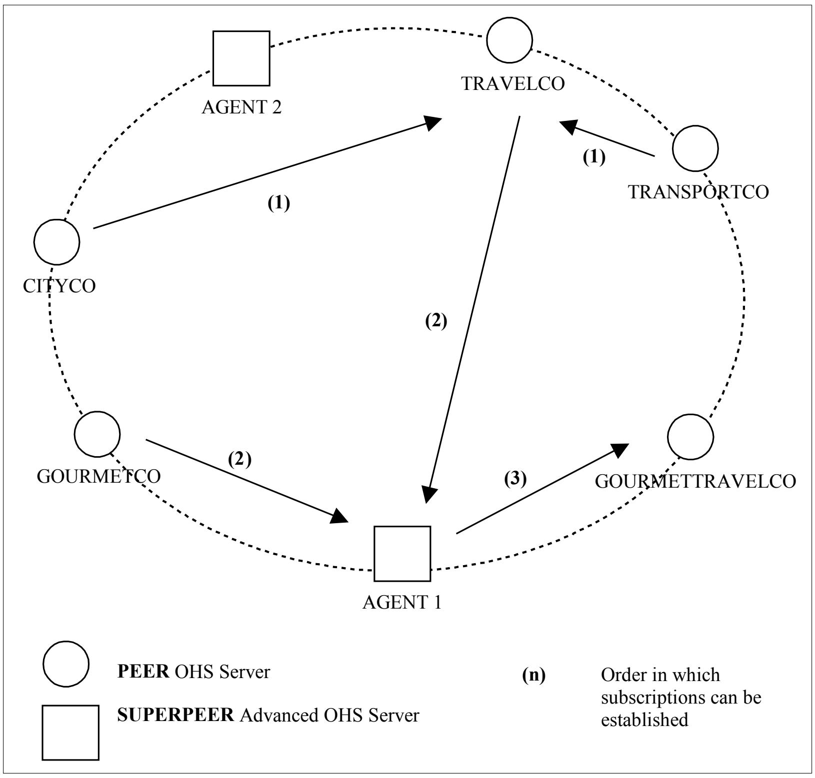
- *The super-peer topology* (Nejdl *et al*., 2004), in which the super-peer backbone can be responsible for message routing and integration/mediation of metadata. This super-peer-based network can support our network, with different metadata schemas and ontologies. Moreover, it can provide better scalability than broadcast-based networks. Based on metadata routing indices (stated in RDF), the super-peer network can support sophisticated routing and distribution strategies. Our super-peers network can also be arranged in a hypercube topology, according to the HyperCuP protocol (Schlosser *et al*., 2002); using this protocol, super-peers can connect to each other. Different kinds of super-peer indices describe the data and schema characteristics of the peers connected to the knowledge network. Super-peer indices exploit the RDF ability to uniquely identify schemas, schema attributes and ontologies, and provide a necessary ingredient for our schema-aware P2P data management infrastructure.

- *The content-based networking model* (Carzaniga *et al*., 2001), in which a message is transmitted from a peer sender to one or more peer receivers without the sender having to address the message to a specific receiver is also relevant to our work. Receiver peers can express interest in the kind of messages they would like to receive, and the network can deliver to the receivers messages matching those interests. Peer receivers express interests through predicate advertisements. In this receiver-driven style of communication, our network could be responsible for efficiently applying predicates to the content of messages so as to minimise the computational and communication costs of the network.

4 An example

To better evaluate the potential impact of a publish-subscribe P2P architecture for OHS in specific domains of real-life activity, an example of applying the proposed architecture in the domain of the tourism industry is illustrated in Figure 2.

Figure 2 An example of the proposed architecture



Assume there is a OHS server/peer (TRAVELCO) that is partly using a tourism-specific ontology for its CM; the CM will also contain elements from other ontologies on hotel information, country/city profiles, transportation information, a list of local restaurants and their menus, *etc*. In order to populate its multimedia content repository, the TRAVELCO server subscribes to two other OHS servers/peers: (1) CITYCO, the server of the authorities to provide the country/city profiles, and (2) TRANSPORTCO, the server of the local transportation services, *etc*.

At the same time, assume there is another OHS server/peer (GOURMETCO) that employs a CM to describe ethnic/local cuisines. The GOURMETCO uses a cuisine-specific ontology that helps describe dishes (*e.g.*, ingredients, country of origin, city of origin, vegetarian classification, *etc*.) as well as wines suitable to accompany each dish.

Finally, assume there is a community that wishes to establish an OHS service/peer GOURMETTRAVELCO for planning food- and wine-tasting tours; they would need to establish their own CM that will probably rely to some extent on a domain-specific ontology (that they will develop) and will also reuse ontologies of other domains (but will match none of them completely).

We can assume that in order to obtain the information necessary to populate their server according to their CM, GOURMETTRAVELCO will need to subscribe to other OHS servers/peers using a complex subscription query that will have to combine four simpler ones:

1 transportation

2 restaurants

3 restaurant menus

4 local dishes and accompanying wines.

For example, the 'Gourmet Tour to Thailand' section of the GOURMETTRAVELCO OHS could be populated by relying on a subscription using the following complex query:

```
∀ (restaurant isLocated 'Thailand')
   ∃ (transport toDestination restaurant) AND
   ∃ (menu isServed restaurant) AND
   ∃ (regionalMenu isServed restaurant)
:
Obtain
   (restaurant, transport, menu, regionalMenu)
```

Each of the simpler queries may be expressed in different ontologies. Each simple query can be served by more than one server but some servers can serve combinations of these queries; TRANSPORTCO will serve simple queries (transportation times), while TRAVELCO can serve both simple queries (transportation times) and complex ones (transportation times to reach a restaurant).

In this case GOURMETTRAVELCO can send a complex subscription query to the P2P network on which all the other servers participate. The P2P network will then be able to establish the optimal set of subscriptions in terms of delivery and processing by employing all the mechanisms described in Section 3.

In particular, a complex query of GOURMETTRAVELCO can be picked up by peers or super-peers in this P2P network. If the query is picked up by an Advanced OHS server (super-peer), the super-peer would in turn be able to invoke the web service operations of TRAVELCO, CITYCO, TRANSPORTCO and GOURMETCO or of other super-peers to obtain the ontologies and the queries served by each of them and then calculate the best possible set of subscriptions; the original query or part of it may need to be reformulated to match different ontologies employed by other peers in the P2P network. As a result, a subscription configuration will be chosen and implemented for the subscription of GOURMETTRAVELCO to multimedia content of interest, which subsequently will be integrated in the GOURMETTRAVELCO OHS server based on the chosen CM and IM.

5 Challenges and open issues

The proposed architecture presents strong potential for enabling truly efficient, up-to-date, scalable and relevant OHS over P2P networks. The proposed approach is novel in comparison to similar work for OHS since it goes beyond the exchange of links among OHS servers; this architecture introduces publish-subscribe mechanisms to improve the performance of OHS and enable exchange of content that is up-to-date. In addition, it employs caching to further ensure content availability. It also envisages peer trust mechanisms. The proposed approach aims to provide a comprehensive framework to enable OHS over P2P by leveraging publish-subscribe mechanisms. However, there are open issues and challenges to be successfully addressed before its full potential can be reached.

Generally, a serious challenge in integrating P2P and ontologies is to design ontology-based peer selection services in order to avoid flooding the peer network with queries (*viz.*, one must ask the 'right' peers to receive the right answers). These peer selection services will rely on mechanisms that will exploit similarity of ontologies for this purpose (Fensel *et al.*, 2003).

An important challenge is the choice of an ontology language and a query language that hold the promise to be established globally. In this sense, RDF(S) and SPARQL (W3C, 2006) can be considered strong candidates at the moment, as they appear to present a good choice for an initial deployment of OHS peers for proof of concept. Although the expressiveness of RDF(S) is low, it is believed that it can demonstrate the value of the proposed framework. Similarly, SPARQL could gain growing acceptance as the language of choice for querying RDF metadata or querying across multiple RDF graphs nameable by URIs.

Another challenge is the application of the CM and IM models for the integration of multimedia content coming from different sources. A potential approach to address this challenge is to start with a specific domain (*e.g.*, tourism) and establish the value of the proposed architecture and of the CM and IM in this context.

In addition, the description of the web services to be exposed by the peers/super-peers of the P2P architecture needs to be stable to enable efficient grown of the proposed P2P framework. The interfaces of the peers need to be well discussed and all the aspects of the operation of the P2P framework need to be researched before a WSDL description is finalised.

There are several aspects of the operation of the P2P network to be detailed and tried in the context of OHS. The availability of efficient algorithms for the optimisation of the P2P framework and the establishment of the necessary metrics in terms of trust and QoS is expected to take significant effort. However, the results of this work present strong potential for contribution to the P2P research community and further fusion of ideas with the OHS research community.

Finally, query resolution techniques need to be further investigated and their efficiency needs to be evaluated in the context of specific query languages such as SPARQL. Taking into account that query languages are still under development and standardisation, the deployment of the proposed architecture has a dependency on this process that needs to be carefully considered.

6 Conclusion

The proposed architecture/framework features two different aspects. The first aspect is the integration of existing innovation in the areas of OHS, P2P, semantic web, and web service technologies for providing efficient information dissemination among OHS peers and the introduction of publish-subscribe mechanisms to provide for improved exchange of hypermedia resources. The second aspect is the additional innovation in the areas of content integration in OHS, semantic query-based routing and P2P networks for multimedia content dissemination taking into account trust and QoS parameters. Both aspects are presented and discussed in this paper.

Based on the proposed architecture, developers can deploy efficient multimedia content exchange in OHS by leveraging P2P architectures, publish-subscribe mechanisms, query resolution techniques and semantic query-based routing.

To establish the potential of the proposed framework, it takes a number of sequential steps of innovation and deployment while establishing value at every step. An example was provided in this paper to illustrate the potential of the architecture. This work outlines a framework before experimentation can be undertaken.

A next step will involve the investigation of the potential of the aspects of the proposed framework in a specific domain. This will enable focusing on the operation of the core publish-subscribe P2P network and trying different algorithms for caching multimedia resources taking into account QoS and trust first. Subsequently, issues of query resolution and advanced query-based routing will be addressed.

In the near future, we will investigate the potential of the proposed framework in dynamic packaging systems of the travel domain. Dynamic packaging enables consumers/travellers (or a booking agent) to build a customised itinerary by assembling multiple components of their travelling choices. A dynamic packaging system produces one reservation, completes the transaction in real time and entails only one payment from the consumer.

Acknowledgements

The authors would like to thank the organising committee and the programme committee of the International Workshop on Architectures, Models and Infrastructures to Generate Semantics in Peer-to-Peer and Hypermedia Systems at the *Seventeenth ACM Conference on Hypertext and Hypermedia (ACM HyperText 2006)* and the unknown referees for their valuable feedback and constructive comments on this work.

References

Anderson, K.M. (2001) 'The extensibility mechanisms of the chimera open hypermedia system', *Journal of Network and Computer Applications*, Vol. 24, No. 1, pp.75–86.

Anderson, K.M., Taylor, R.N. and Whitehead, E.J. (2000) 'Chimera: hypermedia for heterogeneous software development environments', *ACM Transactions on Information Systems*, Vol. 18, No. 3, pp.211–245.

Basney, J., Nejdl, W., Olmedilla, D., Welch, V. and Winslett, M. (2004) 'Negotiating trust on the grid', *Proceedings of the Second Workshop Semantics in Peer-to-peer and Grid Computing*, New York, NY, USA, May, http://drops.dagstuhl.de/opus/volltexte/2005/387/ (accessed April 2007).

Benatallah, B., Hacid, M-S., Paik, H., Rey, C. and Toumani, F. (2006) 'Towards semantic-driven, flexible and scalable framework for peering and querying e-Catalog communities', *Information Systems*, Vol. 31, Nos. 4–5, pp.266–294.

Cai, M., Frank, M., Yan, B. and MacGregor, R. (2004) 'A subscribable peer-to-peer RDF repository for distributed metadata management', *Journal of Web Semantics*, Vol. 2, No. 2, pp.109–130.

Carzaniga, A., Rosenblum, D. and Wolf, A. (2001) 'Design and evaluation of a wide-area event notification service', *ACM Transactions on Computer Systems*, Vol. 19, No. 3, pp.332–383.

Castano, S., Ferrara, A. and Montanelli, S. (2005) 'Ontology-based interoperability services for semantic collaboration in open networked systems', *Proceedings of the First International Conference on Interoperability of Enterprise Software and Applications (INTEROP-ESA)*, Geneva, Switzerland, February, pp.135–146.

Castro, M., Druschel, P., Kermarrec, A-M. and Rowstron, A. (2002) 'Scribe: a large-scale and decentralised application-level multicast infrastructure', *IEEE Journal on Selected Areas in Communications (JSAC)*, Vol. 20, No. 8, pp.100–110.

Chirita, P.A., Idreos, S., Koubarakis, M. and Nejdl, W. (2004) 'Publish/subscribe for RDF-based P2P networks', *Proceedings of the First European Semantic Web Symposium (ESWS)*, Lecture Notes in *Computer Science 3053 Springer 2004*, Heraklion, Crete, Greece, May, pp.182–197.

Dean, M. and Schreiber, G. (Eds.) (2004) *OWL Web Ontology Language Reference (W3C Recommendation)*, 10 February.

Dillon, T., Chang, E. and Hussain, F. (2004) 'Managing the dynamic nature of trust', *IEEE Transaction of Intelligent Systems*, September–October, pp.79–82.

Doan, A.H., Madhavan, J., Dhamankar, R., Domingos, P. and Halevy, A. (2003) 'Learning to map ontologies on the semantic web', *The International Journal on Very Large Data Bases*, Vol. 12, No. 4, pp.303–319.

Fensel, D. and Bussler, C. (2002) 'The web service modeling framework WSMF', *Electronic Commerce Research and Applications*, Vol. 1, No. 2, pp.1–33.

Fensel, D., Staab, S., Studer, R., Van Harmelen, F. and Davies, J. (2003) 'A future perspective: exploiting peer-to-peer and the semantic web for knowledge management', in N.J. Davies, D. Fensel, and F. van Harmelen (Eds.) *Towards the Semantic Web – Ontology-driven Knowledge Management*, London, UK: Wiley, pp.245–64.

Gong, L. (2001) 'Project JXTA: a technology overview', Technical report, SUN Microsystems, http://www.jxta.org/project/www/docs/TechOverview.pdf (accessed April 2007).

Graham, S., Hull, D. and Murray, B. (2006) *Web Services Base Notification 1.3*, OASIS Standard, http://www.oasis-open.org/committees/tc_home.php?wg_abbrev=wsn (accessed April 2007).

Hendler, J. (2001) 'Agents and the semantic web', *IEEE Intelligent Systems*, Vol. 16, No. 2, pp.30–37.

Hussain, F., Chang, E. and Dillon, T. (2004) 'Trustworthiness and CCCI metrics for assigning trustworthiness in P2P communication', *Journal of Computer Systems Science & Engineering*, Vol. 19, No. 4, pp.95–112.

Karousos, N., Pandis, I., Reich, S. and Tzagarakis, M. (2003) 'Offering open hypermedia services to the WWW: a step-by-step approach for developers', *Proceedings of the Twelfth International Conference on World Wide Web (WWW 2003)*, Budapest, Hungary, May, pp.482–489.

Keidl, M., Kreutz, A. and Kemper, A. (2002) 'A publish and subscribe architecture for distributed metadata management', *Proceedings of the Eighteenth International Conference on Data Engineering (ICDE)*, February, pp.309–320.

Lewis, D., Feeney, K., Tiropanis, T. and Courtenage, S. (2004) 'An active, ontology-driven network service for internet collaboration', *Proceedings of the Sixteenth European Conference on Artificial Intelligence (ECAI), Workshop on Application of Semantic Web Technologies to Web Communities*, Valencia, Spain, August.

Madche, A., Staab, S., Studer, R., Sure, Y. and Volz, R. (2002) 'SEAL-tying up information integration and web site management by ontologies', *IEEE-Cs Data Engineering Bulletin (Special Issue on Organizing and Discovering the Semantic Web)*, Vol. 25, No. 1, pp.10–17.

Nejdl, W. and Olmedilla, D. (2004) 'Protecting sensitive data in peer-to-peer networks', *IEEE Intelligent Systems*, Vol. 19, No. 5, pp.82–85.

Nejdl, W., Siberski, W. and Sintek, M. (2003) 'Design issues and challenges for RDF- and schema-based peer-to-peer systems', *ACM SIGMOD Record*, Vol. 32, No. 3, pp.41–46.

Nejdl, W., Wolpers, M., Siberski, W., Schmitz, C., Schlosser, M., Brunkhorst, I., *et al*. (2004) 'Super-peer-based routing strategies for RDF-based peer-to-peer networks', *Journal of Web Semantics*, Vol. 1, No. 2, pp.177–186.

Nürnberg, P.J., Leggett, J.J. and Wiil, U.K. (1998) 'An agenda for open hypermedia research', *Proceedings of the Ninth ACM Conference on Hypertext*, Pittsburgh, Pennsylvania, June, pp.198–206.

Object Management Group (OMG) (2004) 'CORBA Notification Service 1.1', http://www.omg.org/technology/documents/formal/notification_service.htm (accessed April 2007).

OWL-S services coalition (2004) 'OWL-S 1.1 Beta Release', http://www.daml.org/services/owl-s/ (accessed April 2007).

Pandis, I., Karousos, N. and Tiropanis, T. (2005) 'Semantically annotated hypermedia services', *Proceedings of the Sixteenth ACM Conference on Hypertext and Hypermedia*, Salzburg, Austria, September, pp.245–247.

Reich, S., Wiil, U.K., Nürnberg, P.J., Davis, H.C., Groenbaek, K., Anderson, K.M., Millard, D.E. and Haake, J.M. (1999) 'Addressing interoperability in open hypermedia: the design of the open hypermedia protocol', *The New Review of Hypermedia and Multimedia*, Vol. 5, pp.207–248.

Roman, D., Keller, U., Lausen, H., Bruijn, J., Lara, R., Stollberg, M., *et al*. (2005) 'Web service modeling ontology', *Applied Ontology*, Vol. 1, No. 1, pp.77–106.

Sakkopoulos, E., Kanellopoulos, D. and Tsakalidis, A. (2006) 'Semantic mining and web service discovery techniques for media resources management', *Int. Journal of Metadata, Semantics and Ontologies*, Vol. 1, No. 1, pp.66–75.

Schlosser, M., Sintek, M., Decker, S. and Nejdl, W. (2002) 'HyperCuP – hypercubes, ontologies and efficient search on P2P networks', *Proceedings of the First International Workshop on Agents and Peer-to-peer Computing (AP2PC)*, Lecture Notes in *Computer Science 2530 Springer 2003*, Bologna, Italy, July, pp.112–124.

Tam, D., Azimi, R. and Jacobsen, H.A. (2004) 'Building content-based publish/subscribe systems with distributed hash tables', *Proceedings of the First International Workshop on Databases, Information Systems, and Peer-to-Peer Computing (DBISP2P)*, Lecture Notes in *Computer Science 2944 Springer 2004*, Berlin, Germany, September, pp.138–152.

Vdovjak, R., Frasincar, F., Houben, G. and Barna, P. (2003) 'Engineering semantic web information systems in Hera', *Journal of Web Engineering*, Vol. 2, Nos. 1/2, pp.3–26.

W3C (2006) *SPARQL Query Language for RDF*, W3C Candidate Recommendation, 6 April, http://www.w3.org/TR/2006/CR-rdf-sparql-query-20060406/ (accessed April 2007).

Wiil, U.K. (1997) 'Open hypermedia: systems, interoperability and standards', *Journal of Digital Information*, Vol. 1, No. 2, http://journals.tdl.org/jodi/article/view/jodi-8/9 (accessed April 2007).

Yang, B. and Garcia-Molina, H. (2002) 'Improving search in peer-to-peer systems', *Proceedings of the Twenty-second International Conference on Distributed Computing Systems (ICSCS)*, Vienna, Austria, pp.5–14.

Yu, T., Winslett, M. and Seamons, K. (2003) 'Supporting structured credentials and sensitive policies through interoperable strategies in automated trust negotiation', *ACM Transactions on Information and System Security*, Vol. 6, No. 1, pp.1–42.

Zhou, J., Dialani, V., De Roure, D. and Hall, W. (2003) 'A distance based semantic search algorithm for peer-to-peer open hypermedia systems', *Proceedings of the Fourth International Conference on Parallel and Distributed Computing, Applications and Technologies (PDCAT)*, Chengdu, China, August, pp.7–11.

Zhou, J., Hall, W. and De Roure, D. (2004) 'When open hypermedia meets peer-to-peer computing', *Proceedings of the Fourteenth ACM Conference on Hypertext and Hypermedia*, Santa Cruz, California, August, pp.266–267.

Zhou, J., Hall, W., De Roure, D. and Dialani, V. (2007) 'Supporting ad hoc resource sharing on the web: a peer-to-peer approach to hypermedia link services', *ACM Transactions on Internet Technology*, http://eprints.ecs.soton.ac.uk/11023/ (accessed April 2007).