Distributed Ontology – based P2P System Integrating Digital Libraries

Christos E. Alexakos, Kostas I. Paraskevopoulos, Kostas V. Votis, Spiridon D. Likothanassis

Pattern Recognition Laboratory, Dept. Computer Engineering & Informatics, University of Patras
GR 26500 Patras, Hellas (Greece)
+30 2610 997520
{alexakos, paraskev, botis, likothan}@ceid.upatras.gr

ABSTRACT
Distributed Digital Libraries integration is significant for the enforcement of novel searching mechanisms in the internet. The great heterogeneity of systems storing and providing digital content requires the introduction of interoperability aspects in order to resolve integration problems in a flexible and dynamic way. Our approach introduces an innovative service oriented P2P system which initializes a distributed ontology scheme, semantically describing and indexing the digital content stored in distributed Digital Libraries. The proposed architecture enforces the distributed semantic index by defining virtual clusters consisting of nodes with similar or related content in order to provide efficient searching and recommendation mechanisms.

Categories and Subject Descriptors
C.2.1 [Network Architecture and Design]: Distributed networks, Network topology; C.2.4 [Distributed Systems]: Distributed applications, Distributed databases; D.2.11 [Software Architectures]: Software Architectures– Data abstraction, Domain-specific architectures; H.2.5 [Heterogeneous Databases]: Data translation; H.2.4 [Systems]: Multimedia databases; H.2.7 [Database Administration]: Data dictionary/directory; H.3.7 [Information Search and Retrieval]: Search process H.3.7 [Digital Libraries]: Systems issues, Collection, Dissemination, User issues.

General Terms
Design, Performance, Reliability.

Keywords
DL integration, ontology, p2p architecture, SOA, semantic index.

1. INTRODUCTION
The wealth of information available in various Digital Libraries (DLs) accessible over the internet has increased while the ability to search and retrieve relevant information is being reduced. Hypermedia and digital media, being complex information objects, are much more difficult to manage especially when they reside in distributed servers. A lot of approaches have been proposed in order to enable machine-to-machine interaction that in turn will facilitate truly efficient searching.

In this context, searching in distributed and heterogeneous sources was always difficult but it seems that semantics may offer solutions to such unsolved problems [1]. Furthermore, a lot of scientific work has been initialized in the era of “peer-to-peer” architectures promoting solutions that can manage efficiently distributed systems. According to Androutsellis et al. [2] the P2P architectures provide two benefits for the interconnection of distributed computational and storage systems: a) direct exchange of data between the nodes of the architecture and b) ability to treat instability and variable connectivity, automatically adapting to failures in both network connections and computers.

Our approach is adapting a P2P architecture utilizing web services capabilities in order to interconnect the various distributed Digital Libraries towards an integrated approach. The proposed P2P architectures consists of independent DLs and some centralized elements that are used in specific tasks as new node initialization, maintenance of data integrity and system fault-tolerance.

Furthermore, we introduce a distributed multilayered ontology scheme that semantically integrates the description of the digital content residing in each node and partially is utilized as a higher conceptual index supporting the resources discovering mechanism. In order to enforce and improve the search process among the node of the integrated network, we define virtual clusters of DLs based on the description of their content. This virtual classification supports the allocation of the semantic index to more than one nodes accomplishing the efficiency and time improvement of the search process by reducing the number of nodes where a query is executed. Moreover, the distribution of the semantic information provides flexibility and scalability supporting efficient adjustment of the proposed P2P system on potential content and structural changes.

The organization of the paper is as follows: section 2 presents the state-of-the-art technologies utilized by the proposed architecture as well as similar work, section 3 presents the proposed ontology-based P2P DL architecture and its implementation. A discussion and conclusions are provided in section 4.

2. RELATED WORK AND UTILIZED TECHNOLOGIES
The paper addresses the issues of integration of distributed Digital Libraries, which remain extremely challenging despite the different approaches that have been proposed for their handling.
In this section similar work that has been done regarding the above areas is presented. Our proposed architecture is based on the utilization of two mainstream technologies, more specifically web services and ontologies. Thus, an introduction to these two technologies is provided and the current state-of-the-art is presented.

2.1 State-of-the-Art Technologies

2.1.1 Web Services

Web Services (WSs) [3] address a novel paradigm for the utilization of application inherent capabilities that are easily and dynamically exposed via open standard application interfaces and protocols. Such capabilities may be used by other applications through a simple WS invocation, resulting in extremely easy integration. In this context WSs also promote software reusability. In WSs are referred as “software applications identified by a Uniform Resource Identifier (URI), whose interfaces and binding are capable of being defined, described and discovered by XML artifacts and support direct interactions with other software applications using XML-based messages via Internet-based protocols”. WSs are reusable software building blocks distributed over the Web easily accessible via widespread protocols like HTTP and SMTP. They are loosely coupled, communicating through XML based documents.

Although WSs add extreme efficiency regarding interoperability, it is still difficult for WS providers and requesters to have a straightforward and thorough understanding of nontrivial WS relevant statements such as inputs, outputs and constraints. [4] In order for this limitation to be alleviated semantic annotation is necessarily associated with WS description. Thus WSs may be transformed to Semantic WSs defined through their association with service ontologies. The key enabler for WSs technology is XML and the initiatives that form its foundations are SOAP, WSDL, and UDDI. Simple Object Access Protocol (SOAP) [5] is a lightweight protocol enabling structured information exchange in a decentralized, distributed environment. Web Services Description Language (WSDL) [6] provides a model and an XML format for WS description. The last piece of the puzzle is UDDI (Universal Description Discovery & Integration) [7], which supports both design-time and run-time discovery of WSs.

2.1.2 Ontologies

In the last years, the need of knowledge representation combined with a set of rules and concepts has lead the evolution of ontologies as the main tool to describe data contents and their relations in modern information systems. An early definition presented in [8] describes ontologies as: “a hierarchically structured set of terms to describe a domain that can be used as a skeletal foundation for a knowledge base”. An ontology’s main feature is machine readability and understandability which in turn enable automatic cross-communication of different systems. This leads to increased platform independence as well.

The Semantic Web [9] introduction and unanimous approval has driven towards the representation of Ontologies in semantics. For this purpose a variety of semantic markup languages has been developed, based on the dominant XML standard. The most prominent ontology markup languages are DAML+OIL [10] and its successor OWL [11], which build on top of RDF Schema [12].

2.2 Related Work

The wide adoption of the Semantic Web vision together with the prominent peer-to-peer technologies for integrating distribution systems lead to numerous proposed approaches concerning the utilization of ontologies towards to knowledge integration of the data in P2P systems.

Systems described in [13–14] provide solutions about supporting information resource discovery in architectures with multiple ontologies. Furthermore, some approaches [15] deals with the issue of utilizing P2P systems in order to support the semantic web to link semantic definitions. Kokkinidis et al. [16] have exploited the issue of effective and efficient query routing in P2P Digital Libraries Architectures utilizing intentional indexing of DL node views. Moreover, some research efforts [17] are made towards the integration of the different metadata description used in various implementations of Digital Libraries conclude in the utilization of ontologies. Finally, some systems were designed such as EDUTELLA[18], SWAP[19] and Bibster[20] focusing in the knowledge representation of the digital content via semantic based techniques in P2P networks.

3. ONTOLOGY-BASED P2P DL SYSTEM

Our approach introduces an innovative architecture in order to integrate heterogeneous Digital Libraries based on two major aspects: a) the utilization of a scalable P2P Architecture that ensures the stability and the efficiency of the system and b) the system’s enforcement by a distributed ontology scheme that provide a semantic representation of the digital content and a flexible semantic index for efficient and functional search. In the following sub chapters we introduce the proposed architecture and the ontology scheme followed by the presentation of the major functionalities of the system.

3.1 Proposed Architecture

Towards integrating DLs we exhibit the importance of indexing in content retrieval from heterogeneous sources. We propose a distributed ontology-based P2P system in order to adopt indexing techniques that facilitates the structuring of heterogeneous sources by making use of ontologies both as indexing and semantic representation tool. The implementation of this approach requires a specific system architecture that makes it possible to integrate the different semantically annotated DLs in a flexible and interoperable way, the proposed architecture is depicted in Figure 1.

The system architecture consists of the following entities:

- The Digital Library Node (DL Node) represents a library in which a significant proportion of the resources are available in machine-readable format, accessible by means of computers. The digital content may be locally held or accessed remotely via computer networks [21].
- The Virtual Content Cluster is a set of Digital Library nodes that have similar or related content according to the semantic description.
The Cluster Indexing Node is a DL Node belonging to a Virtual Content Cluster and is responsible for maintaining a detailed index of the digital content of the nodes of the Cluster.

The Coordination Entity is a service responsible for the coordination, control and update of network nodes.

The Indexing and Maintenance Mechanism, consists of each Cluster Indexing node, and the Coordination Entity and represents in an abstract way a mechanism responsible for organizing the distributed nodes of our system in a centralized and structured manner.

We will demonstrate now the functionality of each entity in terms of facilitating the integration process of DL Nodes. The DL Nodes store ontologies that semantically describe their digital content, also they keep a local replica of a general ontology, in order to be aware of the current index that denote the content of the other DLs that exist in the system. This ensures a more precise way to grant to the system major P2P prospects.

Virtual Content Cluster comprise of DL Nodes that have similar content, not in terms of creating geographically organized clusters but in terms of content categorization. This prospect accomplishes to organize the distributed content sources in a higher level aiming to add semantic annotation in heterogeneous and geographically distributed sources.

The Cluster Indexing Node provides the necessary connectivity between the nodes that comprising the Virtual Content Cluster, and, by using ontologies, gives a more detailed description of the overall content that is stored in the DL Nodes of the cluster. The main advantage of the Cluster Indexing Node is that reduces the content search effort through a distributed index towards the resources that belong to the corresponding Virtual Content Cluster.

The Coordination Entity is a service that facilitates coordination of the overall architecture and control of the cluster indexing nodes. Moreover, it resolves problems that might occur in the P2P environment. Furthermore, it is used as a repository maintaining all the schemes of the distributed ontology and providing the necessary mechanisms for their retrieval. Also, it is responsible for updating the semantic information residing in each node of the system.

The Indexing and Maintenance Mechanism operates as an entity that works as a centralized structure that implements the content organization of the distributed nodes. The systems belonging to it are responsible for maintaining the coherence of the system, by initializing the appropriate mechanisms for updating the semantic information in the DL Nodes and guide the search process utilizing a distributed index. This effort can be seen as a centralized system module that controls the distribution and indexing of content among the DL Nodes of the architecture.

A service oriented approach is utilized for interoperability and communication between system entities. The utilization of web services technology enables interoperability and facilitation in adoption by existing systems. The utilization of web services is based on the capability of the easy integration that is achieved through an intermediate adapter layer that relays commands and data to the web from the system and vice versa. A lot of implementation has introduced according the exposition of systems capabilities and services to the internet, such .NET framework and Common Object Model (COM) for Microsoft based applications, Enterprise JavaBeans (EJB) for java based applications and PHP classes for web based applications. Another main ability of Web Services is that can facilitate transparency between different implementations that is a desirable feature.

Our proposal consists of five interfaces implemented by using web services protocols:

**Query Web Service.** A web service that is responsible for exchanging SOAP messages in order to transmit the query information through system nodes during the search process.

**Object Retrieval Web Service.** A web service that collects the user requests for resources and retrieves the desirable digital content to the DL Node that is addressed by the user. Similar to the functionality of Query Web Service, Object Retrieval Web Service compiles the request in a transparent to the user manner, locates and finally delivers the objects to him/her.

**Update Web Service.** The proposed system must have the suitable mechanisms for updating the ontologies in order to maintain its dynamic nature. Also it has to allow node and content addition, removal and modification. These updates must be posted from each DL node to the Coordination Entity, allowing the system to maintain its consistency.

**Recommendation Web Service.** An interface between the Cluster Indexing Nodes and the DL Nodes delivers the recommendation results that were found by the Cluster Indexing Nodes to the user, giving the ability to re-query the system to acquire more relative resources.

**Fault-Tolerance Web Service.** An interface supporting the fault-tolerance maintenance mechanism that is administrated by the Coordination Entity and the Cluster Indexing Nodes.

### 3.2 Distributed Ontology Scheme

Our approach introduces a three-layer semantic description and annotation of the digital content that is distributed to the DL Nodes of the proposed P2P Architecture. The distributed ontology scheme is divided in three levels of description in order to provide a more flexible and efficient indexing service to the overall semantic representation of the contents residing in the DL Nodes. Furthermore, it supports the semantic annotation of the metadata.

![Proposed System Architecture](image-url)
supporting the description of the digital content that provided by the distributed DLs.

### 3.2.1 Layers of the Ontology Scheme

The three layers that introduced to the proposed distributed ontology are: a) Global Indexing Ontology layer that describes in a formal conceptual manner the main subject categories of the DLs’ contents, b) Node Indexing Ontology layer that consist of a set of ontologies representing a more detailed description of each subject category and c) the Node Content Ontology layer where the semantic notation of the different metadata description of the distributed DLs lay. The proposed indexing scheme includes additional mapping information between the ontologies in the three layers providing the necessary information to search mechanism in order to navigate inside the ontology-based index. The three layer ontology scheme is depicted in figure 2.

The three-layer approach is utilized in order to distinguish the parts of the ontology that has to be stored in each DL Node of the proposed architecture for supplying a flexible and usable distributed ontology scheme. This aspect is achieved by the utilization of the ontologies that reside in each layer according to the following distribution:

- The Global Indexing Ontology is stored in each DL Node
- The ontologies of the Node Indexing Ontology layer are residing in the associated Indexing Cluster Node.
- Each DL node publishes the metadata of its content according to the corresponding ontologies of the Node Content Ontology layer

The Global Indexing Ontology is comprised of a single ontology describing the main knowledge domains that represent the contents of the integrated DL systems. This layer is primarily used as a “catalogue” for the type of contents and the domain of knowledge that are integrated. Domains are specified in classes and subclasses providing a hierarchical model presenting all the knowledge fields that are included in the hypermedia contents of the distributed servers. There are also a number of properties denoting the relationship between classes.

Figure 3 demonstrates an example of a part of the Global Indexing Ontology consisting of two main classes named as “Persons” and “Publications”. “Publication” class is consists of two subclasses: “Books” and “Magazines” denoting that books and magazines are publications. The relationships between these classes are represented by the object property “isAuthoredBy” and its inverse property “hasAuthored”.

The purpose of the Global Indexing Ontology is to represent in a formal conceptual format the main subject categories of the digital content residing in the distributed DL Nodes and their characteristics while is denoting the relationships between the different terms and entities of the ontology. In addition, it provides a generic indexing and categorization for the digital content of the P2P network. The Global Indexing Ontology is residing in each of the DL node of the P2P network and is being update by the Coordination Entity. Furthermore, each class of the ontology contains information about the location of the associated Node Indexing Ontologies, connecting in that way the first layer with the second one in the proposed distributed ontology scheme.

In the Node Indexing Ontology layer, there is a set of ontologies for the detailed description of each of the subject categories that included in the Global Indexing Ontology. The semantic representation of subject categories follows the same level of detail with the subclasses that stems from the information of the digital content stored in the DL nodes. This ensures that all the subcategories characterizing the content of DL Nodes is conceptually described in the ontologies of Node Indexing Ontology layer providing both the hierarchy of the subcategories and the relationships between them and their characteristics.

Figure 4 presents an example of Node Indexing Ontology describing some of the subcategories that are associated with book publications. The main class of “Book” is related with the class “Historical Period” and the class “Book Type”. The class “Historical Period” consists of four subclasses “Greek Ancient History”, “Roman Period”, “Middle-Ages” and “Modern”. The class “Book Type” consists of three subclasses “Literature”, “Poetry” and “Theatrical”.

Each of the classes in a Node Indexing Ontology can contain the following associations or relationships.

1. A list of nodes that contain the corresponding digital content, i.e. books with poetry. This association ensures the integration of the middle level with the bottom layer of the proposed multilayer ontology.
The semantic annotated metadata schemes that are supported by each DL Node. An ontology in the Node Content Ontology layer is composed of two functional parts, the first part is the ontology scheme and the second is a set of instances of the ontology scheme containing the metadata of the integrated digital data. The distinction of these two parts is similar with the T-Box and A-Box distinction which is drawn in Description Logics [22].

The ontology scheme is primarily used for the conceptual description of the metadata structure of the DLs node digital content. For example, in a DL system that provides books, the metadata is structured by elements as “Book Title”, “Book Author” and “Book Abstract”. According to ontology scheme there is class “Book” with the corresponding metadata elements as properties, namely there are properties of the class “Book” as “hasTitle”, “isAuthoredBy” and “hasAbstract”. This ensures that all terms and their relationships utilized by each digital content provider server separately are included in the ontology scheme.

The set of instances comprises the content of the metadata information provided for the description of the distributed digital data in terms of the ontology scheme. The instances are constructed according to mapping information that depicts the association of the metadata elements to the terms of the ontology scheme. The content of the metadata is transformed by creating instances of the classes of the ontology scheme and fill them with the metadata data according to the mapping information. In the example of books, there is an ontology instance of each book where the property of “Title” takes the values of the metadata element “Book Title”. This transformation simplifies the searching procedure by carrying out the search in ontology instances.

The semantics of the proposed distributed ontology are composed according to the Ontology Web Language (OWL) standard, a recommendation from W3C as a semantic markup language for representing ontologies based on the dominant XML/RDF standard.

3.2.2 Ontology Implementation Issues

As presented before, our approach introduces three different layers of semantic descriptions. In this context, a complete description of digital contents in different DLs with a unified way is possible. What still missing is the implementation issues of the proposed distributed ontology.

The main issue in order to keep the unification of the semantic description is how these layers will be composed in order to maintain the associations with the real content stored in the different DLs. The problem is located in two cases: a) when a new content description appears and b) when an existing semantic description needs update (new classes or relationships). Facing this issue a methodology is proposed comprising two independent procedures, one for each case.

In the first case of the proposed methodology adopt a “top-bottom” approach that comprises of the following steps:

1. Adding the appropriate semantics to the Global Indexing Ontology by adding new classes and subclasses and denoting relationships between new and old classes.
2. Creating new Node Indexing Ontologies giving a more detailed level of description for the additional classes.
3. Creating new Node Content Ontologies according to the metadata that describe the actual digital content of the DL nodes.

In the second case the methodology follows a “bottom-top” approach. When the description of a content in a DL node demands more semantic description, then the following steps are executed:

1. Update the Node Content Ontologies that define the digital content of the DL Node.
2. Update if is needed the associated Node Indexing Ontology.
3. Update if is needed the associated Global Indexing Ontology and inform all the Nodes about the changes.

Another implementation issue that appears in the proposed distributed ontology scheme is the updating of the ontologies that are residing in the distributed nodes. In order to ensure the integrity and unification of the semantic description in the distributed DL Nodes the centralized Coordination Entity is utilized. Furthermore this centralized entity maintains a list of the DL nodes and the corresponding ontology versions. Each time the ontology is changed the proposed centralized entity has the responsibility in order to notify and update the associated ontologies schemes in the DL nodes.

3.3 P2P Supported Functionalities

The dynamic nature, coherency, continuous availability and ability of scaling possessed by the proposed P2P architecture is assured by two main supported functionalities a) New Node Initialization and b) System Fault-Tolerance.

Figure 4. Node Indexing Ontology Example
3.3.1 New Node Initialization
Main objective of supplying the functionality of adding new nodes to the system is the desirable property of system scaling in quality and quantity supplied to the users of the system. The operation of New Node Initialization is implemented by the following six steps:

1. Creating the local Node Content Ontology for the new DL Node to be initialized and integrated into the system. The creation of the local ontology must be made considering the need of detailed semantic representation of the resources that will be added from the DL Node.

2. Finding the corresponding(s) Virtual Content Clusters providing description related to the content of the new node. This process must be completed in respect to the ontologies of the Node Indexing Ontology layer, in order to coordinate the operation of addition and moreover to keep the structure of the upper levels intact.

3. After the new node decides in which Virtual Content Clusters belongs, he must inform the Cluster Indexing Node who is responsible for the indexing inside the Cluster about the upcoming changes in semantic description and indexing that will be made from the incorporation.

4. The Cluster indexing node informs the Coordination entity after the necessary changes are made in order the system to accommodate the new node, maintaining the coherence of the system and updating the semantic description of the overall content.

5. The Indexing and Maintenance Mechanism finally makes the adjustments for the new ontology scheme that indexes and describes the updated content of the system.

6. Finally the Coordination informs all the DL Nodes about the new ontological schemes that have been created and each DL Node updates its corresponding ontologies.

To accomplish the previous procedure the DL Nodes are communicating via SOAP messages and utilizing the Update Web Services interface supplied by the system. Finally the New Node after receiving the appropriate messages configures the Query Web Service and the Object Retrieval Web Service modules which will be used for the interaction with the other entities of the system. This configuration must be completed under the consideration of specificities in storing and representing its own content. The simplicity in the process of adding a new content to the system makes the system capable of supporting the integration of new digital content.

3.3.2 System Fault-Tolerance
The proposed P2P system architecture adopts a virtual centralized approach in coordinating, indexing and organizing the digital content. The system fault-tolerance is assured in the following manner: the coordination entity periodically contacts the Fault Tolerance Web Service of each DL Node in order to verify if the node is operational. If a DL Node has been taken down there are two possible cases:

Case 1: A simple DL Node was taken down. In that case the appropriate changes must be made to the indexing ontology scheme that points to the specific DL Node. A flag is used to denote that the particular DL Node is not available in the corresponding Cluster Indexing Nodes, but the semantic description is still maintained. This approach reduces the effort of deleting and re-initializing the DL Node multiple times. When the node returns in operation uses the System Fault – Tolerance Web Service again to inform about its state.

Case 2: A Cluster Indexing Node was taken down. In that case, the indexing information resides in the failed node does not lost since, the ontology scheme is replicated in the coordination entity. In the proposed architecture the Indexing and Maintenance Mechanism is responsible not only for keeping the indexing information but also assigning the role of Cluster Indexing Node to another DL Node to maintain the P2P prospects, and inform all other nodes about the changes made utilizing System Fault – Tolerance Web Service.

3.4 Content Retrieval
The proposed ontology-based P2P architecture can decrease the time of search by reducing the number of the distributed DL Nodes that accept the search query. This is achieved by an index based search mechanism through the layers of the presented ontology scheme. Furthermore, the semantic representation of the digital content in the P2P network allows an implementation for additional recommendation for digital content based on the relations denoted in the ontology scheme.

3.4.1 Search and Retrieval Process
Our approach is introducing a scalable searching mechanism comprising three steps: a) initialization of the query forms according to the Global Indexing Ontology, b) location of the DL Nodes including the required resources based on the distributed index of the first two layers of the ontology-scheme and c) execution of the query in the Node Content Ontologies of the selected nodes and the retrieval of the results from the query node. The query is forwarded utilizing the web services interfaces through the intermediated nodes of the P2P system resulting to the DL Nodes that contain the required resource. The query results are returned to the initial DL Node that is accessed by the user. The search and retrieval process in the proposed architecture is depicted in figure 5.

The users of the P2P system can access the content in the network through web-based forms supported in each DL Node. These forms can be dynamic generated utilizing server-side includes technologies like PHP, ASP and JSP. The first step of the search process includes the generation of the query forms that will be used by the user in order to define his/her request. These forms are composed according the Global Indexing Ontology residing in each DL node of the network. The Global Indexing Ontology contains the main categories of the digital content stored in the DL Nodes. Moreover, contains properties and relationships that define the classes-categories. These features permit the composition of a query form that includes a dynamic “catalogue” comprising the classes of the Global Indexing Ontology and keywords fields that generated following the properties of these classes. The composed query can be identified by two parts, the first part denote the classes of the Global Indexing Ontology associated with the required information and the second part the query string.
When a user submits his/her request, the DL Node is processing the query consulting the information stored in Global Indexing Ontology. This process includes the recognition of additional query related classes using the existing relationships. Next step is the discovery of the corresponding Indexing Cluster Node using the list residing in the ontology and finally the forwarding of the query through Query Web Service interface to the selected Indexing Cluster Nodes.

The next step of the search process is the identification of the DL Nodes where the required content resides. The first action is to process the query string and isolate the main terms of the query. Thereafter, a parsing to the corresponding Node Indexing Ontology results the matching of the query terms to the ontology terms. The final step comes with the discovery of the list of the query associated DL Nodes that are stored in the Node Indexing Ontology. Utilizing both the list of the related DL Nodes and the Query Web Service, the query is forwarded to the selected DL Nodes.

In the final step there is the execution of the query to the instances of the related Node Content Ontologies that semantically annotate the digital content in each DL Node. Firstly, there is a selection of the appropriate Node Content Ontologies according to the indexing process results. Afterward, a query is composed in according to the RDF Data Query Language (RDQL) [23] querying language for searching in the Node Content Ontologies Instances. RDQL is querying language for RDF structured documents, which is the base language for OWL language. The RDQL queries consists of a graph pattern, expressed as a list of triple patterns and also can have a set of constraints on the values of those variables, and a list of the variables required in the answer set.

Finally there is the elaboration of the results of the RDQL query in order to retrieve the resulted data and sending back to the initial DL Node using the appropriate web services.

Furthermore, the last issue is when a user wishes to retrieve a particular digital content which is included in the results. In that case, the appropriate client scripts are executed in order to invoke the corresponding Object Retrieval Service and acquire the desirable digital content.

3.4.2 Recommendation Service

Another supplementary service to search process, provided by our approach, is the recommendation about related digital content based on the semantic annotation that is integrated in the distributed ontology scheme. A mechanism is proposed for the searching and finding similar resources according to a query that is supplied in the system by a user. This mechanism is based on the semantic representation which has been defined in the middle layer of the proposed ontology scheme, where a more detailed description is presented. This mechanism can recommend to the user additional searching options that follow the query results.

The recommendation initialized in each Indexing Cluster Node in the second step of the presented search process. When the request for a query arrives in an Indexing Cluster Node and the identification of query-related classes is finalized, a mechanism is executed in order to find similar classes utilizing the hierarchy of the entities on the Node Indexing Ontology. The mechanism exploits the fact that ontology hierarchy denotes similar knowledge objects. According to this prospect the recommendation mechanism is searching for subclasses or classes that reside under the same super class. As an example based on the Node Indexing Ontology presented in chapter 3.2.1, the recommendation mechanism can identify that “Greek Ancient History”, “Roman Period”, “Middle-Ages” as subclasses to the “Historical Period” class. Therefore if a request from a user is related to books written in a given period, the recommendation accompanying the results may include the three mentioned subclasses as content subcategories where a more detailed search can be executed as an option. Finally the recommendation results are sending to the initial DL Node utilizing the Recommendation Web Service interface.
4. Conclusions
We have presented an ontology-based a P2P architecture utilizing SOA principles in order to integrate the various distributed Digital Libraries adopting semantic description of the stored digital content. Our approach defines a P2P architecture consisting of distributing and heterogeneous DLs supporting resource retrieval and data exchange between them. Also, additional centralized entities are used in order to execute specific tasks as new node initialization, maintenance of data integrity and system fault-tolerance.

Moreover, a distributed multilayered ontology scheme is introduced in order to facilitate the search process by providing both a well structure index and a ontology based query mechanism on the semantic representation of the metadata describing the actual digital content.

Final, our proposal aims to enforce and improve the search process among the node of the integrated network, by defining virtual clusters of DLs based on their digital content description. Utilizing this virtual classification the allocation of the semantic index to multiple nodes is accomplished assuring the efficiency and time improvement of the search process. Furthermore, the flexibility and stability of the proposed P2P system is supported by the proposed distribution of the semantic information and the associated Indexing and Maintenance Mechanism.

5. REFERENCES