SEA: Introducing the Semantic Exchange Architecture

Thomas Franz, Carsten Saathoff, Olaf Görlitz, Christoph Ringelstein, Steffen Staab
ISWeb Group, Institute for Computer Science
University of Koblenz, Germany
http://isweb.uni-koblenz.de
{franz,saathoff,goerlitz,cringel,staab}@uni-koblenz.de

1. INTRODUCTION

With novel centralized information management systems[1], users benefit from collective data organization in the form of collective information tagging. The centralized architecture of these systems, however, imposes privacy and availability constraints as personal information needs to be handed over to a third party and offline utilization is impossible. The Semantic Exchange Architecture (SEA[2][3]) addresses these issues enabling autonomy of data management and freedom of data organization. SEA ensures privacy and fine grained access control through local data storage without lacking the advantages gained by collective information organization. Distributed information management is achieved through Peer-to-Peer networking in combination with semantic web technologies, while the benefits of collective information tagging are available through anonymous tag distribution in the Peer-to-Peer network. Information can be shared both publicly with all connected peers as well as privately within networks of trust.

As a use case example for SEA, we consider project work in which multiple institutions take part so that confidential data is distributed at multiple locations. The setup of a centralized information management system to share that information increases costs and decreases ease-of-use due to the typical upload and download procedures in centralized systems. SEA enables working group members to easily share their information, e.g. by tagging all relevant information with the name of the working group, and associating group access with that tag.

2. ARCHITECTURE

SEA constitutes a network of decentralized repositories in which information is collectively organized by tags. A repository runs on a local desktop and provides locally stored information as well as a portion of globally shared information. Taggings are used to enable local as well as networked access and exchange of the information distributed over multiple peers.

2.1 Organization of Information

Conventional hierarchical data organization is unable to represent different perspectives onto some data. SEA employs tagging, the association of user defined catchwords with information objects, as a mechanism to allow more flexible data management and retrieval. In contrast to taxonomies, tagging provides more freedom to organize information since it does not impose any relations between tags. An information object can be tagged with multiple tags to represent different perspectives onto the object.

Based on the assumption that multiple users associate the same meaning to a tag, sharing taggings allows further exploitations: First, by requesting all information objects tagged with a tag k, users can retrieve information objects related to k that they are not aware of, however have been tagged with k by other users. Second, users can find information objects that are related to an object o by requesting information objects that share one ore more tags with o.

2.2 Data Model

SEA employs ontologies as meta models (micro models[4]) for the managed data to achieve interoperability and extensibility. We further argue that building novel systems on ontologies from the beginning leverages integration of knowledge, reasoning and further improvements later on. Figure 1 illustrates how we combined ontologies that model tagging, information resources, and access control.

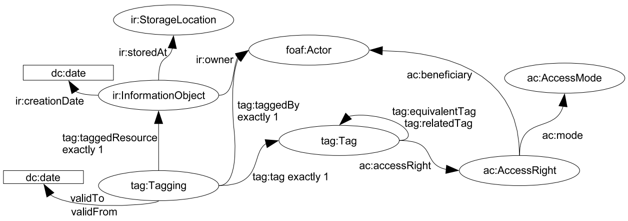

Figure 1: Ontologies in SEA

2.3 Access Control

[1] http://del.icio.us, http://www.flickr.com

[2] http://isweb.uni-koblenz.de/Research/sea

[3] This research was partially supported by the European Commission under contract FP6-001765, aceMedia. The expressed content is the view of the authors but not necessarily the view of the aceMedia project as a whole.

[4] http://esw.w3.org/topic/MicroModels

Demos and Posters of the *3rd European Semantic Web Conference (ESWC 2006)*, Budva, Montenegro, 11th - 14th June, 2006

SEA utililizes tags to organize information but also to define access rights for the shared information. This approach is easy-to-use by users as only the system of tagging needs to be learned in order to both flexibly organize information and maintain access control. Access control is realized based on rules which define the behavior of SEA for non-trivial cases, e.g. if access rights are associated with multiple tags that classify the same information object, or one user associates access rights to a tag, while another user adds the same tag to an information object.

2.4 Data Distribution

Which data is distributed, and how the information retrieval of that data is implemented in SEA depends on how data is shared. We distinguish between public data (shared with everybody), and protected data (shared only with dedicated users). For public data, taggings and object locations are distributed to enable a simple object retrieval as well as exploitations as listed in 2.1. Dealing with protected data is more complex and explained in the following.

2.4.1 *Combining Privacy and Collective Tagging*

Obeying privacy demands and exploiting collective tagging are contradicting goals as privacy demands that data is not publicly shared while exploiting collective tagging demands to share information. SEA supports those exploitations whithout breaking privacy rules by only distributing anonymized taggings (identifiers of information objects and associated tags) for secured data. Such a distribution of taggings allows the identification of information objects by exploitations as listed in 2.1, however, due to the missing location information disallows their retrieval. The retrieval of protected data is based on the consultation of a finite list of peers to which the retrieving user is known, similar to a buddy list in instant messaging software. Consulted peers check whether the retrieving peer has appropriate access rights before providing the requested information. We argue that this solution is sufficient to find information objects that can be accessed by the particular user as owners of protected information objects are expected to know the users for which they grant access and vice versa. If one wants to grant access to users one does not know, public access can be granted.

2.4.2 *Distribution Mechanism*

SEA utilizes a distributed hashtable (DHT) approach to distribute information in the network. Four hashtables are employed to efficiently represent the needed information. Table tab_o contains for each information object id the set of all tags associated with that object and thus allows to retrieve all tags associated to an information object. Another table tab_k allows for querying in the opposite direction, i.e. retrieval of all object ids for a tag. While the computation of tag correlations would be possible by using only tab_o and tab_k, it would require multiple request and thus increase network load. We argue that memory and space costs are lower than those for network bandwith and model tag correlations by an additional DHT tab_{co} that maps each tag k to the set of tags that occur together with k. As we distribute location information for those information objects that are public, that information is contained by the table tab_l that maintains for each information object a set of locations where it is available.

3. IMPLEMENTATION

The main components of SEA are the data repository, SEA core, the peer manager, and the DHT module as depicted in figure 2. The repository is a RDF[5] store that supports SPARQL[6] and stores all metadata available in SEA. SEA Core constitutes the module which is directly accessed by applications using SEA and therefore is responsible for distributing all requests to appropriate modules. The interface of the core is a REST Api, that offers a number of central services: i) Requesting resources with a specific tag, tags of a resource with a specific identifier and tags that are related to another tag, ii) modifying the tags of resources, and iii) authentication of trusted users. The peer manager module is responsible for all operations involving the communication with other peers, i.e. rendevous, authentication and request forwarding. Additionally, it provides peer information for the DHT implementation. Communication with other peers is established via the common interface exposed by every peer. Results of forwarded requests are handed back to the SEA Core for further processing. The DHT module uses a distibuted hashtable implementation to efficiently store general tagging information so that it is available for all peers in the network. SEA is under development[7], efforts are concentrating on the SEA Core implementation and the integration with the Sesame2[8] RDF repository. Additionally, we started implementing a simple file browser that allows to tag arbitrary resources and submits taggings and other information to SEA Core. In parallel we also evaluate possible solutions for the DHT implementation, namely Pastry and Bamboo[9].

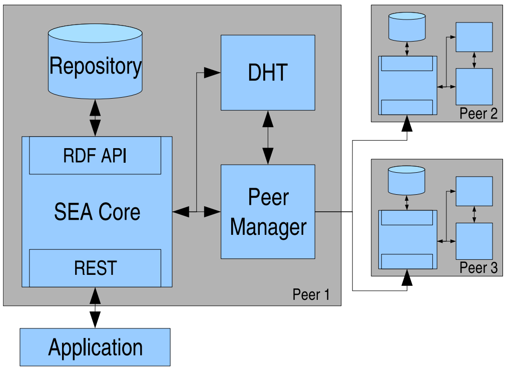

Figure 2: Implementation of SEA

4. CONCLUSION

SEA tackles shortcomings of conventional information sharing platforms by providing secure, collective, and distributed information organization. SEA's open architecture offers easy adoption, extension, and development as it is based on acknowledged standards that are well supported by programming libraries and development tools. Work on SEA contributes to research on P2P systems, Social Network Analysis, and the Semantic Web (in particular the development of the Networked Semantic Desktop).

[5] `http://www.w3.org/TR/rdf-primer/`
[6] `http://www.w3.org/TR/rdf-sparql-query/`
[7] `http://isweb.uni-koblenz.de/Research/sea`
[8] `http://openrdf.org`
[9] `http://bamboo-dht.org/`