LinkZoo: A linked data platform for collaborative management of heterogeneous resources

Marios Meimaris, George Alexiou, George Papastefanatos

Institute for the Management of Information Systems, Research Center "Athena", Greece
{m.meimaris, galexiou, gpapas}@imis.athena-innovation.gr

Abstract. Modern collaborations rely on sharing and reusing heterogeneous resources. The ability to combine different types of information objects in semantically meaningful ways becomes a necessity for the information-intensive requirements of collaborative environments. In this paper we present LinkZoo, a web-based, linked data enabled platform that allows users to create, manage, discover and search heterogeneous resources such as files, web documents, people and events, interlink them, annotate them, exploit their inherent structures, enrich them with semantics and make them available as linked data. LinkZoo easily and intuitively allows for dynamic communities that enable web-based collaboration through resource sharing and annotating, exposing objects on the Linked Data Web under controlled vocabularies and permissions.

Keywords: Linked Data, Semantic Web, Collaborative Environments, Resource Management, Personal Information Management

1   Introduction

The semantic web is being manifested in a large way through the linked data (LD) paradigm, focusing on either the provision of semantic representations of singular entities through web documents or on large datasets created through domain-dependent, often complex publishing workflows. In this sense, the process of creating and publishing linked data has proven to be non-trivial as technical and conceptual expertise is required in order to build tailored LD publishing frameworks, more so when exposing or consuming personal resources in dynamically created collaborative environments.

The co-existence and collaborative management of heterogeneous objects in rich information spaces is a requirement that has not been successfully met. Providing a common representation model for different types of resources, such as files, web documents, persons and events enables their organization under shared contexts and the creation of complex, artefact-oriented aggregations. The reuse of common ontologies and vocabularies provides a rich semantic layer that helps organize, interlink and explore these resources in a multitude of dimensions. LD provides the way for exposing them, publicly or privately, for external reference and processing.

There are several tools in the fields of collaborative semantic editing [2,6], tag-based file systems [1] and semantic desktops [3,4,5]. Most of these either target the editing of resource properties or the semantic representation of physical files; however they do not address adequately the need of non-expert users to be able to leverage LD technology for creating, organizing and exploiting collaborative information spaces containing diverse resources found on the Data Web.

In this paper, we present LinkZoo; a collaborative linked data platform that enables users to reuse established ontologies as well as semantics created on-the-fly in order to annotate and share heterogeneous resources over dynamically defined usage contexts. It provides cloud-based functionality, where users can upload, link and manage various different types of resources, i.e., more than plain files, in collaborative workspaces. It offers the sharing and collaborative authoring of resources, enrichment with properties manually or via external services and their publishing as linked data for citing and machine-based consuming. LinkZoo enables users to create views that organize their resources under very different perspectives and make them available to others. Finally, it offers an intuitive way of searching over private or public resources and exploring them via a faceted browsing functionality.

This paper is organized as follows: in section 2 the main features are presented, in section 3 the design and architecture is described, section 4 deals with the system's implementation and demo and section 5 concludes and discusses future directions.

2 LinkZoo Features

Linked data publishing and sharing. LinkZoo seamlessly integrates the processes of creating and publishing heterogeneous resources as RDF linked data, and offers a toolkit of common actions for their appropriate management. Four resource types are handled at the moment, namely files, web documents (i.e., URLs), persons and directories. Directories are special types of resources that organize into contexts sets of resources and can be enriched and linked with knowledge in the same way as other resources. Extending the list of resource types is a trivial procedure, as we intend to do so with calendar events in the near future. The available user actions include creation of a new resource (file upload, URL import or custom definition via JavaScript), creation of new directories, move, rename and delete resources, sharing of resources and editing/enriching of resource properties, either manually or in an automatic way. A SPARQL RESTful endpoint is deployed in order to be accessed programmatically with the appropriate user credentials.

Collaborative resource management. The created resources can be processed, annotated, enriched and shared by their editors, independently of their type. Currently two user roles are implemented, owner and editor. Owners and editors practically have the same rights, but help keep track of provenance. In the future, the viewer role will be introduced. Furthermore, resources can be private (when their discoverability is defined by the sharing process) or public (when they can be openly discovered and annotated). Shared directories bequeath their sharing status to their contents. Resources can be given properties by any shared editor.

Non-structured, loosely-controlled annotations. Users can assert facts about resources as RDF triples. Some well-known ontologies have been imported in the platform for ease of access (auto-complete mechanism), but users can define their own properties on the fly under their custom namespace or use external ontologies and vocabularies at will. Furthermore, collections of resources in different directories can be aggregated in a drop-zone and assigned with properties in a bulk manner. Enrichment with properties and interlinking with other resources and external LD can be done manually by the users or automatically using external web service APIs.

Keyword search and exploration. LinkZoo offers advanced resource exploration by combining keyword search functionality over resource descriptions with property based filtering. Keyword search is implemented in a close-to-natural-language way, based on the characteristics of a user's available resources. Property filtering is implemented in the form of facets over the list of available properties of the search results. The two methods can be combined and applied in an exploratory "find-as-you-go" manner and ongoing results can be stored in multiple views.

View-centric organization. Resource descriptions provide multiple dimensions of organization due to their properties. The default way of browsing resources is based on an intuitive directory-like manner; still our platform exploits resources' semantics for offering multiple ways for organizing, exploring and searching resources. Users have the ability to organize resources (public or private), based on their characteristics and store the results as linked views. Views can be static or dynamic, where the latter store their search parameters thus enabling the user to refresh the contents of the view based on its definition. Views leverage semantic web by offering intuitive means to users for organizing, searching and discovering new resources either within the platform or the entire LOD cloud.

3   Model and Architecture

LinkZoo uses its own data model implemented as a linked data vocabulary. The model has three parts; the first contains the resources along with their metadata, such as types, titles, descriptions, identifiers, dates and related users, the second contains the set of constructs for user administration and privileges over resources and the third contains the definitions of views and the participation of resources in them.

The main architectural components are shown in Figure 1. LinkZoo employs a quad store for data manipulation; profile manager is used for managing the profile data of each user, resource manager implements all actions applied on resources, view manager is used for defining and update of static and dynamic views whereas the search and exploration module provides keyword search and property filtering. Furthermore, the interlinking and enrichment module uses external APIs in order to retrieve and enrich resources with facts. Currently, Wikipedia, DBPedia and AlchemyAPI are used for knowledge enrichment, and the FalconsAPI for searching external ontologies. Permissions and user privileges are managed separately, as most actions require user authentication.
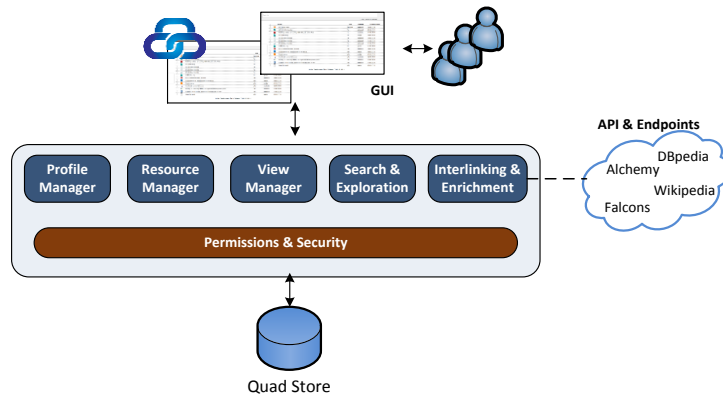
Figure 1. LinkZoo Architecture

4   Implementation and demonstrator

LinkZoo is implemented using open source technologies. Virtuoso 7 open source edition is used as a quad store. The web GUI and application server are deployed on an Apache web server, using PHP for the main functionality.

Resources are given URIs upon creation based on a simple timestamp mechanism. Upon registration, each user is given a dedicated named graph, used in order to store the resources created by each user. When a resource is shared there is no replication of its description, thus allowing users to be able to collaborate on the same data object when manipulating a resource. Hence, the sharing procedure creates two triples on the shared users' graphs, one f their role and one to declare the resource's source graph.

Searching is implemented using auto-complete suggestions over resource types and property values, in combination with a faceted property filtering mechanism. The resulting query strings are visualized as natural-language phrases (e.g. '*find URLs with rdfs:seeAlso dbpedia:Youtube and linkZoo:owner 'John'*'). Search can be limited on a directory, a view or a user's privately shared resources, but can also be done publicly. Public resources are stored in a publicly shared – anonymous - graph.

Views are also named graphs containing references to resource URIs and they can be static or dynamic. A static view is manually created and the user can explicitly add one or more resources to it. A dynamic view is created based on a search operation and the search parameters (SPARQL query) are kept along with the retrieved resources. The user can refresh the view's contents by revaluating the query on the available public or private resources. Views act as workspaces allowing the same resources to be found in many collections at the same time.

Demonstration Scenario. Our demo involves a Human Resources Management scenario. The scenario assumes that companies publish job openings in various formats (documents and websites) and independent recruiters (users of the demo) collaborate in order to find and match candidates to openings, along with related resources such as linkedIn profiles, CVs, referrals, publications, etc. Each recruiter owns his

own portfolio (user graph) with resources and has access to candidate profiles shared to him as well as to public resources.

First, the recruiter discovers available job openings by performing a search on the public workspace for resources posted by certain companies. The results are then explored based on the skills attributed to each job description. Then, the user searches over his personal portfolio for possible candidates and resources that exhibit some of the required skills. The results contain persons, CV documents, user and web pages, which are properly organized into a view per candidate. These are further linked with interview minutes, referrals, and cover letters. Some of the enriched profiles are then shared with other recruiters. Finally, the recruiter filters out candidates based on their skills and save the results as a dynamic view that he refreshes when new possible candidates are found in the shared with him portfolios.

Our platform prototype along with the demo is available at *http://snf-80575.vm.okeanos.grnet.gr/encode2/index.php* and the video presenting the afore-mentioned scenario is at *http://youtu.be/bwQFOr80cZM* .

5   Conclusions and Future Work

We have presented LinkZoo, a LD platform for collaborative management of heterogeneous resources. LinkZoo's social and semantic aspects can be used to facilitate collaboration between non-expert users over diverse domains and scenarios. In the future we intend to extend the coverage of resource types and incorporate data from social networks, such as linkedIn, g+ and youtube in order to build thorough social user profiles. We also intend to study scalability and performance issues concerning large userbases and greedy annotation. Finally, we intend to use the platform as a test-bed for automated integration of information resources on the Data Web.

6   References

1. B. Schandl, N. Popitsch. "Lifting File Systems into the Linked Data Cloud with TripFS." In LDOW. 2010.
2. S. Auer, S. Dietzold, T. Riechert. "OntoWiki - A Tool for Social, Semantic Collaboration". In ISWC 2006: 736-749
3. A. Bernardi, G.A. Grimnes, T. Groza, S. Scerri. "The NEPOMUK Semantic Desktop". Context and Semantics for Knowledge Management 2011: 255-273.
4. T. Franz, S. Staab, R. Arndt. "The X-COSIM integration framework for a seamless semantic desktop." In K-CAP2007.
5. L. Sauermann. "The Gnowsis Semantic Desktop for Information Integration." Wissensmanagement. 2005.
6. D. Quan, D. Huynh, D. R. Karger. "Haystack: A platform for authoring end user semantic web applications." In ISWC 2003. 738-753.