Closing the discovery gap in environmental information resources using semantic annotations: the TaToo Approach

Tomas Pariente Lobo[1], Mauricio Ciprian[1], Gerald Schimak[2], Giuseppe Avellino[3], Sascha Schlobinski[4]

[1] ATOS Origin, Madrid, Spain, mauricio.ciprian@atosresearch.eu, tomas.parientelobo@atosresearch.eu
[2] Austrian Institute of Technology, Seibersdorf, Austria, gerald.schimak@ait.ac.at
[3] Elsag Datamat, Rome, Italy, giuseppe.avellino@elsagdatamat.com
[4] cismet, Saarbrucken, Germany, sascha.schlobinski@cismet.de

Abstract. Internet has allowed the exponential growing of the information available in all knowledge domains. This information is published in a non-coordinated way using different standards and formats. Traditional search engines solve with relative success the needs of most Internet users searching for general purpose information. However, specific users like scientists trying to gather scientific information across boundaries require more specific and refined information retrieval techniques to satisfy their needs. Obtaining this kind of information with traditional search engines is a time-consuming task with no guaranties of success. The environmental domain is not a specific exception to that problem. The TaToo project aims at solving the problem by using a semantic approach i.e. enrich information with semantic annotations in order to improve the discovery of environmental resources.

Keywords: semantic annotation; semantic tagging; model search and discovery; web services; environmental information enrichment.

1 Introduction

The TaToo (Tagging Tools based on a Semantic Discovery Framework) project tries to build tools that allow users to discover and tag resources of interest in order to facilitate their search by other users having similar interests. The tools developed in the TaToo project will be mainly focused on environmental information resources (both data and services).

The tagging process in TaToo will consist of adding semantic meta-information, with the goal of improving the discovery process and also the interpretation and evaluation of its validity. As mentioned in previous papers [1, 2], the addition of axiom-based semantics to environmental data overcomes the limitations of other strategies such as thesauri, dictionaries and schemas due to the fact that the expressivity of the information is increased , and the complexity required to convey is decreased as addressed in [3].

The TaToo tools aim at creating an information enrichment cycle of continuous discovery, enrichment (tagging) and publishing in order to encourage communities to setup, use, extend and promote their knowledge. Fig. 1 reflects the life-cycle of the information enrichment as envisaged in TaToo.

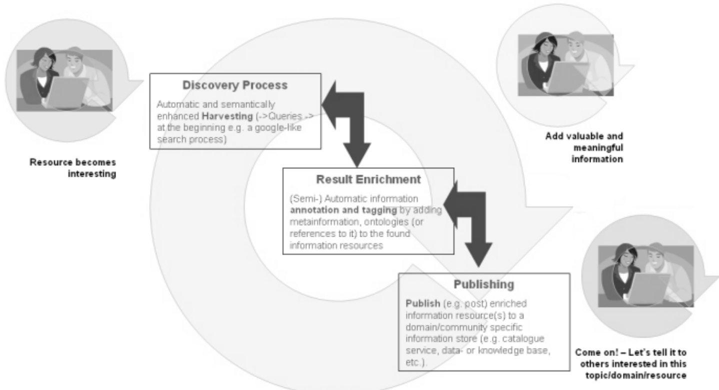

Fig. 1. TaToo's cycles of information enrichment

TaToo foresees to deal with different types of environmental information resources, such as catalogues, environmental models, Web services or Web pages, among others. These resources have to be discovered, tagged, and evaluated. Evaluation and tagging of resources by the end user make possible the information enrichment process; searches will be more and more effective as each time based on a larger amount of available metadata provided by the users.

TaToo has the objective of allowing a cross-domain discovery of resources in the environmental field, meaning that resources annotated with different purposes and possibly different ontologies (defining ontology as a description of an area of knowledge [4]) should be retrieved using a common framework. However, there are two major challenges while trying to solve this semantic heterogeneity issue: Allowing multi-domain annotation schema and implementing an extensible discovery mechanism. In order to solve those problems Wache [5] proposes three ways for integrating ontologies: by using a single ontology, multiple ontologies or a hybrid approach. These ways have been evaluated in a previous paper of TaToo project [1]. We concluded to use the hybrid approach for, evaluating the most widely used ontologies in the environmental field as candidates to be used as basis of the shared ontology. Also, the ontologies used and developed in TaToo project will be based on existing W3C standards, particularly RDF, RDFS and OWL. In the ontology engineering process we will also try to reuse as much shared vocabularies as possible such as DC[6], FOAF[7], SIOC[8]. TaToo will not provide an ontology engineering tool, but it will rely on existing tools for ontology engineering such as the NeOn Toolkit or Protégé. Within TaToo we use the NeOn methodology to build ontology networks and semantic applications.

The TaToo's tools will be applied to solve particular problems in three scenarios embedded in highly complex environmental domains: climate change, agriculture, and anthropogenic impacts of pollution. This last scenario is deeply described in [9]

2 TaToo Architecture

In order to fulfil the TaToo goal of providing functionality for discovery, tagging, searching and evaluating environmental resources, the TaToo framework has to supply a set of services at server side providing the implementation of the functionality, plus a graphical user interface that lets the end users interact with the server interface. In the TaToo architecture it is possible to identify four key building blocks. These building blocks are shown in Fig. 2.

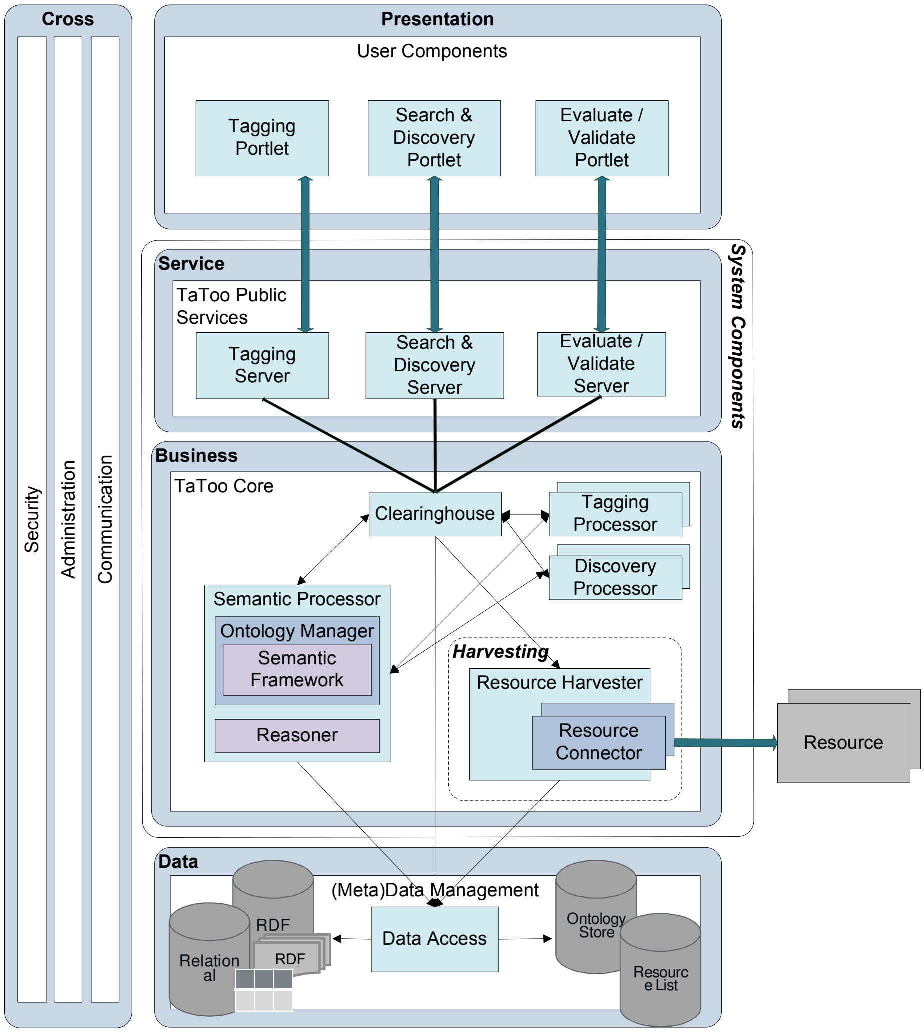

Fig. 2: TaToo Initial Architecture

The 'User Components' building block (Presentation tier) contains three different sub-blocks. Tagging Tool, Search & Discovery Tool, and Evaluate / Validate Tool offer the respective client side components to exploit system corresponding functionality. In general, these components are supposed to be directly used by the end user and can be implemented as portlets, external applications, browser plug-ins or API's.

The 'TaToo Public Services' block (Service tier) contains corresponding sub-blocks at server side. These are generally Web services offering an interface to the underneath TaToo Core components.

The 'TaToo Core' building block (Business tier) is the main part of the architecture and contains 'core' TaToo components implementing the business logic. In particular four components are of major importance:

- The Clearinghouse is a central component for accessing the metadata storage and serves also as an information exchange support between the core system components;
- The Semantic Processor is the fundamental component dealing with Semantics. It uses a set of (pluggable) ontologies (in the environmental domain) to provide functionality based on semantics. In general, it relies on an application framework and a reasoner to provide its functionality. As an application framework provides useful APIs to manipulate RDF, support SPARQL query, and others;
- The Harvester is the component capable of retrieving external resources (and associated metadata) that could be either data or associated metadata stored in catalogues, Web services or information contained in Web pages. The harvester plays the role of retrieving already available resource metadata.

Finally, the '(Meta) Data Management' building block (Data tier) deals with the storage where metadata on resources are stored (both resource owner metadata and metadata provided by end users through tagging).

3. Conclusions

We have presented in this paper the TaToo project and first results obtained so far, including a first definition of TaToo architecture. This architecture will be refined in the cycle of three iterations planned in the project. However, we envisage that the TaToo system will mitigate the burden of adding semantic meta-information to environmental resources in a community driven way facilitating the search and discovery processes.

We expect that at the end of the project, the users of the tools provided in this project will benefit from having semantically-enriched environmental resources shared with their scientific community.

From the technical perspective TaToo will realise an extensible and cross-domain semantic framework able to deal with environmental tagging and discovery. This framework will be validated through the TaToo Validation Scenarios, but will be potentially used in the bigger picture of the environmental domain.

Acknowledgments: The research leading to these results has received funding from the European Community's Seventh Framework Programme (FP7/2007-2013) under Grant Agreement Number 247893.

References

1. G. Schimak et al.: Information enrichment using TaToo's semantic framework; Proceedings of Metadata and Semantics Research (MTSR) Conference 2010, Alcala de Henares, Spain, S. Sanchez-Alonso, I. N. Athanasiadis, E. García-Barriocanal, N. Palavitsinis, http://www.ieru.org/org/mtsr2010/; (2010).
2. A.E. Rizzoli et al.:TaToo: Tagging environmental resources on the web by semantic annotations; Proceedings of International Environmental Modelling and Software Society (iEMSs) 2010 International Congress on Environmental Modelling and Software Modelling for Environment's Sake, Fifth Biennial Meeting, Ottawa, Canada David A. Swayne, Wanhong Yang, A. A. Voinov, A. Rizzoli, T. Filatova (Eds.) http://www.iemss.org/iemss2010/index.php?n=Main.Proceedings; (2010).
3. Villa F., I.N. Athanasiadis, and A.E. Rizzoli, Modelling with knowledge: A review of emerging semantic approaches to environmental modelling, Environmental Modelling and Software, 24(5), 577-587, (2009).
4. W3C, OWL Web Ontology Language Use Cases and Requirements. W3C Recommendation (2004).
5. H. Wache, T. Vögele, U. Visser, H. Stuckenschmidt, G. Schuster, H. Neumann, S. Hübner. Ontology-Based Integration of Information A Survey of Existing Approaches; In IJCAI--01 Workshop: Ontologies and Information Sharing, pp 108-117 (2001).
6. Weibel, S., Kunze, J., Lagoze, C., and Wolf, M.: Dublin Core Metadata for Resource Discovery. RFC. RFC Editor (1998).
7. Li Ding, Lina Zhou, Finin, T., Joshi, A.: How the Semantic Web is Being Used: An Analysis of FOAF Documents. Proceedings of the 38th Annual Hawaii International Conference on System Sciences (2005).
8. J. Breslin, U. Bojars, A. Passant, S. Fernández and S. Decker. SIOC: Content Exchange and Semantic Interoperability between Social Networks. W3C Workshop on the Future of Social Networking, Barcelona, Spain (2009).
9. J. Hrebicek, L. Dusek, M. Kubasek, J. Jarkovsky, K. Brabec, I. Holoubek, L. Kohut and J. Urbanek. TaToo Validation Scenario for Anthropogenic Impact and Global Climate Change. Proceedings of EnviroInfo 2010 International Workshop on Environmental Information Systems and Services - Infrastructures and Platforms. (2010).