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

Tomas Pariente Lobo, Mauricio Ciprian, Gerald Schimak, Giuseppe Avellino, Sascha Schlobinski

<giuseppe.avellino@elsagdatamat.com>
Overview

- The Problem
- Project Overview
- Ontology Development
- High Level Architecture
- Core Components
- Architecture Overview
- Web Portal
- Communities and Validation Scenarios
The Problem

- 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**.
Project Overview

- **TaToo (Tagging Tools based on a Semantic Discovery Framework)** tries to build **Tools** allowing users to **Discover** and **Tag** resources of interest in order to facilitate their discovery
- The focus is on environmental information resources (both data and services)
- **Tagging** process consists of adding **Semantic meta-information**, with the goal of **improving** the **Discovery** process
- TaToo aims at creating an **Information Enrichment Cycle** of continuous resources discovery, enrichment (tagging), and publishing in order to encourage communities to setup, use, extend and promote their knowledge
Project Overview

- 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.
Project Overview

- TaToo foresees to deal with different types of **environmental information resources**, such as **catalogues**, environmental models, Web services, or Web pages.
- **Cross-domain** discovery of resources, meaning that resources annotated with different purposes and possibly different ontologies.
- **Semantic heterogeneity** problem: allowing multi-domain and multilingual annotation schema implementing an extensible discovery mechanism.
Ontology Development

User Needs Analysis

Risk Analysis

Requirements Analysis

Application Requirements

Content Requirements

Application Ontology Development

Content / Domain Ontology Development

Ontologies Integration

Knowledge Base Development

Validation
Ontology Development

- Start from already available ontologies
  - Modelling what is missing (removing what is not required)
  - Changing what does not suite TaToo
- Ontologies Integration: Merging, Alignment, and Mapping
- Identification of how the content should be annotated
- Vocabulary e.g. Dublin Core, FOAF, SIOC
- Ontologies e.g. GeoNames
- Thesaurus e.g. GEMET (multi-linguality)
Ontology Development

- Two kinds of ontologies to be considered and integrated
  - **Application Ontology**
    Describes the classes of contents handled by the application and the RDF properties
  - **Content Ontology** (Domain Ontology)
    Describes possible values for the RDF properties

- TaToo adopts
  - The **NeOn methodology** to model ontologies
  - The **Hybrid approach** for **Ontology Integration**
Ontology Development

Ontology integration approaches overview in Wache, 2001

- Single ontology approach
- Multiple ontology approach
- Hybrid ontology approach
Integrated Ontology

Application Ontology

Bridge Ontology

Content Ontology 1

Content Ontology 2

Content Ontology n

Portlets, Applications, Browser plug-ins / toolbars, …

Data / Services

SOAP/ REST
High Level Architecture

Presentation
- User Components

Service
- TaToo Public Services

Business
- TaToo Core
- Clearinghouse
- Semantic Processor
- Harvester

Data
- RDF
- Relational

Resources
- Models
- Catalogue
- SOAP Service
- REST Service
- XML
Core Components

- 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 a **Semantic Framework** and a **Reasoner** to provide its functionality.

- 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.
Architecture Overview
Web Portal

- TaToo aims at providing accessible and easy to use GUIs in order to facilitate and encourage users in the process of tagging (*taggers*).
- The TaToo Web Portal to take advantage of Web ubiquity
Communities and Validation Scenarios

- **Promote the framework**: relevant **scientific communities** such as the International Environmental and Modelling Society (iEMSs), the International Federation of Information Processing (IFIP), and members of the Central and Eastern European Centre for Persistent Organic Pollutants (CEEPOPsCTR)

- **TaToo results** will be applied to solve particular problems in three **Validation Scenarios** embedded in highly complex environmental domains: **climate change**, **agriculture**, and **anthropogenic impacts of pollution**
Project Partners
Acknowledgement

The research leading to these results has received funding from the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement n° 247893.
TaToo
TAGGING TOOL BASED ON A SEMANTIC DISCOVERY FRAMEWORK

VISION

TaToo, which stands for "Tagging Tool based on a Semantic Discovery Framework", is an FP7 Project sharing the vision of a Single European Information Space for Environment (SIDE). It aims to enable experts as well as general users to share trusted and reliable information, but also to allow easy discovery of information which is already available.

CHALLENGES

A hot issue in environmental modeling is the efficient discovery of high quality and meaningful information. Metainformation attached to current environmental data repositories have been collected, structured and stored in forms that are meant for human users. Yet, it is hard, if not impossible, to machine process such metainformation. As a result, significant human intervention is required for discovering environmental information: (a) The process of locating appropriate sources, querying and exploiting environmental data cannot be automated. (b) Human experts are required to undertake such tasks. Finally (c) the generation of metainformation, just for documenting existing data, turns out to be too heavy a burden for all kinds of users, and more so for data owners.

CLOSE THE GAP

What is needed, is a middleware infrastructure to fill the gap between environmental resources, resources and end-users. This framework must facilitate the life-cycle of environmental information from its collection and persistent storage up to its discovery and purpose-oriented exploitation.

RESULTS

TaToo will provide a semantic framework for discovery and access to environmental resources in a multilingual and multi-domain context. Therefore, it will provide:

- Tools for:
  - Discovery of information
  - Semantic interpretation of information
  - Annotation of information
  - Information Enhancement – Tagging and Annotation
  - Visualisation of information

- Services (based on standards where possible) for:
  - Integration and Processing of Information
    (e.g. a semantic processor)
  - Storing/Archiving of TaToos
    (e.g. semantically enriched information objects)

PROJECT COORDINATES

Project Start: 01.01.2010
Duration: 36 months
Total budget: 3.6 M€ro
Total funding: 2.5 M€ro

ACKNOWLEDGEMENT

The research leading to these results has received funding from the European Community’s Seventh Framework Programme (FP7/2007-2013) under grant agreement no. 247553.

PARTNERS

AIT Austrian Institute of Technology GmbH, Austria
ATOS Origin, Spain
Cismet GmbH, Germany
ELSAG Datamat, Italy
Istituto Dale Mole di Studi sull’Intelligenza Artificiale, Switzerland
Masaryk University, Czech Republic
Joint Research Centre, Italy

CONTACT

Dr. Gerald Schimak
Coordinator TaToo
Phone: +43(1) 50550 - 3125
Fax: +43(1) 50550 - 2813
E-mail: gerald.schimak@ait.ac.at
Web: www.tatoo-project.eu

Giuseppe Avellino, ED
A Finmeccanica Company
BACKUP

- BACKUP
Project Overview

- 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.
Project Overview
Initial Considerations

- Resources (data, services, etc.) have to be **Auto Descriptive** i.e. they have to be [semantically] described by a **proper** set of Metadata

- A **format** for the TaToo Metadata have to be defined
  - Elements belonging to this format have to be identified by URIs
  - Attributes and Values related to Ontology Concepts

- Possible elements have to be well identified and shared (wide accepted dictionary)

- Tags / Annotations are provided by **users** to catch ‘The wisdom of crowds’ (*social tagging ➔ user generated content ➔ Taxonomies - Folksnomies*)

- Tags / Annotations provided by **resource owners** to be considered as well (e.g. external catalogues)