Using ontologies for R functions management

Yuan Lin[a], Pascal Neveu[a], Caroline Domerg[a], Juliette Fabre[a], Alexandre Granier[a],Anne Tireau[a],
Vincent Negre[a], Isabelle Mirbel[b], Olivier Corby[c] and Catherine Faron-Zucker[b]

[a]INRA
Montpellier, France

[b]I3S, Université de Nice
Sophia Antipolis, France

[c]INRIA, Sophia Antipolis Méditerrane
Sophia Antipolis, France

[*]Contact author :
Pascal.Neveu@supagro.inra.fr

The research work of biologists often requires the development of a lot of scripts or functions to manipulate and analyse experimental data. In the laboratory LEPSE specialized on the analysis and modelling of plant responses and adaptation to variable environmental stresses, dozens of R functions are produced every year, concerning various fields such as genetic analyses, high throughput data phenotyping or environmental interactions and involving several databases. As a result, there is an important turn-over of function authors and users which generates different problems like re-using, sharing or understanding of these functions. In this context, in the framework of the DESIR project[1], we have initiated a knowledge management action aiming to capitalize, organize, share and valorize these functions through the development of a knowledge-based repository of these functions.

Given the great diversity of the functions produced, their associated documentation is heterogeneous and it is not pertinent to organise them into packages. We decided instead to index them with some formalized knowledge describing them, in order to retrieve them by formal reasoning. For this purpose, we developed an ontology providing a controlled and structured vocabulary that captures the concepts and properties necessary to describe R functions. This ontology comprises concepts and properties to describe functions - like "Author", "Intention", "Argument", "Value" - as well as the relations between functions - like "hasForRCoreCall", "canBeUsedAfter", "isAdaptedFrom", "looksLike". As a result functions can be retrieved according to a wide range of criteria: their author and/or the graphics produced, their intention(s) (e.g. perform multidimensional exploration), the function(s) they call (e.g. the "lm" R core function or a specific function of the repository) -more generally, it is relevant to generate the call graph of one function to understand it-, the functions from which they are adapted - this makes easier the maintenance of the repository-, the functions after or before which they should be used -this helps to construct chainings of treatments-, their similarity with other functions, etc.

To formalize both the ontology and the annotations of R functions, we adopted the Seman-

---

[1]http://www-sop.inria.fr/edelweiss/projects/desir/wakka.php?wiki=ColorDesirHomePage

tic Web models: The annotations are represented into the Resource Description Framework[2] (RDF) and the ontology in the Ontology Web Language[3] (OWL). As a result we are able to semantically retrieve R functions by expressing queries in the SPARQL language[4]. We developed a semantic web application for the repository and search of annotated R functions. It relies upon the semantic engine Corese [1] dedicated to ontological query answering on the Semantic Web: Corese enables to interpret and process SPARQL queries on RDF annotations and OWL ontologies. Our application provides an environment for (1) storage and annotation: a prototype of Web user interface allows authors to upload R functions (one function per file) and to describe them in a few minutes; and (2) powerful search: users can find and get R functions with a global and accurate understanding and receive suggestions to support their search.

To conclude, we have built a semantic repository of annotated R functions to centralize and share R functions for biologists. It capitalizes expert know-hows that would otherwise oftenly be lost or become non- usable because of a lack of documentation and description. We are convinced that this kind of repository developed for the LEPSE could benefit a much wider community of R function authors and users and be adapted to handle other programming languages.

Références

[1] O. Corby, R. DiengKuntz and C. FaronZucker (2004). Querying the Semantic Web with Corese Search Engine, ECAI 2004, Proc. of the 16th Eureopean Conference on Artificial Intelligence, (Valencia, Spain), August 2004, 705-709.

---

[2] http://www.w3.org/RDF/
[3] http://www.w3.org/2004/OWL/
[4] http://www.w3.org/TR/rdf-sparql-query/