

Importing Knowledge Fragments to CMS-Enabled Data Mining Analytical Reports

Andrej Hazucha, Tomáš Kliegr, and Vojtěch Svátek

University of Economics, Prague,
{xhaza00,tomas.kliegr,svatek}@vse.cz

Descriptive data mining only brings its fruits when the results are provided to the end user in a palatable form. The vehicle for end-user delivery of mining results (and associated information such as data schema, task settings, and domain background knowledge) are so-called *analytical reports*. In order to manage a huge number of reports referring to different mining sessions, we designed a *data mining web portal* based on a *content management system*, together called SEWEBAR-CMS.¹ One of the requirements on the CMS was the ability to interact with *semantic knowledge sources* and other structured data, see [1].

The data analyst who authors an analytical report in the CMS has different possibilities of (semi-)automatically entering structured data into the text.

First, for *locally stored* data such as mining task/result/data descriptions exported from mining tools in PMML (Predictive Model Mark-Up Language), a CMS plugin can pick marked segments of HTML code, produced from PMML using XSLT, and insert them into the report as indicated by the analyst.

Second, sophisticated support for *remote* data/knowledge has been newly added. The infrastructure for this functionality allows to persistently specify

- Links to queryable *resources*
- Template *queries* for these resources (which can be parametrized by the end-user at runtime)
- *XSLT transformations* allowing to insert the results of queries as HTML fragments, either static or *dynamically updated* from the resources.

Currently we experiment with queryable resources in the form of *native XML database* (Berkeley, queried via XQuery), which stores PMML data, and semantic knowledge bases both in the form of *SPARQL endpoint* and *Ontopia Knowledge Suite* (a Topic Maps tool, queried via a Prolog-like language called tolog). Inclusion of further types of resources such as Lucene indices is in progress.

This work has been partially supported by the CSF project no.201/08/0802, and by Grant F4/15/2010 of the University of Economics, Prague.

References

1. Kliegr M., Ralbovský M., Svátek, V, Šimůnek M., Jirkovský V., Nemrava J., Zemánek J.: Semantic Analytical Reports: A Framework for Post-Processing Data Mining Results. In: Proc. ISMIS'09, Springer Verlag, LNCS, 2009, 8898.

¹ SEWEBAR stands for SEmantic WEB and Analytical Reports. More details in <http://sewebar.vse.cz>.