Semantic Analytical Reports: A Framework for Post-processing Data Mining Results

Tomáš Kliegr[1], Martin Ralbovský[1], Vojtěch Svátek[1], Milan Šimůnek[1], Vojtěch Jirkovský[2], Jan Nemrava[1], and Jan Zemánek[1]

[1] University of Economics, Prague, Faculty of Informatics and Statistics, Nám. Winstona Churchilla 4, 130 67 Praha 3, Czech Republic
{tomas.kliegr,ralbovsm,svatek,simunek,nemrava,xzemj22}@vse.cz
[2] Czech Technical University, Dept. of Computer Science and Engineering, Karlovo nám. 13, 121 35 Praha 2, Czech Republic
jirkov1@fel.cvut.cz

Abstract. Intelligent post-processing of data mining results can provide valuable knowledge. In this paper we present the first systematic solution to post-processing that is based on semantic web technologies. The framework input is constituted by PMML and description of background knowledge. Using the Topic Maps formalism, a generic Data Mining ontology and Association Rule Mining ontology were designed. Through combination of a content management system and a semantic knowledge base, the analyst can enter new pieces of information or interlink existing ones. The information is accessible either via semi-automatically authored textual analytical reports or via semantic querying. A prototype implementation of the framework for generalized association rules is demonstrated on the PKDD'99 Financial Data Set.

1 Introduction

Analytical report is a free-text document describing various elements of the data mining task: particularly the data, preprocessing steps, task setting and results. The analyst can also include additional information such as background knowledge, explanation of preprocessing steps and interpretation of the results. Creating analytical reports manually is time-consuming and the output document is not machine-readable, which hinders the possibilities for post-processing – e.g. querying, merging or filtering.

We present a novel framework for semi-automatic generation and processing of analytical reports that addresses these issues through the utilization of semantic web technologies. The framework is developed as part of the SEWEBAR (Semantic Web and Analytical Reports) initiative [11]. The framework is generic and should be suitable for most Data Mining (DM) algorithms. However, a specific implementation of the framework needs to take into account the knowledge representation used by the selected algorithm.

J. Rauch et al. (Eds.): ISMIS 2009, LNAI 5722, pp. 88–98, 2009.

We also present a prototype implementation of the framework for association rules (ARs). As part of the prototype, a data mining ontology for generalized association rules is introduced. To demonstrate the feasibility of the approach, the prototype implementation is used to post-process the output of Ferda [6] association rule mining system on the PKDD'99 Financial Dataset [14].

The rest of the paper is organized as follows. In Section 2 we give an overview of the architecture of the framework and in Section 3 an overview of our prototype implementation. A case study introduced in Section 4 is used to demonstrate the benefits of the framework. The related work is placed towards the end of the paper into Section 5. Section 6 contains conclusions and a plan for future work.

2 Framework Outline

In this section we present an outline of a new framework for post-processing the results of data mining tasks. The framework is based on established standards and seamlessly integrates with existing data mining software as its input is constituted by PMML[1] (Predictive Model Markup Language), which is a widely adopted XML-based standard for definition and sharing of data mining and statistical models. The second part of the framework's input is optional and is constituted by the emerging Background Knowledge Exchange Format (BKEF) specification. While PMML is produced by the DM software, BKEF is created directly based on human input.

PMML and BKEF specifications are stored in a Content Management System (CMS), which allows to merge information contained in one or more specifications with human input by enabling the analyst to include visualizations of BKEF and PMML fragments into the analytical report.

Further in the work flow, PMML and BKEF specifications are transformed into a semantic representation conforming to the Data Mining Ontology and stored in a Knowledge Base (KB), which allows the analysts to append and interlink information in a structured way. KB can be searched using a semantic query language.

An overview of the framework is depicted in Figure 1.

2.1 Input Data Formats

The framework's input is constituted by the description of the data mining models and the description of background knowledge.

For model description the framework uses PMML 3.2, the latest version of the standard at the time of writing. PMML 3.2 has the following components:

- Data Dictionary: database fields that are input for the model
- Mining Schema: database fields used in the model

[1] `http://www.dmg.org/pmml-v3-2.html`

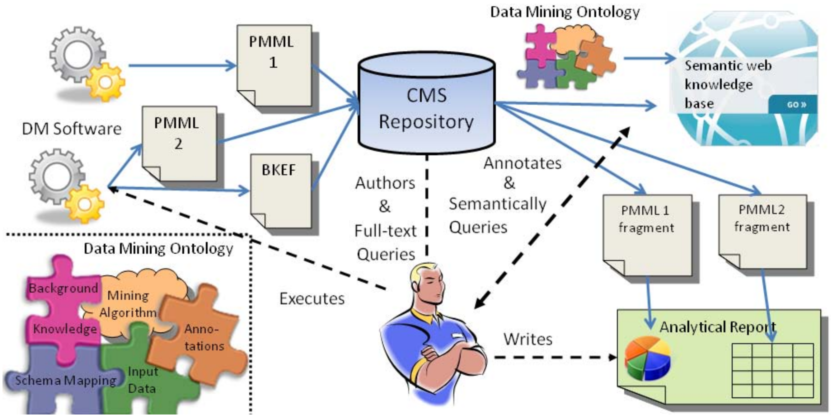

Fig. 1. Framework outline

- Transformation Dictionary: derived fields created through discretization or value mapping of database fields
- Model Description: settings and output specific to the mining algorithm

PMML 3.2 classifies DM algorithms according to knowledge representation into eleven types of Model Descriptions such as Association Rules, Cluster Models and Neural Networks. A framework implementation should support at least one Mining Description.

PMML allows two classes of applications – PMML producers and PMML consumers – to share a model. Export to PMML is supported by all major data mining software implementations acting as PMML producers. There are two major types of PMML consumers – scoring applications and archival (visualization, reference) applications [2]. The framework acts as a PMML consumer that performs the archival function but also reasons over the PMML, effectively establishing a new PMML consumer type.

To the best of our knowledge, there is no established standard analogical to PMML for background knowledge in data mining. Since the availability of such specification is vital, we are working on the BKEF specification. Its purpose is to foster exchange of information between background knowledge producers, such as dedicated user interfaces for acquisition of background knowledge e.g. [10,15], and background knowledge consumers like a CMS or a KB.

2.2 Analytical Report Authoring Support

The next element of the framework is a Content Management System (CMS), an application supporting storage and retrieval of electronic documents. The CMS is intended for storage and authoring of data mining analytical reports and related information. Analytical reports stored in a CMS are free-text documents whose purpose is to interpret the results of data mining tasks in the light of background information, other analytical reports and the expert's knowledge.

From a large part analytical reports hence consist of visualizations of information contained in the PMML (such as data histograms, description of data preprocessing, or model parameters). This allows the analyst to reuse the visualization of the needed bits of PMML in the free-text report so that existing knowledge does not have to be reentered. Additionally, since this generated content preserves the link to its source fragment, the corresponding part of the free-text report can still be resolved to a machine-readable XML representation.

The CMS also handles background knowledge, an example of which are value ranges. Background knowledge can either be created by a dedicated software and exported to the CMS or preferably authored directly in the CMS. In either case, the background knowledge is stored in the BKEF machine-readable format and included into the free-text report in the same way as PMML.

The fact that statements in analytical reports are directly backed by the source data (PMML or BKEF fragments) not only allows to search the reports as structured data but also fosters the credibility of the reports.

2.3 'Semantization' of Analytical Reports in Topic Maps

The CMS described in the previous subsection allows to query the content with full-text search (free-text reports) or XML query languages (PMML, BKEF). However, the structured content can be further 'semantized' by being stored into a KB according to some ontology. This is beneficial when merging heterogeneous data, such as PMML specifications of tasks executed on similar but not same datasets, or when working with background knowledge.

For the interchange and storage of semantic information, the framework relies on semantic web technologies in a broader sense. The resources for knowledge representation designed within the prototype framework implementation follow an ISO/IEC 13250 standard Topic Maps. Topic Maps are in principle interoperable with the semantic web formats RDF/OWL standardized by W3C. We have opted for Topic Maps, since they are simple and document orientated.

Topic Maps represent information through *topics*, *associations* and *occurrences*. A topic is any entity about which a human being can lead a discourse, an association represents a relationship between topics, and an occurrence represents a piece of information relevant to a topic. Types of topics, associations and occurrences used in a topic map constitute its ontology.

The framework defines: i) an ontology that allows to represent pieces of the structured XML content as instances of ontology types, ii) a transformation from structured content to instances.

After the analytical report has been semantized, it can be annotated and interlinked with reports already present in the repository. For example, in an association rule mining task, the analyst can annotate a discovered rule with the degree of its novelty, link it with an already existing rule coming either from a different report or from background knowledge. The prominent desired functionality is the search and reasoning over the resulting KB.

2.4 Data Mining Ontology

The Data Mining Ontology was derived from PMML 3.2 so that all PMML core features[2] can be automatically mapped into the ontology. The ontology consists of the following components: Input Data (including data transformations), Background Knowledge, Schema Mapping, Mining Algorithm and Annotations.

The *Input Data* component comes out of the corresponding components of the PMML standard: Data Dictionary, Mining Schema and Transformation Dictionary. Elements prescribed by the PMML Schema such as DataField were mapped to topic types; enumerations were represented as topic types with their instances representing the enumeration members. Knowledge represented as XML attributes in the PMML Schema, such as interval margins in a discretization, were represented as occurrence types.

The *Background Knowledge* component allows to relate various pieces of background knowledge to a specific data matrix through meta-attributes [9]. Meta-attribute is a generalization of an attribute[3]. The existence of meta-attributes stems from the fact that the same property can be coded in two datasets differently. For example, there can be an attribute `loan` with possible values `A - F` in one dataset, and an attribute `status` with possible values `bad`, `medium`, `good` in another; both attributes referring to loan quality. In the ontology, the meta-attribute provides a common name for the same property (here `loan quality`). A meta-attribute can have several *formats*. In our example, the formats can be named e.g. *Loan Quality [AF-Scale]*, *Loan Quality [Word Scale]*. An attribute in a dataset is a *realization* of a specific format of a meta-attribute.

The ontology supports the pieces of background knowledge introduced in [10]: i) basic value limits, ii) typical interval lengths for discretization, iii) groups of meta-attributes and iv) mutual influence among meta-attributes. These pieces of background knowledge are tied to meta-attributes through formats.

The *Schema Mapping* component allows to align an attribute or derived attribute used in one data mining task with its counterparts in other tasks. This is done through mapping the corresponding meta-attribute to its realizations. The current ontology version allows to express the mapping only in terms of *equivalence* of data values or categories.

The *Mining Algorithm* component is left for further work as it is out of the scope of this work to semantize the eleven Mining Descriptions defined in PMML 3.2. A reference Mining Algorithm component for association rules is, however, introduced as part of the framework's prototype implementation in Section 3.1.

The *Annotations* are the last component of the Data Mining Ontology. Annotations are pieces of knowledge that can be assigned by the analyst to some important concepts in the ontology. For example, an annotation can be assigned to an instance of DiscretizeBin to explain the reason behind the discretization.

Not considering the undefined Mining Algorithm component, basically all core PMML features are incorporated into the ontology. The current version of the ontology does not, however, support some PMML 3.2 features such as

[2] As core we consider features required by the PMML 3.2 XML Schema.

[3] Also referred to as *data field* or *database column*.

model composition. We are not, however, aware of any issue that would prevent extending the ontology so that it encompasses the remaining features.

3 Framework Prototype

This section describes a reference implementation of the framework for asssociation rules, which acts as a PMML consumer for two academic data mining programs Ferda and LISp-Miner[4]. Both use GUHA method to generate rules.

3.1 GUHA Method and GUHA-Based AR Ontology

GUHA method is one of the first methods of exploratory data analysis, developed in the mid-sixties in Prague. It is a general framework for retrieving interesting knowledge from data. The method has firm theoretical foundations based on observational logical calculi and statistics [5]. GUHA is realized by GUHA procedures, such as 4FT procedure for mining association rules. GUHA association rules extend mainstream association rules (as defined in [8]) in two ways:

– Boolean attributes are allowed for antecedent and consequent. *Boolean attributes* are recursive structures that enable conjunctions, disjunctions and negations of combinations of individual items. Details can be found in [7].
– A more general kind of dependency between antecedent and consequent than confidence and support or a specific interest measure is allowed. We call these dependencies *4ft-quantifiers*. The generalized association rule can be written in form $\varphi \approx \psi$, where φ and ψ are *Boolean attributes* and $\approx$ is a *4ft-quantifier*.

It has been shown in [8] that GUHA association rules are a generalization of mainstream association rules. Hence an association rule mining ontology based on GUHA allows to express the setting and discovered rules not only for the GUHA algorithm but also for other AR mining algorithms such as those generated by the popular *apriori* algorithm.

Utilizing these facts we have proposed the Association Rule Mining Ontology based on GUHA (GUHA AR Ontology). GUHA AR Ontology is interoperable with the core features of the Model Description specification of ARs in PMML in addition to supporting the GUHA-specific extensions. This ontology is designed so that it can be used in place of the Mining Algorithm component of the Data Mining Ontology.

3.2 Framework Implementation

This section gives an overview of the steps necessary to implement the framework on the example of our GUHA-based prototype.

In order to meet the framework's input data requirement, we used the PMML Extension mechanism [2] to incorporate GUHA features into the PMML Association Model thus creating a GUHA AR PMML model. Both our DM tools had

[4] `http://lispminer.vse.cz`, `http://ferda.sourceforge.net`

to be made conformable with the adapted model. Based on our prior experience with background knowledge [9,10,15], the first version of BKEF was drafted.

For the analytical report authoring support we used the PHP-based open source Joomla! (`http://www.joomla.org`) CMS, one of the most popular open source CMS systems as of time of writing. Joomla's advantages include object-oriented architecture, thousands of available extensions and active developer community. We extended Joomla! with an XSLT transformation[5] plug-in and defined transformations from PMML to HTML and from BKEF to HTML. This visualization is used by another new Joomla! extension, which allows the analyst to include visualized PMML fragments into the analytical report.

The semantization of analytical reports is based on the GUHA AR Ontology introduced in Subsection 3.1. Ontopia Knowledge Suite[6] (OKS) was used as the topic map repository and knowledge base. OKS can be either interactively browsed through using the Omnigator application or queried using *tolog*, a query language based on Prolog and SQL.

4 Case Study: Financial Dataset

The goal of this case study is to evaluate the usability of the prototype implementation in operation and to demonstrate the potential of our approach. In the experiment we went through the process of designing a data mining task, executing it, generating a PMML model, uploading it to the Joomla! CMS, inputing background knowledge, authoring an analytical report, semantizing the information via the OKS knowledge base and, finally, executing sample tolog queries. We used the Financial Data Set introduced in PKDD'99 [14].

4.1 Data Mining Task

The Financial Dataset consists of 8 tables describing the operations of bank customers. Among the 6181 customers we aim to identify subgroups with high occurrence of bad loans. For the mining schema we used the columns `duration`, `status` and `district`; all columns come from the `Loans` table.

Since there is no background knowledge specified in the PKDD'09 task setting, the following was introduced for case study purposes:

– *Background Knowledge 1 (BK1)* The quality of loans in Central Bohemia, the richest region of the country, is generally good. In other regions, it is in average lower. If area is expressed in terms of cities rather than regions then the value `Central Bohemia` maps to the value `Prague`.
– *Background Knowledge 2 (BK2)* If loan quality is expressed in terms of values `A - D` then the value `A` maps to `good`, `B` to `medium` and `C`,`D` to `bad`.
– *Background Knowledge 3 (BK3)* If loan duration is expressed in months then three bins should be created $< 0; 12 >$, $< 13; 23 >$ and $< 24; inf >$.

[5] An XML technology for translating between different knowledge representations [1].
[6] `http://www.ontopia.net`

All these pieces of background knowledge can be input in a structured way directly into the data mining systems: [10] introduces the mechanisms needed for BK1 and [15] for BK2 and BK3. For example, using the convention introduced in [9], BK1 can be input into LISp-Miner's KB module:

$$Region(CentralBohemia) \rightarrow^{+} LoanQuality(good). \quad (1)$$

Using BK2 and BK3, the data were preprocessed in the following way: the `duration` column was discretized into `1 year`, `13-23 months` and `two years+` categories. A `statusAggregated` derived field was created from the `status` column by mapping the status values `A` to `Good`, `B` to `Medium` and `C`, `D` to the category `Bad`. The derived field `district` was created by $1:1$ mapping from the `district` column, which has a granularity of municipality.

In Ferda, we formulated the following task:

$$duration(SS[1-1])\&district(Praha) \Rightarrow statusAggregated(I[2-2])$$

Here, $duration(SS[1-1])$ means that all subsets of the attribute of minimal and maximal length equal to 1 are created. Derived boolean attribute setting $statusAggregated(I[2-2])$ means that intervals of `status` (viewed as cardinal domain) of maximal and minimal length 2 are created. The *4ft-quantifier* used was *above average dependence* with parameters $p = 0.1$ and $minSup = 0.1$. This quantifier can be verbally interpreted as *Among object satisfying antecedent, there are at least 100p per cent more objects satisfying consequent than among all observed objects and there are at least minSup percent observed objects satisfying both antecedent and consequent.*

AR 1: $duration(13-23months) \Rightarrow statusAggregated(medium, bad)$
AR 2: $duration(1year) \Rightarrow statusAggregated(good, medium)$
AR 3: $duration(2y+)\&district(Prague) \Rightarrow statusAggregated(good, medium)$

The three rules listed above are the strongest among the 7 found.

4.2 Handling Data Mining Knowledge in the Prototype Framework

The knowledge relating to the DM task was then input to the CMS system. The mining model was input automatically via PMML while the background knowledge was input manually, since automated import for BKEF is not yet available. Using tools offered by the CMS, the analyst authored the report; the analyst's productivity increased through the possibility to reuse the HTML visualization of the structured content that was generated by the XSLT transformation.

In the report the analyst stated that AR 1 is not interesting – despite its strength concerning quantifier values. Rule AR 2 was, in turn, found surprising and useful. The analyst did not comment on Rule AR 3. The result of the analyst's work is a human-readable report, which is interlinked with its source data – primarily the PMML. However, since the report is written in free text, the information added by the analyst, such as information pertaining to the novelty of the association rules, is not machine-readable.

This is solved by the next step in the framework. The structured content is 'semantized' through the conversion to the Data Mining Ontology. The semantization can be done e.g. with an XSLT transformation, however, in our experiment this was done by manually reentering the data to OKS.

The GUHA AR Ontology introduced in Subsection 3.1 was used in place of the Mining Algorithm component of the ontology. The attributes referred to by background knowledge were expressed in terms of meta attributes *Area*, *Loan Quality* and *Loan Duration* and their formats *Area [Region]*, *Loan Quality [Alphabetical]* and *Loan Duration [Months]*. Formats were mapped to realizations using the schema mapping component of the ontology.

The resulting topic map was stored in a knowledge base created in OKS. The knowledge base allows the analyst to input new pieces of knowledge: machine-readable annotations. In this way, AR1 is annotated as 'Not Interesting' while AR2 is annotated as 'Surprising and Useful'.

4.3 The Added Value of Semantization

The knowledge base allows for sophisticated search with the tolog language. The analyst can choose which 'vocabulary' to use in the query. The analyst can e.g. decide between querying in terms of background knowledge and/or in terms of a specific dataset. Two example queries are listed below.

In the first query, the analyst wants to find all discovered rules that are annotated as surprising and have a good loan quality in the consequent. The primary ontology elements exploited by the corresponding tolog query are schema mapping and meta-attributes. The rule found is (as expected)

AR 2: $duration(1year) \Rightarrow statusAggregated(good, medium)$.

The analyst's second query uses the tolog's inference engine to find discovered rules whose antecedent subsumes the background knowledge rule BK 1: $Region(CentralBohemia) \rightarrow^{+} LoanQuality(good)$. The result of the query is AR3. Although the analyst had initially failed to notice that AR 3 corresponds to BK 1, the system was able to infer this automatically.

5 Related Work

The current work follows up on initial attempts for data mining analytic reporting published in [11]. However, this early work did not explicitly consider ontologies or even mark-up languages such as PMML. We are not aware of any other initiative for sharing data mining results, from any domain, over the semantic web. The impact of semantic web technology on data mining has typically been perceived as shift to 'knowledge-intensive' mining, in which ontologies serve as prior knowledge [3,4,16] or as a means to (locally) interpret the results [12]. We also cover this aspect to some degree, although we put more stress on the data integration and inferencing (querying) aspects.

Ontologies have already been suggested as formal models of the data mining field. The applications were in data mining work flows [3,16], grid-based data mining [4] or parameterizing a specific mining tool [13]. Most importantly, these DM ontologies were only applied in the phases of the DM process before or during the actual mining disregarding the problem of post-processing.

PMML has been used by various subjects in industry and research. However, search and aggregation applications over PMML documents have not been significantly reported. Use of this powerful mark-up language has typically been restricted to model exchange among mining tools. Combination of PMML with truly semantic resources has not been mentioned so far.

The setup of the prototype framework was complicated by the lack of open source semantic-aware CMS systems that would be in production-ready stage of development. This fact caused the separation of the framework into two independent systems – the Joomla! CMS and the OKS knowledge base.

To the best of our knowledge, there are no such systems based-upon the Topic Map standard existing or under active development. However, what concerns the W3C RDF/OWL standards, there is some support emerging. For example, the community around Drupal CMS is working on RDF and SPARQL support[7]. For this reason, we are considering a reimplementation of our framework to technologies conforming to the W3C standards.

6 Conclusions and Future Work

This paper introduces, to the best of our knowledge, the first systematic solution to post-processing data mining results that exploits semantic web technologies. The framework is built upon proven standards and technologies such as XML technologies and content management systems. The proposed data mining ontology is designed with respect to the industry standard PMML format, which should foster adoption of the framework among data mining practitioners.

Using the PKDD'99 Financial Dataset, we have shown how the framework can ease routine tasks such as authoring an analytical report. Its main strength lies, however, in the possibility to benefit from the querying and data integration capabilities given by the use of semantic web technologies. Using a topic-map-driven knowledge base, we have executed two example queries that used background knowledge, semantic annotation and schema mapping.

Since the input of the framework is constituted by PMML, the prototype implementation can be easily adapted to consume results from other DM tools such as Weka or SPSS[8]. The ontology, Joomla! extensions, BKEF specification, and other resources are available online[9].

The future work should focus on the completion of the Data Mining Ontology and of the BKEF specification. There is a work-in-progress on a new Joomla! extension for elicitation of background knowledge from the experts. We are also

[7] `http://drupal.org/project/sparql`

[8] XSLT transformations need to be customized to fit the required PMML Mining Model and the possible DM tool's extensions to PMML.

[9] `http://keg.vse.cz/sewebar/`

considering reimplementation of the knowledge base part of our framework to technologies conforming to the W3C RDF/OWL standards.

Acknowledgement. The work described here has been supported by Grant No. 201/08/0802 of the Czech Science Foundation, by Grant No. ME913 of Ministry of Education, Youth and Sports, of the Czech Republic, and by Grant No. IGA 21/08 of the University of Economics, Prague.

References

1. W3C: XSL Transformation (1999), `http://www.w3.org/TR/xslt`
2. DMG: PMML 3.2 Specification, `http://www.dmg.org/pmml-v3-2.html`
3. Bernstein, A., Provost, F., Hill, S.: Toward intelligent assistance for a data mining process: An ontology-based approach for cost-sensitive classification. IEEE Trans. on Knowl. and Data Eng. 17(4), 503–518 (2005)
4. Cannataro, M., Comito, C.: A data mining ontology for grid programming. In: Proceedings of the 1st Int. Workshop on Semantics in Peer-to-Peer and Grid Computing (SemPGrid 2003), pp. 113–134 (2003)
5. Hájek, P., Havránek, T.: Mechanizing Hypothesis Formation (Mathematical Foundations for a General Theory). Springer, Heidelberg (1978)
6. Kováč, M., Kuchař, T., Kuzmin, A., Ralbovský, M.: Ferda, New Visual Environment for Data Mining. In: Znalosti 2006, Czech Rep., pp. 118–129 (2006) (in Czech)
7. Ralbovský, M., Kuchař, T.: Using Disjunctions in Association Mining. In: Perner, P. (ed.) ICDM 2007. LNCS (LNAI), vol. 4597, pp. 339–351. Springer, Heidelberg (2007)
8. Rauch, J.: Logic of Association Rules. Applied Intelligence 22, 9–28 (2005)
9. Rauch, J., Šimůnek, M.: Dealing with Background Knowledge in the SEWEBAR Project. In: ECML/PKDD Workshop: Prior Conceptual Knowledge in Machine Learning and Data Mining, Warszaw 2007, pp. 97–108 (2007)
10. Rauch, J., Šimůnek, M.: LAREDAM Considerations on System of Local Analytical Reports from Data Mining. Toronto 20.05.2008 – 23.05.2008. In: Foundations of Intelligent Systems, pp. 143–149. Springer, Berlin (2008)
11. Rauch, J., Šimůnek, M.: Semantic Web Presentation of Analytical Reports from Data Mining – Preliminary Considerations. In: Web Intelligence 2007, pp. 3–7. IEEE Computer Society, Los Alamitos (2007)
12. Svátek, V., Rauch, J., Ralbovský, M.: Ontology-Enhanced Association Mining. In: Ackermann, M., Berendt, B., Grobelnik, M., Hotho, A., Mladenič, D., Semeraro, G., Spiliopoulou, M., Stumme, G., Svátek, V., van Someren, M., et al. (eds.) EWMF 2005 and KDO 2005. LNCS (LNAI), vol. 4289, pp. 163–179. Springer, Heidelberg (2006)
13. Suyama, A., Yamaguchi, T.: Specifying and Learning Inductive Learning Systems using Ontologies. In: AAAI 1998 Workshop on the Methodology of Applying Machine Learning, pp. 29–36 (1998)
14. Workshop Notes on Discovery Challenge (workshop at PKDD 1999), Prague, Czech Rep. (August 1999)
15. Zeman, M., Ralbovský, M., Svátek, V., Rauch, J.: Ontology-Driven Data Preparation for Association Mining. In: Znalosti 2009, Czech Republic, pp. 270–283 (2009)
16. Žáková, M., Křemen, P., Železný, F., Lavrač, N.: Planning to Learn with a Knowledge Discovery Ontology. In: Planning to Learn Workshop at ICML 2008 [CD-ROM], pp. 29–34. Omnipress, Madison (2008)