Challenges in Ontology-Based Association Rules Mining

Eya Ben Ahmed
ISG, Tunisia
eya.benahmed@gmail.com

ABSTRACT: *Data mining has emerged to address the problem of drawing interesting knowledge from data. Among the most used data mining techniques, we concentrate on association rules which lead to the derivation of useful associations and correlations within data.*

In parallel, the advance of the ontology which is one of the most important concepts in knowledge representation, has speedily altered the way of information structuring and sharing.

Recently, the area of coupling association rules and ontology has been a focus for several researchers, and hence, the introduced strategies are rising very rapidly. This survey aims to discuss the ontology-based association rules mining methods, and to provide a vision for future work in such a promising issue.

Keywords: Data Mining, Association Rules, Ontologies

Received: 11 November 2013, Revised 30 November 2013, Accepted 2 December 2013

1. Introduction

In recent years, the exceptional growth in the amount of data was a key factor towards a significant effort within the knowledge discovery in databases process. In this respect, to guarantee an efficient knowledge management, many researches advocate the extraction of frequent patterns.

One of the most important patterns in data mining is to discover association rules from a database. An association rule expresses a correlation between itemsets [1]. Such knowledge is very useful in making better decisions [3] and may be drawn from several data sources such as databases, data warehouses,…etc.

In parallel, ontologies are becoming more common and extensively used. In fact, they are applied to define the semantics of a field through depicting conceptual models [4]. They play a significant role in offering a generally agreed comprehension of a domain. They are intended to sum up the semantics of such a definite field. Because of that, ontologies have become crucial tools for information representation and processing at the semantic level.

Facing the semantic wealth of ontologies and the significance of data mining techniques, especially association rules, such coupling of them, allow, on the one hand, the semantic knowledge extraction, on the other hand, it may permit the domain knowledge evaluation through comparing such a derived knowledge and the expert prerequisites.

In this present study, we present an overview of related work on coupling ontology and association rules.

Several approaches dedicated to mining association rules from ontologies have been proposed in the literature. According to related sources, we can distinguish two main trends, namely: (i) *Ontology as a tool for association rules mining approaches* and (ii) *Ontology as a goal for association rules mining approaches*.

The main thrust of this paper is to scrutinize of different works related to ontology-based association rules mining and to introduce a comparative study between those works.

The paper is organized as follows: section 2 gives preliminary background and presents main notions. We discuss association rules mining using ontologies in Section 3.

We present some future challenges in Section 4. Finally¸ the conclusion and future work are finally presented in Section 5.

2. Background

In the sequel, we present our basic background. First, we sketch the ontology concept. Then, we scrutinize the association rules.

2.1 Ontology

According to Gruber [4], ontology is a conceptualization of a specification. Such a conceptualization consists in a corpse of formally represented knowledge, i.e., the objects, the concepts, and the relationships that hold among them in a domain of interest. A conceptualization is a simplified sight of the world that we want to describe to achieve a given goal.

The concept of ontology is increasingly employed to offer a shared understanding of an interest domain aiming to improve communication among humans and computers.

Indeed, the detection of the concepts and the relations between concepts is fundamental to build valuable domain ontology.

2.2 Association rules

In data mining, association rules are among the most widespread research issues. This technique expresses the correlation between sets of items in a series of transactions [1]. Indeed, such rules are derived from sets of itemsets that occur jointly at least according to given frequency i.e., support. In addition, the association rule is an "*IF antecedent THEN consequent*" rule that promises, with a definite probability i.e., confidence threshold which whenever the antecedent occurs, the consequent will happen. These pattern classes are generated using the Apriori algorithm [2]. Considering a fixed confidence value, the setting of the support threshold will determine the extracted patterns.

3. Association Rules Mining Using Ontologies

Several approaches dedicated to mining association rules from ontologies have been proposed in the literature. According to related sources, we can distinguish two main trends, as depicted in figure 1, namely: (i) *Ontology as a tool for association rules mining approaches* and (ii) *Ontology as a goal for association rules mining approaches*.

3.1 Ontology as a tool for association rules mining approaches

It is noteworthy that such trend of approaches considers the ontology as a tool to assist the mining process.

3.1.1 Tseng et al proposal

Tseng et al. address the association rules mining using ontological knowledge covering classification and composition relations [16]. For example, for the purchased item "*Sony VAIO*", "*PC*" and "*Desktop*" are its generalizations, while "60 *GB Segate*" and "512 *MB RAM*" are its components. Then, a redefinition of the support computation is presented and two algorithms AROC and AROS are introduced.

Thus, these methods derive associations that can boost distinct hierarchical levels between composition and classification relations in used ontology.

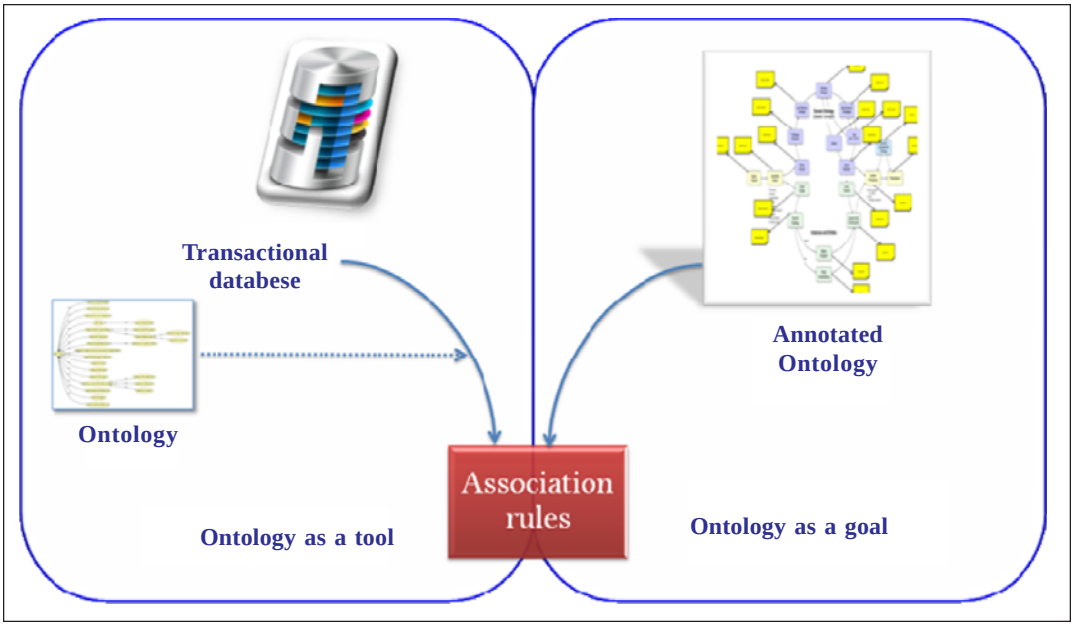

Figure 1. Ontology-based association rules principle

New transactions are regularly added to the database, and the ontology of items likewise evolves [17]. Thus, the discovery of interesting association rules is an iterative process which requires to repeatedly adjusting the support and the confidence thresholds. In this perspective, an algorithm called MIFO is introduced for incremental maintenance of frequent discovered patterns taking into account the evolution of several factors, namely the updated data, the evolved ontology, and quality metrics changing, i.e., support and confidence.

3.1.2 Wu et al. contribution

Wu et *al*. proposed an ontology-based framework for multidimensional association rules mining in order to assist decisionmakers in accurate queries formulation with reduced system resources consumption [18, 19].

Such use of ontology is motivated by the limits of hierarchical attributes representation in modeling star data warehouse schema without any exploitable semantics. Indeed, several ontologies are used such as queries history ontology, schema ontology, schema constraints ontology and domain ontology.

Its main idea is when any user launches a query, the system checks the used syntax using the history ontology. Then, it checks its semantics using schema constraints ontology. Finally, the search engine undertakes for the useful knowledge extraction.

Based on these different ontologies, the mining platform closely guide the extraction process of association rules for deriving effective and useful results consistent with the expectations of the user. The authors demonstrated an intelligent assistance system during their derivation targeted knowledge.

3.2 Ontology as a goal for association rules mining approaches

Under this category, we classify the approach of Manda et *al*. [11]. It is obvious that Gene Ontology (GO) has become an internationally recognized standard representing the function, process and neighborhoods genes. The richness of GO annotations is a valuable source of implicit knowledge between these data as illustrated in figure 2.

Manda et *al*. introduce a new method for association rules mining from sub-ontologies in several levels of abstraction.

Indeed, they propose a generalization procedure called bottom-up Cross-Ontology Data Mining-Level by Level taking into account the structure and semantics of GO. Such a strategy generates generalized transactions from the annotated data and derives association rules inter-ontologies at several abstraction levels.

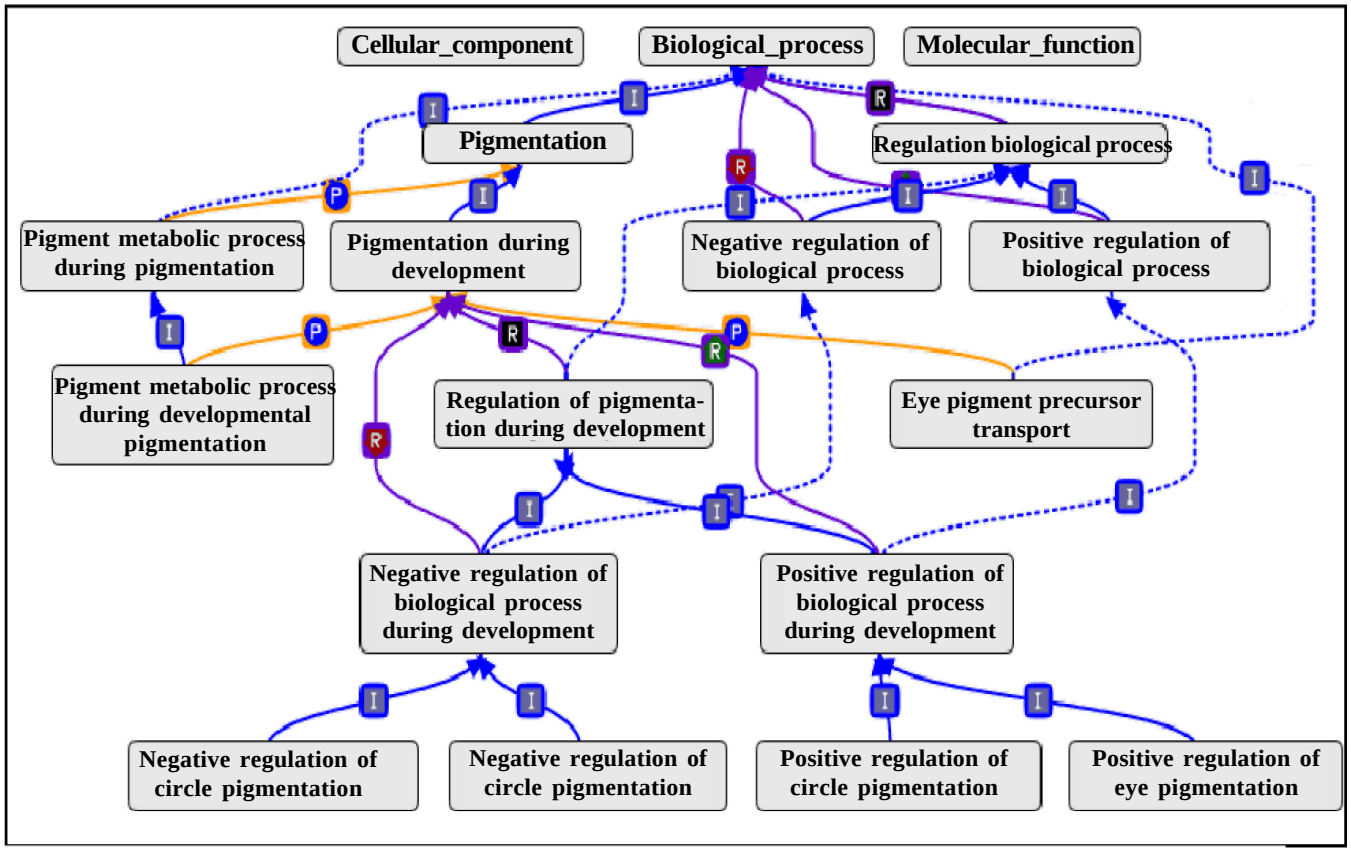

Figure 2. Gene Ontology

GO annotations in form of transactions are usually established at several levels in the GO hierarchy. GO is parsed and loaded into tables of relational databases.

Initially, the T Level is the set of transactions where the level is the depth of the deepest annotation throughout the operation. The Apriori algorithm is applied to the initial set of transactions to generate a set of rules. All the same rules under ontology will be pruned.

Then COLL replaces all present GO annotations at the current level by their immediate parent (s) already linked by "*is -a*" or " *part-of* " relationship to outline a new data set of *T* Level- 1transactions. The COLL applies the Apriori transactions for 1 level and adds new rules to all interesting rules. COLL outputs a non-redundant set of rules that satisfies the measurement thresholds specified interest, $R_{Interesting} = \{R_1, R_2, ... R_p\}$ where R_i contains a term of the antecedent and the term of a different GO sub-ontology as a consequence.

The method was applied to the data annotations chicken and mouse datasets. These experiments demonstrated that COLL generates a better quality of results compared to other methods. Rules biologically interesting discovered reveal unexpected knowledge about the co-occurrence of GO terms.

4. Discussion

The comparison between these approaches seems crucial. We rely on eight criteria classified into two categories such as: (*i*) class of ontology; (*ii*) ontological relations; *(iii*) constraint-based association rules; (*iv*) incremental aspect; (*v*) quality measures ; (*vi*) levels; (*vii*) application domain validation and (*viii*) conducting experimental evaluation.

Table 1 shows a comparison between the different strategies of ontology-based association rules based on these criteria. From the ontological perspective, Tseng et *al*., Wu et *al*., and Manda et *al*., used domain ontologies.

In addition, Tseng et *al*. and Wu et *al*., consider classification and composition ontological relations. However, Manda et *al*. integrate more complex relations.

From the mining side, the majority of these approaches use a constraint-based mining of association rules while Wu et *al*. neglect the constraint criteria on association rules derivation. Besides, except the work of Tseng et *al*. [17] all the rest of proposals ignore the incremental aspect on extracting knowledge.

Moreover, the quality metrics of generated patterns are exclusively redefined by Tseng et *al*. [17]. With respect to the hierarchies of classes, the hierarchical association rules [14] are the most common class of derived patterns.

In respect to the validation facet, only Manda et *al.*, consider a specific application domain for experiments. Moreover, only Wu et al., discard the experimental evaluation of their study.

At the sight of this comparative analysis, we can infer that both of used ontologies and association rules classes are generally limited. Such potential extensions are the main subject of the next section.

		Ontology							Association rules						Validation			
		Class of ontology		*Ontological relations*			*Constraints*		*Incremental aspect*		*Quality measures*		*Levels*		*Application domain*		*Experimental evaluation*	
		Domain	Application	Classification	Composition	other relations	Without constraints	With constraints	Static	Incremental	Classic support & confidence	Adapted support & confidence	Uni-level	Multi-levels	specific	general	Yes	No
Ontology as a tool	Tseng et al., [16]	X		X	X		X		X		X			X		X	X	
	Tseng et al., [17]	X		X	X		X			X		X		X		X	X	
	Wu et al., 2011 [18, 19]	X		X	X			X	X		X			X		X		X
Ontology as a goal	Manda et al, 2012 [11]	X		X	X	X	X		X		X			X	X		X	

Table 1. Comparative table of ontology-based association rules mining

5. Futures directions

This section identifies areas and approaches that require further research to produce improved ontology-based association rules extraction.

As shown in figure 2, the ontology-based pattern mining can be categorized with respect to the types of data and involved

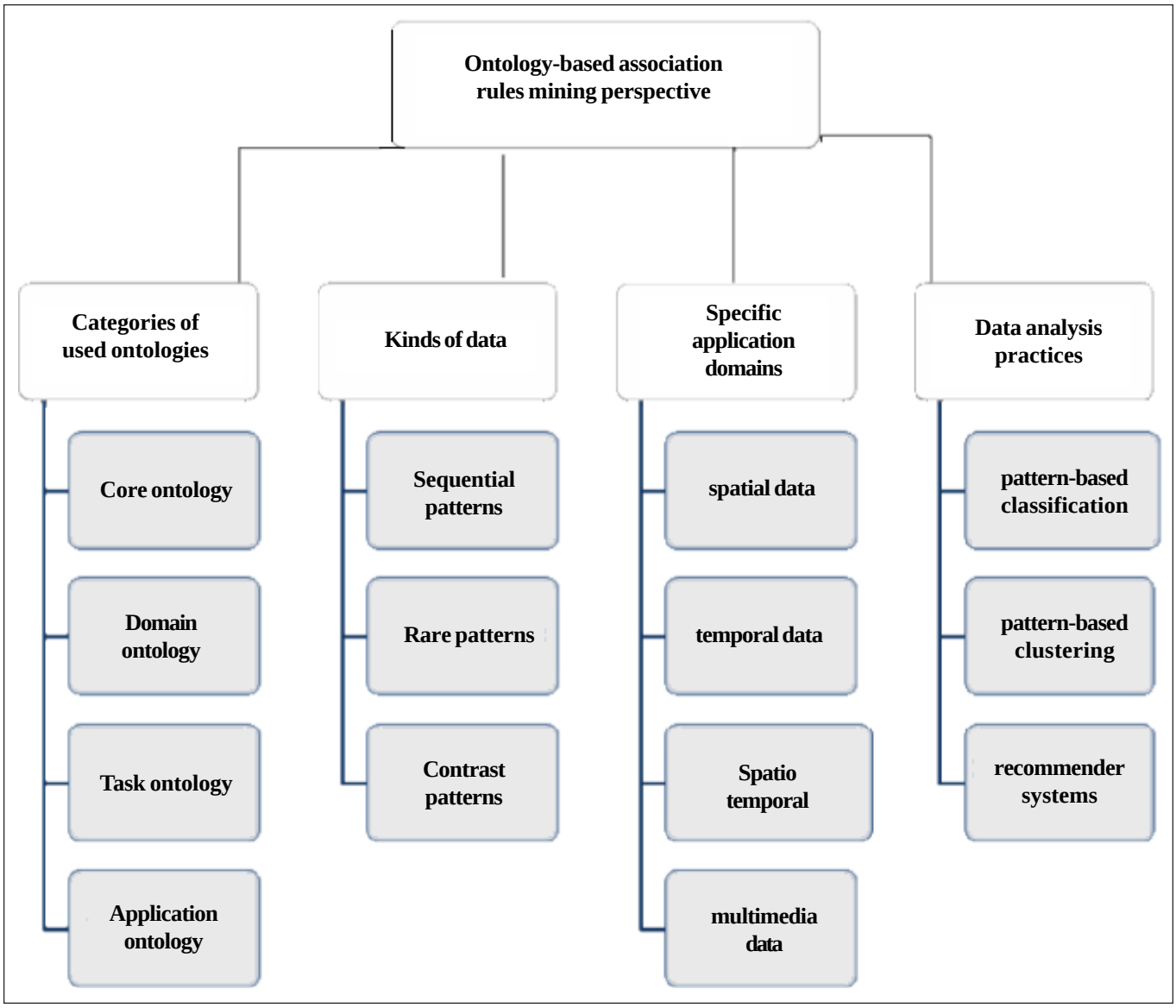

Figure 3. Ontology-based association rules perspectives

ontology, through the following criteria:

(i). **Kinds of data and features to be extracted**: Several types of knowledge may be derived such as *sequential patterns* [15], expressing a sequence of ordered actions or *rare patterns* [10] describing infrequent associations or *contrast association rules* [12] which may be used to check the inconsistencies that may exist between a corpus and reference ontology for example;

(ii). **Categories of used ontologies :** Different categories of ontologies can be employed:

a) **Core ontology:** It defines the generic concepts required to understand the other concepts [7], from which generic association rules may be drawn.

b) **Domain ontology:** Models a particular domain through defining a set of vocabularies and concepts that describe the target world [5].

Concrete association rules may be derived based on domain ontology.

c) **Task ontology [13]:** Tends to conceptualize specific tasks in the systems, such as planning, design or simulation tasks. Using such trend of ontology, active rules in form of Event Condition Action (ECA) can be extracted [8].

d) **Application ontology:** Offers a terminological structure to suit the requirements of an explicit activity.

(iii). **Specific application domains**: Various categories of application data incorporate spatial data, temporal data, spatiotemporal data and multimedia data. This range can lead to radically varied patterns extraction. For example, from spatial ontology, respectively temporal one, we can extract spatial association rules [9] respectively temporal [6].

(iv). **Data analysis practices**: Ontology-based frequent pattern mining commonly operates as an intermediate stage for enhanced data comprehension.

For instance, it can be employed as a feature extraction step for classification in pattern-based classification or pattern-based clustering. Likewise, pattern analysis can also be employed in recommender systems, which suggest information items.

6. Conclusion

In this paper, we first draw an overview of ontology-based association rules mining. Current approaches were confronted according to several comparative criteria. Such analytical study allows us to sketch the future work in such a context.

References

[1] Agrawal, R., Imielinski, T., Swami, A. (1993). Mining association rules between sets of items in large databases. *In*: Proc. of the ACM SIGMOD Conference on Management of Data, p. 207-216, Washington, D. C., May.

[2] Agrawal, R., Srikant, R. (1994). Fast algorithms for mining association rules. *In*: Proc. of the VLDB Conference, Santiago, Chile, September. Expanded version available as IBM Research Report, RJ9839, June.

[3] Fayyad, U., Piatesky-Shapiro, G., Smith, P. (1996). From data mining to knowledge discovery in databases. AI Mag, 17 (3) 37-54.

[4] Gruber, T. (1993). A Translation Approach to Portable Ontology Specifications, Knowledge Acquisition, 5 (2) 199-221. Elsevier, Amsterdam.

[5] Jean, S., Dehainsala, H., Nguyen Xuan, D., Pierra, G., Bellatreche, L., Aït-Ameur, Y. (2007). OntoDB: It is Time to Embed your Domain Ontology in your Database. *In*: Proceedings of the 12th International Conference on Database Systems for Advanced Applications (DASFAA'07) , v. 4443 of Lecture Notes in Computer Science, p. 1119- 1122.

[6] Juan, M., Ale, G., Rossi, H. (2000). An approach to discovering temporal association rules. *In*: Proc. of the 2000 ACM Symposium on Applied Computing, Villa Olmo, Via Cantoni 1, 22100 Como, Italy, p. 294–30.

[7] Kassel, G., Vers une, G. (2009). ontologie formelle des artefacts, IC 2009 : Actes des 20 Journées Francophones d'Ingénierie des Connaissances, Hammamet, Tunisia, May 25-29.

[8] Klaus, R., Dittrich, R., Gatziu, S., Geppert, A. (1995). The Active Database Management System Manifesto: A Rulebase of ADBMS Features. Rules in Database Systems, V. 985 of Lecture Notes in Computer Science, Springer, p. 3-20.

[9] Koperski, K., Han, J. (1995). Discovery of spatial association rules in geographic information databases, Advances in Spatial Databases Lecture Notes in Computer Science V. 951, p. 47-66.

[10] Liu, B., Hsu, W., Ma, Y. (1999). Mining Association Rules with Multiple Minimum Supports. SIGKDD Explorations ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD-99), San Diego, CA, USA.

[11] Manda, P., McCarthy F., Bridges, F. S. (2013). Interestingness measures and strategies for mining multi-ontology multi-level association rules from gene ontology annotations for the discovery of new GO relationships. J Biomed Inform.

[12] Minaei-Bidgoli, B., Tan, P., Punch, W. (2004). Mining interesting contrast rules for a web-based educational system, In International conference on machine learning applications, Los Angeles, California, p.1–8.

[13] Mizoguchi, R., Kozaki, K., Sano, T., Kitamura, Y. (2000). Construction and Deployment of a Plant Ontology, The 12th International Conference, EKAW, p. 113-128.

[14] Srikant, R., Agrawal, R. (1995). Mining generalized association rules. In: Proc. of the 21st International Conference on Very Large Databases. Zurich, Switzerland, p. 407- 419.

[15] Srikant, R., Agrawal, R. (1966). Mining sequential patterns: generalizations and performance improvements. *In*: Proceeding of the 5th international conference on extending database technology (EDBT'96), Avignon, France, p. 3–17.

[16] Tseng, M., Lin, W., Jeng, R. (2007). Mining Association Rules with Ontological Information, Second International Conference on Innovative Computing, *Information and Control*, (ICICI '07).

[17] Tseng, M., Lin, W., Jeng, R. (2007). Incrememtal Maintenance of Ontology-Exploiting Association Rules, International Conference on Machine Learning and Cybernetics, p. 2280 – 2285.

[18] Wu, C., Lin, W., Wu, C. (2007). Ontology-Assisted Query Formulation in Multidimensional Association Rule Mining, IEEE International Conference on Granular Computing. (GRC).

[19] Wu, C., Lin, W., Jiang, C., Wu, C. (2011). Toward intelligent data warehouse mining: An ontology-integrated approach for multidimensional association mining. Expert Syst. Appl. 38 (9) 11011-11023.