# Towards Ad-Hoc Rule Semantics for Gene Expression Data

Marie Agier[1,2], Jean-Marc Petit[2], and Einoshin Suzuki[3]

[1] DIAGNOGENE, 15000 Aurillac, France
[2] LIMOS, UMR 6158 CNRS, Univ. Clermont-Ferrand II, 63177 Aubière, France
[3] Electrical And Computer Engineering, Yokohama National University
79-5 Tokiwadai, Hodogaya, Yokohama 240-8501, Japan

**Abstract.** The notion of rules is very popular and appears in different flavors, for example as association rules in data mining or as functional (or multivalued) dependencies in databases. Their syntax is the same but their semantics widely differs. In this article, we focus on semantics for which Armstrong's axioms are sound and complete. In this setting, we propose a unifying framework in which any "well-formed" semantics for rules may be integrated. We do not focus on the underlying data mining problems posed by the discovery of rules, rather we prefer to emphasize the expressiveness of our contribution in a particular domain of application: the understanding of gene regulatory networks from gene expression data. The key idea is that biologists have the opportunity to choose - among some predefined semantics - or to define the meaning of their rules which best fits into their requirements. Our proposition has been implemented and integrated into an existing open-source system named MeV of the TIGR environment devoted to microarray data interpretation.

## 1 Introduction

Microarray technology provides biologists with the ability to measure the expression levels of thousands of genes in a single experience. It is believed that genes of similar function yield similar expression patterns in microarray experiences [1]. As data from such experiences accumulates, it is essential to have accurate means for assigning functions to genes. Also, the interpretation of large-scale gene expression data provides opportunities for developing novel mining methods for selecting for example good drug candidates (all genes are potentially drug targets) from among tens of thousands of expression patterns [2, 3].

However, one real challenge lies in inferring important functional relationships from these data. Beyond the cluster analysis [4], a more ambitious purpose of genetic inference is to find out the underlying regulatory interactions from the expression data, using efficient inference procedures.

Rules between genes are a promising knowledge to reveal regulatory interactions from gene expression data. The conjecture that association rules could be a model for the discovery of gene regulatory networks has been partially validated in [5, 6, 7, 8, 9, 10, 11]. Nevertheless, we believe that many different kinds of rules could be useful to cope with

M.-S. Hacid et al. (Eds.): ISMIS 2005, LNAI 3488, pp. 494–503, 2005.
© Springer-Verlag Berlin Heidelberg 2005

different biological objectives and the restricted setting of association rules could not be enough.

Clearly, the notion of *rules* is very popular and appears in different flavors, the two more famous examples being association rules in data mining and functional dependencies in databases. A simple remark can be done on these rules: their syntax is the same but their semantics i.e. their meaning widely differs.

In this paper, we propose a unifying framework in which any "well-formed" semantics for rules may be integrated. The key features of our approach are the following:

1. Given a dataset, defining a semantics in collaboration with domain experts (e.g. biologists and physicians).
2. Verifying if the semantics fits into our framework, i.e. if Armstrong's axiom system is sound and complete for this semantics [12].
3. Discovering the rules from the dataset, more precisely a cover for exact rules [13, 14] and a cover for approximate rules [15, 16].
4. Computing in a post-processing step, several quality measures for the obtained rules.

Note that we do not focus on the underlying data mining problems posed by the discovery of rules, rather we prefer to emphasize the expressiveness of our contribution in a particular application domain: the understanding of gene regulatory networks from *gene expression data*.

Due to space limitation, we introduce only one semantics based on a pairwise comparison of experiences to analyze *variations* of gene expression levels [5]. This semantics is close to the semantics of functional dependencies extended to deal with gene expression data. For others semantics, the reader is referred to [10, 17].

Our proposition has been implemented in a friendly graphical user interface to make it useful by biologists. We chose to integrate it as a module into a microarray data analysis open-source software: MeV. This tool is a part of an application suite, called TM4, developed by The Institute for Genomic Research (TIGR) [18].

*Paper Organization.* In Section 2, the framework of our approach is given. In Section 3, one semantics for rules is detailed and its compliance with the framework is shown in Section 4. Implementation details are given in Section 5 and we conclude in Section 6.

## 2 Framework of Our Approach

Our approach is based on the notion of *rule*, also called *implication*. A *rule* is an expression of the shape $X \rightarrow Y$ i.e. "X implies Y" and the *semantics* of the rule is the signification one wants to give to this implication. For example, association rules in data mining or functional dependencies in databases are two types of semantics.

In this paper, we focus on special kinds of rules, which exhibit nice properties, i.e. Armstrong's axiom system is sound and complete for the considered semantics. Such a semantics for rules is called "well-formed" in the sequel. We have chosen to focus on Armstrong's axioms since they apply obviously for functional dependencies but also for *implications* defined on a closure system [19] and thus turn out to have many practical applications (see examples given in [19]).

Practical interests of a well-formed semantics are twofold:

- Firstly, we can perform some kind of *reasoning* on rules from the Armstrong's axioms: From a set of rules $F$, it is possible to know if a rule is *implied* by this set of rules [20]. This problem is known as the implication problem and a linear time algorithm does exist for this problem. Thus, if there is a relation $r$ which satisfies $F$ then we know that all the rules that can be deduced from $F$ thanks to the Armstrong's axioms will be satisfied in this relation.
- We can also work on "small" *covers* of rules [21, 22] and propose a discovery process specific to the considered cover, but applicable to *all* well-formed semantics. It is also possible to propose covers for non-satisfied rules [23].

The theoretical framework that we propose to use for the generation of rules defined with well-formed semantics, comes from the inference of functional dependencies [24, 14]. Basically, since by definition the Armstrong's axioms apply for any well-formed semantics, the augmentation axiom implies a *monotone property*: given an attribute $A$, $X \rightarrow A \Rightarrow \forall\, Y \supset X, Y \rightarrow A$.
That is to say that the predicate "X implies A" is monotone with respect to set inclusion, thus the predicate "X does not imply A" is anti-monotone. So well known characterization may be used to produce the rules [25].
In other words, the largest left-hand sides not implying $A$ constitute the *positive border* of the predicate "X does not imply A" and the smallest left-hand sides implying $A$ constitute its *negative border*. Consequently, this negative border gives a subset of the canonical cover (i.e. rules with minimal left-hand sides and $A$ as right-hand side) while the positive border gives a subset of the Gottlob and Libkin cover [23] (i.e. rules with maximal left-hand sides and $A$ as right-hand side).

Details on the generation of rules are out of the scope of this paper, interested readers are referred to [13, 14, 25].

Moreover, an important key point of our approach is to take into account characteristics of gene expression data. Indeed, from the microarray analysis domain, two underlying "constraints" have to be understood: firstly, the number of experiences is small (a few hundreds at most) whereas the number of genes is large (several thousands). Such a constraint differs widely from those usually held in databases or data mining where the number of tuples can be huge whereas the number of attributes (i.e. genes in the context of this paper) remains rather small. That is why for example our approach does not take into account a minimum support threshold as usual for association rules. Statistical measures are computed *a posteriori* on the discovered rules.

Secondly, data pre-processing steps on gene expression data are not fully understood yet and therefore, we have to take into account noisy data. Thus microarray technology delivers numerical values with a relatively small confidence on these values, biologists have to interpret the data, for example as levels of expression, which implies a discretization step. In this setting, we propose to deal with noise in data not as an explicit pre-treatment step but implicitly within the semantics of the rules.

## 3 Example of Semantics for Rules

Our approach consists to interact with biologists in order to establish a semantics for the rules which fits into their objectives and requirements.

In the sequel, we restrict ourselves to one semantics studying similar variations of gene expression levels. In [17], we propose two others semantics for rules specially defined for gene expression data.

In many cases, it does make sense to compare experiences in a pairwise fashion to find some regularities between experiences. Such kind of reasoning is well known in the database community through the notion of *functional dependencies*. However, in our context, the FD satisfaction - its meaning - has to be relaxed to take into account noise in gene expression data. Since a crisp FD $X \rightarrow Y$ can be rephrased as "equal X-values correspond to equal Y-values", we would like to obtain something like "close X-values correspond to close Y-values". Thus, instead of requiring strong equality between attribute values, we admit an error less or equal to the absolute value of the difference (obviously, other norms should have been taken). This leads to the following definition.

**Definition 1.** *(pairwise comparison semantics) Let $X, Y \subseteq \mathbb{G}$, be two sets of genes and $r$ a relation over $\mathbb{G}$. A rule $X \rightarrow Y$ is satisfied in $r$ with the semantics pc defined with two thresholds $\epsilon_1$ and $\epsilon_2$, denoted by $r \models_{pc} X \rightarrow Y$, if and only if $\forall t_1, t_2 \in r$, if $\forall g \in X, \epsilon_1 \leq |t_1[g] - t_2[g]| \leq \epsilon_2$ then $\forall g \in Y, \epsilon_1 \leq |t_1[g] - t_2[g]| \leq \epsilon_2$.*

Classical satisfaction of functional dependencies is achieved when $\epsilon_1 = \epsilon_2 = 0$.

Thus, $X \rightarrow Y$ can be interpreted in our context as follows: for each gene $g$ of $X$, each time $g$ has a similar expression level in two experiences of $r$, then for each gene $g$ of $Y$, $g$ has also a similar expression level in those experiences.

*Example 1.* Let us consider a running example made of a set of 6 experiences ($t_1, t_2, t_3, t_4, t_5$ and $t_6$) over a set of 8 genes ($g_1, g_2, g_3, g_4, g_5, g_6, g_7$ and $g_8$) as depicted in Table 1.

**Table 1.** A running example

| r | $g_1$ | $g_2$ | $g_3$ | $g_4$ | $g_5$ | $g_6$ | $g_7$ | $g_8$ |
|---|---|---|---|---|---|---|---|---|
| $t_1$ | 1.9 | 0.4 | 1.4 | -1.5 | 0.3 | 1.8 | 0.8 | -1.4 |
| $t_2$ | 1.7 | 1.5 | 1.2 | -0.3 | 1.4 | 1.6 | 0.7 | 0.0 |
| $t_3$ | 1.8 | -0.7 | 1.3 | 0.8 | -0.1 | 1.7 | 0.9 | 0.6 |
| $t_4$ | -1.8 | 0.4 | 1.7 | 1.8 | 0.6 | -0.4 | 1.0 | 1.5 |
| $t_5$ | -1.7 | -1.4 | 0.9 | 0.5 | -1.8 | -0.2 | 1.2 | 0.2 |
| $t_6$ | 0.0 | 1.9 | -1.9 | 1.7 | 1.6 | -0.5 | 1.1 | 1.3 |

Let us suppose that the biologists are interested in low variations of expression levels between experiences. Thresholds should be defined as follows: $\epsilon_1$ = 0.0 and $\epsilon_2$ = 0.2. The hypothesis is that a gene does not vary for two experiences if the difference of the expression levels is between 0.0 and 0.2.

The expression levels of the genes $g_6$ and $g_7$ are plotted in Figure 1. In that case, the rule $g_6 \rightarrow g_7$ is satisfied in the relation $r$ (on the other hand the rule $g_7 \rightarrow g_6$ is not satisfied because of the variation between the experiences $e_3$ and $e_4$).

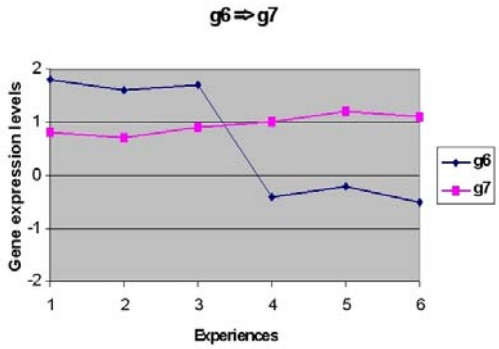


**Fig. 1.** Expression levels of the genes $g_6$ and $g_7$

The rule $g_6 \rightarrow g_7$ is interpreted in the following way: for some two experiences, if the expression level of the gene $g_6$ does not vary then the expression level of the gene $g_7$ does not vary neither.

## 4 Well-Formed Semantics

The second step of the process is to verify the well-formedness of the semantics defined by domain experts, i.e. verify that this semantics can be used within our framework.

This step is very important since many semantics could be defined, some of them verifying these requirements, others not (cf Theorem 2).

**Definition 2.** *A semantics $s$ is* well-formed *if Armstrong's axiom system is sound and complete for $s$.*

Let us recall the Armstrong's axiom system for a set of rules $F$ defined over a set of attributes (i.e. genes in our context) $\mathbb{G}$:

1. (reflexivity) if $X \subseteq Y \subseteq \mathbb{G}$ then $F \vdash Y \rightarrow X$
2. (augmentation) if $F \vdash X \rightarrow Y$ and $W \subseteq \mathbb{G}$, then $F \vdash XW \rightarrow YW$
3. (transitivity) if $F \vdash X \rightarrow Y$ and $F \vdash Y \rightarrow Z$ then $F \vdash X \rightarrow Z$

The notation $F \vdash X \rightarrow Y$ means that a proof of $X \rightarrow Y$ can be obtained using Armstrong's axiom system from $F$. Moreover, given a semantics $s$, the notation $F \models_s X \rightarrow Y$ means that for all relations $r$ over $\mathbb{G}$, if $r \models_s F$ then $r \models_s X \rightarrow Y$.

As expected, the semantics previously introduced verify these requirements.

**Theorem 1.** *The semantics $pc$ is* well-formed*.*

We need to show that Armstrong's axiom system is sound and complete for $pc$.

**Lemma 1.** *Armstrong's axiom system is sound for $pc$.*

*Proof.* Let $F$ be a set of rules. We need to show that if $F \vdash X \rightarrow Y$ then $F \models_{pc} X \rightarrow Y$.

Let $r$ be a relation over a set of genes $\mathbb{G}$.

1. (reflexivity) evident.
2. (augmentation) Let $t_1, t_2 \in r$ such that $\forall g \in X \cup W, \epsilon_1 \leq |t_1[g] - t_2[g]| \leq \epsilon_2$. We need to show that $\forall g \in Y \cup W, \epsilon_1 \leq |t_1[g] - t_2[g]| \leq \epsilon_2$, which implies that $r \models_{pc} XW \rightarrow YW$. By assumption $F \vdash X \rightarrow Y$, then we have $\forall g \in Y, \epsilon_1 \leq |t_1[g] - t_2[g]| \leq \epsilon_2$. The result follows.
3. (transitivity) Let $t_1, t_2 \in r$ such that $\forall g \in X, \epsilon_1 \leq |t_1[g] - t_2[g]| \leq \epsilon_2$. We need to show that $\forall g \in Z, \epsilon_1 \leq |t_1[g] - t_2[g]| \leq \epsilon_2$, which implies that $r \models_{pc} X \rightarrow Z$. By assumption, $F \vdash X \rightarrow Y$ and $F \vdash Y \rightarrow Z$, then $\forall g \in Y, \epsilon_1 \leq |t_1[g] - t_2[g]| \leq \epsilon_2$ and $\forall g \in Z, \epsilon_1 \leq |t_1[g] - t_2[g]| \leq \epsilon_2$ respectively. The result follows.

**Lemma 2.** *Armstrong's axiom system is complete for pc.*

*Proof.* We need to show that if $F \models_{pc} X \rightarrow Y$ then $F \vdash X \rightarrow Y$ or equivalently, if $F \nvdash X \rightarrow Y$ then $F \not\models_{pc} X \rightarrow Y$. As a consequence, assuming that $F \nvdash X \rightarrow Y$, it is enough to give a counter-example relation $r$ such that $r \models_{pc} F$ but $r \not\models_{pc} X \rightarrow Y$.

Let $r$ over $\mathbb{G}$ be the relation shown in Table 2, with $\epsilon_1 = 0.0$ and $\epsilon_2 = 0.2$.

**Table 2.** Counter-example

| $X^+$ | | | $U - X^+$ | | |
|---|---|---|---|---|---|
| 0.1 | ... | 0.1 | 0.1 | ... | 0.1 |
| 0.2 | ... | 0.2 | 1.2 | ... | 1.2 |

Firstly, we have to show that $r \models_{pc} F$. We suppose the contrary that $r \not\models_{pc} F$ and thus, $\exists V \rightarrow W \in F$ such that $r \not\models_{pc} V \rightarrow W$. It follows by the construction of $r$ that $V \subseteq X^+$ and $\exists A \in W$ such that $A \in U - X^+$. Since $V \in X^+$, we have $F \vdash X \rightarrow V$ and since $F \vdash V \rightarrow W$, we have $F \vdash V \rightarrow A$. Thus, by the transitivity rule, $F \vdash X \rightarrow A$ and thus $A \in X^+$. This leads to a contradiction since $A \in W$, and thus $r \models_{pc} F$.

Secondly, we have to show that $r \not\models_{pc} X \rightarrow Y$. We suppose the contrary that $r \models_{pc} X \rightarrow Y$. It follows by the construction of $r$ that $Y \subseteq X^+$ and thus $F \vdash X \rightarrow Y$. It leads to a contradiction since $F \nvdash X \rightarrow Y$ was assumed, and thus $r \not\models_{pc} X \rightarrow Y$.

As an example of semantics which does not fit into our framework, let us consider the following semantics, noted $pc'$, which extends the semantics $pc$ with an additional constraint:

**Definition 3.** *(pc') Let $X, Y \subseteq \mathbb{G}$, be two sets of genes and $r$ a relation over $\mathbb{G}$. A rule $X \rightarrow Y$ is satisfied in $r$ with the semantics $pc'$ defined with two thresholds $\epsilon_1$ and $\epsilon_2$, denoted by $r \models_{pc'} X \rightarrow Y$, if and only if $\forall t_1, t_2 \in r$, if $\forall g \in X$, $\epsilon_1 \leq |t_1[g] - t_2[g]| \leq \epsilon_2$ then $\forall g \in Y, \epsilon_1 \leq |t_1[g] - t_2[g]| \leq \epsilon_2$ AND $\exists t_1, t_2 \in r$ such that $\forall g \in X, \epsilon_1 \leq |t_1[g] - t_2[g]| \leq \epsilon_2$.*

We have the following result:

**Theorem 2.** *The semantics $pc'$ is not* well-formed.

*Proof.* Let $F$ be a set of rules and $X \subseteq \mathbb{G}$, we have $F \vdash X \rightarrow X$ by the reflexivity axiom. Nevertheless, $F \not\models_{pc'} X \rightarrow X$. Let us consider the following counter-example made of 4 experiences ($t_1, t_2, t_3$ and $t_4$) over a set of 2 genes ($g_1$ and $g_2$) as depicted in Table 3.

**Table 3.** Counter-example for $pc'$

| r | $g_1$ | $g_2$ |
|---|---|---|
| $t_1$ | -1.8 | 1.8 |
| $t_2$ | -1.7 | 0.2 |
| $t_3$ | 0.2 | -1.4 |
| $t_4$ | 0.3 | -1.8 |

Let us consider the thresholds $\epsilon_1 = 0.0$ and $\epsilon_2 = 0.2$. We can see that $r \not\models_{pc'} g_2 \rightarrow g_2$ because $\nexists\, t_1, t_2 \in r$ such that $0.0 \leq |t_1[g_2] - t_2[g_2]| \leq 0.2$. By the way, the result is proved since the reflexivity axiom is not sound.

## 5 Implementation

We have implemented the generation of rules as a C++/STL modules integrated into an open-source freeware devoted to microarray data analysis: MeV (MultiExperimentViewer) [18]. This tool is a part of an application suite, called TM4, developed by The Institute for Genomic Research (TIGR). These tools devoted to microarray data propose various functions such as storing the data, image analysis, normalization, interpretation of the results.

MeV is the application devoted to the analysis of gene expression data. Furthermore, MeV takes in input several file formats resulting from various image analysis software, has an important number of functionalities already integrated and is based on a GUI easy to use for biologists.

For the interface, we chose to limit as much as possible the options proposed to the users to make it easier. An example of the graphical user interface developed on top of MeV is presented in Figure 2.

The software was tested on several datasets and we were naturally interested in the post-treatment of rules. Without being exhaustive, four quality measures (support, confidence, dependence and lift) are computed to be able to sort the rules following these criterion. We plan to integrate some other quality measures of rules [26]. The user looks at only the rules he considers interesting according to these various indications.

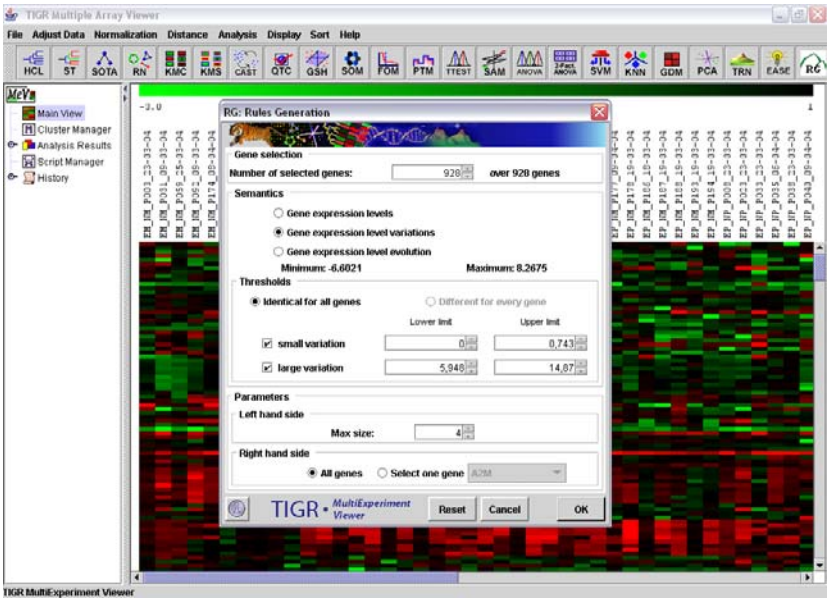


**Fig. 2.** Graphical user interface of the software

An application has been performed on expression profiles of a sub-sample of genes from breast cancer tumors. The results are presented in [17].

## 6 Conclusion

In order to attempt a reverse engineering of gene regulatory networks from gene expression data, we have proposed an on-going work aiming at defining different semantics of rules between genes, fitting in the same theoretical framework. Such rules form a complementary and hopefully new knowledge with respect to classical unsupervised techniques used so far [4].

The framework proposed in this paper, based on Armstrong's axiom system, is able to deal with different kinds of semantics in a unified manner. The semantics proposed for gene expression data were implemented as an extension of a free software dedicated to the analysis of microarray data (MeV of TIGR Institute).

For the time being, soundness and completeness of the Armstrong's axiom system have to be proved for every new semantics. We are currently working on a generic definition of a semantics which ensures that a semantics is well-formed if and only if it complies with this definition.

Moreover we are working on a more interactive process of discovery of rules, which would consist in requiring to the biologists some "templates" for the rules they are interested in, and then determining the semantics for which these rules are satisfied.

## References


1. Shena, M., al.: Quantitative monitoring of gene expression patterns with a cdna microarray. Science (1995) 467–470
2. Fuhrman, S., Cunningham, M., Wen, X., Zweiger, G., Seilhamer, J., Somogyi, R.: The application of shannon entropy in the prediction of putative drug targets. BioSystems **55** (2000) 5–14
3. Scherf, U., al.: A gene expression database for the molecular pharmacology of cancer. Nature Genetics **24** (2000) 236–244
4. Eisen, M., Spellman, P., Brown, P., Botstein, D.: Gene expression profiling predicts clinical outcome of breast cancer. Proc Natl Acad Sci **95** (1998) 14863–14868
5. Aussem, A., Petit, J.M.: $\epsilon$-functional dependency inference: application to dna microarray expression data. In Pucheral, P., ed.: Bases de données avancées (BDA'02), Evry, France (2002)
6. Berrar, D.P., Granzow, M., Dubitzky, W.: A Practical Approach to Microarray Data Analysis. Kluwer Academic Publishers (2002)
7. Becquet, C., Blachon, S., Jeudy, B., Boulicaut, J.F., Gandrillon, O.: Strong-association-rule mining for large-scale gene-expression data analysis: a case study on human sage data. Genome Biology **3** (2002)
8. Icev, A., Ruiz, C., Ryder, E.F.: Distance-enhanced association rules for gene expression. In: BIOKDD'03, in conjunction with ACM SIGKDD, Washington, DC, USA. (2003)
9. Creighton, C., Hanash, S.: Mining gene expression databases for association rules. Bioinformatics **19** (2003) 79–86
10. Agier, M., Chabaud, V., Petit, J.M., Sylvain, V., D'Incan, C., Vidal, V., Bignon, Y.J.: Towards meaningful rules between genes from gene expression data. In: poster, MGED'03, Aix en Provence. (2003)
11. Cong, G., Xu, X., Pan, F., A.K.H.Tung, Yang, J.: Farmer: Finding interesting rule groups in microarray datasets. In: SIGMOD. (2004)
12. Armstrong, W.W.: Dependency structures of data base relationships. In: Proc. of the IFIP Congress 1974. (1974) 580–583
13. Mannila, H., Räihä, K.J.: Algorithms for Inferring Functional Dependencies from Relations. DKE **12** (1994) 83–99
14. Demetrovics, J., Thi, V.: Some remarks on generating Armstrong and inferring functional dependencies relation. Acta Cybernetica **12** (1995) 167–180
15. Eiter, T., Gottlob, G.: Identifying the minimal transversals of a hypergraph and related problems. SIAM Journal on Computing **24** (1995) 1278–1304
16. Lopes, S., Petit, J.M., Lakhal, L.: Functional and approximate dependencies mining: Databases and FCA point of view. JETAI **14** (2002) 93–114
17. Agier, M., Petit, J.M., Chabaud, V., Pradeyrol, C., Bignon, Y.J., Vidal, V.: Vers différents types de règles pour les données d'expression de gènes-Application à des données de tumeurs mammaires. In: Actes du Congrès INFORSID'04, Biarritz. (2004)
18. Saeed, A., Sharov, V., White, J., Li, J., Liang, W., Bhagabati, N., Braisted, J., Klapa, M., Currier, T., Thiagarajan, M., Sturn, A., Snuffin, M., Rezantsev, A., Popov, D., Ryltsov, A., Kostukovich, E., Borisovsky, I., Liu, Z., Vinsavich, A., Trush, V., Quackenbush, J.: TM4: a free, open-source system for microarray data management and analysis. Biotechniques **34** (2003) 374–78
19. Ganter, B., Wille, R.: Formal Concept Analysis. Springer-Verlag (1999)
20. Beeri, C., Berstein, P.: Computational problems related to the design of normal form relation schemes. ACM TODS **4** (1979) 30–59
21. Maier, D.: Minimum covers in the relational database model. JACM **27** (1980) 664–674

22. Guigues, J.L., Duquenne, V.: Familles minimales d'implications informatives résultant d'un tableau de données binaires. Math. Sci. Humaines **24** (1986) 5–18
23. Gottlob, G., Libkin, L.: Investigations on Armstrong relations, dependency inference, and excluded functional dependencies. Acta Cybernetica **9** (1990) 385–402
24. Mannila, H., Räihä, K.J.: Algorithms for inferring functional dependencies from relations. DKE **12** (1994) 83–99
25. Mannila, H., Toivonen, H.: Levelwise search and borders of theories in knowledge discovery. DMKD **1** (1997) 241–258
26. Tan, P.N., Kumar, V., Srivastava, J.: Selecting the right objective measure for association analysis. Information Systems **29** (2004) 293–313