Moonis Ali   Jeng-Shyang Pan
Shyi-Ming Chen   Mong-Fong Horng (Eds.)

Modern Advances in Applied Intelligence

27th International Conference
on Industrial Engineering and Other Applications
of Applied Intelligent Systems, IEA/AIE 2014
Kaohsiung, Taiwan, June 3-6, 2014
Proceedings, Part II

Springer

Volume Editors

Moonis Ali
Texas State University, San Marcos, TX, USA
E-mail: ma04@txstate.edu

Jeng-Shyang Pan
National Kaohsiung University of Applied Sciences, Taiwan
E-mail: jspan@cc.kuas.edu.tw

Shyi-Ming Chen
National Taiwan University of Science and Technology, Taipei, Taiwan
E-mail: smchen@mail.ntust.edu.tw

Mong-Fong Horng
National Kaohsiung University of Applied Sciences, Taiwan
E-mail: mfhorng@ieee.org

ISSN 0302-9743 e-ISSN 1611-3349
ISBN 978-3-319-07466-5 e-ISBN 978-3-319-07467-2
DOI 10.1007/978-3-319-07467-2
Springer Cham Heidelberg New York Dordrecht London

Library of Congress Control Number: 2014939236

LNCS Sublibrary: SL 7 – Artificial Intelligence

Typesetting: Camera-ready by author, data conversion by Scientific Publishing Services, Chennai, India

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Deriving Processes of Information Mining Based on Semantic Nets and Frames

Sebastian Martins, Darío Rodríguez, and Ramón García-Martínez

Information Systems Research Group. National University of Lanus. Argentina
PhD Program on Computer Science. National University of La Plata. Argentina
rgarcia@unla.edu.ar

Abstract. There are information mining methodologies that emphasize the importance of planning for requirements elicitation along the entire project in an orderly, documented, consistent and traceable manner. However, given the characteristics of this type of project, the approach proposed by the classical requirements engineering is not applicable to the process of identifying the problem of information mining, nor allows to infer from the business domain modelling, the information mining process which solves it. This paper proposes an extension of semantic nets and frames to represent knowledge of the business domain, business problem and problem of information mining; and a methodology to derive the information mining process from the proposed knowledge representations is introduced.

Keywords: information mining, requirement engineering, deriving processes of information mining, semantic nets, frames.

1 Introduction

In [1] we have proposed five processes for information mining related to the following business intelligence problems: discovery of behavior rules, discovery of groups, discovery of significant attributes, discovery of group-membership rules and weighting of behavior or group-membership rules, and the identification of information-systems technologies that can be used for the characterized processes.

The proposed processes are based on intelligent systems-based methods [2-3] such as: TDIDT algorithms (Top Down Induction Decision Trees), self-organizing maps (SOM) and Bayesian networks. TDIDT algorithms allow the development of symbolic descriptions of the data to distinguish between different classes [4]. Self-organizing maps can be applied in the construction of information cluster [5]. Bayesian networks can be applied to identify discriminative attributes in large information bases and detect behavior patterns in the analysis of temporal series [6]. Our research work lays on the hypothesis that is emerging a new discipline [7] called Information Mining Engineering (IME).

One of the early stages of IME is to understand which are the requirements that business intelligence poses to an Information Mining Project [8]. The requirement elicitation process is addressed by most commonly used IME methodologies [9-11].

A. Moonis et al. (Eds.): IEA/AIE 2014, Part II, LNAI 8482, pp. 150–159, 2014.

IME methodologies mention the necessity of business understanding as starting point for any IME project development. In general, business understanding phase may be decomposed in the following sub phases: determine business objectives (in this paper the business problem), assess situation, determine IME project goals (in this paper the problem of information mining) and produce project plan.

In this paper we address two problems: [a] how to represent: the business domain, the business problem and the problem of information mining, and [b] how to derive the processes of information mining from the selected representation.

Knowledge Engineering provides formalisms that allow capturing the business domain, the business problem and the problem of information mining. We have explored the use of frames [12] and semantic networks [13] for the capturing process.

In this context, this paper introduces knowledge representation formalisms based on frames and semantic networks to represent knowledge in an IME Project (section 2), proposes a methodology to derive the processes of information mining (section 3), presents a case study exemplifying how to derive the process of information mining from the proposed knowledge representation formalisms (section 4), and some conclusions are drawn (section 5).

2 Knowledge Representation Formalisms for Business Domain, Business Problem and Problem of Information Mining

In this section we present: proposed notation for modelling domain problem (section 2.1) and problem of information mining (section 2.2) with semantic networks; and the frames used to capture concept definitions and relations among them (section 2.3).

2.1 Proposed Notation for Modelling Domain Problem with Semantic Networks

Representation rules that apply to the modelling of the business domain, the business problem and the problem of information mining are: [a] the concepts are represented by ovals with solid lines (Fig. 1.a), [b] the attributes and values are represented by ovals with dashed lines (Fig. 1.b), [c] the identifier of an instance of the concept is defined within the oval concatenating the type of concept with a dash "-" between them (Fig. 1.c), [d] the l value or reference of the label of the arcs depend on the relationship that this represents: [d.1] if it joins an attribute with its value, it is labelled with the word "value" (Fig. 1.d), [d.2] if it denotes a relationship instance of a concept it is labelled with the word "instance" (Fig. 1.e), [d.3] if it liaises between a concept and its attribute, it is labelled with one or more words that represent the connection between the two (Fig. 1.f), [d.4] if it specifies an inheritance relationship of a concept, it is labelled with the word "subclass" (Fig. 1.g). Figure 1.d indicates a "Brand equipment" attribute, whose value is “Nokia”. Figure 1.d shows that the concept of " client -1" (with 1 being the client identifier as explained in Figure 1.c) belongs to the general concept client therefore get all the features that this represents. Figure 1.f indicates a membership relation between the concept plan and the attribute length indicating that any plan has an estimated duration. Figure 1.g represents

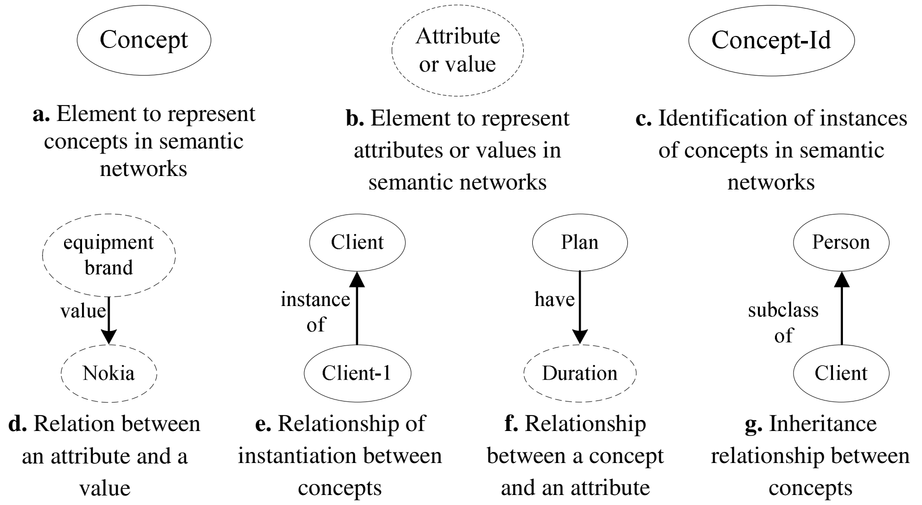

Fig. 1. Proposed Notation for Modelling Domain Problem with Semantic Networks

an inheritance relationship , in which the class inherits all the characteristics that define the parent concept, in this case the Customer concept inherits all the features of person, because a customer is a person, according to this relationship.

It is essential to emphasize that in cases presented by figures 1.d, 1.e and 1.g the possible value of the label is defined, no other value accepted for these relations, while in case presented by figure 1.f, an example is given value, because depending on the content of the concepts, attributes that relate the value of the label of the arc will vary.

2.2 Proposed Notation for Modelling Problem of Information Mining with Semantic Networks

In the Semantic Network of Problem of Information Mining, it must be identified the flow of the described problem, through the identification of elements of input and output. To provide this information additional elements are incorporated /figure 2).

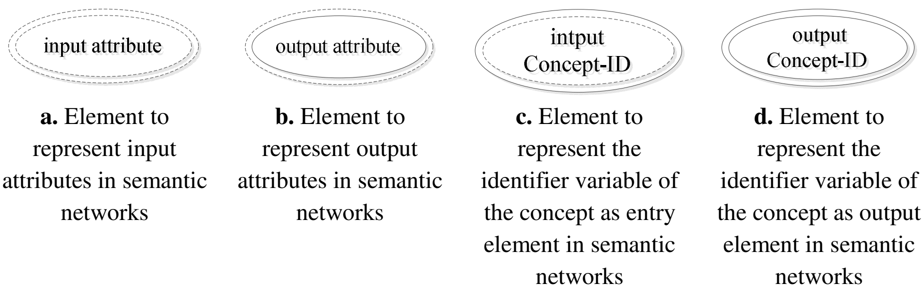

Fig. 2. Proposed Notation for Modelling Problem of Information Mining with Semantic Networks

These additional notations are: [a] the input attributes are represented with double oval with dashed lines (figure 2.a), [b] the output attributes are represented with double oval, taking the inner oval with solid line and the outer oval with intermittent line (figure 2.b), [c] the concepts instance identifiers, which are input element are represented by double oval, having its outer oval with solid line and its internal oval with dotted line (figure 2.c), and [d] the identifiers of instances of concepts, which are output element are represented by double oval whose lines are continuous (figure 2.d).

2.3 Frames Used to Capture Definitions of Concepts and Relations

The proposed frames are used to capture different aspects (concepts and relations) of business domain and information mining domain. They are:

Frame Term-Category-Definition of Business Domain: which aims to identify and classify all relevant elements within the business domain.

Frame Concept-Attribute-Relation-Domain Value of Business Domain: which aims is defining the structure of the elements of the business.

Frame Concept- Domain_Relation of Business Domain: which aims to identify the relationships between the concepts that define the business model.

Frame Term-Category-Definition of Problem of Information Mining: which aims to identify and classify the relevant elements of the problem of information mining

Extended Frame Concept-Attribute-Relation-Domain_Value of Problem of Information Mining: Which aims is to define the structures of the elements to the problem of information mining, identifying the input and output elements thereof.

Frame Concept-Problem_Relation of Problem of Information Mining: which aims to identify the relationships between the concepts that define the problem of information mining.

3 Proposed Methodology to Derive the Processes of Information Mining

We have developed a methodology (shown in figure 3) to derive the processes of information mining from frames and semantic nets. The methodology has three phases: “Analysis of Business Domain”, “Analysis of the Problem of Information Mining”, and “Analysis of the Process of Information Mining”.

The phase “Analysis of Business Domain” develops three tasks: “Identification of the Elements and Structure of the Business Domain”, “Identification of Relationships Between Concepts of Business Domain”, and “Conceptualization of the Business Domain”. The task “Identification of the Elements and Structure of the Business Domain” has as input the “Business Domain Description” and the “Business Domain Data”; and produces as output the “Frame Term-Category-Definition of Business Domain” and the “Frame Concept-Attribute-Relation-Domain_Value of Business Domain”. The task “Identification of Relationships Between Concepts of Business Domain” has as input the “Frame Term-Category-Definition of Business Domain” and the “Frame Concept-Attribute-Relation-Domain_Value of

Business Domain"; and produces as output the "Frame Concept-Domain_Relation of Business Domain". The task "Conceptualization of the Business Domain" has as input "Frame Concept-Domain_Relation of Business Domain" and the "Frame Concept-Attribute-Relation-Domain_Value of Business Domain"; and produces as output the "Semantic Net of the Business Domain Model".

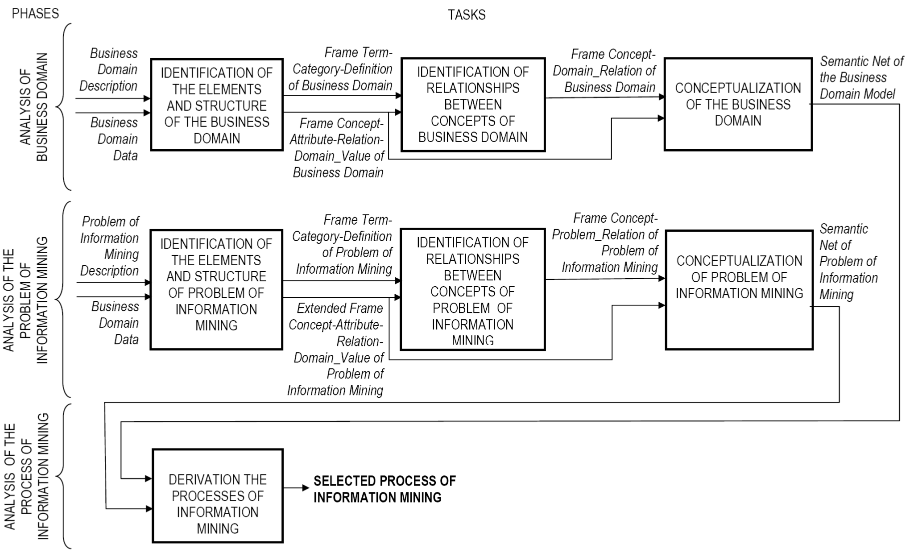

Fig. 3. Methodology to derive the processes of information mining

The phase "Analysis of the Problem of Information Mining" develops three tasks: "Identification of the Elements and Structure of Problem of Information Mining", "Identification of Relationships Between Concepts of Problem of Information Mining" and "Conceptualization of Problem of Information Mining". The task "Identification of the Elements and Structure of Problem of Information Mining" has as input the "Problem of Information Mining Description" and the "Business Domain Data"; and produces as output the "Frame Term-Category-Definition of Problem of Information Mining" and the "Extended Frame Concept-Attribute-Relation-Domain_Value of Problem of Information Mining". The task "Identification of Relationships Between Concepts of Problem of Information Mining" has as input the "Frame Term-Category-Definition of Problem of Information Mining" and the "Extended Frame Concept-Attribute-Relation-Domain_Value of Problem of Information Mining"; and produces as output the "Frame Concept- Problem_Relation of Problem of Information Mining". The task "Conceptualization of Problem of Information Mining" has as input the "Extended Frame Concept-Attribute-Relation-Domain_Value of Problem of Information Mining" and the "Frame Concept- Problem_Relation of Problem of Information Mining"; and produces as output the "Semantic Net of Problem of Information Mining".

The phase "Analysis of the Process of Information Mining" develops one task: "Derivation the Processes of Information Mining" which has as input the "Semantic

Net of the Business Domain Model" and the "Semantic Net of Problem of Information Mining"; and produces the "Selected Process of Information Mining".

4 Case Study

In this section a proof of concept on the derivation procedure of process of information mining from the business domain modelling is presented. It takes the following case [14]:

"...Description of the Business and Business Problem:
"Mobile Services Argentina SA" is a company belonging to the telecommunications sector that operates nationwide. It offers products and services in various trade and generic brands, offers a wide variety of rate plans (personal or corporate use) according to the needs of customers. It is currently promoting customized retention campaigns of clients throughout the country in order to offer new products and services according to the characteristics of each customer, to preempt competitors who are also growing nationwide. The business objective is to characterize customers in different geographical regions of coverage of the company, in order to facilitate the definition of marketing campaigns aimed at maintaining existing customers and obtaining new customers under the common preferences of each region, which will also help improve the profitability of the company. With this, it also seeks to determine the behavior of its customers and improve the knowledge we have of them. "Mobile Services Argentina SA" has the goal of customer satisfaction, which makes by offering products and quality services, dividing all mobile operations in the country in five regions: Coast, Cuyo, Pampas/Center, Patagonia and North.
Problem of Information Mining: Determine the rules, based on the variables status, type of plan and its duration, equipment brand and customer type, identifying the behavior of a set of customers according to their geographical location (region, state, city).
Data description:
The following are the relevant data that stores the company during the course of their activities:

- *Client Code: the company uses to identify him. Example: 99861598.*
- *Track Client: example: AAA (active client), BA (thin client).*
- *Date Added: belongs to the client. Example: 10/03/2003.*
- *Code of the city: to the customer's residence. Example: 379.*
- *Code of the province to which the city belongs. Example: 3.*
- *Contracted Plan: rate plan code that engages the customer. Example: U21.*
- *Duration: number of months of customer engagement. Example: 12.*
- *Customer type: example: PR / PO (prepaid customer), FF (client billing).*
- *Equipment Brand: belonging to the client. Example: NOKIA.*
- *Geographic Region: indicates the region to which each province belongs (Table 1)…"*

Table 1. Code of regions by provinces

REGION	DESCRIPTION	PROVINCES MEMBERS
1	Coast	Misiones, Corrientes, Formosa, Chaco, Santa Fe
2	Cuyo	Mendoza, San Luis, San Juan
3	Pampas Center	Buenos Aires, Federal Districl, Córdoba, La Pampa
4	Patagonia	Neuquén, Río Negro, Chubut, Santa Cruz, Tierra del Fuego
5	North	Jujuy, Salta, Tucumán, Catamarca, La Rioja, Santiago del Estero

When task "Identification of the Elements and Structure of the Business Domain" is applied, we obtain "Frame Term-Category-Definition of Business Domain" (Table 2) and the "Frame Concept-Attribute-Relation-Domain_Value of Business Domain" (Table 3). When task "Identification of Relationships Between Concepts of Business Domain" is applied, we obtain "Frame Concept-Domain_Relation of Business Domain" (Table 4).

Table 2. Frame Term-Category-Definition of Business Domain

Term	Category	Definition
belongs	Relationship	A client belongs to a region
belongs	Relationship	A client belongs to a province
belongs	Relationship	A client belongs to a city
city	Attribute	city where the client live
Client	Concept	A person who hire the service
client code	Attribute	unique client identification code
customer type	Attribute	Kind of customer by contract
date Added	Attribute	Date when the client belongs to the company
duration	Attribute	Number of months of hiring
equipment brand	Attribute	Client's equipment brand
has	Relationship	A client has a state with the company
has	Relationship	A client has a certain brand of telephone equipment
has	Relationship	A plan has a certain duration
hire	Relationship	A client hire a plan
identifies	Relationship	The client code identifies a specific client
identifies	Relationship	The plan code identifies a specific plan
is	Relationship	A customer is a certain type of customer
Plan	Concept	Kind of contract
plan code	Attribute	unique plan identification code
province	Attribute	Province where a client live
region	Attribute	Region where a client live
state	Attribute	Client's state
was registered	Relationship	A client was registered at a date

Table 3. Frame Concept-Attribute-Relation-Domain_Value of Business Domain

Concept	Attribute	Relationship	Value
Client	Code	identifies	Numeric
	state	has	alphabetical EG: AAA, BA.
	date added	was registered	dd/mm/AAAA
	province	belongs	Numeric EG: 379
	city	belongs	Numeric EG: 3
	region	belongs	1 to 5
	type	is	alphabetical EG: PR/PO, FF.
	equipment brand	has	alphabetical EG: Nokia
Plan	Code	identifies	Alphanumeric EG: U21.
	duration	has	Numeric

Table 4. Frame Term-Category-Definition of Business Domain

Concept	Concept associated	Relationship	Description
Client	Plan	hire	A client hire a plan

When task "Conceptualization of the Business Domain" is applied, we obtain the "Semantic Net of the Business Domain Model" (Figure 4). When task "Identification of the Elements and Structure of Problem of Information Mining" is applied, we obtain "Frame Term-Category-Definition of Problem of Information Mining" (Table 5) and "Extended Frame Concept-Attribute-Relation-Domain_Value of Problem of Information Mining" (Table 6). When task "Identification of Relationships Between Concepts of Problem of Information Mining" is applied, we obtain "Frame Concept- Problem_Relation of Problem of Information Mining" (Table 7). When task "Conceptualization of Problem of Information Mining" is applied, we obtain the "Semantic Net of Problem of Information Mining" (Figure 5).

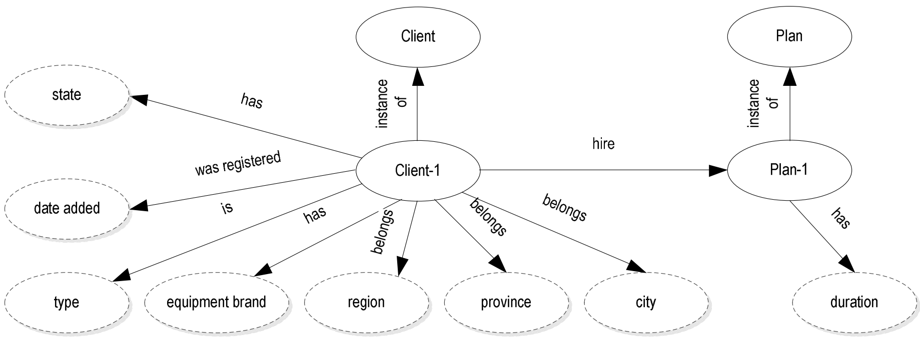

Fig. 4. Semantic Net of the Business Domain Model

Table 5. Frame Term-Category-Definition of Problem of Information Mining

Term	Category	Definition
by	Relationship	A client belongs to a group by his city
by	Relationship	A client belongs to a group by his province
by	Relationship	A client belongs to a group by his region
city	Attribute	city where the client live
Client	Concept	A person who hire the service
composed	Relationship	A group is composed by clients
Customer type	Attribute	Kind of customer by contract
define	Relationship	A group is defined by a rule
duration	Attribute	Number of months of hiring
equipment brand	Attribute	Client's equipment brand
Group	Concept	group of clients
group code	Attribute	unique group identification code
has	Relationship	A plan has a certain duration
identifies	Relationship	The plan code identifies a specific plan
identifies	Relationship	The group code identifies a specific group
identifies	Relationship	The rule code identifies a specific rule
Plan	Concept	Kind of contract
plan code	Attribute	unique plan identification code
province	Attribute	Province where a client live
region	Attribute	Region where a client live
rule	Concept	Variables that define the assignment of a client to a particular group
rule code	Attribute	unique rule identification code
state	Attribute	Client's state
subset	Relationship	The situation may be a variable that defines the mapping of the customer in a group
subset	Relationship	The plan may be a variable that defines the mapping of the customer in a group
subset	Relationship	The equipment brand may be a variable that defines the mapping of the customer in a group
subset	Relationship	The customer type brand may be a variable that defines the mapping of the customer in a group

Table 6. Extended Frame Concept-Attribute-Relation-Domain_Value of Problem of Information Mining

Concept	Attribute	Relation-ship	Input/ Output	Value
Rule	code	identifies	Output	Numeric
	state	subset		alphabetical EG: AAA, BA.
	Equip-ment brand	subset		alphabetical EG: Nokia
	type	subset		alphabetical EG: PR/PO, FF.
Plan	plan code	identifies		Alpha-numeric EG: U21.
	duration	has		Numeric
Group	group code	identifies	Output	Numeric
Client	province	by	Input	Numeric EG: 379
	city	by	Input	Numeric EG: 3
	region	by	Input	1 to 5

Table 7. Frame Concept- Problem_Relation of Problem of Information Mining

Concept	Concept associated	Relationship	Description
Plan	Rule	subset	The plan may be a variable that defines the mapping of the customer in a group
Rule	Group	define	A group is defined by a rule
Group	Client	composed	A group is composed by clients

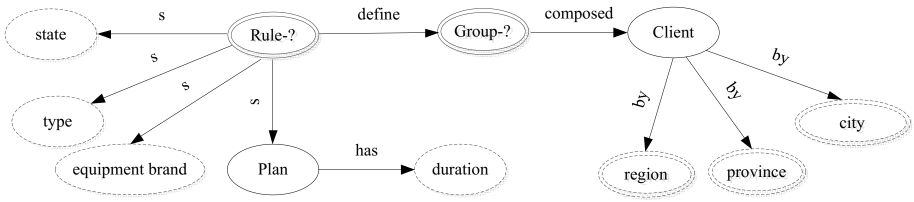

Fig. 5. Semantic Net of the Business Domain Model

When task "Derivation the Processes of Information Mining" is applied, we obtain the "Processes of Information Mining" which for the case study is "Discovery of group-membership rules" as has been proposed in [1].

5 Conclusions

This research started from the premise that the tools involved in software engineering processes do not apply to information mining projects. The authors belong to a research group that during ten years have been developing a body of knowledge for Information Mining Engineering, trying to help improve productivity of Latin-American SMEs (small and medium enterprises) in the software industry.

In this context, this paper introduced formalisms based on knowledge representation techniques (frames and semantic networks) to model the domain of business and information mining process domain. In line with the tradition of Knowledge Engineering, the process of transformation that involves the formalization allows to find explicit and implicit concepts (and relations among them) in the description of the domain and the business problem.

Based on the explanation of the concepts and relationships, a procedure to derive the exploitation process information from the proposed formalism is presented. A proof of concept that illustrates the application of the proposed procedure was developed. As future research, it is planned to develop the validation of the proposed methodology in selected information mining.

Acknowledgments. The research reported in this paper was funded by Grant-BENTR13 of the Committee on Scientific Research of the Province of Buenos Aires (CIC, Argentina), Grant-UNLa-SCyT-33A105 and partially by Project UNLa-33A167 of the Secretary of Science and Technology of National University of Lanus (Argentina).

References

1. García-Martínez, R., Britos, P., Rodríguez, D.: Information Mining Processes Based on Intelligent Systems. In: Ali, M., Bosse, T., Hindriks, K.V., Hoogendoorn, M., Jonker, C.M., Treur, J. (eds.) IEA/AIE 2013. LNCS (LNAI), vol. 7906, pp. 402–410. Springer, Heidelberg (2013)
2. Kononenko, I.Y., Cestnik, B.: Lymphography Data Set. UCI Machine Learning Repository (1986), `http://archive.ics.uci.edu/ml/datasets/Lymphography` (Último acceso 29 de Abril del 2008)
3. Michalski, R.: A Theory and Methodology of Inductive Learning. Artificial Intelligence 20, 111–161 (1983)
4. Quinlan, J.: Learning Logic Definitions from Relations. Machine Learning 5, 239–266 (1990)
5. Kohonen, T.: Self-Organizing Maps. Springer (1995)
6. Heckerman, D., Chickering, M., Geiger, D.: Learning bayesian networks, the combination of knowledge and statistical data. Machine Learning 20, 197–243 (1995)
7. García-Martínez, R., Britos, P., Pesado, P., Bertone, R., Pollo-Cattaneo, F., Rodríguez, D., Pytel, P., Vanrell, J.: Towards an Information Mining Engineering. In: Software Engineering, Methods, Modeling and Teaching. Edited by University of Medellín, pp. 83–99 (2011) ISBN 978-958-8692-32-6
8. Britos, P., Dieste, O., García-Martínez, R.: Requirements Elicitation in Data Mining for Business Intelligence Projects. In: Avison, D., Kasper, G.M., Pernici, B., Ramos, I., Roode, D. (eds.) Advances in Information Systems Research, Education and Practice. IFIP International Federation for Information, vol. 274, pp. 139–150. Springer, Boston (2008)
9. Chapman, P., Clinton, J., Keber, R., Khabaza, T., Reinartz, T., Shearer, C., Wirth, R.: CRISP-DM 1.0 Step by step BIguide. Edited by SPSS (2000)
10. Pyle, D.: Business Modeling and Business intelligence. Morgan Kaufmann (2003) ISBN 1-55860-653-X
11. SAS, SAS Enterprise Miner: SEMMA (2008), `http://www.sas.com/technologies/analytics/datamining/miner/semma.html` (last access November 24, 2013)
12. Minsky, M.: A framework for Representing Knowledge. AI-Memo 306. MIT Artificial Intelligence Laboratory (1974), `http://hdl.handle.net/1721.1/6089`
13. Sowa, J.: Semantic Networks. In: Shapiro, S. (ed.) Encyclopedia of Artificial intelligence, 2nd edn., Wiley & Sons (1992) ISBN 978-0471503071
14. Basso, D.Y., Vegega, C.: Case Study: Cell Phones. Business Intelligence Seminar Final Research Paper. Master Program on Information Systems Engineering. Graduate School, Regional School Buenos Aires, National Technological University, Argentine (2013)