Semantic interoperability of large systems through a formal method: Relational Concept Analysis

Mickael Wajnberg, Mario Lezoche, Alexandre Blondin Masse, Petko Valtchev, Hervé Panetto, Louise Tyvaert

Semantic interoperability of large systems through a formal method: Relational Concept Analysis

Mickaël Wajnberg*,**. Mario Lezoche**. Alexandre Blondin-Massé*.
Petko Valchev*. Hervé Panetto**. Louise Tyvaert**

* Département d'informatique, Université du Québec à Montréal, CP 8888, Montréal (Québec) H3C 3P8
(e-mail: wajnberg.mickael@courrier.uqam.ca, blondin_masse.alexandre@uqam.ca, valtchev.petko@uqam.ca).
** Research Centre for Automatic Control (CRAN UMR 7039), Université de Lorraine,
CNRS, Boulevard des Aiguillettes B.P.70239, 54506 Vandoueuvre-lès-Nancy, France
(e-mail: wajnberg.mickael@univ-lorraine.fr, mario.lezoche@univ-lorraine.fr, herve.panetto@univ-lorraine.fr)

Abstract: Interoperability is a major stake for industry, and in general for all the systems, of any dimension, that need to share contents in every shape. It provides that the exchanges between different parts of different entities perform in a perfect way. Various problems could arise and let the interoperation difficult or impossible. One of those problems could be the presence of implicit knowledge in the systems models. This kind of problems can be faced through knowledge formalisation strategies.

The Formal Concept Analysis (FCA) is a mathematical tool to represent the information in a structured and complete way. In this scientific work, we present an extension of the FCA, the Relational Concept Analysis, to reveal tacit knowledge hidden in multi contexts systems.

Keywords: Semantic Interoperability, Large system, Formal Concept Analysis, Relational Concept Analysis.

1. INTRODUCTION

In order to reduce time-to-market and product costs while improving product quality, industrial companies face many challenges. Some of these challenges include managing the increasing diversity and complexity of products, and enhancing collaborative and integrated engineering. It is possible to represent the enterprises process through their information systems. Enterprise systems (ES), in fact, are large-scale application software packages that support business processes, information flows, reporting, and data analytics in complex organizations (Wu et al, 2014). They are large and complex systems that need internal and external interoperation to efficiently work and produce (Chen et al, 2013). This underline the need for interoperability as well as the difficulty in dealing with lossless model transformations occurring at each level of the system boundaries. Interoperability can be defined as the ability of two or more systems to share, to understand and to consume information (Lezoche at al., 2012). Our research focuses on the conceptual level of interoperability, namely the ability to understand the exchanged information. But the exchanged information is not the complete information present in the systems and in the models representing those systems. A key activity is the ability of extracting tacit knowledge from the models to represent all the knowledge possessed by the systems to let them interoperate at the maximum possible rate.

1.1 Scientific problem

The knowledge is the key point to let different entities semantically interoperate. Large systems, like ES, are composed by multiple sub systems and they are usually modelled to be created, optimised, managed and to identify their most important characteristics and properties to interoperate with other systems (Wu et al, 2014). But modelling a system is an operation usually domain dependent and, often modeller dependent, so it happens that some knowledge remains implicitly hidden in the structure of the model. The focus of this research is to make explicit all the implicit knowledge hidden in the structure of the models that represents the systems.

The formalisation of the knowledge is one of the various approaches that exist to make explicit the tacit knowledge a model owns in its structure..The three core elements of this approach are: the concepts, the relationships and the properties. Those elements are subjected to formal constraints to create the knowledge in the form we know it.

2. FORMAL TOOLS

A concept is characterized both by a set of properties and a set of objects. For instance, the concept of "car" can be described by all the cars in the world, or by the set of all the properties of a car: has wheels, engine, doors, seats, etc... The formal concept analysis (FCA) consists in organizing formal

concepts into a hierarchy from a formal context, *id est* a set of objects with the properties of each object. This hierarchy will be presented under the shape of a lattice.

FCA allows to make explicit knowledge from a single-context system, and in a certain way, can be seen as a clustering method, where each concept is a cluster, and from each concept, knowledge in the form of association rules can be extracted.

But FCA is limited to the knowledge extraction from a single context at a time. Relational Concept Analysis (RCA) has been introduced as an extension of the FCA paradigm. It deals with multi-context systems and it is used to extract the tacit knowledge that exists in the crossover of the system's contexts.

We will, in this section, briefly describe how these two methods operate: in the first part, we will present the FCA basics, in the second section we will present the RCA.

2.1 Formal Concept Analysis

FCA has been introduced in (Wille, 1982), (Carpineto et al., 2004) and (Ganter et al., 2012). It is a method that aims to organize information from a context in a way that knowledge, such as association rules on a dataset, can be easily extracted.

To summarize we can present FCA as follows.

Given a set of objects O, and a set of attributes A, we can easily represent ourselves a cross-table that gives for each object $o \in O$ the set of attributes $A \subset A$ that o has. Such a cross table is called a formal context. The Table 1 is a formal context.

Table 1. Formal Context

/	A_0	A_1	A_2	A_3
O_0			×	×
O_1	×	×		
O_2	×		×	
O_3		×	×	×

As said earlier, FCA can be seen as a clustering method. Moreover, FCA tries to answer the following questions:

- How to regroup the objects of such a context considering their attributes?

- How to regroup the objects in multiple clusters such that we get a hierarchy/a granularity in the groups?

- How to present the information in a compact and intelligible way?

A cluster in formal concept analysis is defined by the name formal concept. A formal concept is a pair $(X, Y) \subset O \times A$ where every element $o \in X$ carries every single attribute a

present in Y and such that there are no other objects of O than these of X that carries simultaneously every attributes of Y.

FCA aims to define from the context K set of all concept C_K. It, also, gives to this set of concepts an order relation: a concept C_2 is greater than a concept C_1 if and only if the set of object of C_1 is a subset of the set of object of C_2.

$$\forall c_1 = (X_1, Y_1) \in C_K, \forall c_2 = (X_2, Y_2) \in C_K, c_1 \leq c_2 \Leftrightarrow X_1 \subset X_2$$

Provided with that order relation, $(C_K, \leq)$ defines a Partially Ordered Set (POSET), and in particular a complete lattice. Numerous algorithms have been developed in the literature to conceive such a lattice from the formal context (Stumme, 2009), (Kuznetsov, 2002).

The lattice structure allows an easy reading to extract knowledge from the clusters: any concept in the lattice gives an information of the shape "all the objects in those concepts share the attributes in the concept and they are the only one". Also, if we take any inferior concept, it will give a more specific information. And of course, a superior concept will give a more general observation.
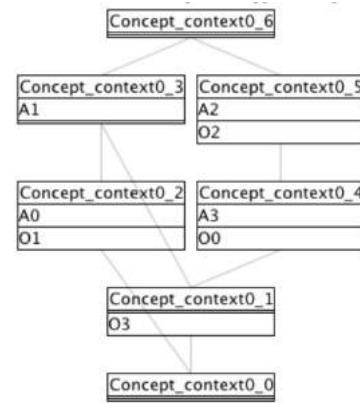
Figure 1. Lattice $\mathcal{L}$

Also, we can notice that any concept provides an association rule. Let $C_r = (O_r, A_r)$ be a concept from the reduced lattice, and $C = (O, A)$ the same concept but in the extended version lattice. We have: $A_r \rightarrow A \setminus A_r$ an association rule of confidence 100%. The support is of course $\frac{|O|}{|O|}$. These extracted rules represent the most important knowledge contained in a concept lattice.

FCA is a powerful formal tool to easily extract knowledge from a context, but is limited to one context at a time. Therefore, FCA by itself is not suited as a process for semantic interoperability. RCA is an extension to the FCA paradigm that aims to cover that lack. In addition to the FCA feature, it extracts the tacit knowledge that exists only in the implicit relationships between diverse contexts.

2.2 Relational Concept Analysis

How to use the interaction between different sets of objects present in different formal contexts to discover implicit knowledge? That's the main question RCA aims to answer.

Let $\mathfrak{K}$ be a set of formal contexts. A set $\mathfrak{R}$ of relations handles the links between the contexts of $\mathfrak{K}$: for each $R_{i,j} \in \mathfrak{R}$ there exists $K_i, K_j \in \mathfrak{K}$ and $R_{i,j}$ is defined by a triple of (O_i, O_j, I) where O_i is the set of objects of K_i, O_j is the set of objects of K_j and $I \subset O_i \times O_j$ and denotes all the links between the objects of O_i and the objects of O_j. (Rouane-Hacene et al., 2007), (Rouane-Hacene et al., 2013)

RCA consists in enriching the lattices that every $K \in \mathfrak{K}$ produces through FCA using the relations $\mathfrak{R}$. Until convergence, and for each relation $R_{i,j} \in \mathfrak{R}$ and a scaling operator, RCA enriches the source context K_i using the information contained in the target lattice $\mathcal{L}_j$ produced using FCA on the context K_j. When every context has been updated, the corresponding lattices have to be updated. And since, the lattices have been updated, new concepts might have been created, so another iteration takes place. The following part gives details for the context enrichment procedure through an example.

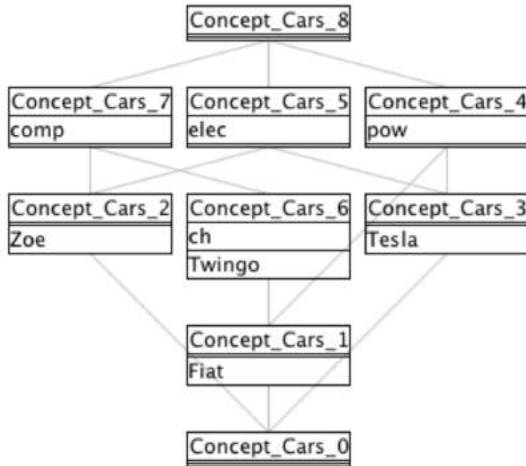Let's $K_1 = Cars$ and $K_2 = People$ be two contexts as follows:

Table 2. Formal Context Cars

Cars	electrical	powerful	compact	cheap
Twingo			X	X
Tesla3	X	X		
ZOE	X		X	
Fiat500		X	X	X

Table 3. Formal Context Persons

Persons	male	female	tall	little	rich	modest
Albert	X		X		X	
Benjamin	X			X		X
Chloe		X		X	X	
Damien	X		X			X
Elodie		X	X			X

and their respective Lattices:

Figure 2. Lattice $\mathcal{L}_1$
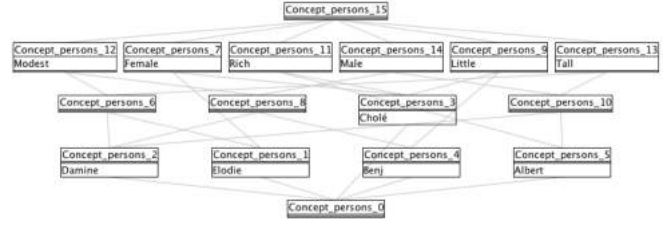
Figure 3. Lattice $\mathcal{L}_2$

Then let's have $R_{2,1} = Buy$ the relation between K_1 and K_2

Table 4. Formal Context Buy

Buys	Twingo	Tesla	Zoe	Fiat
Albert		X		
Benjamin	X			
Chloe		X		X
Damien			X	
Elodie			X	

To enrich context K_1 through the relation $R_{1,2}$ and the context K_2 we first need to define a scaling operator. The scaling operator can be summarized as the rules to apply to the relation $R_{2,1}$ to extract information. In the present example, we choose the existential operator $\exists$, then we proceed as follow:

Table 5. Scaling operator

Persons	$\exists buy : C_0$	$\exists buy : C_1$	$\exists buy : C_2$	$\exists buy : C_3$	$\exists buy : C_4$	$\exists buy : C_5$	$\exists buy : C_6$	$\exists buy : C_7$	$\exists buy : C_8$
Albert				X	X	X			X
Benjamin							X	X	X
Chloe		X		X	X	X	X	X	X
Damien		X				X		X	X
Elodie		X				X		X	X

- we add to K_1 a set of column: one column for each concept in the target lattice;

- for each column, we consider the objects of the referred concept, for instance concept C_2 has the objects "Zoe";

- for each line of the column we refer to the relation to decide if we put a cross or not. For instance, in the line of the object "Damien" to decide if there is a cross we check (the query shape depends on the scaling operator) if there exists an element in the image of "Damien" by $R_{1,2}$ (the set $\{Zoe\}$ in the target concept extension;

- we proceed to such a reasoning for every column and line of the table.

Of course, many scaling operators can be defined, but the main idea stands. Since new columns had been added to a formal context, lattices have to be updated, and new concepts may have emerged, and since new concepts have appeared in the lattices, then the operation can be processed again until a fixed point is reached, such a fixed point is guaranteed (Rouane-Hacene et al., 2013). With new concepts comes new explicit knowledge.

In the next section, we will present how RCA can be used as an interoperability technique on heterogeneous data. We chose an application in the healthcare domain since a neurologist could provide a large amount of real world data. This data comes from f-MRI and EEG, two clinical examination producing results that can't be interpreted altogether a priori.

3. CASE STUDY

In neurology, the brain activity is measured, among other techniques, by electroencephalography (EEG) and functional Magnetic Resonance Imagery (f-MRI).

Each area of the brain (cortical areas) emits different signal shapes over time to activate different functions of the brain. During the sleep, the brain produces a signal called sleep spindles, an amplified activity within a short period of time. Such discrete event occurs multiple times during the sleep and can be well delimited in time. The spindle makes the whole brain activates, each cortical area possibly presents a different activity during a single spindle. A spindle can be studied through EEG and f-MRI.

EEG consists in putting electrical captors on the head of a patient, the captors make the acquisition of the electrical signal (in millivolts) in function of the time. The temporal precision is such that the electrical brain activity, sleep spindles in particular, can be measured and visualized for each captor
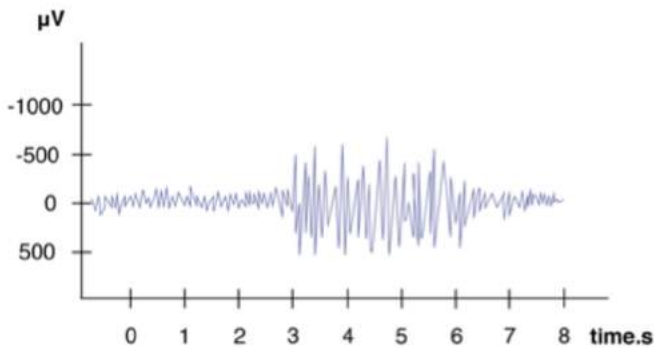

Figure 4. Sleep spindle signal

But since EEG is taken with surface captors, the spatial localisation is rough and cannot allow us to determine for a captor where the signal exactly come from.

On the other hand, f-MRI studies the brain activity through the blood flow: an activated area consumes oxygen, and oxygenated blood has a different magnetism than deoxygenated blood, and such an oxygen concentration can be measured through Magnetic Resonance Imagery. So, brain activity can be measured with an excellent spatial precision through this technique (precision to a Voxel -meaning 3D pixel- roughly a $1mm^3$). But, the oxygenation reaction is measured with a high temporal latency due to the bio-physical restriction to the system (oxygen is driven through the blood which is not an immediate operation), and the reaction takes three to nine seconds to appear.

The neurologists emitted, and search confirmation, to the hypothesis that the cortical areas in which a spindle occurs (it can appear at different places at the same time) depends directly on the signal shape. To prove right or wrong such hypothesis, a possibility to analyse in concert the shape of the electrical signal with the exact areas activities would be a first step into such an analysis. To do so, we propose here RCA has as the tool to make interoperate EEG data with f-MRI data and that, by combining the two clinical examinations emulates a tool that has the spatial precision of f-MRI and temporal precision of EEG. Such a combined used of these clinical exams, have never be studied, according to the neurologists.

The global process would follow the process in *Figure 5*.
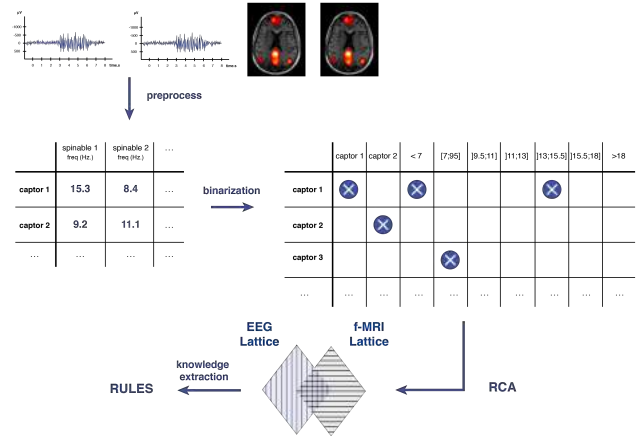
Figure 5. Rule definition process through RCA method

Starting from the two clinical examinations operated at the same time, the data are pre-processed to numerical values: for each spindle, each captor and voxel gets a highly sampled signal over time. Then an analysis brings for each voxel the highest level reached in response to a spindle. After that the application of the Fourier Transformation to the electrical signal allows to retain the most represented frequencies that are in the spindle (frequency with the highest value on the Fourier transformation diagram).

After the pre-process, a process to make the data in binary format is needed. For each spindle, the highest response voxel is designated and spatial local maximums too, also for each captor is recorded if the signal has a high frequency response (frequency with the highest value on the Fourier Transformation diagram is between 15 and 22Hz and) or a low frequency response (between 6 and 13 Hz)

From this process can be set the contexts and relations, and a RCA process can be launched. Lattices are then built and we want to show that from those results we can extract meaningful knowledge that would allow to get the spatial and temporal interesting aspects of both clinical examination.

3.1 Relational Concept Analysis Model

In this section, we present a toy example to show the model we will use when all the real data will be collected, pre-processed and made in binary tables.

First, we have the basic objects: three spindles, three EEG captors, five Voxels presented in Table 6. The interesting

information is contained in the relation linking these trivial contexts.

Table 6. Left to right, up to down: formal context for: spindles, captors and voxels

K(Spindles)	S1	S2	S3
Spindle1	X		
Spindle2		X	
Spindle3			X

K(Captors)	C1	C2	C3
Captor1	X		
Captor2		X	
Captor3			X

K(Voxels)	V1	V2	V3	V4	V5
Voxel1	X				
Voxel2		X			
Voxel3			X		
Voxel4				X	
Voxel5					X

The first information retained is structural, it links a Voxel to a Captor. The relations cover holds true for a couple $(Captor\,C, Voxel\,V)$ if the captor C could receive the electric signal of the occurence of a spindle in the location of the brain characterized by the voxel V.

Table 7. Cover Relation: Captors - Voxels

R(Covers)	Voxel1	Voxel2	Voxel3	Voxel4	Voxel5
Captor1	X	X			
Captor2			X		
Captor3				X	X

Then we consider the relations between the spindles and the captors. For each spindle, each captor gets a signal that can be of value high or low frequency as explained in the previous section.

Table 8. Frequency Relations: Spindle – Captors

R(High)	Captor1	Captor2	Captor3	R(Low)	Captor1	Captor2	Captor3
Spindle1	X			Spindle1		X	X
Spindle2	X	X		Spindle2			X
Spindle3		X	X	Spindle3	X		

Finally, we need to introduce the relations between the spindles and the voxels. There are two, the first one *Abs Max* holds true for a couple $(Spindle\,S, Voxel\,V)$ if for the spindle S the Voxel V is the voxel that has the highest output through the f-MRI. The second relation *Loc Max* gives states if a voxel is spatially a local maximum for the f-MRI output.

Table 9. Abs Max Relation: Highest output in f-MRI

R(AbsMax)	Voxel1	Voxel2	Voxel3	Voxel4	Voxel5
Spindle1		X			
Spindle2		X			
Spindle3			X		

Table 10. Loc Max Relation: Local maximum in f-MRI

R(LocMax)	Voxel1	Voxel2	Voxel3	Voxel4
Spindle1		X		
Spindle2		X		X
Spindle3	X		X	X

By using an RCA process in the less restrictive configuration (every context is used, every relation is used in both directions and only the existential scaling operator is used), three lattices are generated.

Figure 6. Lattice $\mathcal{L}_1$

Figure 7. Lattice $\mathcal{L}_3$

4. RESULTS

It is to be highlighted that RCA is a procedure that creates more knowledge than the one that can be only extracted through the FCA procedure: relational attributes link the lattices together. To extract such knowledge, we suggest the following pseudo-algorithm that extract the knowledge contained in a lattice node in the form of an RDF graph. It basically is a breadth search strategy applied to the structure of a lattice family in the case where no relation links a lattice to itself.

Data Input : Starting node N_0 from lattice $\mathcal{L}$, the lattice family
Result Output : RDF graph G

Let a node have id of N_0 and objects of N_0.extension;
Add N to G;
While{not all nodes in G are marked}
    Take a non-marked node and make it the current node N;
    Mark N;
        For{each relation R for which $\mathcal{L}$ is not a target}
            list the target nodes;
            keep only minimal elements of the lattice;
            If{list is empty}
                continue to next relation;
            Else
                For{each element E in the reduced list}
                    If{exists in G a target node T of id E}
                        create an edge from N to T with label R;
                    Else
                        create a node T with id of E and objects of E.extension;
                        add T to G;
                        create an edge from N to T with label R;
Return G;
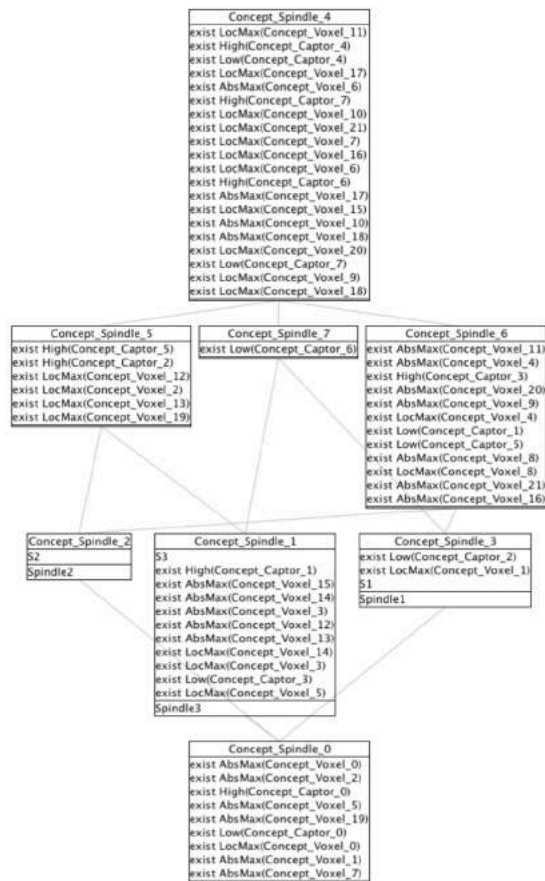
Figure 8. Lattice $\mathcal{L}_2$

Though this procedure, by starting from node *Concept_Spindle_6* we generated the following graph in figure 9. We can easily read informations such as if captor C3 receives a low frequencies signal and C1 a high frequencies signal, then the absolute maximal response is in the cortical area characterized by voxel V2 and another local maximum will appear in one of V4 or V5 voxel.
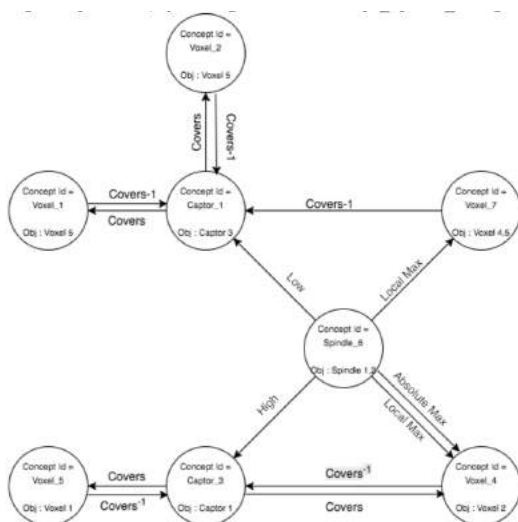

Figure 9. RDF Graph containing the resulting knowledge

4. CONCLUSIONS

This scientific work focuses on the semantic interoperability of large systems and propose a methodology to extract the structural tacit knowledge and to express it in an explicit and formal way. The mathematical approach called Formal Concept Analysis create the basis of the work but the multi contextual reality of the large systems needs some extensions that can take care of the various relationships between the different contexts owned by each system taken in account. RCA works on the background defined by FCA but provides the possibility to extract the tacit knowledge contained in the relations between the different contexts of a multi-context system. The proposed solution of the full process has as result a RDF graph containing the knowledge, in an explicit way, extrapolated from the existing relationships between the instances of the models.

The next step is to validate this approach with the complete set of real data. Another great issue is to focus on the automatic building of constrains rules that would optimise the cluster selection of the composed Lattices.

REFERENCES

Claudio Carpineto and Giovanni Romano. Concept data analysis: Theory and applications. John Wiley & Sons, 2004.

Chen, Wen; Iqbal, Asif; Abdrakhmanov, Akbar; Parlar, Jay; George, Chris; Lawford, Mark; Maibaum, Tom; Wassyng, Alan; Large-Scale Enterprise Systems: Changes and Impacts, Enterprise Information Systems: 14th International Conference, ICEIS 2012, Wroclaw, Poland, Revised Selected Papers, 2013, Springer Berlin Heidelberg, isbn 978-3-642-40654-6.

Bernhard Ganter and Rudolf Wille. Formal concept analysis: mathematical foundations. Springer Science & Business Media, 2012.

Sergei O. Kuznetsov and Sergei A. Obiedkov. Comparing performance of algorithms for generating concept lattices. Journal of Experimental & Theoretical Artificial Intelligence, 14(2-3):189–216, 2002.

Mario Lezoche, Esma Yahia, Alexis Aubry, Hervé Panetto, Milan Zdravković. Conceptualising and structuring semantics in Cooperative Enterprise Information Systems Models. Computers in Industry, Elsevier, 2012, 63 (8), pp.775-787.

Mohamed Rouane-Hacene, Marianne Huchard, Amedeo Napoli, and Petko Valtchev. Extraction de concepts et de relations en analyse relationnelle de concepts (arc). 2007.

Mohamed Rouane-Hacene, Marianne Huchard, Amedeo Napoli, and Petko Valtchev. Relational concept analysis: mining concept lattices from multi-relational data. Annals of Mathematics and Artificial Intelligence, 67(1):81–108, 2013.

Gerd Stumme. Formal concept analysis. In Handbook on ontologies, pages 177–199. Springer, 2009.

Rudolf Wille. Restructuring Lattice Theory: An Approach Based on Hierarchies of Concepts, pages 445–470. Springer Netherlands, Dordrecht, 1982.

Wu He, Li Da Xu, "Integration of Distributed Enterprise Applications: A Survey", Industrial Informatics IEEE Transactions on, vol. 10, pp. 35-42, 2014, ISSN 1551-3203.