Mining Description Logics Concepts With Relational Concept Analysis

Marianne Huchard, Amedeo Napoli, Amine Mohamed Rouane Hacene, Petko Valtchev

Mining Description Logics Concepts With Relational Concept Analysis

Marianne Huchard[1], Amedeo Napoli[2], Mohamed Rouane Hacene[2], and Petko Valtchev[3]

[1] LIRMM, 161 rue Ada, 34392 Montpellier Cedex 5, France, *huchard@lirmm.fr*
[2] LORIA, Campus Sciences, BP 239, 54506 Vandœuvre-lès-Nancy Cedex France, {*napoli, rouanehm*}*@loria.fr*
[3] Dépt. d'informatique, UQAM, CP 8888, succ. Centre-Ville, Montréal, Canada, H3C 3P8, *valtchev.petko@uqam.ca*

Abstract. *Symbolic objects* were originally intended to bring both more structure in data and more intelligibility in final results to statistical data analysis. We present here a framework of similar motivation, i.e., combining a data analysis method, — the concept analysis (FCA) — with a knowledge description language inspired by description logic (DL) formalism. The focus is hence on proper handling of relations between individuals in the construction of formal concepts. We illustrate the relational concept analysis (RCA) framework which complements standard FCA with a dedicated data format, a set of scaling operators, an iterative process for lattice construction, and translations to and from a DL language.

1 Introduction

Symbolic objects (SO) (Diday (1998)) were designed to meet the urgent need for processing of more realistically structured data, i.e., beyond mere real number vectors, in statistical data analysis, while representing the final results in a more intelligible manner. On data formats, beside the variety of value domains of the descriptive variables (taxonomic, interval, histogram, etc.), higher-level structure is also provided for, e.g., in *hordes* which provide for nesting of individuals. In the broader field of knowledge discovery from data, structure and intelligibility have been pursued through a symbiosis with knowledge representation (KR) (Brachman and Anand (1996))

Formal concept analysis (FCA) (Ganter and Wille (1999)) as data analysis paradigm also endorsed KR concerns. In fact its target FCA structure, the concept lattice, represents a natural framework for both taxonomies and conceptual hierarchies. While the standard FCA framework barely admits structure in the input datasets, recent trends targeted the complexly structured data. For example, a first trend admits explicit inter-individual links which, once expressed as first-class objects within a *power context family*, are dealt in a straightforward way, i.e., grouped into formal concepts representing new, and compound, relations (Prediger and Wille (1999)). Independently, and somewhat closer to the SO approach to structure, logic-based KR has been tentatively introduced in the *conceptual scaling* mechanism which enables the processing of non-binary data in FCA. Thus, in (Prediger and Stumme (1999)), a language of the description logic DL family (Baader et al. (2003)) was used to express conditions involving domain concepts and relations, which were then applied to individuals as binary attributes. It is noteworthy that the symbiosis of SO and FCA, i.e., the concept analysis of symbolic datasets, has been investigated as well (Polaillon (1998)).

Our own study on concept analysis of complex datasets is motivated by the rapidly growing need for interoperability between mining mechanism and modern KR environments, especially in the wake of the Semantic Web launch. In simple terms, this means mining tools must be able to process data expressed in languages, such as OWL and SWRL, and output the discovered knowledge in equally compatible formats. In this respect, our concerns combine, on the one hand, the adequate clustering of relational datasets, as logically-founded languages describe individuals by means of both unary predicates (*concepts*) and binary ones (relations, or *properties*), and, on the other hand, the design of compound expressions to intentionally describe the discovered clusters. As an approach for the concept analysis of relational data, we proposed a dedicated framework, called *relational concept analysis* (RCA), which offers simple solutions to both concerns. Moreover, the framework relies on three original components: a data format inspired by the entity-relationship conceptual data model, a scaling method applying various policies in the translation of inter-individual links into binary attributes, and an iterative lattice construction process allowing many separate individual sorts to be analyzed simultaneously.

The present paper summarizes the RCA theoretical foundations and illustrates its *modus operandi* using a small-size, albeit realistically structured dataset. The following Section 2 provides minimal background on FCA and then RCA, and briefly examines the composition of a DL language. Section 3 introduces the sample dataset, which is then analyzed w.r.t. two different scaling policies. The analysis processes based on wide and on narrow scaling are followed in Section 4 and Section 5, respectively.

2 From FCA to RCA

The following is a brief presentation of the RCA framework. Details may be found in (Huchard et al. (2007)) while an implementation is available within the GALICIA platform[1].

2.1 Standard FCA

FCA is the process of abstracting conceptual descriptions from a set of individuals described by attributes (Ganter and Wille (1999)). Formally, a *context* $\mathcal{K}$

[1] `http://sourceforge.net/projects/galicia/`

associates a set of objects (O) to a set of attributes (A) through an incidence relation $I \subseteq O \times A$, i.e., $\mathcal{K} = (O, A, I)$. For example, in Section 3, a context is presented where objects are scientific publications (e.g., monographs, journal articles, conference papers, theses, etc.), whereas attributes are general topics (e.g., software engineering, lattice theory, etc.). The represented incidence relation is therefore to be interpreted as "speaks about" or "deals with".

In this settings, FCA focuses at the way objects group together on grounds of shared attributes. Intuitively, each subset of objects is examined together with the respective set of shared attributes (e.g., a set of publications determines a list of all common topics). Among all object sets, only maximal ones are kept, i.e., sets comprising *all* objects incident to the shared attributes. This is formalized by two applications mapping object sets to attribute ones and *vice versa*, both denoted $'$ hereafter. For instance, on objects, the $'$ application is defined as follows: $' : \mathcal{P}(O) \rightarrow \mathcal{P}(A); \quad X' = \{a \in A \mid \forall o \in X, oIa\}$.

A basic result states that maximal sets of objects, called *extents* in FCA, are in one-to-one correspondence to maximal sets on attributes, or *intents*. Furthermore, the pairs $(X, Y) \in \mathcal{P}(O) \times \mathcal{P}(A)$, of mutually corresponding sets, i.e., such that $X = Y'$ and $Y = X'$, called *(formal) concepts*, form a complete lattice with respect to the inclusion of the extents, i.e., the X part. Extracting the concept lattice $\mathcal{L}$ of a context $\mathcal{K}$ is the key task in FCA. Fig. 2 shows, on its right-hand side, the concept lattice of the publication context which is itself embedded in the table on the left-hand side (only the first four columns).

The classical FCA apparatus is limited to datasets that either originally represent binary relations or can be easily, i.e., with no significant precision loss, transformed to such relations. Indeed, the *conceptual scaling* mechanism translating non-binary attributes (e.g., numerical or nominal) into binary ones, amounts to replacing attribute values by predicates on them. For instance, the domain of *nbOfPages* attribute in publications could be split into the ranges *short*, *standard*, and *long* (paper), each of them expressed as a predicate (e.g, *nbOfPages*$\leq$ 6 for short one). Observe that the definition of the predicates precedes the scaling process and is usually the charge of a domain expert.

Unsurprisingly, the data stored in a relational database remains well beyond the reach of the above approach, and for some good reasons. First, the underlying entity-relationship (ER) conceptual data model admits several *entities*, i.e., sorts of individuals, that are connected by *relationships*, i.e., n-ary predicates on entities, whereas FCA typically focuses on a single set of individuals (although these may generate a family of contexts) and yields a single concept lattice. As an illustration, imagine a database modeling a collection of scientific publications, researchers, topics, author-to-paper links, references among publications, etc. Moreover, a natural way of analyzing such data would be to form concepts that reflect commonalities both in individual properties and in their links to other individuals, following, for instance, the way DL concepts are defined (see Section 2.3). Although approaches dealing with relations have been studied in FCA, none of them allows links and properties to be mixed in concept intents. To bridge the gap, we have proposed a relational FCA framework, called *relational concept analysis* (RCA), that basically adds a new data format, a set of scaling mechanisms for relational links and an iterative method for the simultaneous construction of a set of concept lattices.

2.2 RCA summary

The RCA data format, a *relational context family* (RCF), combines FCA and ER as it consists of a set of contexts and a set of binary relations, each involving the objects from two contexts of the RCF.

Definition 1. A *relational context family* $\mathcal{R}$ is a pair $(\mathbf{K}, \mathbf{R})$, where $\mathbf{K}$ is a set of contexts $\mathcal{K}_i = (O_i, A_i, I_i)$, $\mathbf{R}$ is a set of relations $r_k \subseteq O_i \times O_j$, where O_i and O_j are the object sets of the formal contexts $\mathcal{K}_i$ and $\mathcal{K}_j$.

Let now a relation r (e.g., authoring of papers by researchers) link objects from a context $\mathcal{K}_i$, the *domain* of r, to those of $\mathcal{K}_j$, its *range*. In order to scale upon r so that one can use the information it conveys in the concept analysis upon $\mathcal{K}_i$, we consider the conceptual structure, i.e., all (known) formal concepts, of $\mathcal{K}_j$. The concepts are turned into binary predicates just as in classical scaling. The key difference is that in assigning such a predicate to an object o_i from $\mathcal{K}_i$, instead of comparing an attribute value to a range of such values, a set of objects, i.e., the links of type r for o_i, denoted $r(o_i)$, is compared to the extent of a concept c_j on $\mathcal{K}_j$. For instance, to describe researchers with respect to the authored papers, these will be compared to the extents of the formal concepts on the entire papers collection (e.g., journal papers on statistics). Various relationships between $r(o_i)$ and the extent of c_j (e.g., inclusion, non-empty intersection, intersection of a certain size, etc.) may be required in order for o_i to acquire the corresponding attribute, invariably denoted by $r : c_j$. These are discussed in the next paragraph.

Relational scaling opens the way to lattice construction. However, the global analysis process is not one-shot, it rather proceeds iteratively, i.e., by successive steps alternating scaling and concept formation. Indeed, as no restriction is imposed in the relational structure of a RCF, there may well be circuits in the way contexts are related by relations, hence the mutual dependence between such contexts in the sense that each of them requires the other(s) to be processed first in order to provide the formal concepts required for scaling. To break the deadlock, a bootstrapping step is performed in the beginning of each RCA process, in which all object sorts get the lattice corresponding exclusively to their local properties (from the underlying contexts). In the subsequent steps, scaling is used to translate the already available structure, i.e., formal concepts, from the range context of a relation to the domain one. More precisely, the current lattices are first used to scale upon the relations of the RCF thus generating new attributes in the respective domain contexts. The lattices of the extended contexts are then constructed, possibly triggering a new scaling/construction step. Indeed, as the new attributes may yield new extents, the lattices and hence the scales they represent may evolve, hence the need to re-scale in order to keep the domain contexts in line with the evolution. The global process of iterative lattice construction, called MULTI-FCA, nevertheless converges to a set of lattices representing a fixed-point. Section 4 and Section 5 illustrate the way MULTI-FCA unfolds.

2.3 Description logics and relational scaling

Description logics (DL) are KR formalisms rooted in first order predicate logic that offer means to structure the otherwise flat logical representation, namely in terms of *concepts*, *roles*, and *individuals* (Baader et al. (2003)). DL languages allow expressions, or *descriptions*, to be composed out of other descriptions up to an arbitrary depth. A DL language is built on top of a collection of *primitive concept* and *role* names which denote the meaningful concepts and relations from a domain (e.g., Human, Female, Doctor, child, father, etc.), individual names (e.g., Ann) and constants ($\top$ and $\bot$).

Concepts are interpreted as sets of individuals (their *instances*) and roles as sets of individual pairs[2]. Further concepts and roles are defined by combining concept and role names, either primitive or already defined, via a set of constructors, e.g., conjunction ($\sqcap$), disjunction ($\sqcup$), negation ($\neg$). By definition, a role has a domain and a range concept and is inherited by the sub-concepts of the domain concept. It may be further restricted for every concept it applies to, for instance, by applying universal or existential quantifiers to the set of links. Thus, given a role r and a concept C, the following concept expressions can be composed: (*i*) $\forall \mathsf{r}.\mathsf{C}$ (*value restriction*), (*ii*) $\exists \mathsf{r}.\mathsf{C}$ (*full existential quantification*), and (*iii*) $\exists \mathsf{r}.\top$ (*limited existential quantification*). All these work as filters on the individuals: (*i*) collects those whose links of type r, if any, point exclusively to instances of the concept denoted by the expression C, (*ii*) those with at least one r link to such an instance, and (*iii*) those with at least one r link, regardless of the underlying concept. As an illustration, consider the expression of the concept of "all fathers *and* all parents of a female child whose children are all doctors" in DL:

$$\mathsf{Male} \sqcap \exists \mathsf{child}.\top \sqcup \mathsf{Human} \sqcap \exists \mathsf{child}.\mathsf{Female} \sqcap \forall \mathsf{child}.\mathsf{Doctor}$$

Individuals are represented in a DL language as constants (e.g., Ann) and characterized by a set of ground predicates, unary for the concepts they belong to and binary for the roles they possess (e.g., Human(Ann), Female(Ann), child(Ann, Mary)). Consequently, the translation of a collection of DL individuals into an RCF is immediate: First, each individual is assigned a unique

[2] See (Baader et al. (2003)) for formal definitions for DL syntax and semantics.

concept to express its very nature (e.g., Human) in the same way entities within an ER schema do. Such concepts are translated as contexts while their instances become the respective formal objects. Next, the remainder of the concepts an individual belongs to are translated as binary attributes and attached to the underlying context (e.g., Female for the context modeling Human). Finally, all roles become relations in the RCF whose domain and range contexts are determined following the individuals in the role pairs and the contexts comprising their respective translations.

The DL formalism has a direct impact on the RCA scaling mechanism as well. Indeed, as mentioned previously, given a relation r that connects objects from $\mathcal{K}_i$ to those from $\mathcal{K}_j$, the various ways to assign an attribute $r : c_j$, where c_j is a concept on $\mathcal{K}_j$, to objects o_i from $\mathcal{K}_i$ follow restriction constructors from DL. More precisely, we defined several scaling policies, termed *encoding schemes*, including a value-restriction-like scheme, called *strict narrow*, a full existential-like one, or *wide*, and a third one, called simply *narrow*, that amounts to a combination of both. Indeed, while strict narrow scheme only requires $r(o_i) \subseteq extent(c_j)$, the narrow adds the condition $r(o_i) \neq \emptyset$. The latter condition is implied by the requirement of a wide scheme, i.e., $r(o_i) \cap extent(c_j) \neq \emptyset$. The way narrow and wide encoding scheme work is illustrated below.

Given the forward translation from a DL language to an RCF and the above scaling policies, the reverse translation of the formal concepts yielded by RCA into a DL knowledge base is immediate.

3 Running example

The sample RCF is made of a single context and two binary relations. The *Papers* context assigns publications, as objects, to the topics they refer to — software engineering (*se*), lattice theory (*lt*) and man machine interface (*mmi*) — as attributes. The relation *cites* models citations while *develops* connects a long publication, e.g., a thesis, to a paper whose key ideas the former extensively develops. Fig. 2 depicts the RCF both as a conceptual schema and as a data graph made of links and individuals (to whom codes are assigned for subsequent use in the text). In order to eases the tracking of

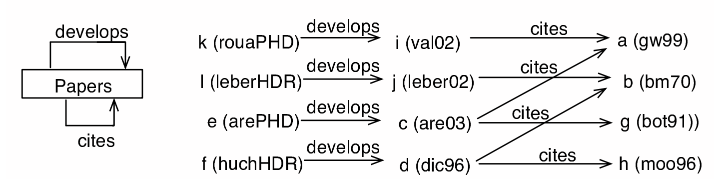

Fig. 1. Sample RCF. **Left:** As UML schema; **Right:** As data graph.

the gradual emergence of formal concepts, the example was stripped of a large number of papers and citation links. It nevertheless shows a complex, three-level link structure: Indeed, a set of four papers (level one) are substantial developments of four other papers (level two) which cite papers on level three. Moreover, though cycles in links are avoided, these are dealt with in much the same way. The RCF corresponds to a DL knowledge base with two roles (develops and cites) and four concepts, i.e., Papers, AboutLatticeTheory, AboutSoftwareEngineering, and AboutManMachineInterface.

With an object set $O = \{a..l\}$ and attribute set $A = \{lt, mmi, se\}$ the information content of the RCF can be summarized as follows (see Fig. 2):

- $I \subseteq O \times A$; $I = \{(a, lt), (b, lt), (g, mmi), (h, se)\}$,
- $cites \subseteq O \times O$; $cites = \{(c, a), (c, g), (d, b), (d, h), (i, a), (j, b)\}$,
- $develops \subseteq O \times O$; $develops = \{(e, c), (f, d), (k, i), (l, j)\}$.

Thus, initially, only level-three papers share descriptions and hence form concepts, e.g., a and b share the lt topic and therefore form the lattice theory publication concept. The lattice yielded by the paper context, regardless of the existing links, is given in Fig. 2 (on the right). Obviously, the aforementioned concept $c0 = (\{a, b\}, \{lt\})$ is the only non-trivial one. This lattice, once translated into binary attributes by scaling, enables new groupings, e.g., of c, d, i, j which cite at least one paper on lattices. The resulting concept trig-

		lt	mmi	se
gw99	a	x		
bm70	b	x		
are03	c			
dic96	d			
arePHD	e			
hucHDR	f			
bot91	g		x	
moo96	h			x
val03	i			
leber02	j			
rouaPHD	k			
leberHDR	l			

Kpapers

	a	b	g	h
a				
b				
c	x		x	
d		x		x
e				
f				
g				
h				
i	x			
j		x		
k				
l				

cites

	c	d	i	j
a				
b				
c				
d				
e	x			
f		x		
g				
h				
i				
j				
k			x	
l				x

develops

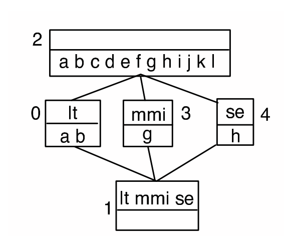

Fig. 2. Left: Initial RCF on papers; **Right:** Lattice 1 on papers.

gers yet further sharing, this time at level one, and the whole process goes on. The way the analysis process unfolds and its final result depend on the exact scheme used for scaling, as shown by the next two sections.

4 Narrow scaling-based RCA

Intuitively, the narrow scheme favors compact lattices as potentially less objects will get the new attributes due to the stronger requirements.

Step 1 Narrow scaling upon *cites* and w.r.t. lattice in Fig. 2 adds five new attributes of the type *cites:c* to the context. However, given the non-empty citation link sets ($cites(c) = \{a, g\}$, $cites(d) = \{b, h\}$, $cites(i) = \{a\}$, $cites(j) = \{b\}$), only two of them, i.e., *cites:c2* and *cites:c0*, are effectively assigned to a paper. Thus, all level-two papers, i.e., c, d, i, j, get the attribute *cites:c2* in the scaled context as *c2* comprises the entire dataset, whereas only i and j get *cites:c0* as well. Correspondingly, the relation I is extended with the pairs (*c,cites:c2*), (*d,cites:c2*), (*i,cites:c2*), (*j,cites:c2*), (*i,cites:c0*), and (*j,cites:c0*) (see Fig. 3).

					cites				develops
		lt	mmi	se	2	0	3	4	2
gw99	a	x							
bm70	b	x							
are03	c				x				
dic96	d				x				
arePHD	e								x
hucHDR	f								x
bot91	g		x						
moo96	h			x					
val03	i				x	x			
leber02	j				x	x			
rouaPHD	k								x
leberHDR	l								x

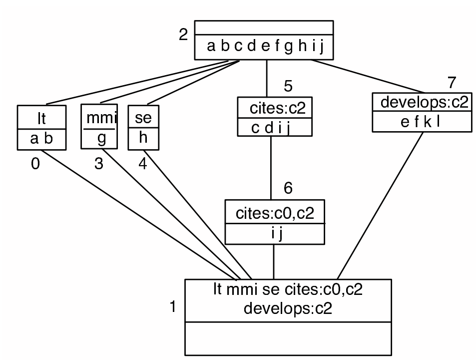

Fig. 3. Narrow scaling, step 1. **Left:** The scaled context; **Right:** Lattice 2.

Narrow scaling upon *develops* ranges only over papers having such links, i.e., e, f, k, and l. As none of the developed papers, i.e., c, d, i, and j, belongs to a non-trivial concept in the scaling lattice (i.e., other than *c2*), level-one papers only get the attribute *develops:c2*. The resulting scaled context and its lattice are given in Fig. 3. Three new concepts appear in the lattice:

- *c5=({c,d,i,j},{cites:c2})* – papers citing only papers of the RCF,
- *c6=({i,j},{cites:c0, cites:c2})* – papers citing only papers about lattice theory, i.e., in *c0*, as *c2* is redundant,
- *c7=({e,f,k,l},{develops:c2})* – developments of papers citing papers of the RCF.

Step 2 Given Lattice 2, richer than the initial one, narrow scaling is applied again upon *cites* and *develops*. While scaling upon *cites* does not add anything new, *develops* makes new incidences appear. First, as all the developed papers belong to the extent of*c5*, all the level-on papers also get the *develops:c5* attribute. Moreover, *k* and *l* also get *develops:c6*. The scaling yields a new RCF and its corresponding lattice, both given in Fig. 4.

					cites				develops		
		lt	mmi	se	2	0	3	4	2	5	6
gw99	a	x									
bm70	b	x									
are03	c				x						
dic96	d				x						
arePHD	e								x	x	
hucHDR	f								x	x	
bot91	g		x								
moo96	h			x							
val03	i				x	x					
leber02	j				x	x					
rouaPHD	k								x	x	x
leberHDR	l								x	x	x

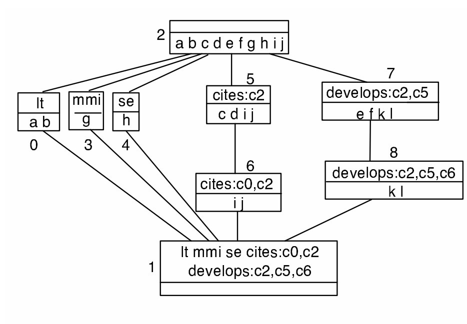

Fig. 4. Narrow scaling, step 2. **Left:** The scaled context; **Right:** Lattice 3.

The only new abstraction discovered at this stage is *c8* comprising publications that develop papers citing only papers about lattice theory. Step four terminates the analysis process, as no new concepts will be produced by further scaling. The interpretations of the formal concepts from the final lattice and their respective translations into DL are provided in Table 1.

5 Wide scaling-based RCA

The trace of the process with a wide scaling scheme starts immediately after the basic step of lattice construction on the unscaled context (see Fig. 2).

Step 1 When the inital lattice is used to scale upon *cites*, only the descriptions of papers c, d, i, j evolve. Hence all level-two papers get the attributes *cites:c0* and *cites:c2*, whereas c gets *cites:c3* as well, and d *cites:c4*. The result is to be seen in the scaled context in Fig. 5.

Applying wide scaling to *develops* and the initial lattice yields the same results as in the identical step of the narrow scaling-based process. Thus, *c2* being the only one whose extent comprises developed papers, all level-one papers get the attribute *develops:c2*. This yields the RCF depicted in Fig. 5 together with its lattice. The newly constructed concepts *c5*, *c6*, and *c7* represent, respectively, papers citing at least one paper about lattice theory,

Id	textual interpretation	translation into DL
$c0$	papers on lattice theory	AboutLatticeTheory
$c1$	papers having all properties	$\bot$
$c2$	all papers of the dataset	Paper
$c3$	papers on man machine interface	AboutManMachineInterface
$c4$	papers on software engineering	AboutSoftwareEngineering
$c5$	papers citing papers of the dataset	$C5 \equiv \exists$cites.$\top \sqcap \forall$cites.Paper
$c6$	papers citing only papers on lattice theory	$C6 \equiv \exists$cites.$\top \sqcap \forall$cites.AboutLatticeTheory
$c7$	papers developing only papers that cite only papers of the dataset	$\exists$develops.$\top \sqcap \forall$develops.$C5$
$c8$	papers developing only papers that cite only papers on lattice theory	$\exists$develops.$\top \sqcap \forall$develops.$C6$

Table 1. Narrow scaling. Interpretation of the mined concepts.

papers citing at least one paper both about man machine interface and lattice theory, and papers citing at least one paper both about software engineering and lattice theory. Furthermore, *c8* represents papers that develop at least one paper of the experiment.

		lt	mmi	se	cites 2	cites 0	cites 3	cites 4	develops 2
gw99	a	x							
bm70	b	x							
are03	c				x	x	x		
dic96	d				x	x		x	
arePHD	e								x
hucHDR	f								x
bot91	g		x						
moo96	h			x					
val03	i				x	x			
leber02	j				x	x			
rouaPHD	k								x
leberHDR	l								x

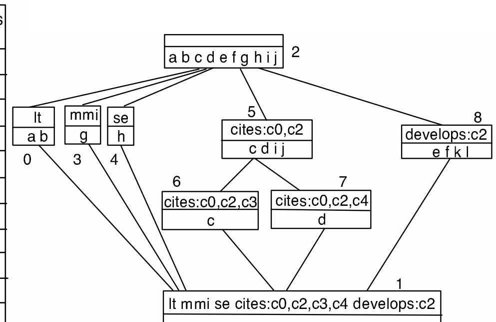

Fig. 5. Wide scaling. **Left:** RCF 2; **Right:** Lattice 2 on papers.

Step 2 Applying wide scaling to *cites* and the lattice of Fig. 5 does not bring any new incidence pair to the context. In contrast, scaling upon *develops* creates new attributes out of the concepts discovered at the previous step, i.e., *c5* to *c8*, and hence abstractions. Thus, all level-one papers get the *develops:c5* attribute as the extent of *c5* comprises all level-two concepts. In addition, e gets *develops:c6* and f *develops:c7*. The RCF of step 2 is drawn in Fig. 6 together with its lattice.

					cites				develops			
		lt	mmi	se	2	0	3	4	2	5	6	7
gw99	a	x										
bm70	b	x										
are03	c				x	x	x					
dic96	d				x	x		x				
arePHD	e								x	x	x	
hucHDR	f								x	x		x
bot91	g		x									
moo96	h			x								
val03	i				x	x						
leber02	j				x	x						
rouaPHD	k								x	x		
leberHDR	l								x	x		

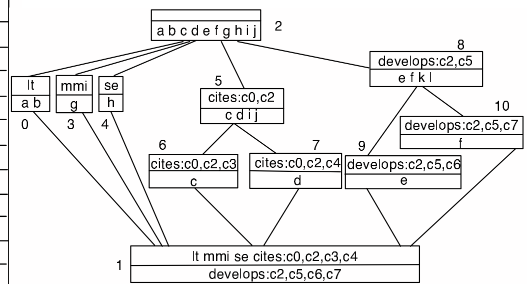

Fig. 6. Wide scaling. **Left:** RCF 3; **Right:** Lattice 3 on papers.

The newly formed concepts *c9* and *c10* represent papers that develop papers citing: papers about man machine interface and lattice theory and papers about software engineering and lattice theory, respectively. Moreover, the already existing concept *c8* gets more focused as it turns out to represent papers that develop papers citing work on lattice theory. The interpretation and the translation into a DL format of all the concepts from the final lattice is presented in Table 2.

Id	textual interpretation	translation into DL
$c5$	papers citing one+ paper on lattice theory	∃cites.AboutLatticeTheory
$c6$	papers citing one+ paper on lattice theory and one+ on man machine interface	∃cites.AboutLatticeTheory ⊓ ∃cites.AboutManMachineInterface
$c7$	papers citing one+ paper on lattice theory and one+ paper on software engineering	∃cites.AboutLatticeTheory ⊓ ∃cites.AboutSoftwareEngineering
$c8$	papers developing one+ paper that cites one+ paper on lattice theory	∃develops.∃cites.AboutLatticeTheory
$c9$	papers developing one+ paper that cites one+ paper on both lattice theory and man machine interface	∃develops.∃cites.(AboutLatticeTheory ⊓ AboutManMachineInterface)
$c10$	papers developing one+ paper that cites one+ paper on both lattice theory and software engineering	∃develops.∃cites.(AboutLatticeTheory ⊓ AboutSoftwareEngineering)

Table 2. Wide scaling: interpretation of concepts (only unseen in Table 1).

6 Conclusion

The RCA framework illustrated here is a first step towards the complete interoperability between data mining and KR tools. Indeed, its input is fully compatible with the standard data models, e.g., the relational one, while its results are easily expressible in terms of a DL language. Therefore, the knowledge mined from the input data is directly available for reasoning and problem-solving.

Many issues with RCA are yet to be tackled: First, the scalability is still an open issue, since the size of lattices grows rapidly w.r.t. the growth of relations between contexts. Various tracks for preventing combinatorial explosion are currently explored, e.g. using reduced structures such as iceberg lattices or Galois sub-hierarchies. Next, algorithmic aspects are among primary concerns. For instance, efficiency could be further improved by replacing construction from scratch by incremental lattice maintenance. Finally, we are currently studying further scaling policies, e.g., the quantified existential restrictions providing upper/lower limits of the number of links to lay in a concept.

References

BAADER, F., CALVANESE, D., MCGUINNESS, D., NARDI, D., and PATEL-SCHNEIDER, P. (Eds.) (2003), *The Description Logic Handbook*, Cambridge University Press.

BRACHMAN, R. J. and ANAND, T. The Process of Knowledge Discovery in Databases. In: U. Fayyad, G. Piatetsky-Shapiro, P. Smyth, and R. Uthurusamy (Eds): *Advances in Knowledge Discovery and Data Mining*. AAAI/MIT Press, 37–57.

DIDAY, E. (1998), Symbolic Data Analysis: a Mathematical Framework and Tool for Data Mining. In: A. Rizzi, M. Vichi, and H.-H. Bock (Eds.): *Advances in Data Science and Classification*. Springer, Berlin, 409–416.

GANTER, B. and WILLE, R. (1999), *Formal Concept Analysis, Mathematical Foundations*: Springer, Berlin.

HUCHARD, M., NAPOLI, A., ROUANE H., M., andVALTCHEV, P. (2007), A proposal for combining formal concept analysis and description logics. In: S. Kuznetsov and S. Schmidt (Eds.): *Proc. of the 5th Intl. Conf. on Formal Concept Analysis*. Springer, Berlin.

POLAILLON, G. (1998), *Organisation et interprétation par les treillis de Galois de données de type multivalué, intervalle ou histogramme*, Thèse de doctorat, Université Paris IX Dauphine.

PREDIGER, S. and STUMME, G. (1999), Theory-driven Logical Scaling: Conceptual Information Systems Meet Description Logics. In: E. Franconi and M. Kifer (Eds.): *Proc. of the 6th Intl. Workshop on Knowledge Representation meets Databases*, Linköping, Sweden, 46–49.

PREDIGER, S. and WILLE, R. (1999), The Lattice of Concept Graphs of a Relationally Scaled Context. In: W. Tepfenhart and W. Cyre (Eds) *Proc. of the 7th Intl. Conf. on Conceptual Structures*, Springer, Berlin. 401–414.