A Database Browser based on Pattern Concepts

Jens Kötters Heinz W. Schmidt [1]

[1]Computer Science & IT, RMIT University, Melbourne, Australia

Workshop "Formal Concept Analysis meets Information Retrieval", 2013

Part I

A digest of the paper

Kötters, J.: Concept Lattices of a Relational Structure. 

Example Database

female
c0
Anne
Dora
Emily

male
c0
Bob
Chris

Parent	
c0	c1
Anne	Bob
Anne	Chris
Bob	Dora
Bob	Emily

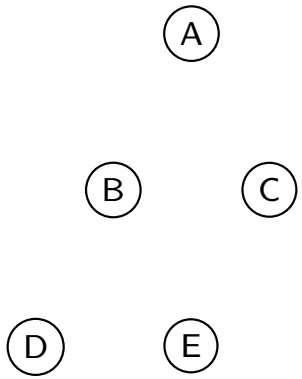
We want to build concepts from database content.
Consider an example database with three tables.

Example Database

female
c0
Anne
Dora
Emily

male
c0
Bob
Chris

Parent	
c0	c1
Anne	Bob
Anne	Chris
Bob	Dora
Bob	Emily



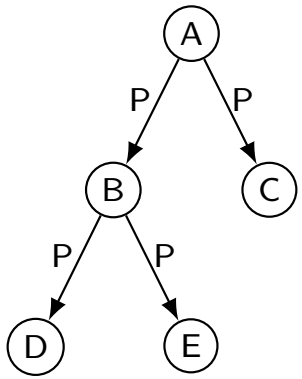
The database is visualized as a graph.
Nodes represent objects (the table entries).

Example Database

female
c0
Anne
Dora
Emily

male
c0
Bob
Chris

Parent	
c0	c1
Anne	Bob
Anne	Chris
Bob	Dora
Bob	Emily



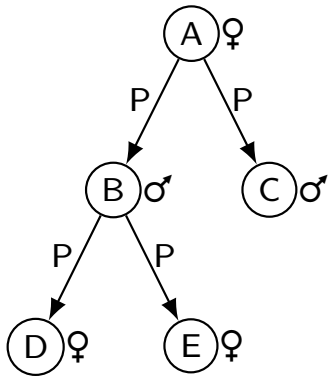
Each n-column table states an n-ary relation.
The parent relation is represented by P-edges.

Example Database

female
c0
Anne
Dora
Emily

male
c0
Bob
Chris

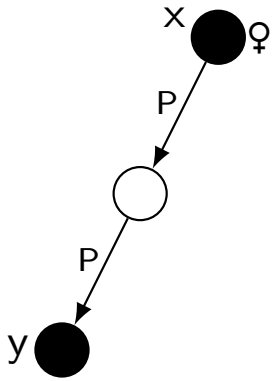
Parent	
c0	c1
Anne	Bob
Anne	Chris
Bob	Dora
Bob	Emily



Unary relations are represented by node labels. Formally, the graph is a structure **D** of signature $\Sigma := \{P, ♂, ♀\}$.

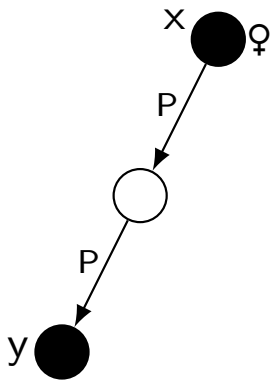
Example Query
(asking for the grandmother relation)

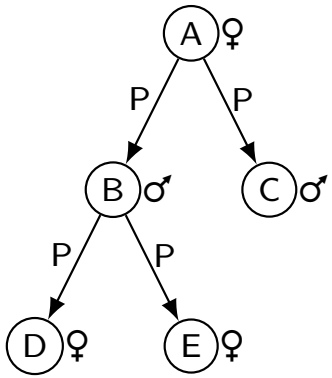
Graphs with black nodes represent conjunctive queries; black nodes indicate distinguished variables (here: x and y).



Example Query
(asking for the grandmother relation)

Queries are thus patterns, to be matched against **D**.
(Formally, a match is a Σ-structure homomorphism.)

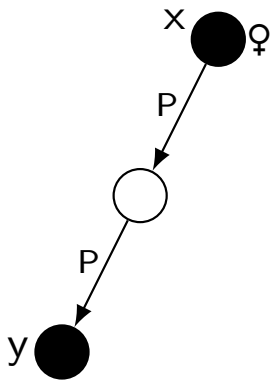


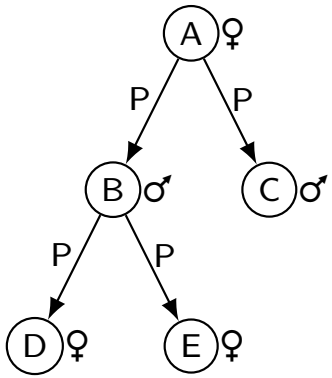


Example Query
(asking for the grandmother relation)

The black nodes' targets constitute the actual solutions.
Each solution is a row in the result table.

x	y
A	D
A	E





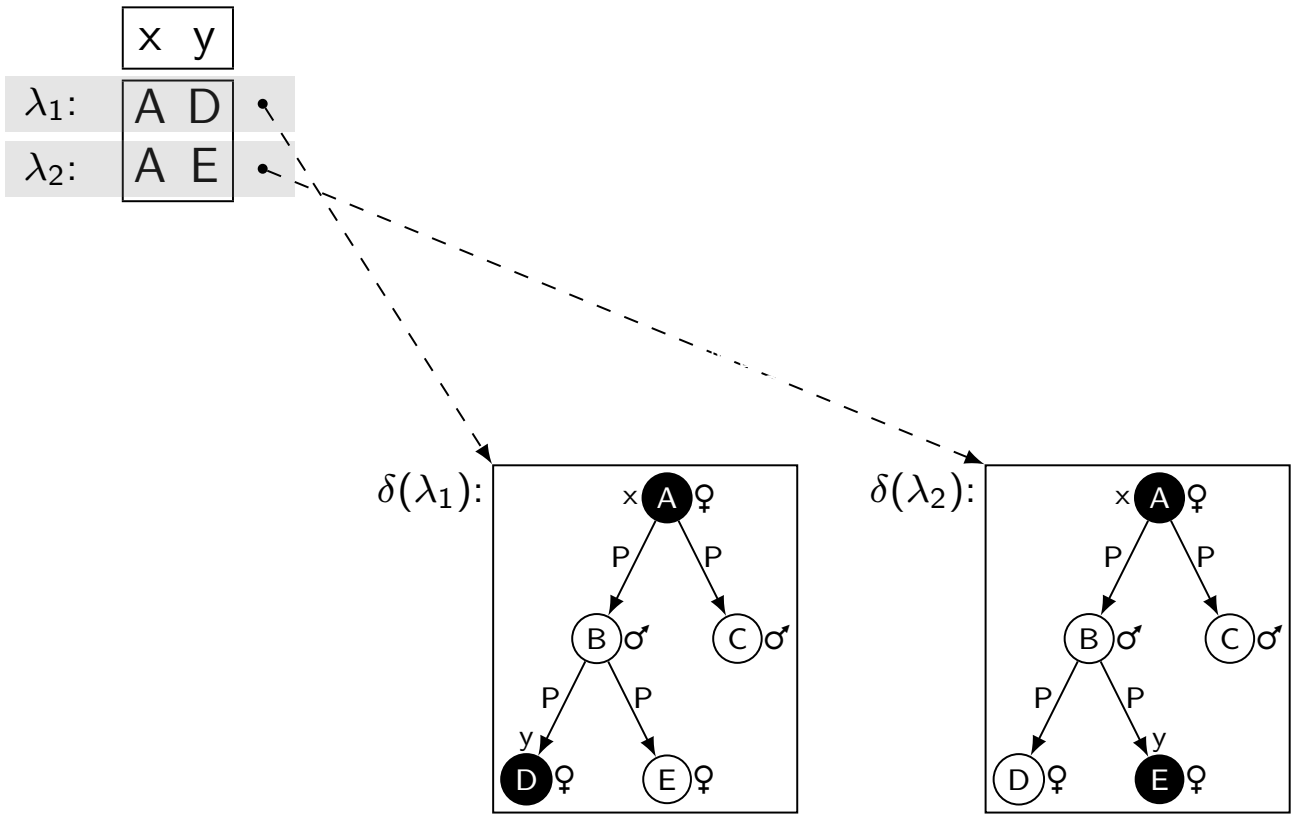
Existence of "the" Most Specific Query (MSQ)

A concept is a result table (*extent*), along with a most specific query producing it (*intent*). How to find the query?

x	y
A	D
A	E

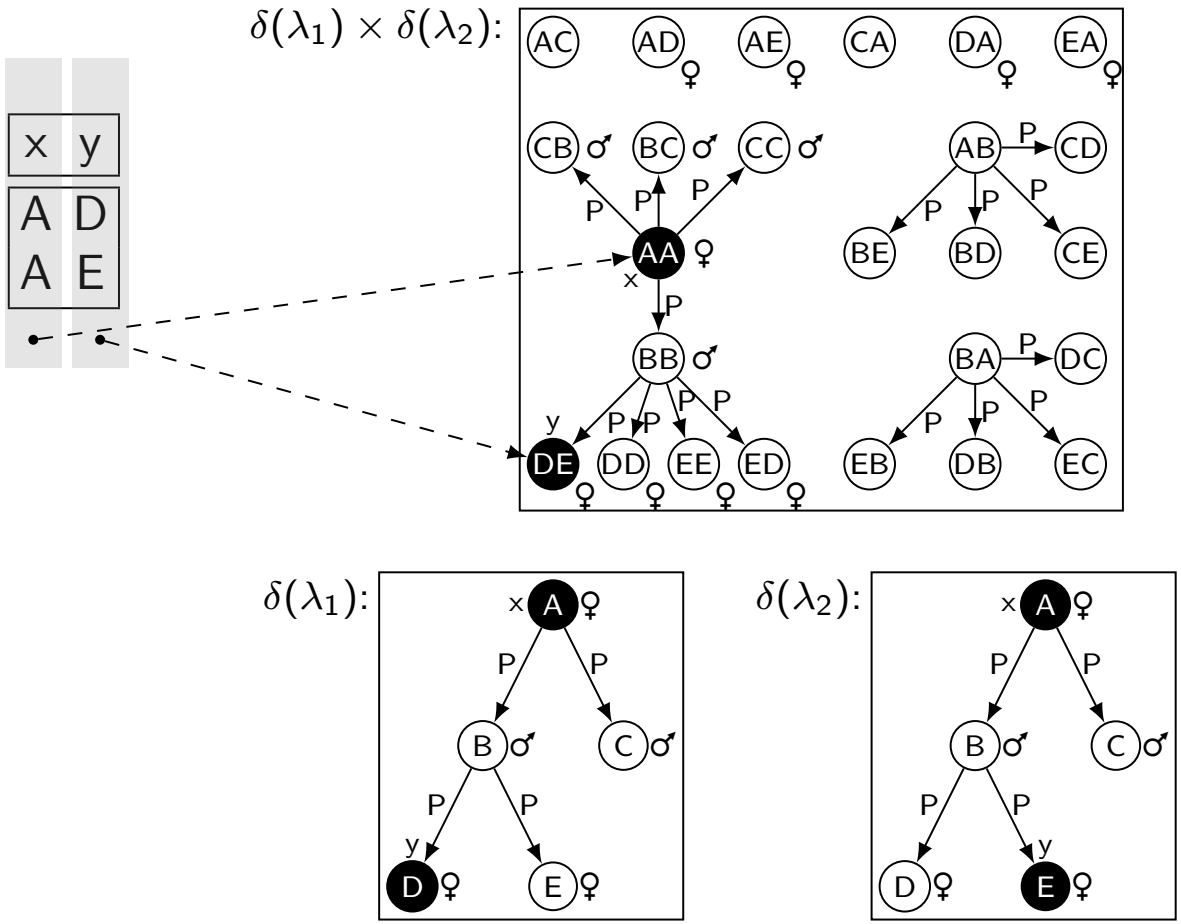
Existence of "the" Most Specific Query (MSQ)

The description of a row λ is "**D** regarded as a pattern" (with the row's objects as distinguished variables).



Existence of "the" Most Specific Query (MSQ)

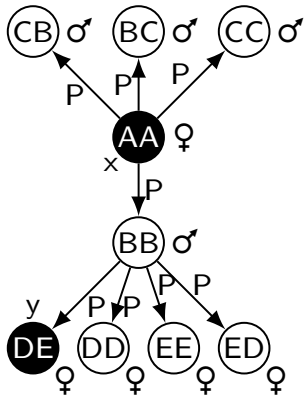
The product of structures [Gr08] realizes an MSQ.
Distinguished variables are given by the table's columns.



Example Concept

The components with black nodes form the actual pattern.
A connected pattern describes a concept.

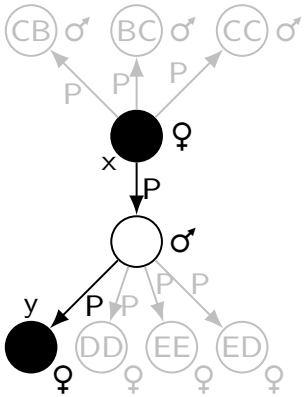
x	y
A	D
A	E



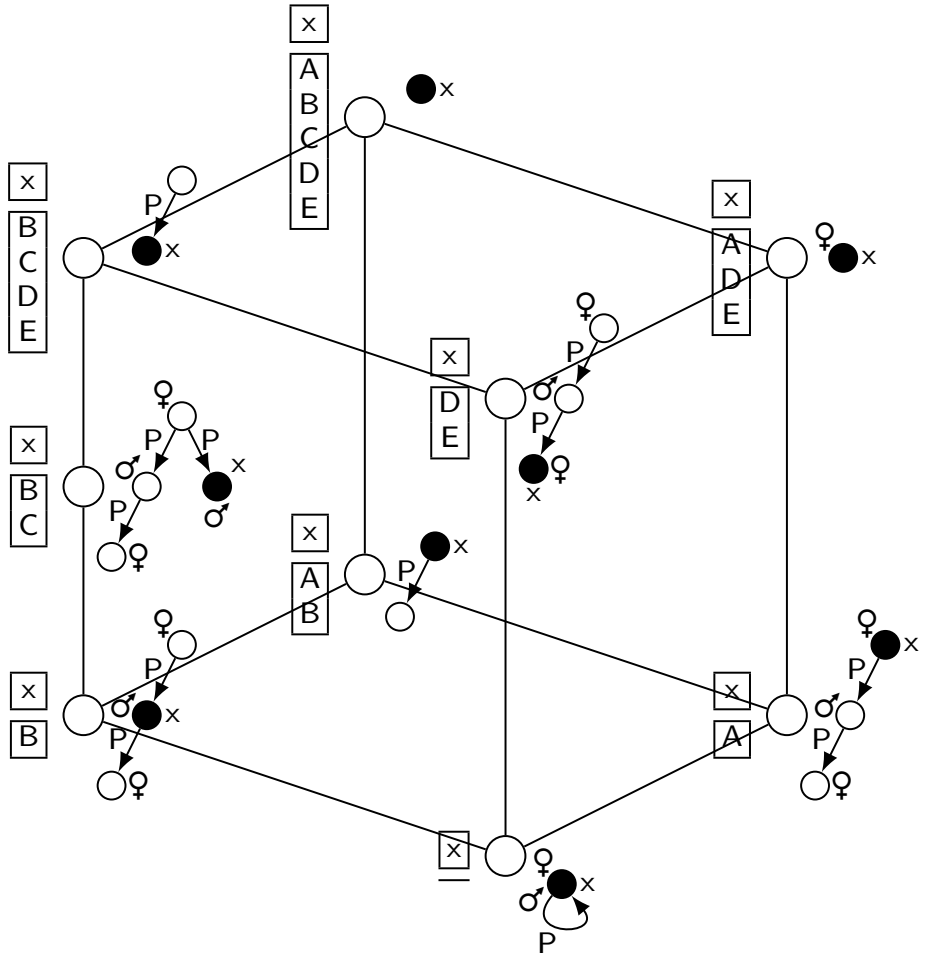
Example Concept

A minimal equivalent subpattern can be obtained by folding.
It is unique up to isomorphism.

x	y
A	D
A	E



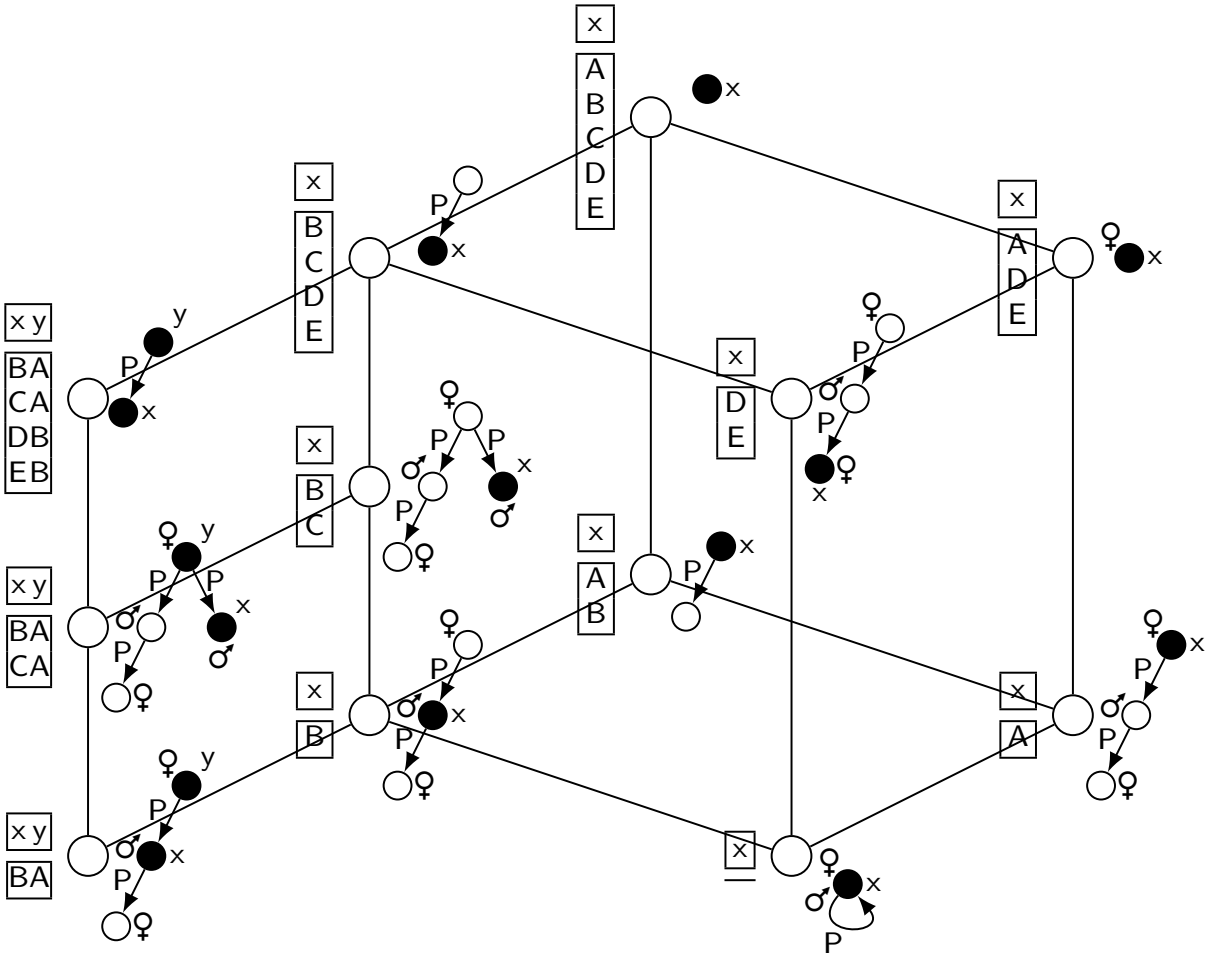
Concept Lattice $\mathfrak{C}_{\mathbf{D}}[\{x\}]$



$\mathfrak{C}_{\mathbf{D}}[X]$ is the set of concepts expressed using the set X of distinguished variables.

Concept Lattice $\mathfrak{C}_{\mathbf{D}}[\{x\}]$

... and a few concepts from $\mathfrak{C}_{\mathbf{D}}[\{x, y\}]$



The join defines the table order: $T_1 \leq T_2 \Leftrightarrow T_1 = T_1 \bowtie T_2$.

Remark: The ordered set $(\mathfrak{C}_{\mathbf{D}}, \leq)$ of concepts is not a lattice!

Theorem

Dropping "pattern connectedness" yields the completion $\mathfrak{L}_\mathbf{D}$.

Theorem

The ordered set $(\mathfrak{L}_\mathcal{D}, \leq)$ is a complete lattice. Infimum and supremum are given by

$$\bigwedge_{i\in I}(T_i, W_i) = \left(\bigbowtie_{i\in I} T_i, (\coprod_{i\in I} W_i)''\right) , \tag{1}$$

$$\bigvee_{i\in I}(T_i, W_i) = \left((\bigtimes_{i\in I} T_i)'', (\prod_{i\in I} W_i)''\right) . \tag{2}$$

For all $X \subseteq \mathrm{Var}$, the suborders $(\mathfrak{L}_\mathcal{D}[X] \cup \{\top\}, \leq)$ and $(\mathfrak{C}_\mathcal{D}[X] \cup \{\top\}, \leq)$, where $\top$ denotes the maximum of $(\mathfrak{L}_\mathcal{D}, \leq)$, are $\bigwedge$-sublattices of $(\mathfrak{L}_\mathcal{D}, \leq)$.

Part II

Pattern Structures

Both $\mathfrak{L}_{\mathbf{D}}$ and lattices $\mathfrak{L}_{\mathbf{D}}[X]$ can be obtained from suitable pattern structures [GK01]. For details see our paper.

Extensions

female
c0
Anne
Dora
Emily

male
c0
Bob
Chris

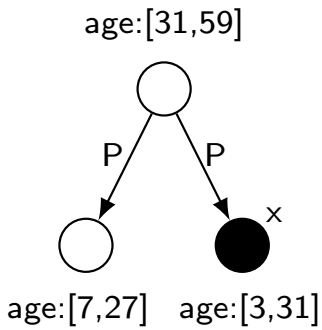
Parent	
c0	c1
Anne	Bob
Anne	Chris
Bob	Dora
Bob	Emily

age	
c0	c1
Anne	59
Bob	31
Chris	27
Dora	7
Emily	3

META		
table	column	type
Parent	c0	person
Parent	c1	person
male	c0	person
female	c0	person
age	c0	person
age	c1	number

The approach can be generalized to many-sorted signatures and structures to account for column types.

Extensions



"Augmented structures" account for simple type-specific signature extensions.

Browsing Session


```
00| DATABASE BROWSER (press 'h' for help)
01| Concept#0>top
02| Concept#0>intent
03| x :  age(x,n0) AND 3<=n0<=59
04| Concept#0>specialize
05| Concept#1>intent
06| x :  age(x,n0) AND female(x) AND 3<=n0<=59
07| Concept#1>extent
08|
09| x
10| -----
11| Anne
12| Dora
13| Emily
14|
15| Concept#1>generalize
16| Concept#0>specialize
17| Concept#3>intent
18| x :  age(p0,n0) AND age(p4,n4) AND age(x,n5) AND parent(p0,p4) AND parent(p0,x)
19| AND 31<=n0<=59 AND 7<=n4<=27 AND 3<=n5<=31
20|
21| Concept#3>specialize
```

Browsing Session (continued)


```
22| Concept#14>intent
23| y,x :  age(y,n0) AND age(p3,n3) AND age(x,n2) AND parent(y,p3) AND parent(y,x)
24| AND 31<=n0<=59 AND 7<=n3<=27 AND 3<=n2<=31
25|
26| Concept#14>extent
27|
28| y | x
29| ---------
30| Anne| Bob
31| Anne| Chris
32| Bob | Dora
33| Bob | Emily
34|
35| Concept#14>specialize
36| ->Concept#46***
37| ->Concept#18
38|
39| ------------------------------------------------------------------
40| y,x :  age(y,n0) AND age(p3,n3) AND age(x,n4) AND parent(y,p3) AND
41| parent(y,x) AND 31<=n0<=59 AND 7<=n3<=27 AND 3<=n4<=27
```

Numbers of Concepts

#(concepts)	DB relations	+constants	+numeric comparison
x:person	9	12	18
x,y:person	26	59	85

The table shows the number of computed concepts for one or two variables of `person` type, depending on query signature.

Selected References

Ganter, B., Kuznetsov, S.O.:
Pattern structures and their projections.
In: Delugach, H.S., Stumme, G. (eds.) Proceedings of ICCS 2001. LNCS, vol. 2120, pp. 129–142. Springer (2001)

Grätzer, G.:
Universal algebra.
Springer (2008)

Kötters, J.:
Concept Lattices of a Relational Structure.
In: Pfeiffer, H.D., Ignatov, D.I., Poelmans, J., Gadiraju, N. (eds.) Proceedings of ICCS 2013. LNCS, vol. 7735, pp. 301–310. Springer (2013)