The theory and practice of coupling formal concept analysis to relational databases

Jens Kötters Peter W. Eklund [†]

[†]School of Information Technology, Deakin University, Geelong, Australia

FCA4AI 2018
Stockholm, 13 July 2018

Background Theory

FCA + Database Theory

Previous work [Koe13] relates FCA with database theory. A table of analogies:

Standard FCA	FCA + Database Theory
Formal context	Relational Structure [Koe13], Power context family [Koe16]
Set of Objects	Table
Set of Attributes	Conjunctive query
Concept lattice	Conjunctive-query lattice

Lattices of n-ary concepts

The conjunctive-query lattice can be decomposed into sublattices $\mathfrak{L}[\{x_1, \ldots, x_n\}]$ of n-ary concepts described by variables $x_1, \ldots, x_n$. All sublattices of n-ary concepts are isomorphic (irrespective of variable names), so we can speak of *the* lattice of n-ary concepts. The extents are n-ary relations.

The lattice $\mathfrak{C}[\{x_1, \ldots, x_n\}]$ contains the concepts of $\mathfrak{L}[\{x_1, \ldots, x_n\}]$ where intents correspond to connected graphs.

Database Scaling

Example: Literature Database

Book

title	author	publication_date
Alice in Wonderland	1	1865-11-26
To the Lighthouse	2	1927-05-05
The Hitchhiker's Guide to the Galaxy	3	1979-10-12
Trigger Warning	4	2015-02-03
Harry Potter and the Deathly Hallows	5	2007-07-21
The Casual Vacancy	5	2012-09-27
The Shining	6	1977-01-28
Doctor Sleep	6	2013-09-24
The Da Vinci Code	7	2003-03-18
Inferno	7	2013-03-14

Author

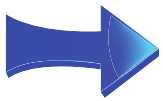
id	first_name	last_name	nationality	date_of_birth
1	Lewis	Carroll	British	1832-01-27
2	Virginia	Woolf	British	1882-01-25
3	Douglas	Adams	British	1952-03-11
4	Neil	Gaiman	British	1960-11-10
5	J. K.	Rowling	British	1965-07-31
6	Stephen	King	American	1947-09-21
7	Dan	Brown	American	1964-06-22

Conceptual scaling of a many-valued context (1)

Author table

id	first_name	last_name	nationality	date_of_birth
1	Lewis	Carroll	British	1832-01-27
2	Virginia	Woolf	British	1882-01-25
3	Douglas	Adams	British	1952-03-11
4	Neil	Gaiman	British	1960-11-10
5	J. K.	Rowling	British	1965-07-31
6	Stephen	King	American	1947-09-21
7	Dan	Brown	American	1964-06-22

DOB context



DOB	19C	20C	21C
Lewis Carroll	×		
Virginia Woolf	×		
Douglas Adams		×	
Neil Gaiman		×	
J. K. Rowling		×	
Stephen King		×	
Dan Brown		×	

Centuries scale

Centuries	19C	20C	21C
1832-01-27	×		
1865-11-26	×		
1882-01-25	×		
1927-05-05		×	
1947-09-21		×	
1952-03-11		×	
1960-11-10		×	
1964-06-22		×	
1965-07-31		×	
1977-01-28		×	
1979-10-12		×	
2003-03-18			×
2007-07-21			×
2012-09-27			×
2013-03-14			×
2013-09-24			×
2015-02-03			×

Conceptual scaling of a many-valued context (2)

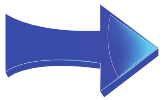
Author table

id	first_name	last_name	nationality	date_of_birth
1	Lewis	Carroll	British	1832-01-27
2	Virginia	Woolf	British	1882-01-25
3	Douglas	Adams	British	1952-03-11
4	Neil	Gaiman	British	1960-11-10
5	J. K.	Rowling	British	1965-07-31
6	Stephen	King	American	1947-09-21
7	Dan	Brown	American	1964-06-22

Nationalities scale

Nationalities	British	American	French	Russian
British	×			
American		×		
French			×	
Russian				×

Nationality context



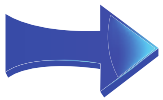
nat	British	American	French	Russian
Lewis Carroll	×			
Virginia Woolf	×			
Douglas Adams	×			
Neil Gaiman	×			
J. K. Rowling	×			
Stephen King		×		
Dan Brown		×		

We say that the Nationalities scale is bound to the nationality column (and the Centuries scale was bound to the date_of_birth column).

Conceptual scaling of a many-valued context (3)

Author table

id	first_name	last_name	nationality	date_of_birth
1	Lewis	Carroll	British	1832-01-27
2	Virginia	Woolf	British	1882-01-25
3	Douglas	Adams	British	1952-03-11
4	Neil	Gaiman	British	1960-11-10
5	J. K.	Rowling	British	1965-07-31
6	Stephen	King	American	1947-09-21
7	Dan	Brown	American	1964-06-22



Derived context

Authors	DOB:19C	DOB:20C	DOB:21C	nat:British	nat:American	nat:French	nat:Russian
Lewis Carroll	×			×			
Virginia Woolf	×			×			
Douglas Adams		×		×			
Neil Gaiman		×		×			
J. K. Rowling		×		×			
Stephen King		×			×		
Dan Brown		×			×		

We consider the subcontexts obtained from the scales as facets

Higher-arity scales: Foreign keys

Author table

id	first_name	last_name	nationality	date_of_birth
1	Lewis	Carroll	British	1832-01-27
2	Virginia	Woolf	British	1882-01-25
3	Douglas	Adams	British	1952-03-11
4	Neil	Gaiman	British	1960-11-10
5	J. K.	Rowling	British	1965-07-31
6	Stephen	King	American	1947-09-21
7	Dan	Brown	American	1964-06-22

Book

title	author	publication_date
Alice in Wonderland	1	1865-11-26
To the Lighthouse	2	1927-05-05
The Hitchhiker's Guide to the Galaxy	3	1979-10-12
Trigger Warning	4	2015-02-03
Harry Potter and the Deathly Hallows	5	2007-07-21
The Casual Vacancy	5	2012-09-27
The Shining	6	1977-01-28
Doctor Sleep	6	2013-09-24
The Da Vinci Code	7	2003-03-18
Inferno	7	2013-03-14

Equality scale

Equality	=
(1,1)	×
(1,2)	
(1,3)	
(2,1)	
(2,2)	×
(2,3)	
(3,1)	
(3,2)	
(3,3)	×
...	

Binary context "wrote"

wrote	id=author
(Lewis Carroll,Alice in Wonderland)	×
(Virginia Woolf,To the Lighthouse)	×
(Douglas Adams,Hitchhiker's Guide)	×
(Neil Gaiman,Trigger Warning)	×
(J.K. Rowling,Harry Potter 7)	×
(J.K. Rowling,The Casual Vacancy)	×
(Stephen King,The Shining)	×
(Stephen King,Doctor Sleep)	×
(Dan Brown,The Da Vinci Code)	×
(Dan Brown,Inferno)	×

The first parameter of the Equality scale is bound to *Author.id*, the second parameter is bound to *Book.author*.

Higher-arity scales: Measuring distance

Distance scales can be used to measure spatial distance between objects or the time span between events.
To measure at what age an author wrote a particular book, we instead use the foreign key condition as a domain expression (which defines the object set of a derived context) and use a distance scale (not the one below !!) on top of this.

Distance scale

Distance	$=0$	≤ 1	≤ 2
(1,1)	×	×	×
(1,2)		×	×
(1,3)			×
(2,1)		×	×
(2,2)	×	×	×
(2,3)		×	×
(3,1)			×
(3,2)		×	×
(3,3)	×	×	×

Binary context "wrote"

wrote	wrote	age$\leq$30	age$\leq$40	age$\leq$50
(Lewis Carroll, Alice in Wonderland)	×		×	×
(Virginia Woolf, To the Lighthouse)	×			×
(Douglas Adams, Hitchhiker's Guide)	×	×	×	×
(Neil Gaiman, Trigger Warning)	×			
(J. K. Rowling, Harry Potter 7)	×			×
(J. K. Rowling, The Casual Vacancy)	×			×
(Stephen King, The Shining)	×	×	×	×
(Stephen King, Doctor Sleep)	×			
(Dan Brown, The Da Vinci Code)	×		×	×
(Dan Brown, Inferno)	×			×

Power Context Family

The contexts for each facet can be assembled in a power context family.

0	sort: Author	sort: Book
Lewis Carroll	×	
Virginia Woolf	×	
Douglas Adams	×	
Neil Gaiman	×	
J. K. Rowling	×	
Stephen King	×	
Dan Brown	×	
Alice in Wonderland		×
To the Lighthouse		×
Hitchhiker's Guide		×
Harry Potter 7		×
The Casual Vacancy		×
Trigger Warning		×
The Shining		×
Doctor Sleep		×
The Da Vinci Code		×
Inferno		×

1	nationality: GB	nationality: USA	DOB: 19C	DOB: 20C	DOB: 21C	pubdate: 19C	pubdate: 20C	pubdate: 21C
Lewis Carroll	×		×					
Virginia Woolf	×		×					
Douglas Adams	×			×				
Neil Gaiman	×			×				
J. K. Rowling	×			×				
Stephen King		×		×				
Dan Brown		×		×				
Alice in Wonderland						×		
To the Lighthouse							×	
Hitchhiker's Guide							×	
Harry Potter 7								×
The Casual Vacancy								×
Trigger Warning								×
The Shining							×	
Doctor Sleep								×
The Da Vinci Code								×
Inferno								×

2	wrote: wrote	wrote: age≤30	wrote: age≤40	wrote: age≤50
(Lewis Carroll, Alice in Wonderland)	×		×	×
(Virginia Woolf, To the Lighthouse)	×			×
(Douglas Adams, Hitchhiker's Guide)	×	×	×	×
(Neil Gaiman, Trigger Warning)	×			
(J. K. Rowling, Harry Potter 7)	×			×
(J. K. Rowling, The Casual Vacancy)	×			×
(Stephen King, The Shining)	×	×	×	×
(Stephen King, Doctor Sleep)	×			
(Dan Brown, The Da Vinci Code)	×		×	×
(Dan Brown, Inferno)	×			×

Conjunctive Queries

Formalizations of Conjunctive Queries

In popular use:

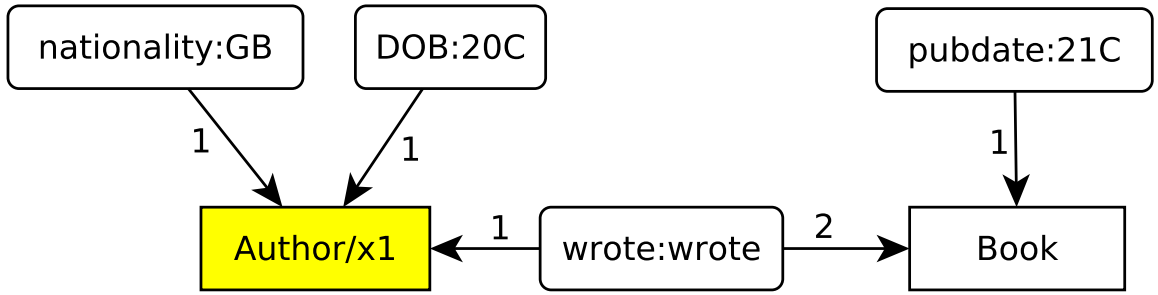
- Tableaux
- Logical Formulas
- Datalog Rules

Other formalizations in selected literature:

- Relational Structures [CM77]
- Windowed Relational Structures [Koe13]
- Windowed Power Context Families [Koe16]
- Windowed Intension Graphs [Koe16]

Windowed Intension Graph

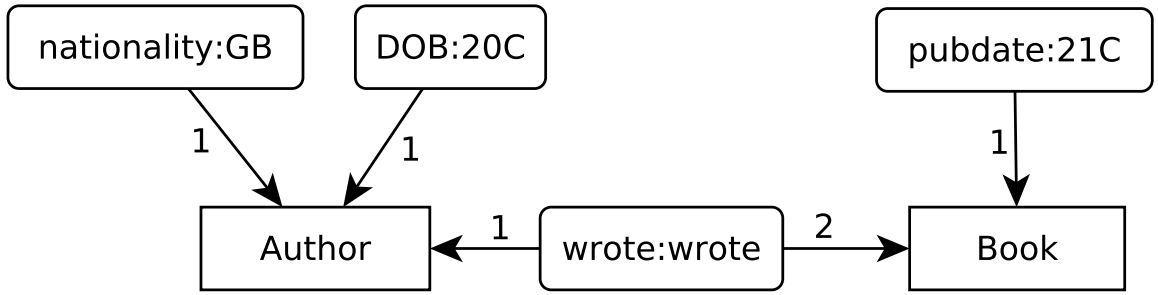
"20th-century-born British authors who published in the 21st century"



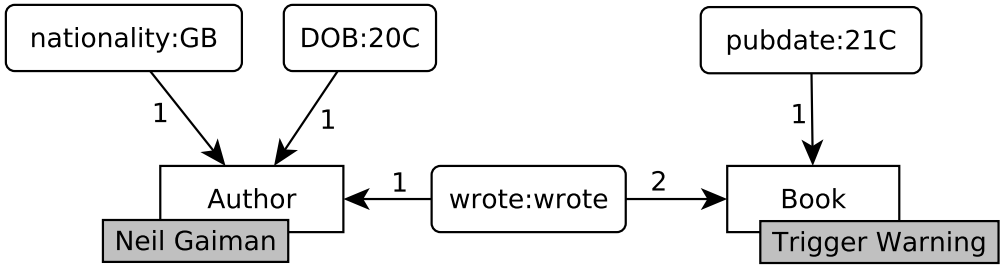
Terminology: object node, relation node, subject node, label, marker, window

Intension Graph

The underlying intension graph.



Solution



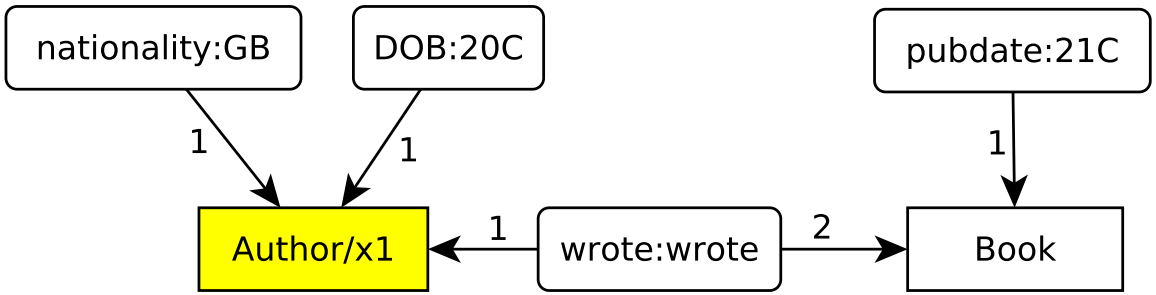
There are two more solutions in the below power context family.

0	sort: Author	sort: Book
Lewis Carroll	×	
Virginia Woolf	×	
Douglas Adams	×	
Neil Gaiman	×	
J. K. Rowling	×	
Stephen King	×	
Dan Brown	×	
Alice in Wonderland		×
To the Lighthouse		×
Hitchhiker's Guide		×
Harry Potter 7		×
The Casual Vacancy		×
Trigger Warning		×
The Shining		×
Doctor Sleep		×
The Da Vinci Code		×
Inferno		×

1	nationality: GB	nationality: USA	DOB: 19C	DOB: 20C	DOB: 21C	pubdate: 19C	pubdate: 20C	pubdate: 21C
Lewis Carroll	×		×					
Virginia Woolf	×		×					
Douglas Adams	×			×				
Neil Gaiman	×			×				
J. K. Rowling	×			×				
Stephen King		×		×				
Dan Brown		×		×				
Alice in Wonderland						×		
To the Lighthouse							×	
Hitchhiker's Guide							×	
Harry Potter 7								×
The Casual Vacancy								×
Trigger Warning								×
The Shining							×	
Doctor Sleep								×
The Da Vinci Code								×
Inferno								×

2	wrote: wrote	wrote: age≤30	wrote: age≤40	wrote: age≤50
(Lewis Carroll, Alice in Wonderland)	×		×	×
(Virginia Woolf, To the Lighthouse)	×			×
(Douglas Adams, Hitchhiker's Guide)	×	×	×	×
(Neil Gaiman, Trigger Warning)	×			
(J. K. Rowling, Harry Potter 7)	×			×
(J. K. Rowling, The Casual Vacancy)	×			×
(Stephen King, The Shining)	×	×	×	×
(Stephen King, Doctor Sleep)	×			
(Dan Brown, The Da Vinci Code)	×		×	×
(Dan Brown, Inferno)	×			×

Result Table (Concept Extension)

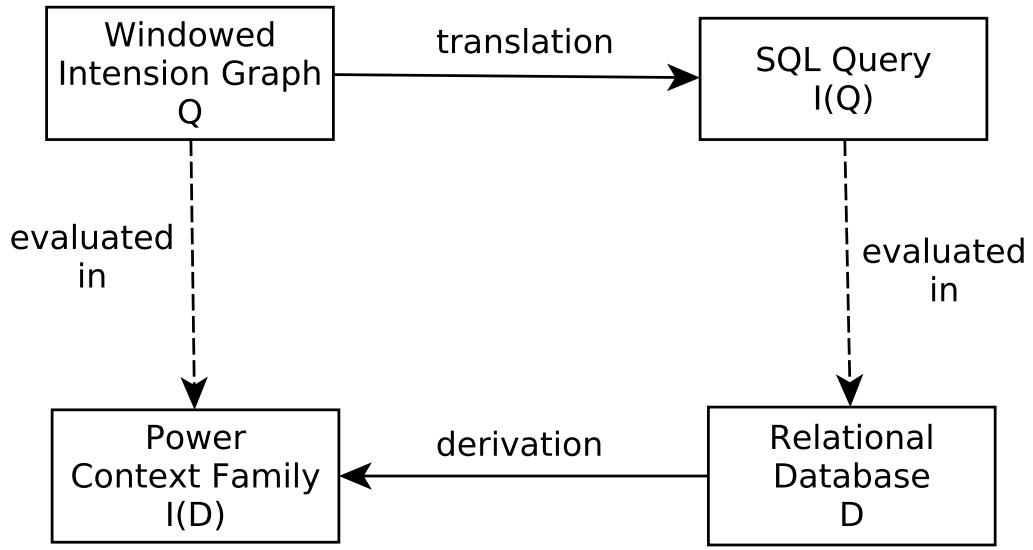


Result Table

x1
Neil Gaiman
J. K. Rowling

SQL Interpretation

Syntactic Interpretation



$$R_{I(D)}(Q) = R_D(I(Q))$$

Database Scales

Centuries	19C	20C	21C
1832-01-27	×		
1865-11-26	×		
1882-01-25	×		
1927-05-05		×	
1947-09-21		×	
1952-03-11		×	
1960-11-10		×	
1964-06-22		×	
1965-07-31		×	
1977-01-28		×	
1979-10-12		×	
2003-03-18			×
2007-07-21			×
2012-09-27			×
2013-03-14			×
2013-09-24			×
2015-02-03			×

A database scale assigns an SQL definition to each attribute. The corresponding scale context (left side) can be derived if so desired.

$$\sigma_{\text{Centuries}}(19\text{C}) \equiv z_1 \text{ BETWEEN "1800-01-01" AND "1899-12-31"}$$
$$\sigma_{\text{Centuries}}(20\text{C}) \equiv z_1 \text{ BETWEEN "1900-01-01" AND "1999-12-31"}$$
$$\sigma_{\text{Centuries}}(21\text{C}) \equiv z_1 \text{ BETWEEN "2000-01-01" AND "2099-12-31"}$$

Database Facets

DOB	19C	20C	21C
Lewis Carroll	×		
Virginia Woolf	×		
Douglas Adams		×	
Neil Gaiman		×	
J. K. Rowling		×	
Stephen King		×	
Dan Brown		×	

Similarly, a facet provides SQL definitions of its attributes. It is obtained by a variable substitution in the underlying scale's SQL definition, according to the binding. (here: $z_1 \rightarrow \mathrm{t}_1.\mathrm{date\_of\_birth}$)

$\Phi_{\mathrm{DOB}}(19\mathrm{C}) \equiv \mathrm{t}_1.\mathrm{date\_of\_birth}$ BETWEEN "1800-01-01" AND "1899-12-31"

$\Phi_{\mathrm{DOB}}(20\mathrm{C}) \equiv \mathrm{t}_1.\mathrm{date\_of\_birth}$ BETWEEN "1900-01-01" AND "1999-12-31"

$\Phi_{\mathrm{DOB}}(21\mathrm{C}) \equiv \mathrm{t}_1.\mathrm{date\_of\_birth}$ BETWEEN "2000-01-01" AND "2099-12-31"

Database Facets

pubdate	19C	20C	21C
Alice in Wonderland	×		
To the Lighthouse		×	
Hitchhiker's Guide		×	
Harry Potter 7			×
The Casual Vacancy			×
Trigger Warning			×
The Shining		×	
Doctor Sleep			×
The Da Vinci Code			×
Inferno			×

Thereby, a relation between values is translated into a relation between objects. The scales encode the actual logic; they should be generic and reusable.

$\Phi_{\text{pubdate}}(19C) \equiv t_1.\text{publication\_date BETWEEN "1800-01-01" AND "1899-12-31"}$

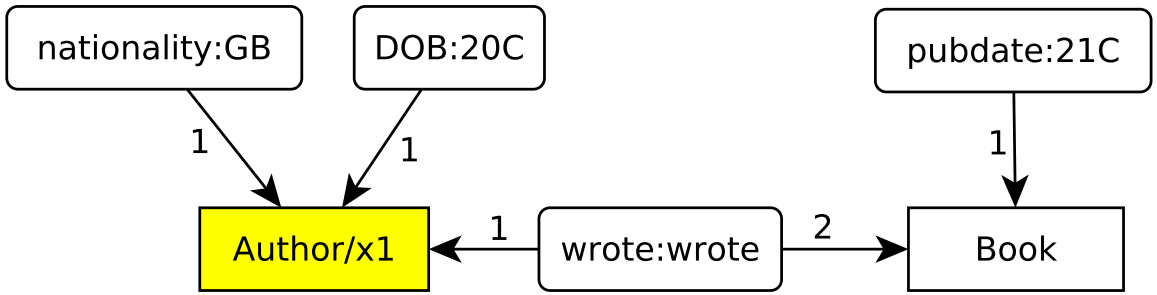
$\Phi_{\text{pubdate}}(20C) \equiv t_1.\text{publication\_date BETWEEN "1900-01-01" AND "1999-12-31"}$

$\Phi_{\text{pubdate}}(21C) \equiv t_1.\text{publication\_date BETWEEN "2000-01-01" AND "2099-12-31"}$

SQL Translation

$$\begin{aligned}
\text{SELECT DISTINCT } & \Omega_{\text{sort}(u_1)}(u_1) \text{ AS } x_1 \text{ , ..., } \\
& \Omega_{\text{sort}(u_m)}(u_m) \text{ AS } x_m \\
\text{FROM } & \text{sort}(v_1) \text{ AS } v_1 \text{ , ..., } \\
& \text{sort}(v_n) \text{ AS } v_n \\
\text{WHERE } & \Phi_{c_1}(a_1)(v_{11}, ..., v_{1n_1}) \text{ AND ...} \\
\text{AND } & \Phi_{c_k}(a_k)(v_{k1}, ..., v_{kn_k})
\end{aligned}$$

Example: SQL Translation



SELECT DISTINCT CONCAT(v1.first_name," ",v1.last_name) AS x_1

FROM Author AS v_1 ,

Book AS v_2

WHERE v_1.nationality="GB"

AND v_1.date_of_birth BETWEEN "1900-01-01" AND "1999-12-31"

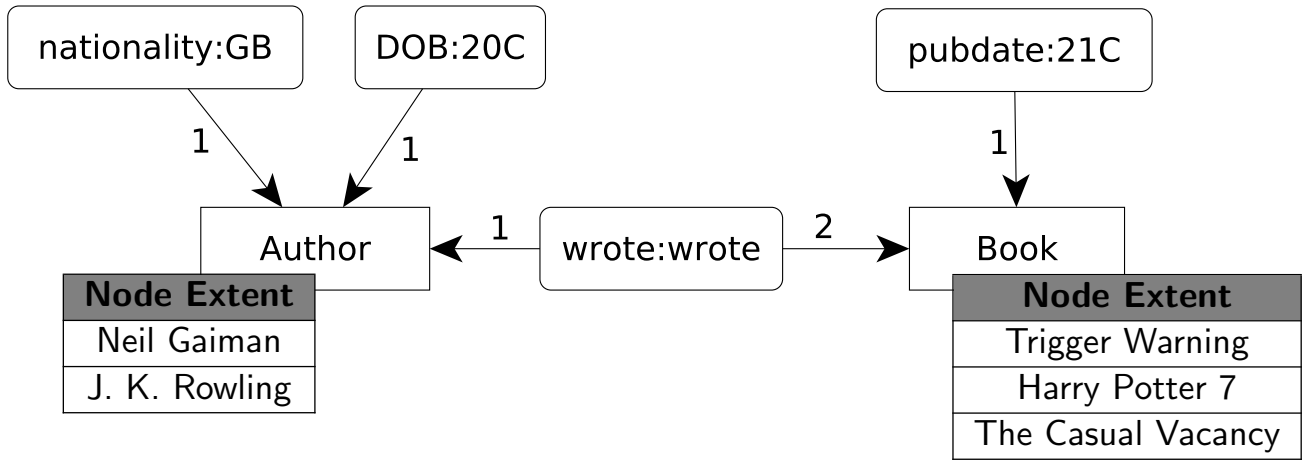
AND v_1.id = v_2.author

AND v_1.publication_date BETWEEN "2000-01-01" AND "2099-12-31"

Navigation

Projectional Concept Graph

In a projectional concept graph, each node is considered as a unary concept in a system of interrelated concepts. The node extent is a unary concept extent in the conjunctive-query lattice. However, we do not compute the graph closure (i.e. the intent in the conjunctive-query lattice).



Projectional Concept Graph

Definition

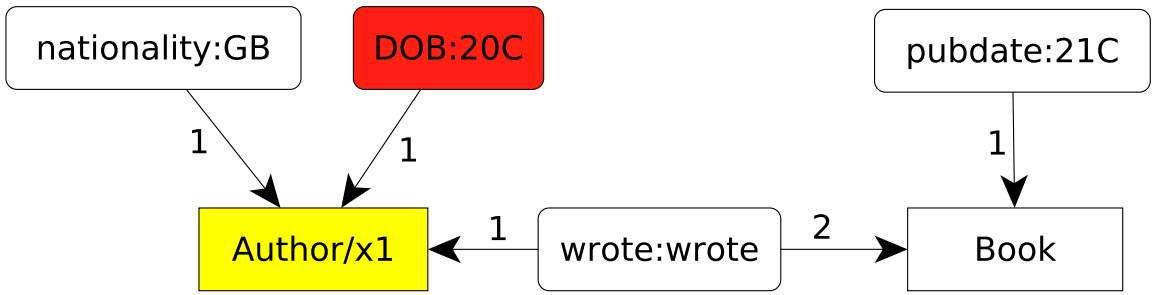
A projectional concept graph is a 5-tuple $(V, E, \nu, \kappa, \mathrm{ext}_{\vec{\mathbb{K}}})$ comprised of an intension graph $\mathcal{G} := (V, E, \kappa, \nu)$ and its extension map

$$\mathrm{ext}_{\vec{\mathbb{K}}}(v) := \{\ \varphi(v) \mid \varphi \in \mathcal{S}(\mathcal{G}, \vec{\mathbb{K}})\ \}$$

for a given power context family $\vec{\mathbb{K}}$ with $\mathcal{S}(\mathcal{G}, \vec{\mathbb{K}}) \neq \emptyset$. We call $\mathrm{ext}_{\vec{\mathbb{K}}}(v)$ the node extent of v.

Projectional Concept Graph

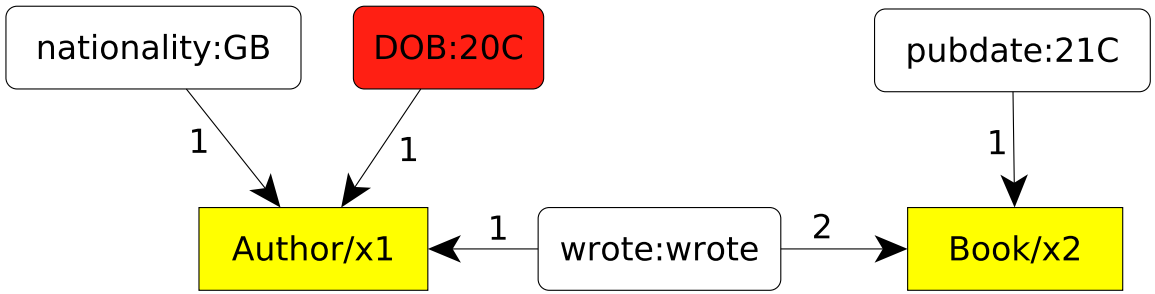
We envision relation nodes as controls in a user interface to show/hide associated value columns.



x1	x1.date_of_birth
Neil Gaiman	1960-11-10
J. K. Rowling	1965-07-31

Projectional Concept Graph

Showing only projections eliminates combinatorial explosion in result tables. But windows of size ≥ 2 are still supported, if the actual combinations are of interest.



x1	x1.date_of_birth	x2
Neil Gaiman	1960-11-10	Trigger Warning
J. K. Rowling	1965-07-31	Harry Potter 7
J. K. Rowling	1965-07-31	The Casual Vacancy

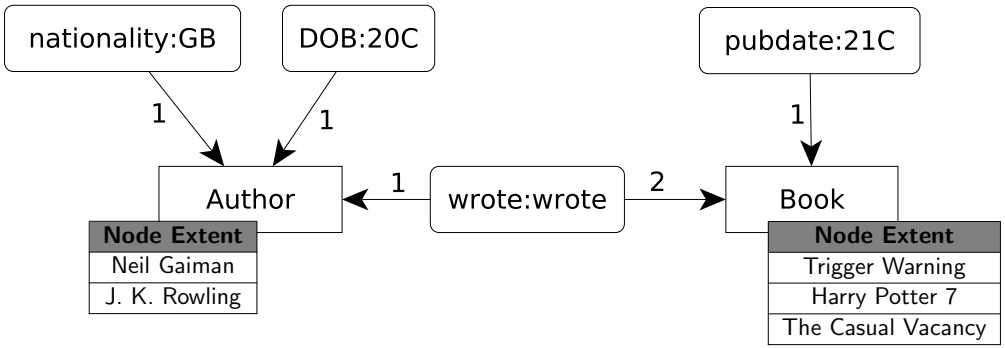
Refinement Triple

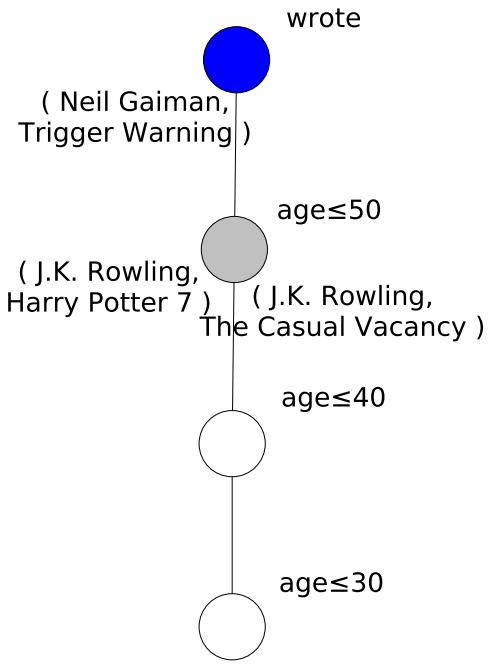
With each projectional concept graph, there is an associated refinement triple $(E^+, \kappa^+, \theta^+)$.

- E^+: Associates with each object node v a list $E^+(v)$ of facets. Each facet corresponds to a new relation node that can be connected to v.
- κ^+: Provides for each object or relation node u a list $\kappa^+(u)$ of scale intents (or equivalently, scale concepts), which can replace the current label $\kappa(u)$.
- θ^+ : A list of pairs of object nodes that can be merged.

Each refinement option leads to another projectional concept graph that at least one solution.

Example: Refinement Triple





The label refinements in $\kappa^+(u)$ correspond, for each facet, to concepts in the facet's concept lattice.

Summary

Summary

- Revisited: Relational scaling
- Application: Building query vocabulary around a relational database
- Application: Faceted navigation in power context families
- Proposal: A new class of concept graphs

Selected References I

[CM77] Ashok K. Chandra and Philip M. Merlin

Optimal implementation of conjunctive queries in relational databases.
In: Proceedings of the 9th annual ACM symposium on theory of computing, pp. 77–90 (1977)

[DE09] Jon R. Ducrou and Peter W. Eklund

Faceted document navigation using conceptual structures.
In: Conceptual Structures in Practice. Chapman & Hall/CRC (2009)

[GE99] Bernd Groh and Peter W. Eklund

Algorithms for creating relational power context families from conceptual graphs.
In: Proceedings of ICCS 1999. LNAI, vol. 1640, pp. 389–400. Springer (1999)

[Her02] Joachim Hereth

Relational scaling and databases.
In: Proceedings of ICCS 2002. LNAI, vol. 2393, pp. 62–76. Springer (2002)

[HRV02] Marianne Huchard, Cyril Roume and Petko Valtchev

When concepts point at other concepts: the case of UML diagram reconstruction.
In: Proceedings of FCAKDD 2002, pp. 32–43.

[Koe11] Jens Kötters

Object configuration browsing in relational databases.
In: Proceedings of ICFCA 2011. LNCS, vol. 6628, pp. 151–166. Springer (2011)

[Koe13] Jens Kötters

Concept lattices of a relational structure.
In: Proceedings of ICCS 2013. LNCS, vol. 7735, pp. 301–310. Springer (2013)

Selected References II

[Koe16] Jens Kötters

Intension graphs as patterns over power context families.
In: Proceedings of CLA 2016. CEUR Workshop Proceedings, vol. 1624 (2016)

[PW99] Susanne Prediger and Rudolf Wille

The lattice of concept graphs of a relationally scaled context.
In: Proceedings of ICCS 1999. LNAI, vol. 1640, pp. 401–414. Springer (1999)

[Pri00] Uta Priss

Lattice-based information retrieval.
In: Knowledge Organization, vol. 27(3), pp. 132–142 (2000)

[Wil97] Rudolf Wille

Conceptual Graphs and Formal Concept Analysis.

In: Proceedings of ICCS 1997. LNCS, vol. 1257, pp. 290–303. Springer (1997)