



HAL
open science

Analyse Formelle de Concepts : Une approche pour fouiller des ensembles de données multi-relationnels, et quelques applications au génie logiciel

Marianne Huchard

► **To cite this version:**

Marianne Huchard. Analyse Formelle de Concepts : Une approche pour fouiller des ensembles de données multi-relationnels, et quelques applications au génie logiciel. 2012. lirmm-00808686

HAL Id: lirmm-00808686

<https://hal-lirmm.ccsd.cnrs.fr/lirmm-00808686>

Submitted on 5 Jan 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Analyse Relationnelle de Concepts: Une approche pour fouiller des ensembles de données multi-relationnels, et quelques applications au Génie Logiciel

*Université de Montréal
Colloque du Département d'Informatique et de Recherche Opérationnelle*

Marianne Huchard

November 1, 2012

Brief presentation of FCA – Formal Concept Analysis

A methodology for:

- ▶ data analysis, data mining
- ▶ knowledge representation
- ▶ unsupervised learning






Roots:

- ▶ lattice theory, Galois correspondences (Birkhoff, 1940; Barbut & Monjardet, 1970)
- ▶ concept lattices (Wille, 1982)

Brief presentation of FCA – Formal Concept Analysis

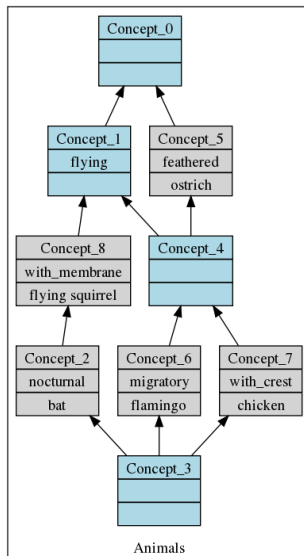
Contexts and concepts

- ▶ Handled data
 - ▶ entities with characteristics
 - ▶ provided with a Formal Context (a binary table)

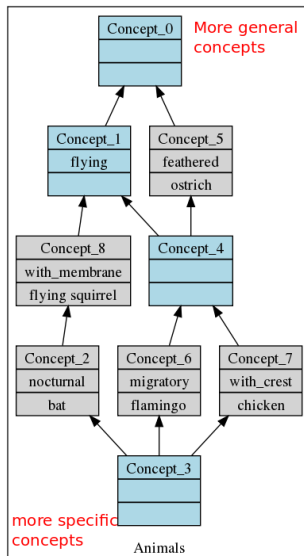
		flying	nocturnal	feathered	migratory	with_crest	with_membrane
	flying squirrel	×					×
	bat	×	×				×
	ostrich			×			
	flamingo	×		×	×		
	chicken	×		×		×	

- ▶ Concept : maximal group of entities sharing characteristics
- ▶ Concept lattice : concepts with a partial order relation

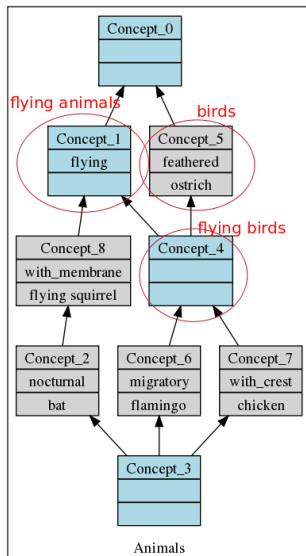
Brief presentation of FCA – Formal Concept Analysis



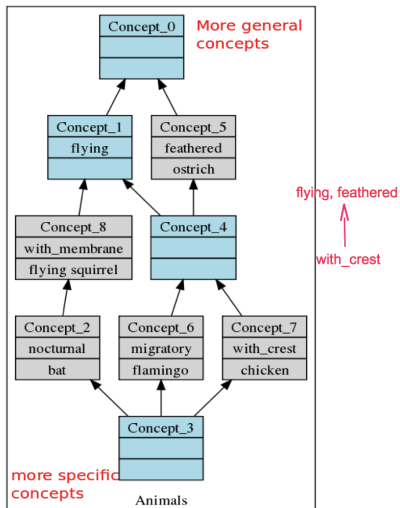
Brief presentation of FCA – Formal Concept Analysis



Brief presentation of FCA – Formal Concept Analysis



Brief presentation of FCA – Formal Concept Analysis



FCA and complex data

- ▶ many-valued contexts (integers, floats, terms, structures, symbolic objects, etc.) (Ganter et Wille, Polaillon, ...)
- ▶ fuzzy descriptions (Yahia et al., Belohlavek, ...)
- ▶ hierarchies on values (Godin et al., Carpineto et Romano, ...)
- ▶ logical description (Chaudron et al., Ferré et al., ...)
- ▶ graphs (Liquière, Prediger et Wille, ...)
- ▶ **linked objects** (Priss, Hacène-Rouane et al., ...)
- ▶ etc.

Relational Concept Analysis (RCA)

- ▶ Extends the purpose of FCA for taking into account object categories and links between objects
- ▶ Main principles:
 - ▶ a relational model based on the entity-relationship model
 - ▶ integrate relations between objects as *relational* attributes
 - ▶ iterative process
- ▶ RCA provides a set of interconnected lattices
- ▶ Produced structures can be represented as ontology concepts within a knowledge representation formalism such as description logics (DLs).

Joint work with:

A. Napoli, C. Roume, M. Rouane-Hacène, P. Valtchev

Relational Context Family (RCF)

A simple entity-relationship model to introduce RCA

Relational Context Family

- ▶ **object-attribute contexts**
 - ▶ Pizza
 - ▶ Ingredient
- ▶ **object-object context**
 - ▶ $\text{has-topping} \subseteq \text{Pizza} \times \text{Ingredient}$

Relational Context Family (RCF)

A RCF \mathcal{F} is a pair (K, R) with:

- ▶ K is a set of object-attribute contexts $K_i = (O_i, A_i, I_i)$
- ▶ R is a set of object-object contexts $R_j = (O_k, O_l, I_j)$,
 - ▶ (O_k, O_l) are the object sets of formal contexts $(K_k, K_l) \in K^2$
 - ▶ $I_j \subseteq O_k \times O_l$
 - ▶ K_k is the *source/domain context*, K_l is the *target/range context*.
 - ▶ we may have $K_k = K_l$.

Relational Context Family (RCF) / object-attributes contexts

Pizza	thin	thick	calzone
okonomi			×
alberginia		×	
margherita	×		
languedoc	×		
four-cheeses	×		
three-cheeses	×		
frutti-di-mare	×		
quebec		×	
regina	×		
hawai		×	
lorraine	×		
kebab			×

Ingredient	fruit-vegetable	meat	fish	dairy	cereal-leguminous	veg-oil
tomato-sauce	×					
cream				×		
tomato	×					
basilic	×					
olive	×					
olive oil						×
soy	×					
mushroom	×					
eggplant	×					
onion	×					
pepper	×					
anas	×					
mozza				×		
goat-cheese				×		
emmental				×		
fourme-ambert				×		
squid			×			
shrimp			×			
mussels			×			
ham		×				

Relational Context Family (RCF) / object-object context / part 1

	tomato-sauce	cream	tomato	basilic	olive	olive oil	soy	mushroom	eggplant	onion	pepper	anas
has-topping												
okonomi	x					x	x	x				
alberginia	x					x	x		x	x		
margherita	x		x	x	x	x						
languedoc	x		x	x	x	x				x	x	
four-cheeses		x										
three-cheeses		x										
frutti-di-mare	x				x	x						
quebec	x											
regina	x							x				
hawai	x											x
lorraine		x								x		
kebab	x		x		x					x		

Relational Context Family (RCF) / object-object context / part 2

	mozza	goat-cheese	emmental	fourme-ambert	squid	shrimp	mussels	ham	bacon	chicken	maple-sirup	corn
has-topping												
okonomi												
alberginia												
margherita	x											
languedoc	x											
four-cheeses	x	x	x	x								
three-cheeses	x	x	x									
frutti-di-mare	x				x	x	x					
quebec	x							x			x	x
regina	x								x			
hawai	x							x				
lorraine			x						x			
kebab			x							x		

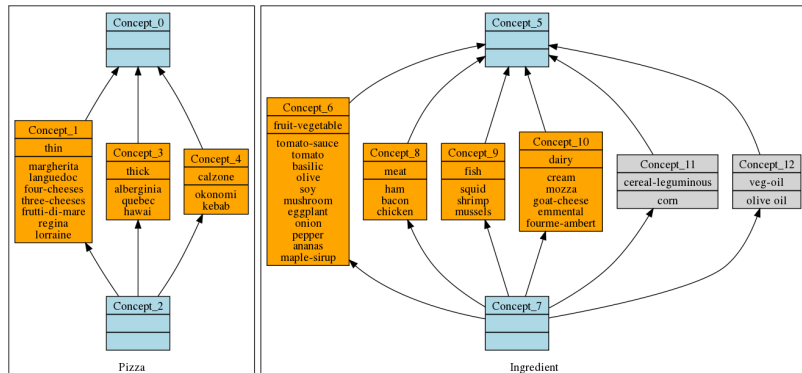
Data patterns we would like to extract

Using a classification on ingredients by their categories of topping (fruit-vegetable, dairy, etc.)

- ▶ All pizzas, even different, except four-cheese and three-cheese, contain at least one topping which is a vegetable
- ▶ Two pizzas (four-cheese and three-cheese) have all their topping in dairy ingredients
- ▶ For pizzas: have meat \Rightarrow have dairy
- ▶ For pizzas: being thin \Rightarrow have at least dairy
- ▶ For pizzas: have only dairy \Rightarrow being thin

RCA - Initial Lattice building

At the beginning, only the object-attribute contexts are used to build the foundation of the concept lattice family



RCA - Introducing relations as relational attributes

Given an object-object context $R_j = (O_k, O_l, I_j)$,

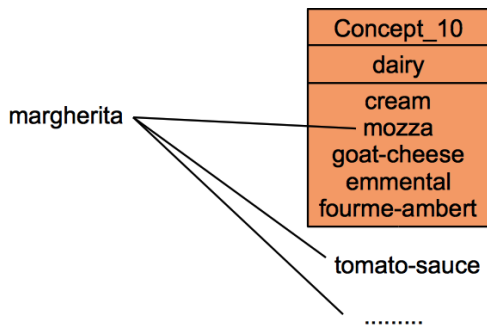
There are different notable schemas between an object of domain O_k and concepts formed on O_l .

E. g.

- ▶ **Existential**: an object is linked (by R_j) to at least one object of the extent of a concept
- ▶ **Universal**: an object is linked (by R_j) only to objects of the extent of a concept

RCA - Existential relational attributes

margherita has one topping in **Concept_10** extent: **mozza**.
It has other links to other concept extents.



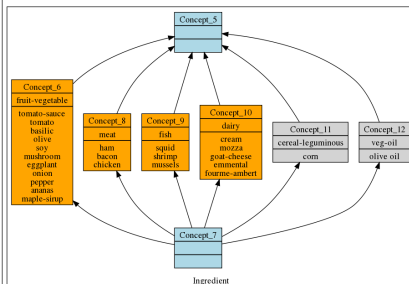
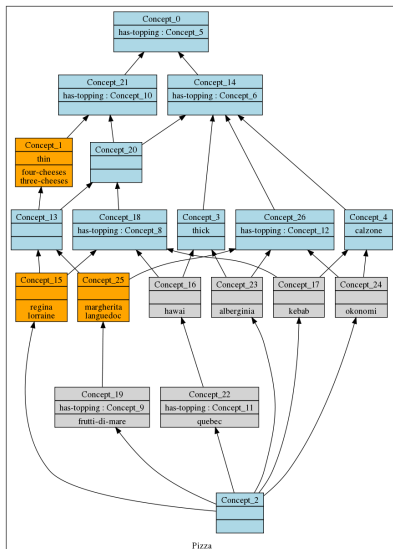
\exists has-topping.Concept_10 is assigned to **margherita**

RCA - Existential relational attributes

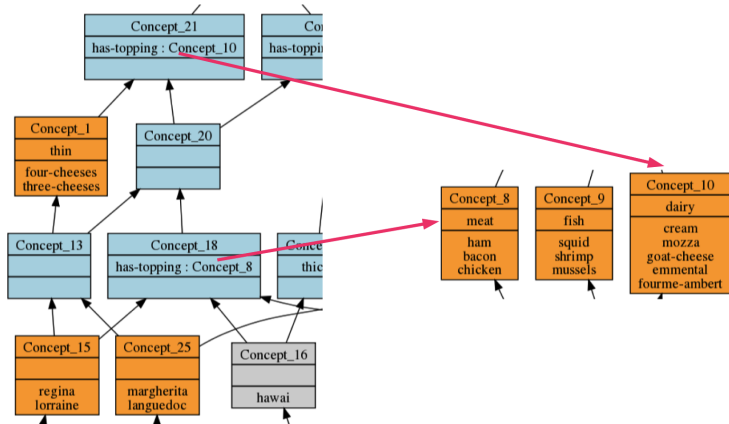
Scaled relations with domain O_i are concatenated to K_i , the object-attribute context on O_i

Pizza	thin	thick	calzone		has-topping. Concept_7	has-topping. Concept_5	has-topping. Concept_6	has-topping. Concept_8	has-topping. Concept_9	has-topping. Concept_10	has-topping. Concept_11	has-topping. Concept_12
okonomi			x	has-topping	\exists							
alberginia		x		okonomi		x	x					x
margherita	x			alberginia		x	x					x
languedoc	x			margherita		x	x			x		x
four-cheeses	x			languedoc		x	x			x		x
three-cheeses	x			four-cheeses		x				x		
frutti-di-mare	x			three-cheeses		x				x		
quebec		x		frutti-di-mare		x	x		x	x		x
regina	x			quebec		x	x	x		x	x	
hawai		x		regina		x	x	x		x		
lorraine	x			hawai		x	x	x		x		
kebab			x	lorraine		x	x	x		x		
				kebab		x	x	x		x		

Relational Concept Family / exists



Relational Concept Family / exists



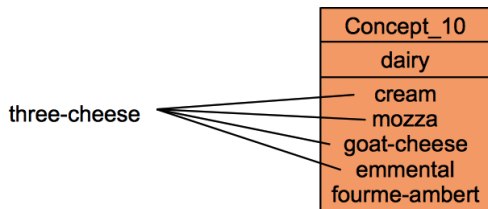
Concept_21: pizzas with at least one topping in dairy

Concept_18: pizzas with at least one topping in meat

have at least one meat topping \Rightarrow have at least one dairy topping

RCA - Universal relational attributes

three-cheese has topping in and only in **Concept_10** extent.



$\forall \exists$ has-topping. **Concept_10** is assigned to **three-cheese**

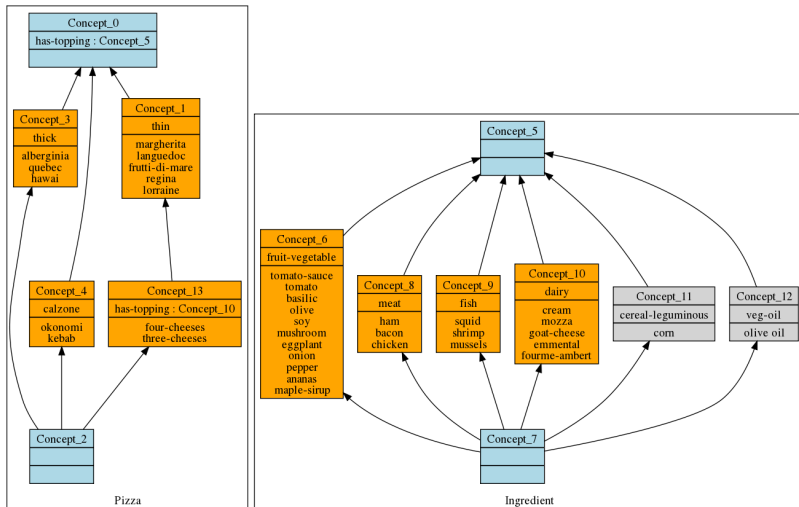
RCA - Universal relational attributes

Scaled relations with domain O_i are concatenated to K_i , the object-attribute context on O_i

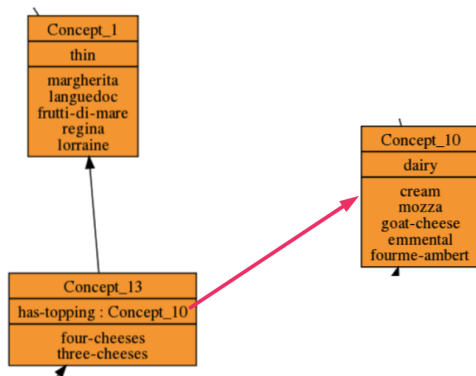
Pizza	thin	thick	calzone
okonomi			x
alberginia		x	
margherita	x		
languedoc	x		
four-cheeses	x		
three-cheeses	x		
frutti-di-mare	x		
quebec		x	
regina	x		
hawai		x	
lorraine	x		
kebab			x

has-topping	$\forall \exists$ has-topping. Concept_7	$\forall \exists \forall$ has-topping. Concept_5	$\forall \exists \forall$ has-topping. Concept_6	$\forall \exists \forall \forall$ has-topping. Concept_8	$\forall \exists \forall \forall \forall$ has-topping. Concept_9	$\forall \exists \forall \forall \forall \forall$ has-topping. Concept_10	$\forall \exists \forall \forall \forall \forall \forall$ has-topping. Concept_11	$\forall \exists \forall \forall \forall \forall \forall \forall$ has-topping. Concept_12
okonomi		x						
alberginia		x						
margherita		x						
languedoc		x						
four-cheeses		x						
three-cheeses		x						
frutti-di-mare		x						
quebec		x						
regina		x						
hawai		x						
lorraine		x						
kebab		x						

Relational Concept Family / forall



Relational Concept Family / forall



Concept_13: pizzas with only dairy topping

Concept_1: thin pizzas

have only dairy topping \Rightarrow thin

RCA - Introducing relations as relational attributes

Relational scaling is the process by which links are established between objects and concepts.

For each relational context $R_j = (O_k, O_l, I_j)$, a scaled context $R_j^* = (O_k, A, I_j)$ is created.

- ▶ A is a set of relational attributes $a = S R_j.C$, where C is in the concept set of a lattice built on objects of O_l , denoted by \mathcal{L}_l^n
- ▶ I_j contains (o, a) iff $S(R_j(o), Extent(C))$ is true.

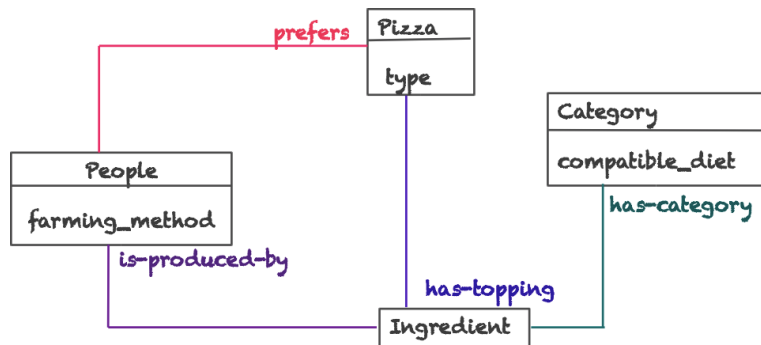
S is a *scaling* operator, the most used are:

- ▶ $S_{\exists}(R_j(o), Extent(C))$ is true iff $R_j(o) \cap Extent(C) \neq \emptyset$.
- ▶ $S_{\forall\exists}(R_j(o), Extent(C))$ is true iff $R_j(o) \subseteq Extent(C) \wedge \exists x \in R_j(o), x \in Extent(C)$

Scaling operators

Operator	Attribute form	Condition
Universal (narrow)	$\forall r.c$	$r(o) \subseteq \text{Ext}(c)$
Covers	$\supseteq r.c$	$r(o) \supseteq \text{Ext}(c)$
Existential (wide)	$\exists r.c$	$r(o) \cap \text{Ext}(c) \neq \emptyset$
Universal strict	$\forall \exists r.c$	$r(o) \subseteq \text{Ext}(c)$ and $r(o) \neq \emptyset$
Qualified cardinality restriction	$\geq n r.c$	$r(o) \subseteq \text{Ext}(c)$ and $ r(o) \geq n$
Cardinality restriction	$\geq n r.\top_{\mathcal{L}}$	$ r(o) \geq n$

General Entity-Relationship diagram



General ER diagram may present cycle/circuits between classes/objects

The RCA schema

Input

$RCF = (K, R)$: n object-attribute contexts, m object-object contexts

Initialization step

build, for i in $1..n$, $L^0[i]$ the concept lattice of the context \mathcal{K}_i

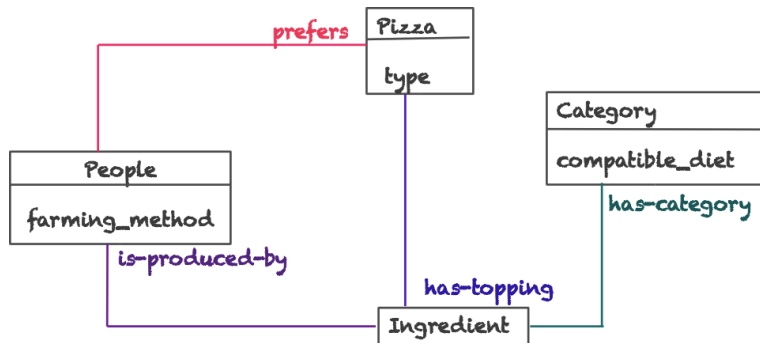
Step p

- ▷ Apply relational scaling to all object-object contexts R_j , using the lattices of step $p - 1$ and the chosen scaling operator
- ▷ concatenate \mathcal{K}_i with the scaled R_j^* whose domain is O_j
- ▷ update lattices of step $p - 1$ to build, for i in $1..n$, the lattice $L^p[i]$ for the context \mathcal{K}_i concatenated as explained previously

Output (fix point)

The concept lattice family obtained when no new concepts are added

Analysis of pizza data



\exists prefers $\forall \exists$ has-topping $\forall \exists$ has-category $\forall \exists$ is-produced-by

Analysis of pizza data - object-attribute contexts

Pizza	thin	thick	calzone
forest		×	
occitane			×
three-cheese	×		
four-cheese	×		
lorraine	×		
arctic		×	

People	organic-farmer	conventional-farmer
Amedeo	×	
Amine	×	
Cyril		×
Marianne	×	
Petko		×

Ingredient
tomato-sauce
cream
onion
bacon
salmon
soy-cream
mozza
goat-cheese
emmental
fourme-ambert
eggplant
mushroom

Category	mediterranean	vegan	vegetarian
fruit-vegetable	×	×	×
meat			
fish	×		
dairy	×		×

Analysis of pizza data - object-object contexts

prefers	forest	occitane	three-cheese	four-cheese	lorraine	arctic
Amedeo	×					
Amine		×				
Cyril				×	×	
Marianne			×			×
Petko						×

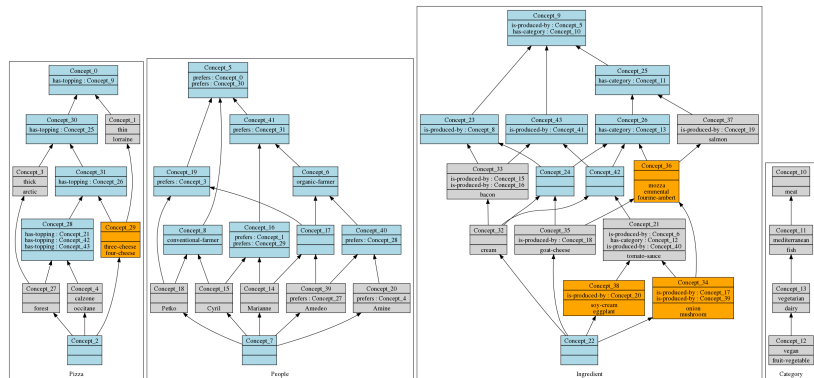
	tomato-sauce	cream	onion	bacon	salmon	soy-cream	mozza	goat-cheese	emmental	fourme-ambert	eggplant	mushroom
has-topping												
forest						×						×
occitane	×		×								×	
three-cheese	×						×	×	×			
four-cheese	×	×					×	×	×	×		
lorraine		×	×	×			×					
arctic	×	×			×		×					

Analysis of pizza data - object-object contexts

is-produced-by	Amedeo	Amine	Cyril	Marianne	Petko
tomato-sauce	×	×			
cream			×		
onion	×				
bacon			×		
salmon				×	×
soy-cream		×			
mozza				×	×
goat-cheese					×
emmental				×	×
fourme-ambert				×	×
eggplant		×			
mushroom	×				

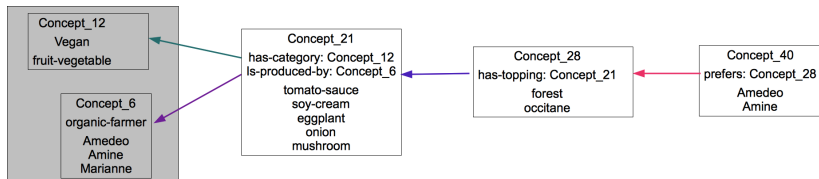
has-category	fruit-vegetable	meat	fish	dairy
tomato-sauce	×			
cream				×
onion	×			
bacon		×		
salmon			×	
soy-cream	×			
mozza				×
goat-cheese				×
emmental				×
fourme-ambert				×
eggplant	×			
mushroom	×			

Concept lattice family



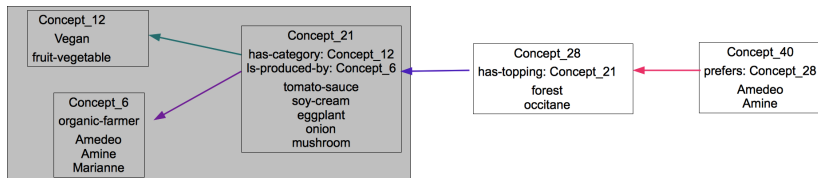
\exists prefers $\forall \exists$ has-topping $\forall \exists$ has-category $\forall \exists$ is-produced-by

Concept lattice family



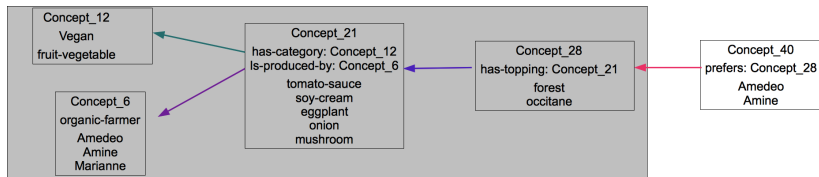
Step 0

Concept lattice family



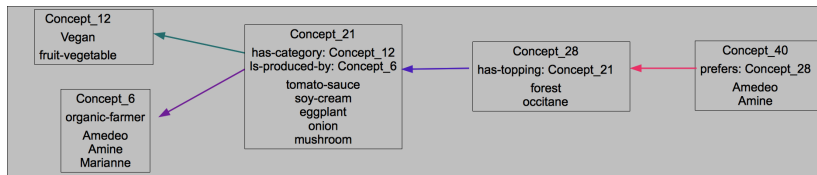
Step 1

Concept lattice family



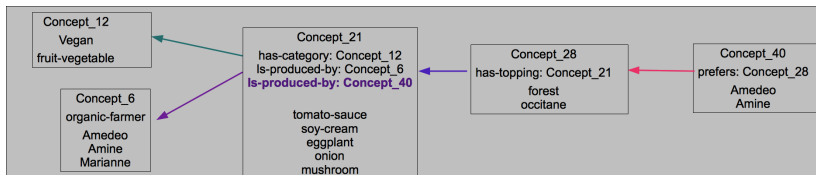
Step 2

Concept lattice family



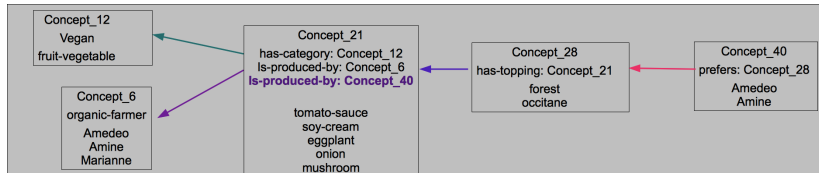
Step 3

Concept lattice family



Step 4

Concept lattice family



- ▶ People: $\exists \text{prefers. Concept_28} \Rightarrow \text{organic farmer}$
- ▶ Ingredient: $\forall \exists \text{has-category. Concept_12} \Leftrightarrow \forall \exists \text{is-produced-by. Concept_6}$ (organic farmers)
- ▶ Amedeo/Amine prefer at least one pizza with only vegan topping ingredients and produced only by organic farmers

A synthesis on RCA

- ▶ an iterative method to produce interconnected classifications
- ▶ converges after a number of iterations that depends on the structure
- ▶ variations on scaling can be done

Tools

- ▶ Galicia: <http://galicia.sourceforge.net/>
- ▶ eRCA: <http://code.google.com/p/erca/>

RCA - Current tracks

- ▶ Querying the concept lattice family
- ▶ Exploratory RCA: select, divide, step-by-step
- ▶ Metrics for guiding the process and filtering concepts
- ▶ Build Galois sub-hierarchy (AOC-poset)