



# Analyse Relationnelle de Concepts

Marianne Huchard

## ► To cite this version:

Marianne Huchard. Analyse Relationnelle de Concepts. Journée Charles Hermite "Gestion de la connaissance à l'aide des méthodes formelles", Dec 2014, Nancy, France. lirmm-01154430

**HAL Id: lirmm-01154430**

**<https://hal-lirmm.ccsd.cnrs.fr/lirmm-01154430>**

Submitted on 21 May 2015

**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.

# Relational Concept Analysis (RCA)

*Journée Charles Hermite  
Gestion de la connaissance à l'aide des méthodes formelles*

*Introduction to RCA by Marianne Huchard*

*Followed by Tool presentation by Xavier Dolques*

Nancy, 10 décembre 2014

# Brief presentation of FCA – Formal Concept Analysis

A methodology for:






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

Roots:

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

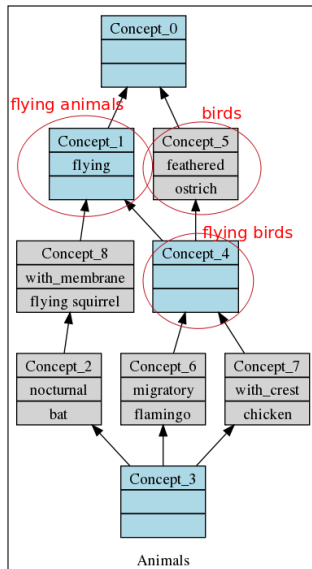
# Brief presentation of FCA – Formal Concept Analysis

- ▶ Handled data
  - ▶ entities (objects) with characteristics (attributes)
  - ▶ 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

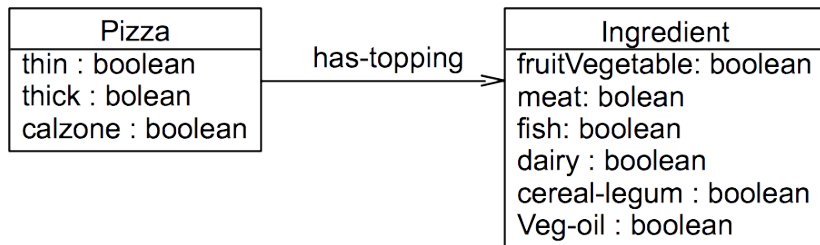
# Brief presentation of FCA – Formal Concept Analysis



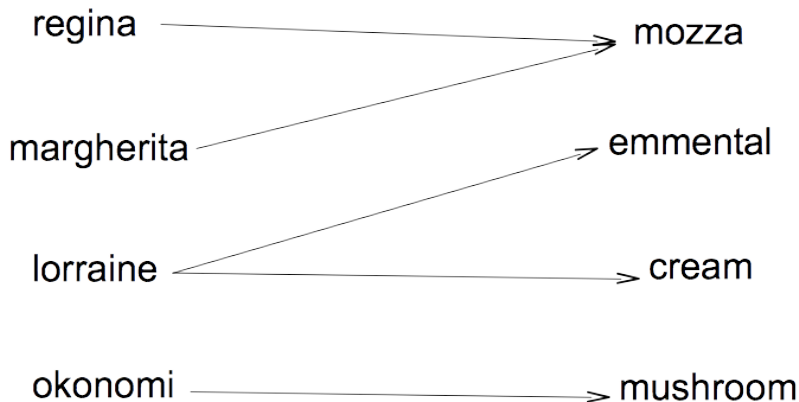
# FCA and complex data

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

# A flavor of Relational Concept Analysis

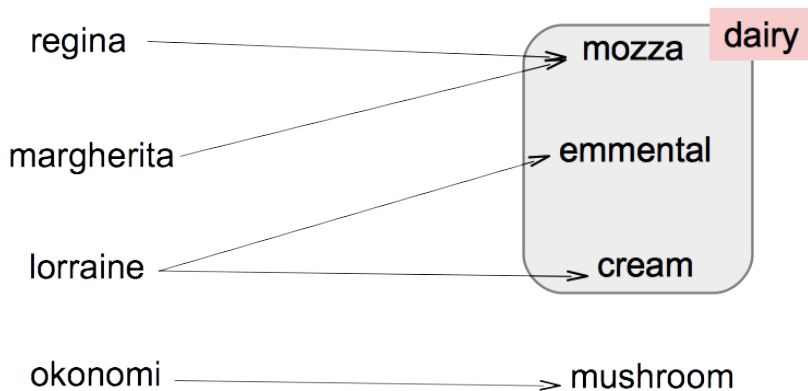


## A flavor of Relational Concept Analysis

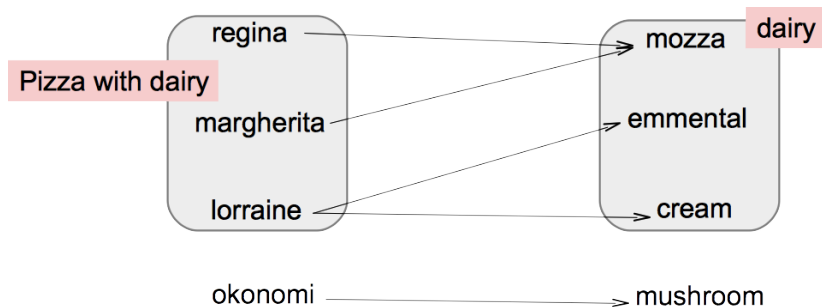




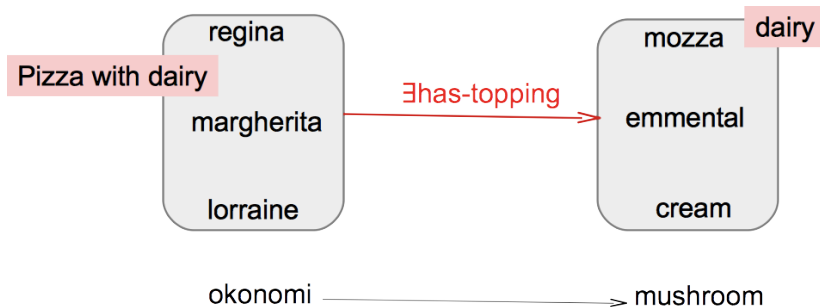
# A flavor of Relational Concept Analysis



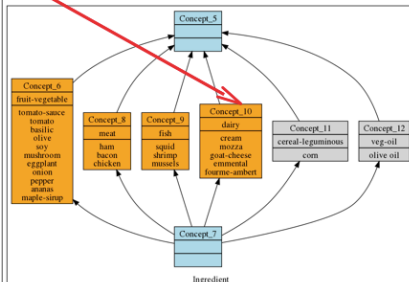
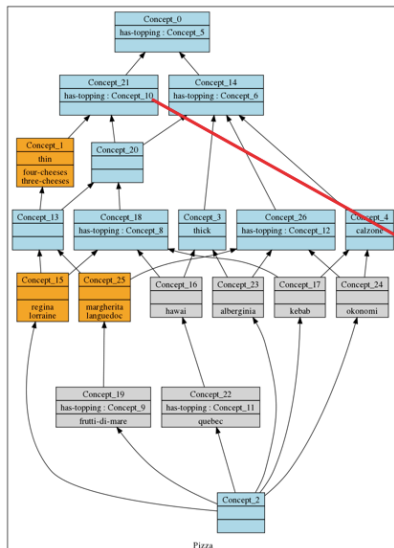
# A flavor of Relational Concept Analysis



# A flavor of Relational Concept Analysis



# A flavor of Relational Concept Analysis



# 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

Joint formalization work with:

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

*M. Rouane Hacene, M. Huchard, A. Napoli, P. Valtchev. Relational concept analysis: mining concept lattices from multi-relational data, Ann. Math. Artif. Intell. 67(1), pp. 81-108, 2013*

# Relational Context Family (RCF)

## Relational Context Family

A simple populated Entity-Relationship model

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

# 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-vege.	meat	fish	dairy	cereal-legum.	veg-oil
tomato-sauce	×					
cream				×		
tomato	×					
basilic	×					
olive	×					
olive oil						×
soy	×					
mushroom	×					
eggplant	×					
onion	×					
pepper	×					
ananas	×					
mozza				×		
goat-cheese				×		
emmental				×		
fourme-ambert				×		
squid			×			
shrimp			×			
mussels			×			
ham		×				
bacon		×				
chicken		×				
maple-sirup	×					
corn					×	

# RCF / object-object context / part 1

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



## RCF / object-object context / part 2

	mozza	goat-cheese	emmental	fourme-ambert	squid	shrimp	mussels	ham	bacon	chicken	maple-sirup	corn
has-topping												
ekonomi												
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		

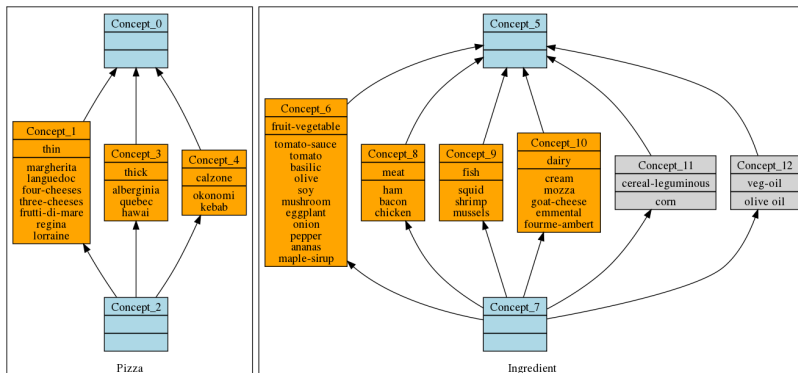
# Data patterns we would like to extract

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

- ▶ create groups and classifications
  - ▶ The group of pizzas that contain at least one topping which is a vegetable
  - ▶ The group of pizzas (four-cheese and three-cheese) that have all their topping in dairy ingredients
- ▶ find implications
  - ▶ 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 possible schemas between an object of domain  $O_k$  and concepts formed on  $O_l$ .

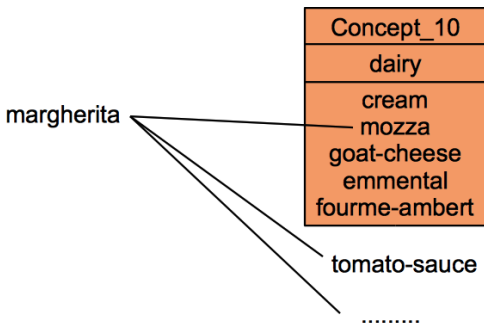
*E. g.*

- ▶ **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

$\exists$  and  $\forall$  are **scaling operators**

# RCA - Existential relational attributes

**margherita** has one topping in **Concept\_10** extent: **mozza**.  
It has other links to other concept extents.



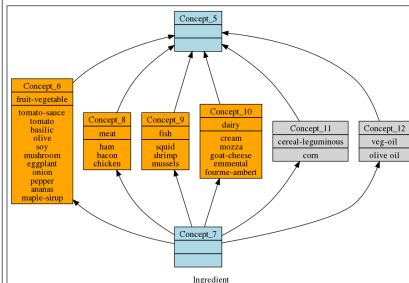
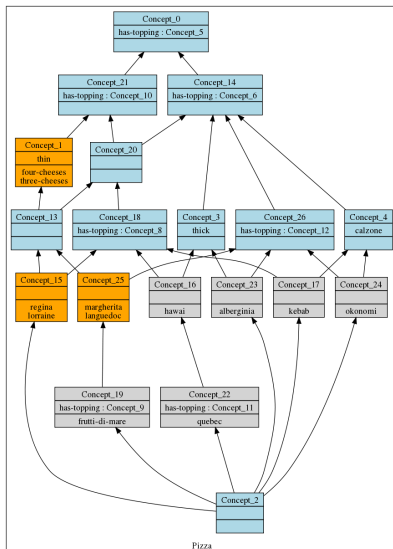
$\exists$ has-topping.**Concept\_10** is assigned to **margherita**

# RCA - Relational extension

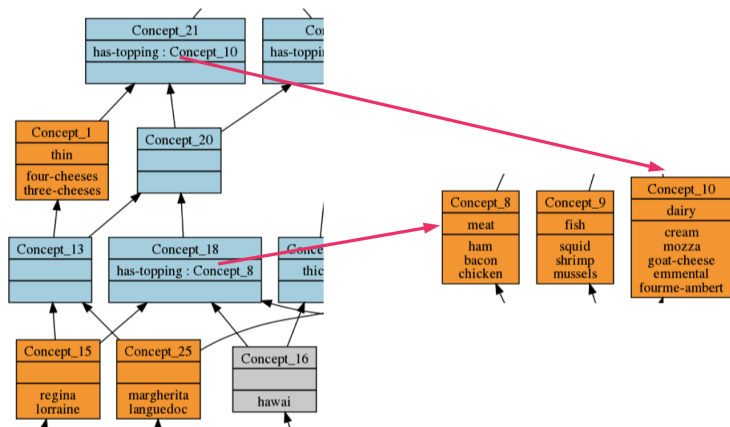
Scaled relations with domain  $O_i$  are concatenated to  $K_i$ , the object-attribute context on  $O_i$

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

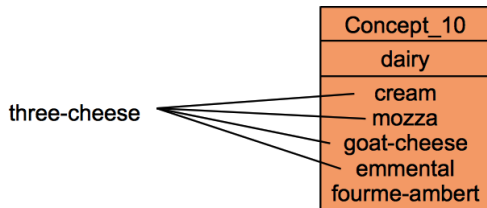
Concept\_18: pizzas with at least one topping in meat

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.



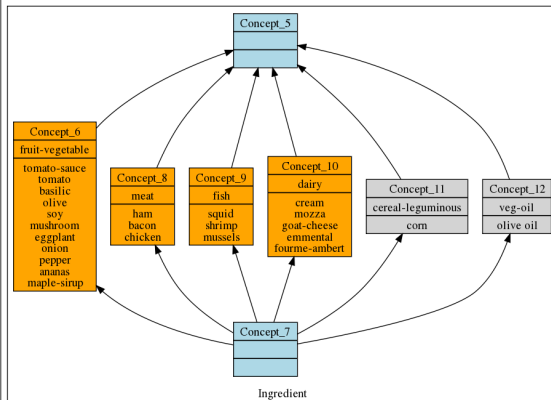
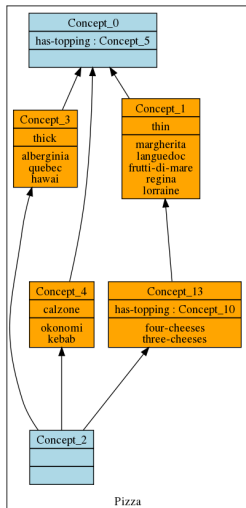
$\forall e \in \text{has-topping.Concept\_10}$  is assigned to **three-cheese**

## RCA - Relational extension

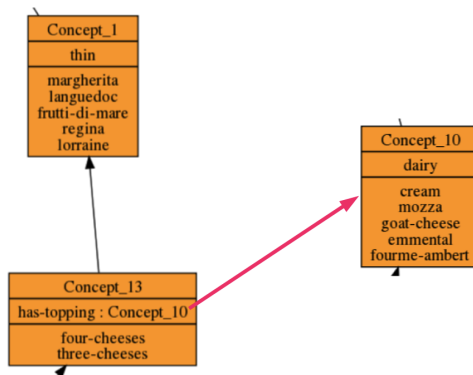
Scaled relations with domain  $O_i$  are concatenated to  $K_i$ , the object-attribute context on  $O_i$

Pizza	thin	thick	calzone
ekonomi			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

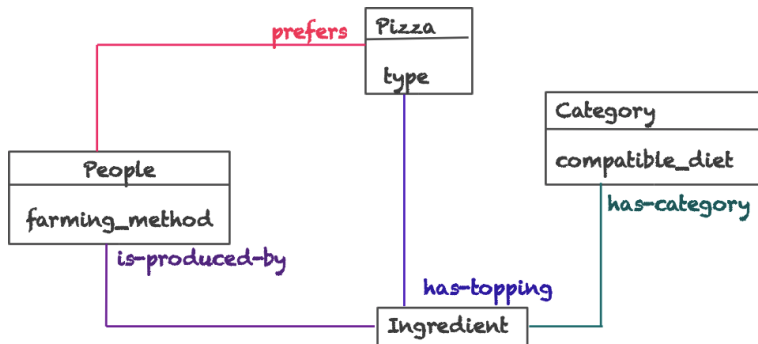


Concept\_13: pizzas with only dairy topping

Concept\_1: thin pizzas

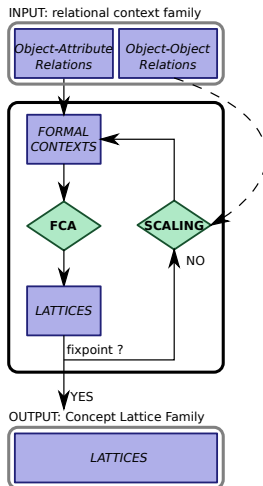
have only dairy topping  $\Rightarrow$  thin

# General Entity-Relationship diagram may have circuits



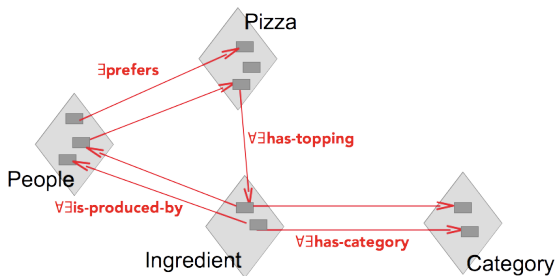
$\exists \text{ prefers } \forall \exists \text{ has-topping } \forall \exists \text{ has-category } \forall \exists \text{ is-produced-by}$

# RCA schema



Credit X. Dolques

# Interconnected lattices



# 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						×

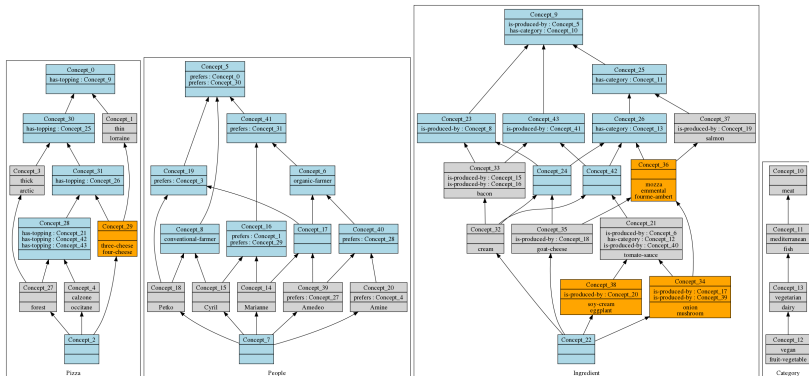
has-topping	tomato-sauce	cream	onion	bacon	salmon	soy-cream	mozza	goat-cheese	emmental	fourme-ambert	eggplant	mushroom
forest						×						×
occitane	×		×								×	
three-cheese	×						×	×	×			
four-cheese	×	×					×	×	×	×		
lorraine		×	×	×			×					
arctic	×	×			×		×					

# 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	×				

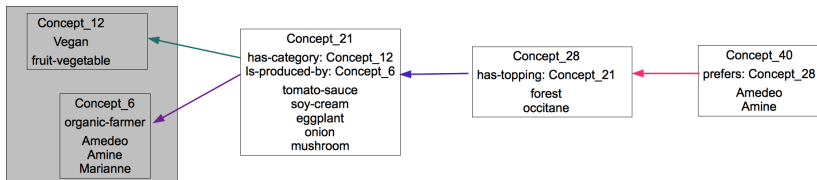
has-category	fruit-vegetable	meat	fish	dairy
tomato-sauce	×			
cream				×
onion	×			
bacon		×		
salmon			×	
soy-cream	×			
mozza				×
goat-cheese				×
emmental				×
fourme-ambert				×
eggplant	×			
mushroom	×			

# Concept lattice family



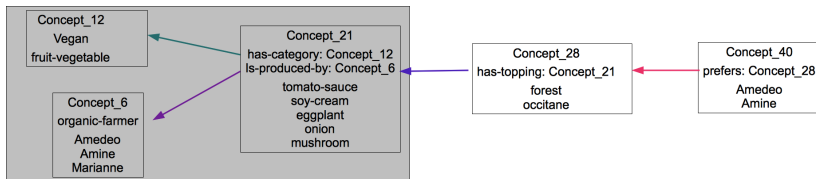
$\exists \text{ prefers } \forall \exists \text{ has-topping } \forall \exists \text{ has-category } \forall \exists \text{ is-produced-by}$

# Concept lattice family



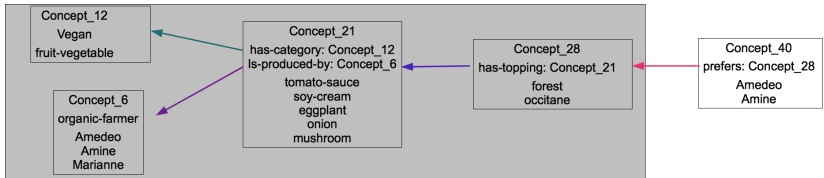
Step 0

# Concept lattice family



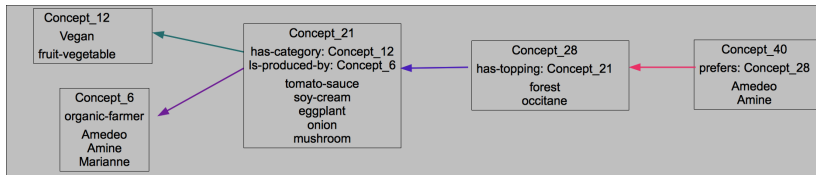
Step 1

# Concept lattice family



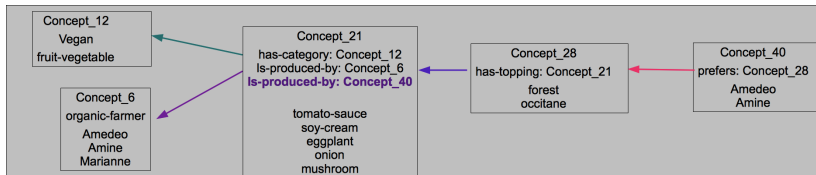
Step 2

# Concept lattice family



Step 3

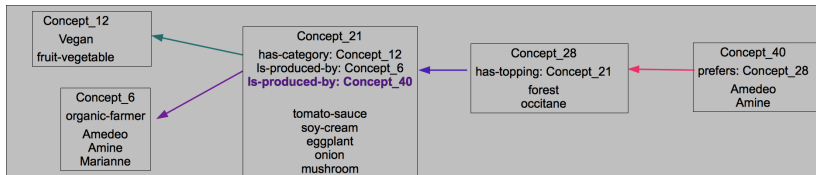
# Concept lattice family



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

- ▶ Ontology construction (R. Bendaoud, M. Rouane Hacene, Y. Toussaint, B. Delecroix, A. Napoli)
- ▶ Ontology restructuring (M. Rouane-Hacene, R. Nkambou and P. Valtchev)
- ▶ Discovering hidden user profiles in a semantic actors-activities network (Z. Azmeh, I. Mirbel)

# FRESQUEAU project (ANR11\_MONU14)

Joint work with: A. Braud, X. Dolques, C. Grac, F. Le Ber, C. Nica

<http://engees-fresqueau.unistra.fr>

- ▶ Develop new methods to study, compare and exploit the whole set of available parameters describing the state of watercourses
- ▶ Extraction of implication rules (with premise of size 1)  
*Presence of taxons of size from 0 to 2 cm and with a lifetime of 1 month implies presence of SO4*
- ▶ RCA-AOC
- ▶ Connection with propositionnalization approach
- ▶ Ongoing work on temporal RCA

## Reengineering of existing software, by building new software artefacts

- ▶ UML class diagram refactoring (M. Dao, M. Huchard, M. Rouane Hacene, C. Roume, P. Valtchev, G. Arévalo, J.-R. Falleri, C. Nebut)
- ▶ UML Use case diagram refactoring (X. Dolques, M. Huchard, C. Nebut, P. Reitz)
- ▶ Blob design defect correction (N. Moha, M. Rouane Hacene, P. Valtchev, Y.-G. Guéhéneuc)
- ▶ Extracting architectures in object-oriented software (A.-E. El Hamdouni, A. Seriai, M. Huchard)

# Applications to Software engineering

## Learning from model transformation examples, and inferring transformation rules

- ▶ Learning model Transformation patterns in MDE (H. Saada, X. Dolques, M. Huchard, C. Nebut, H. A. Sahraoui)

## Classification of software artefacts

- ▶ Classification of web services (Z. Azmeh, M. Driss, F. Hamoui, M. Huchard, N. Moha, C. Tibermacine)

## Software analysis

- ▶ Analysis of the evolution of class diagrams (A. Osman-Guédi, A. Miralles, B. Amar, M. Huchard, T. Libourel and C. Nebut)

# Conclusion

RCA: an opportunity for analyzing more deeply dataset composed of objects and relations

- ▶ an iterative method to produce interconnected classifications
- ▶ converges after a number of iterations that depends on the structure
- ▶ a variety of scaling operators
- ▶ reduced structures can be used instead of lattices:  
AOC-posets, iceberg lattices
- ▶ can be mixed with other FCA extension (to numerical data for example)

- ▶ A querying mechanism and navigation tools
- ▶ Compare systematically AOC-poset, iceberg lattices and concept lattices in the applications
- ▶ Exploration-based methodology (method convergence)