Analyzing inheritance hierarchies through Formal Concept Analysis

A 22-years walk in a landscape of conceptual structures

Marianne Huchard

LIRMM, CNRS and Univ. Montpellier, France

MASPEGHI@ECOOP2015

Prague, 2015 july, 5

Motivation

Class models, inheritance hierarchies

- Capture domain knowledge
- Focus on classification and reuse

Normal form

- No redundancy
- All abstractions are created
- All specialization links are present
- Most compact structure with these properties

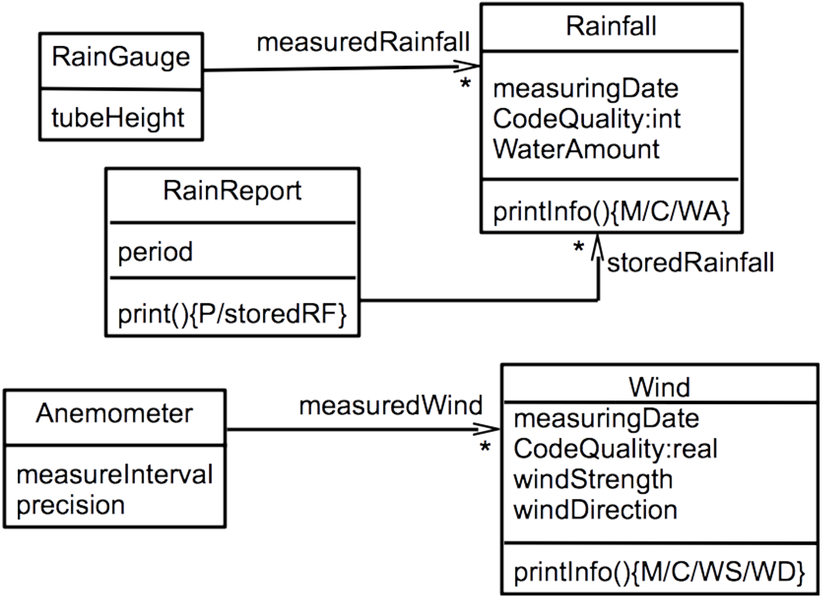
→ To restructure the current model or just to analyze it

Show historical steps in normalization

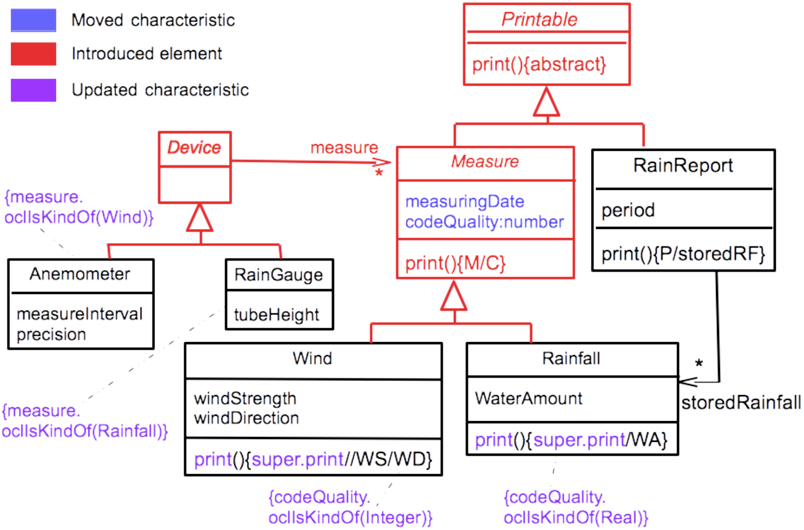
Outline

- Normalization illustration
- Formal Concept Analysis *(Ganter and Wille 1999)*
  Galois lattices *(Barbut and Monjardet 1970)*
  - Flat characteristics *(Godin et al., 1993)*
  - Hierarchical characteristics *(Godin et al., 1993)*
- Relational Concept Analysis *(Huchard et al. 2007)*
  - Reified characteristics *(Roume et al. 2004)*
  - Clustered reified characteristics *(unpublished)*
- Ongoing work (terms, exploration)

Normalization objective: initial class model



Normalization objective: normal form



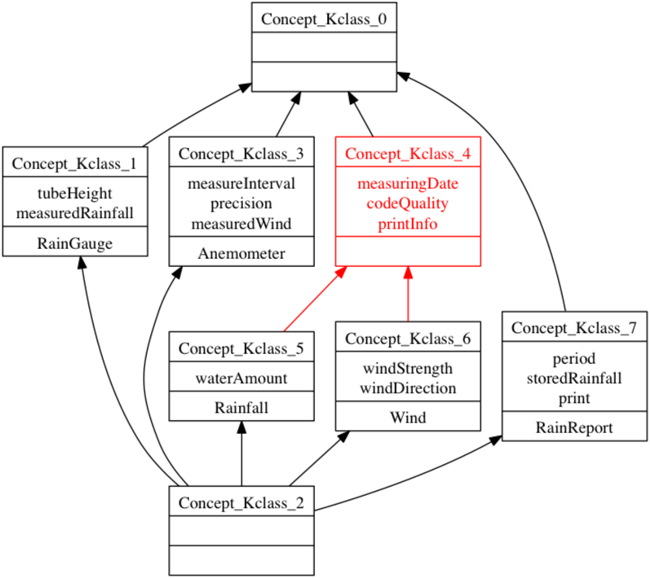
Flat characteristics: Formal context (Godin and Mili, 1993)

Kclass	tubeHeight	measureInterval	precision	measuringDate	codeQuality	waterAmount	windStrength	windDirection	period	measuredRainfall	measuredWind	storedRainfall	print	printInfo
RainG.	×									×				
Anem.		×	×								×			
Rainf.				×	×	×								×
Wind				×	×		×	×						×
RainR.									×			×	×	

Flat characteristics: Concept

Kclass	tubeHeight	measureInterval	precision	measuringDate	codeQuality	waterAmount	windStrength	windDirection	period	measuredRainfall	measuredWind	storedRainfall	print	printInfo
RainG.	×									×				
Anem.		×	×								×			
Rainf.				×	×	×								×
Wind				×	×		×	×						×
RainR.									×			×	×	

Flat characteristics: concept lattice



Concept_Kclass_0

Concept_Kclass_1
tubeHeight
measuredRainfall
RainGauge

Concept_Kclass_3
measureInterval
precision
measuredWind
Anemometer

Concept_Kclass_4
measuringDate
codeQuality
printInfo

Concept_Kclass_5
waterAmount
Rainfall

Concept_Kclass_6
windStrength
windDirection
Wind

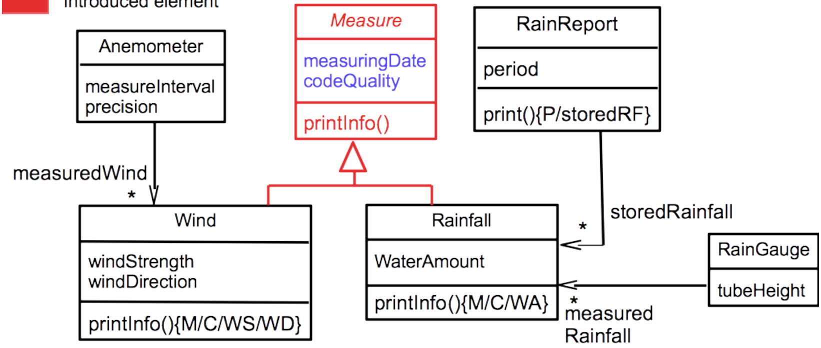
Concept_Kclass_7
period
storedRainfall
print
RainReport

Concept_Kclass_2

Flat characteristics: revisited model

🟦 Moved characteristic

🟥 Introduced element



Anemometer
measureInterval
precision

measuredWind
*

Measure
measuringDate
codeQuality
printInfo()

RainReport
period
print(){P/storedRF}

storedRainfall
*

Wind
windStrength
windDirection
printInfo(){M/C/WS/WD}

Rainfall
WaterAmount
printInfo(){M/C/WA}

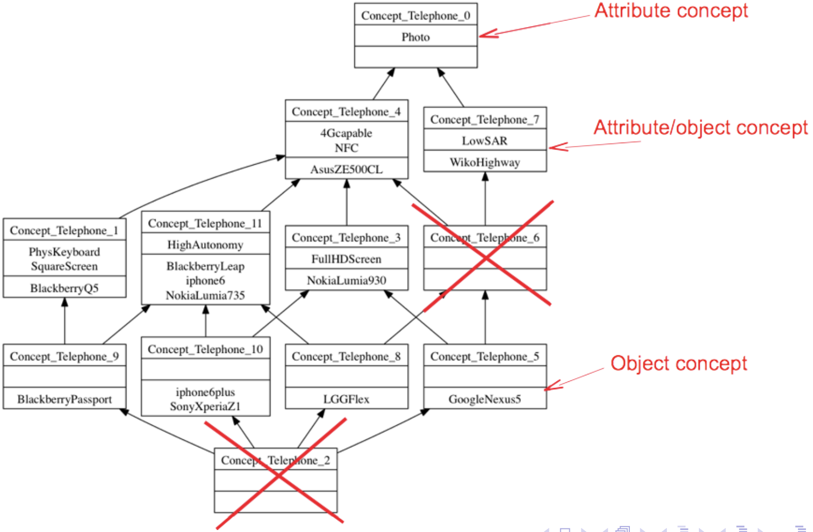
RainGauge
tubeHeight

*
measured
Rainfall

Parenthesis: AOC-poset (Telephone classification)

With O the described entities, A the characteristics

#concepts in lattice $\leq 2^{min(|O|,|A|)}$, #concepts in AOC-posets $\leq |O| + |A|$



Concept_Telephone_0
Photo

Concept_Telephone_4
4Gcapable
NFC
AsusZE500CL

Concept_Telephone_7
LowSAR
WikoHighway

Concept_Telephone_1
PhysKeyboard
SquareScreen
BlackberryQ5

Concept_Telephone_11
HighAutonomy
BlackberryLeap
iphone6
NokiaLumia735

Concept_Telephone_3
FullHDScreen
NokiaLumia930

Concept_Telephone_6

Concept_Telephone_9
BlackberryPassport

Concept_Telephone_10
iphone6plus
SonyXperiaZ1

Concept_Telephone_8
LGGFlex

Concept_Telephone_5
GoogleNexus5

Concept_Telephone_2

Attribute concept

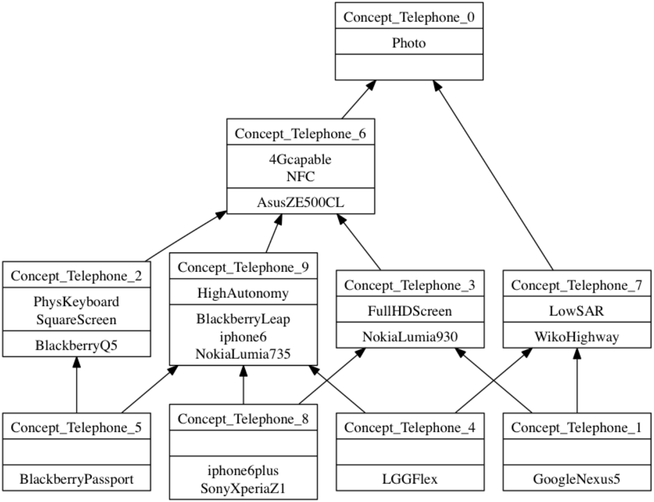
Attribute/object concept

Object concept

Parenthesis: AOC-poset (Telephone classification)

With O the described entities, A the characteristics

#concepts in lattice $\leq 2^{min(|O|,|A|)}$, #concepts in AOC-posets $\leq |O| + |A|$



Concept_Telephone_0
Photo

Concept_Telephone_6
4Gcapable
NFC
AsusZE500CL

Concept_Telephone_2
PhysKeyboard
SquareScreen
BlackberryQ5

Concept_Telephone_9
HighAutonomy
BlackberryLeap
iphone6
NokiaLumia735

Concept_Telephone_3
FullHDScreen
NokiaLumia930

Concept_Telephone_7
LowSAR
WikoHighway

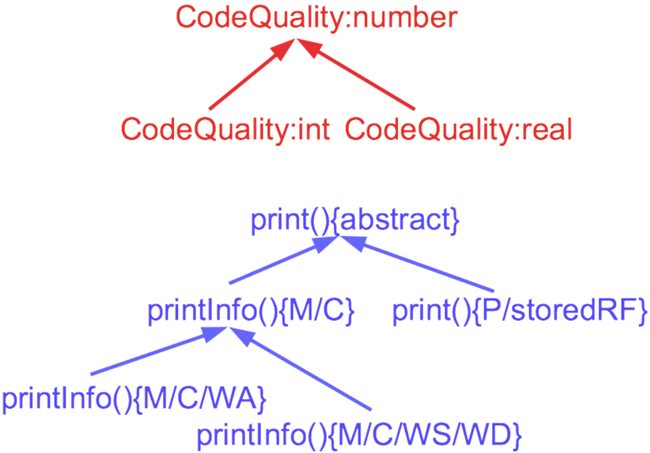
Concept_Telephone_5
BlackberryPassport

Concept_Telephone_8
iphone6plus
SonyXperiaZ1

Concept_Telephone_4
LGGFlex

Concept_Telephone_1
GoogleNexus5

Hierarchical characteristics (Godin and Mili, 1993)



CodeQuality:number

CodeQuality:int CodeQuality:real

print(){abstract}

printInfo(){M/C} print(){P/storedRF}

printInfo(){M/C/WA}

printInfo(){M/C/WS/WD}

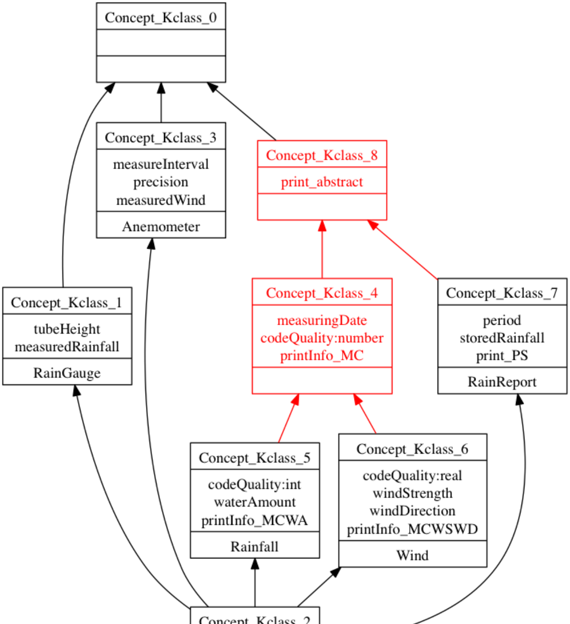
Hierarchical characteristics: Formal context (part 1)

Kclass	tubeHeight	measureInterval	precision	measuringDate	codeQuality:number	codeQuality:int	codeQuality:real	waterAmount	windStrength	windDirection	period	measuredRainfall	measuredWind
RainGauge	×											×	
Anemometer		×	×										×
Rainfall				×	×	×		×					
Wind				×	×		×		×	×			
RainReport											×		

Hierarchical characteristics: Formal context (part 2)

Kclass	measuredWind	storedRainfall	print_abstract	printInfo_MC	printInfo_MCWA	printInfo_MCWSWD	print_PS
RainGauge							
Anemometer	×						
Rainfall			×	×	×		
Wind			×	×		×	
RainReport		×	×				×

Hierarchical characteristics: Concept Lattice



Concept_Kclass_0

Concept_Kclass_3

measureInterval
precision
measuredWind

Anemometer

Concept_Kclass_8

print_abstract

Concept_Kclass_1

tubeHeight
measuredRainfall

RainGauge

Concept_Kclass_4

measuringDate
codeQuality:number
printInfo_MC

Concept_Kclass_7

period
storedRainfall
print_PS

RainReport

Concept_Kclass_5

codeQuality:int
waterAmount
printInfo_MCWA

Rainfall

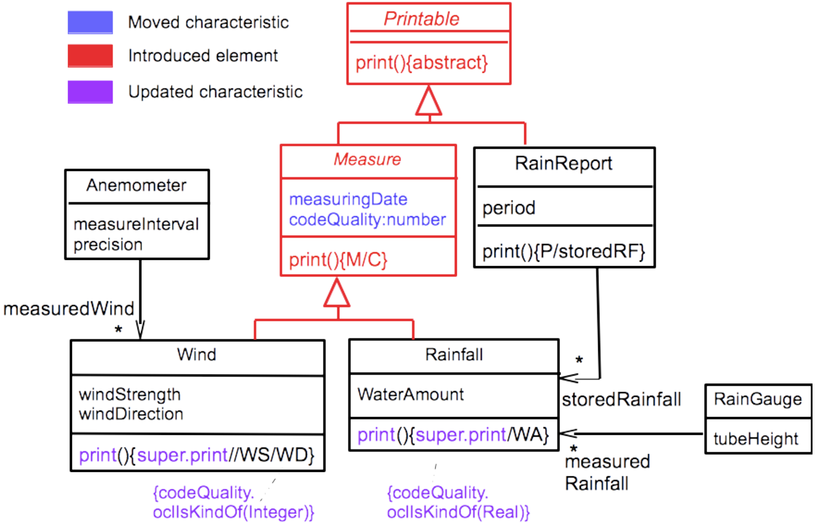
Concept_Kclass_6

codeQuality:real
windStrength
windDirection
printInfo_MCWSWD

Wind

Concept_Kclass_2

Hierarchical characteristics: revisited model



Relational Concept Analysis - going further with reification

Dataset

- Several entity categories (OA-contexts): classes, attributes, operations, roles
- Relations between entities (OO-contexts): hasAttribute, hasRole, hasOperation, hasTypeEnd

Principle

- Builds one concept lattice for each entity category
- Iterate on lattice construction and integration of relational attributes (encoding OO-contexts) in OA-contexts
- Until a fix-point

Reified characteristics: Object-Attribute contexts

Kclass
RainGauge
Anemometer
Rainfall
Wind
RainReport

Krole	measuredRainfall	measuredWind	storedRainfall
RG::measuredRainfall	×		
RR::storedRainfall			×
A::measuredWind		×	

Koperation	print	MC	MCWA	MCWSWD	PstoredRF
RR::print	×				×
R::printinfo	×	×	×		
W::printinfo	×	×		×	

Kattribute	tubeHeight	measureInterval	precision	measuringDate	codeQuality	waterAmount	windStrength	windDirection	period	number	int	real
RG::tubeHeight	×											
A::measureInterval		×										
A::precision			×									
R::measuringDate				×								
W::measuringDate				×								
R::codeQuality					×					×	×	
W::codeQuality					×					×		×
R::waterAmount						×						
W::windStrength							×					
W::windDirection								×				
RR::period									×			

Reified characteristics: Object-Object contexts

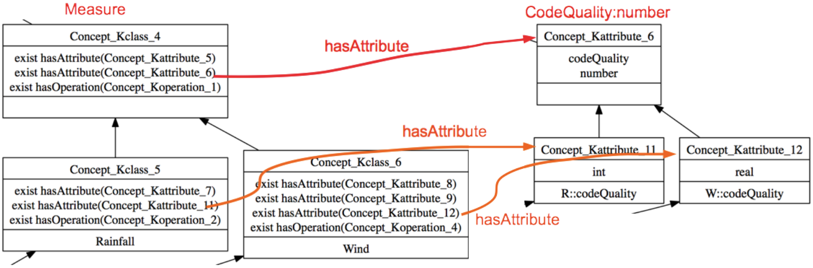
hasAttribute	RG::tubeHeight	A::measureInterval	A::precision	R::measuringDate	W::measuringDate	R::codeQuality	W::codeQuality	R::waterAmount	W::windStrength	W::windDirection	RR::period
RainGauge	×										
Anemometer		×	×								
Rainfall				×		×		×			
Wind					×		×		×	×	
RainReport											×

hasRole	RG::measuredRainfall	RR::storedRainfall	A::measuredWind
RainGauge	×		
Anemometer			×
Rainfall			
Wind			
RainReport		×	

hasOperation	RR::print	R::printinfo	W::printinfo
RainGauge			
Anemometer			
Rainfall		×	
Wind			×
RainReport	×		

hasTypeEnd	Rainfall	Wind
RG::measuredRainfall	×	
RR::storedRainfall	×	
A::measuredWind		×

Reified characteristics: detail on class lattice and attribute lattice



Measure

Concept_Kclass_4

exist hasAttribute(Concept_Kattribute_5)
exist hasAttribute(Concept_Kattribute_6)
exist hasOperation(Concept_Koperation_1)

hasAttribute

CodeQuality:number

Concept_Kattribute_6

codeQuality
number

hasAttribute

Concept_Kclass_5

exist hasAttribute(Concept_Kattribute_7)
exist hasAttribute(Concept_Kattribute_11)
exist hasOperation(Concept_Koperation_2)

Rainfall

Concept_Kclass_6

exist hasAttribute(Concept_Kattribute_8)
exist hasAttribute(Concept_Kattribute_9)
exist hasAttribute(Concept_Kattribute_12)
exist hasOperation(Concept_Koperation_4)

Wind

Concept_Kattribute_11

int

R::codeQuality

Concept_Kattribute_12

real

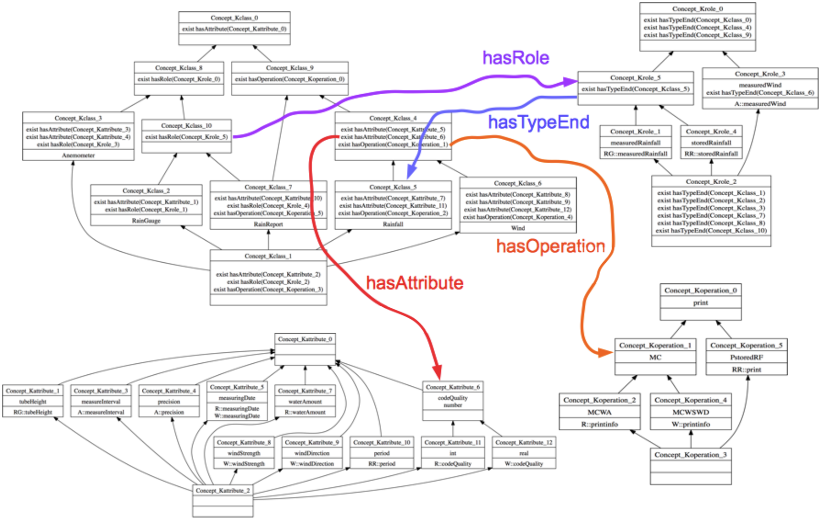
W::codeQuality

hasAttribute

Class Lattice

Attribute Lattice

Reified characteristics: all lattices



Concept_Kclass_0
exist hasAttribute(Concept_Kattribute_0)

Concept_Kclass_8
exist hasRole(Concept_Krole_0)

Concept_Kclass_9
exist hasOperation(Concept_Koperation_0)

Concept_Kclass_3
exist hasAttribute(Concept_Kattribute_3)
exist hasAttribute(Concept_Kattribute_4)
exist hasRole(Concept_Krole_3)
Anemometer

Concept_Kclass_10
exist hasRole(Concept_Krole_5)

Concept_Kclass_4
exist hasAttribute(Concept_Kattribute_5)
exist hasAttribute(Concept_Kattribute_6)
exist hasOperation(Concept_Koperation_1)

Concept_Kclass_2
exist hasAttribute(Concept_Kattribute_1)
exist hasRole(Concept_Krole_1)
RainGauge

Concept_Kclass_7
exist hasAttribute(Concept_Kattribute_10)
exist hasRole(Concept_Krole_4)
exist hasOperation(Concept_Koperation_5)
RainReport

Concept_Kclass_5
exist hasAttribute(Concept_Kattribute_7)
exist hasAttribute(Concept_Kattribute_11)
exist hasOperation(Concept_Koperation_2)
Rainfall

Concept_Kclass_6
exist hasAttribute(Concept_Kattribute_8)
exist hasAttribute(Concept_Kattribute_9)
exist hasAttribute(Concept_Kattribute_12)
exist hasOperation(Concept_Koperation_4)
Wind

Concept_Kclass_1
exist hasAttribute(Concept_Kattribute_2)
exist hasRole(Concept_Krole_2)
exist hasOperation(Concept_Koperation_3)

Concept_Krole_0
exist hasTypeEnd(Concept_Kclass_0)
exist hasTypeEnd(Concept_Kclass_4)
exist hasTypeEnd(Concept_Kclass_9)

Concept_Krole_5
exist hasTypeEnd(Concept_Kclass_5)

Concept_Krole_3
measuredWind
exist hasTypeEnd(Concept_Kclass_6)
A::measuredWind

Concept_Krole_1
measuredRainfall
RG::measuredRainfall

Concept_Krole_4
storedRainfall
RR::storedRainfall

Concept_Krole_2
exist hasTypeEnd(Concept_Kclass_1)
exist hasTypeEnd(Concept_Kclass_2)
exist hasTypeEnd(Concept_Kclass_3)
exist hasTypeEnd(Concept_Kclass_7)
exist hasTypeEnd(Concept_Kclass_8)
exist hasTypeEnd(Concept_Kclass_10)

Concept_Kattribute_0

Concept_Kattribute_1
tubeHeight
RG::tubeHeight

Concept_Kattribute_3
measureInterval
A::measureInterval

Concept_Kattribute_4
precision
A::precision

Concept_Kattribute_5
measuringDate
R::measuringDate
W::measuringDate

Concept_Kattribute_7
waterAmount
R::waterAmount

Concept_Kattribute_6
codeQuality
number

Concept_Kattribute_8
windStrength
W::windStrength

Concept_Kattribute_9
windDirection
W::windDirection

Concept_Kattribute_10
period
RR::period

Concept_Kattribute_11
int
R::codeQuality

Concept_Kattribute_12
real
W::codeQuality

Concept_Kattribute_2

Concept_Koperation_0
print

Concept_Koperation_1
MC

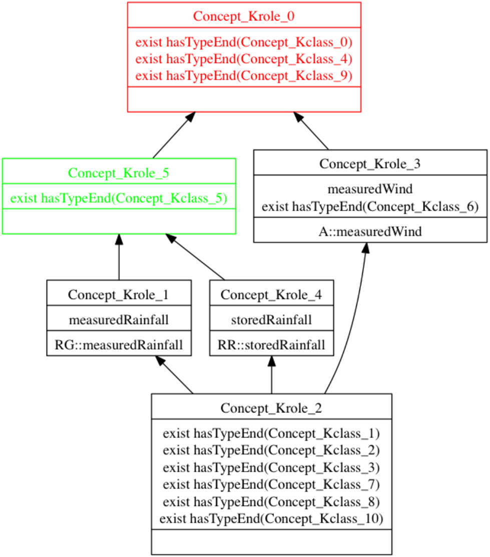
Concept_Koperation_5
PstoredRF
RR::print

Concept_Koperation_2
MCWA
R::printinfo

Concept_Koperation_4
MCWSWD
W::printinfo

Concept_Koperation_3

Reified characteristics: Role lattices



Concept_Krole_0

exist hasTypeEnd(Concept_Kclass_0)
exist hasTypeEnd(Concept_Kclass_4)
exist hasTypeEnd(Concept_Kclass_9)

Concept_Krole_5

exist hasTypeEnd(Concept_Kclass_5)

Concept_Krole_3

measuredWind
exist hasTypeEnd(Concept_Kclass_6)

A::measuredWind

Concept_Krole_1

measuredRainfall

RG::measuredRainfall

Concept_Krole_4

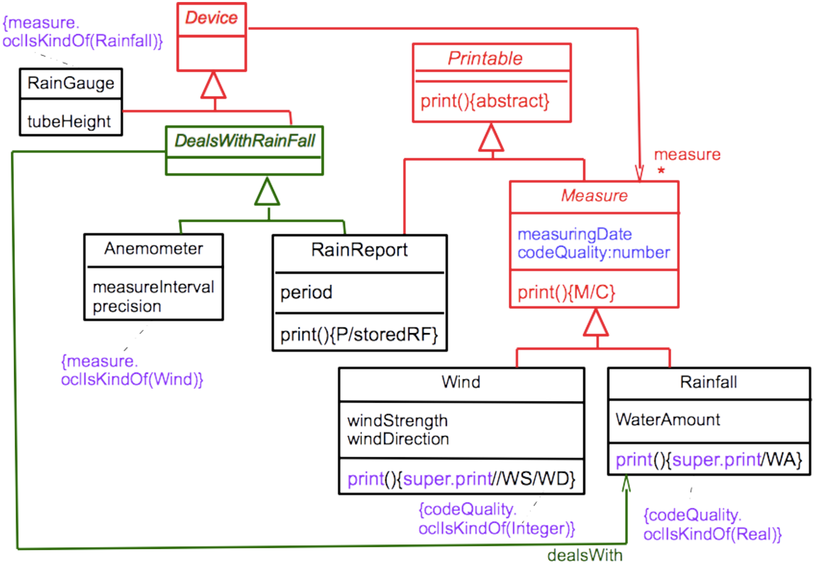
storedRainfall

RR::storedRainfall

Concept_Krole_2

exist hasTypeEnd(Concept_Kclass_1)
exist hasTypeEnd(Concept_Kclass_2)
exist hasTypeEnd(Concept_Kclass_3)
exist hasTypeEnd(Concept_Kclass_7)
exist hasTypeEnd(Concept_Kclass_8)
exist hasTypeEnd(Concept_Kclass_10)

Reified characteristics: revisited model



Clustered reified characteristics

Krole1	measuredRainfall	measuredWind
RG::measuredRainfall	×	
A::measuredWind		×

Krole2	storedRainfall
RR::storedRainfall	×

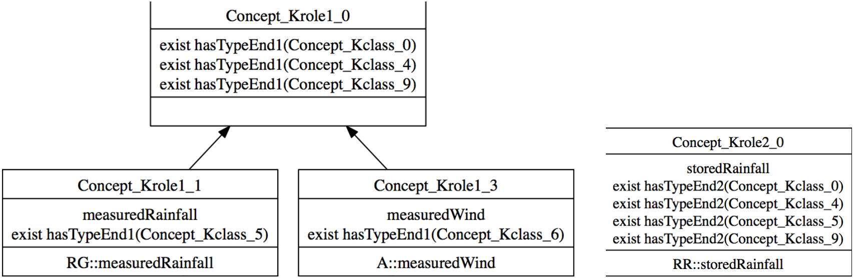
hasTypeEnd2	Rainfall	Wind
RR::storedRainfall	×	

hasRole1	RG::measuredRainfall	A::measuredWind
RainGauge	×	
Anemometer		×
Rainfall		
Wind		
RainReport		

hasRole2	RR::storedRainfall
RainGauge	
Anemometer	
Rainfall	
Wind	
RainReport	×

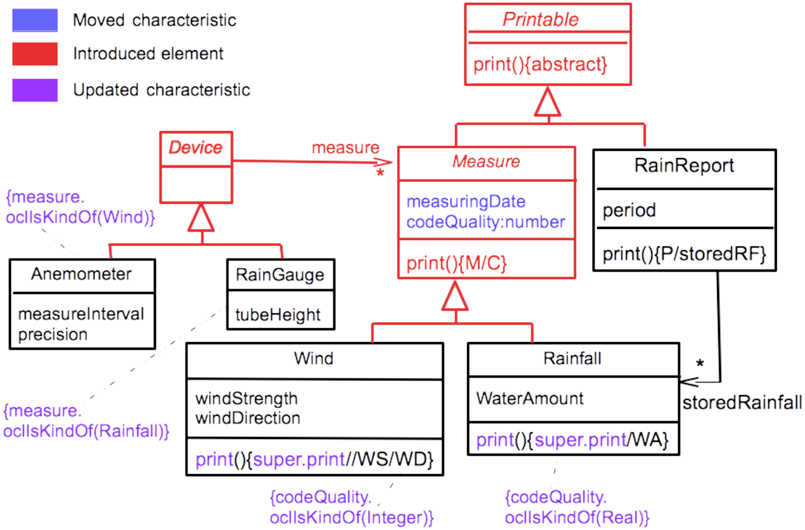
hasTypeEnd1	Rainfall	Wind
RG::measuredRainfall	×	
A::measuredWind		×

Clustered characteristics: Separated role lattices

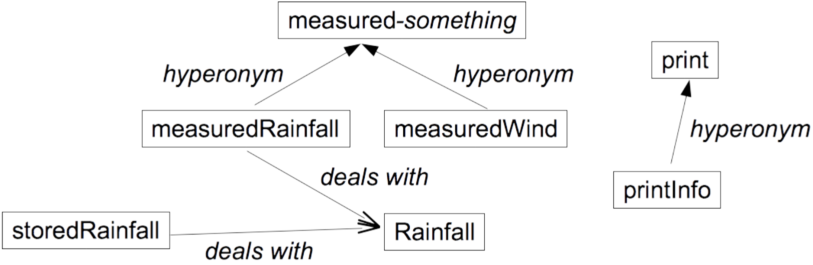

Concept_Krole1_0
exist hasTypeEnd1(Concept_Kclass_0)
exist hasTypeEnd1(Concept_Kclass_4)
exist hasTypeEnd1(Concept_Kclass_9)
Concept_Krole1_1
measuredRainfall
exist hasTypeEnd1(Concept_Kclass_5)
RG::measuredRainfall
Concept_Krole1_3
measuredWind
exist hasTypeEnd1(Concept_Kclass_6)
A::measuredWind
Concept_Krole2_0
storedRainfall
exist hasTypeEnd2(Concept_Kclass_0)
exist hasTypeEnd2(Concept_Kclass_4)
exist hasTypeEnd2(Concept_Kclass_5)
exist hasTypeEnd2(Concept_Kclass_9)
RR::storedRainfall

The same clustering has to be applied to attributes and operations to avoid "over-generalization"

Clustered characteristics: Normal form

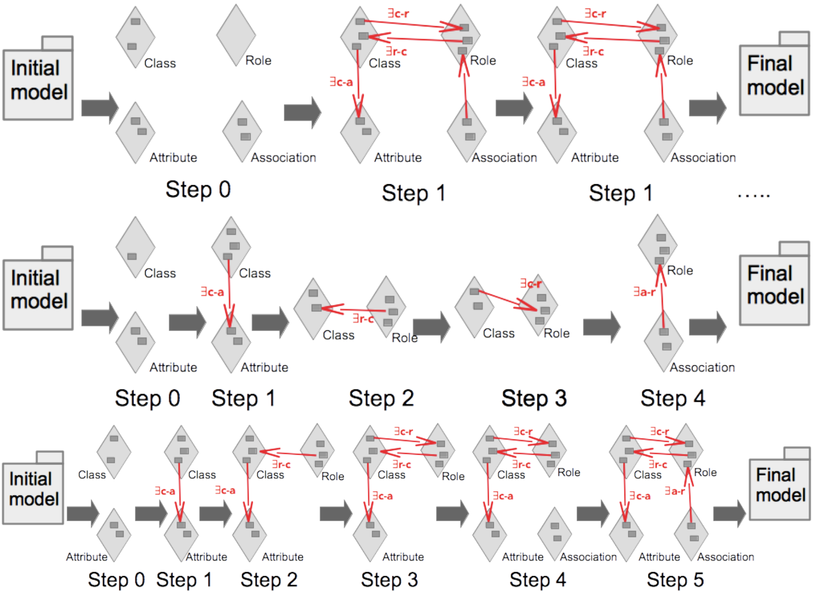


Ongoing work: Lexical resources (Falleri et al. 2010)



RCA input is tuned depending lexical resource

Ongoing work: Exploratory approaches (Miralles et al. 2015)

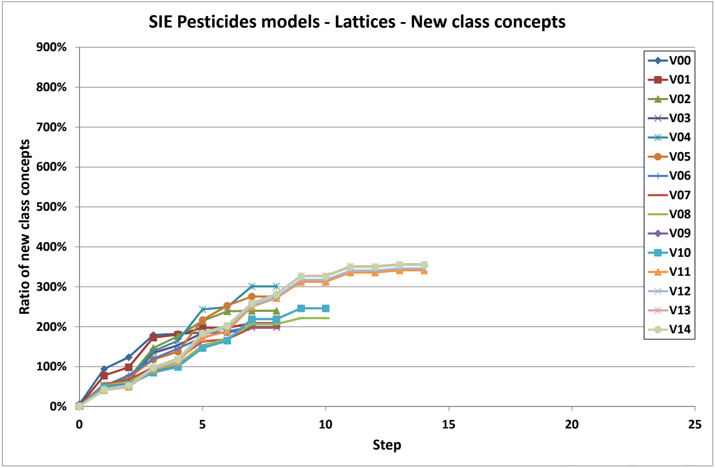


Ongoing experiment on Pesticides model (Miralles et al.)

34 to 170 classes (254 to 552 model elements)

Lattice-based RCA: 165% to 253% of new classes

→ 208 to 345 potential new factorization classes to be analyzed

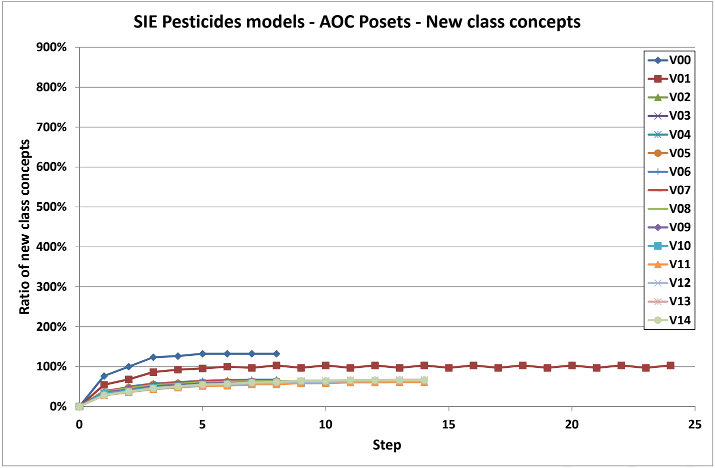


Ongoing experiment on Pesticides model (Miralles et al.)

34 to 170 classes (254 to 552 model elements)

AOC-poset-based RCA: 56% to 132% of new classes

→ 44 to 70 potential new factorization classes to be analyzed



Conclusion/Perspectives

Conclusion

- A similar approach for restructuring ontologies [Valtchev et al.]
- RCA is used for other datasets and usages (data in environmental domain, software components/services classification, software analysis for extracting assets and feature models for product lines)

Perspectives

- Towards a unique, generic tool
- Improving lexical resource acquisition and analysis
- Work on methodological aspects (exploration, guiding metrics)

Questions?

Tools: Objecteering modules[1], RCAexplore[2], Talend workflow[3], Java packages

Acknowledgements: G. Arevalo, M. Dao, H. Dicky, X. Dolques, C. Dony, J.-R. Falleri, A. Miralles, M. Lafourcade, F. Le Ber, H. Leblanc, T. Libourel, C. Nebut, A. Osman-Guédi, V. Prince, M. Rouane-Hacène, C. Roume, P. Valtchev

[1] http://www.objecteering.com/

[2] http://dolques.free.fr/rcaexplore.php

[3] http://www.talend.com/