Formal Concept Analysis, A framework for knowledge structuring and exploration. Applications to service directories and product lines.

Marianne Huchard

Formal Concept Analysis

A framework for knowledge structuring and exploration
Applications to service directories and product lines

Marianne Huchard
Montpellier University, LIRMM, France

13th IEEE Joint Conferences and workshops

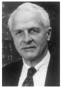
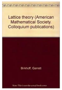
IEEE SOSE2019, BigDataService2019, MobileCloud2019, AiTest2019, DAPPCON2019, JCC2019, CCR2019, MHBDT2019, WREHE2019, SeTM2019

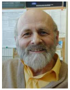
San Francisco East Bay, USA
April 5th, 2019

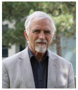
Sommaire

Introduction

Fields: symbolic AI, symbolic knowledge engineering





- lattice theory, Galois connections
  (Birkhoff, 1940; Barbut & Monjardet, 1970)
- concept lattices
  (Wille, 1982, Ganter & Wille, 1999)

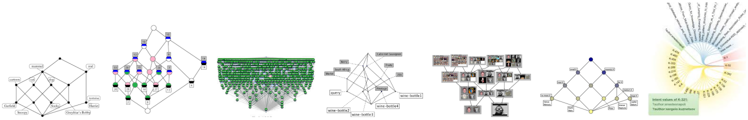
Formal concepts are "a natural feature of information representation which is as fundamental to hierarchies and object/attribute structures as set theory or relational algebra are for relational databases".

Uta Priss. 40th anniv. vol. of Annual Review of Inf. Sc. and Tech., 2006

Introduction

Knowledge structuring and exploration

- data analysis, data mining, hierarchical multi-clustering
- knowledge representation (e.g. ontology construction)
- classification, indexation (information retrieval)
- unsupervised learning (based on examples description)
- supervised learning (adding classes in description)



Credit to U. Priss, G. Greene, K. Bertet, A. Napoli, M. Alam, T. Tilley, ... et al.

Sommaire

Formal Concept Analysis

Formal Context: Simplest form: entities with characteristics

Drone	Gimbal	GPS	GLONASS	Avoidance	Headless	Altitude Hold	FT l 10	FT ge 10	FT ge 20
Syma X4S Assault 					×		×		
Syma X8G 					×		×		
Parrot Bebop 		×				×		×	
DJI Ryze Tello 						×		×	
Hubsan X4 H502S 		×			×	×		×	
Aosenma CG035 GPS FPV 	×	×			×	×		×	
DJI Mavic Air 	×	×	×	×	×	×			×
Yuneec Typoon H Pro 	×	×	×	×	×	×			×

https://www.thedronechart.com  Hum, typo, here. Typoon in World of Warcraft? Teaspoon?

Concept

A maximal group of objects (object closed set, extent)
sharing a maximal group of attributes (attribute closed set, intent)

Drone	Gimbal	GPS	GLONASS	Avoidance	Headless	Altitude Hold	FT l 10	FT ge 10	FT ge 20
syma X4S Assault					×		×		
Syma X8G					×		×		
Parrot Bebop 		×				×		×	
DJI Ryze Tello 						×		×	
Hubsan X4 H502S 		×			×	×		×	
Aosenma CG035 GPS FPV 	×	×			×	×		×	
DJI Mavic Air	×	×	×	×	×	×			×
Yuneec Typoon H Pro	×	×	×	×	×	×			×

Concept

Concept_Drone_8 — Concept identifier

Altitude Hold
FT ge 10 — Intent

DJI Ryze Tello
Parrot Bebop
Hubsan X4 H502S
Aosenma CG035 GPS FPV — Extent

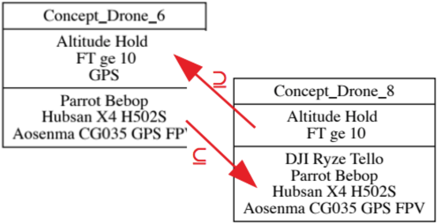
Drone	Gimbal	GPS	GLONASS	Avoidance	Headless	Altitude Hold	FT l 10	FT ge 10	FT ge 20
syma X4S Assault					×		×		
Syma X8G					×		×		
Parrot Bebop 		×				×		×	
DJI Ryze Tello 						×		×	
Hubsan X4 H502S 		×			×	×		×	
Aosenma CG035 GPS FPV 	×	×			×	×		×	
DJI Mavic Air	×	×	×	×	×	×			×
Yuneec Typoon H Pro	×	×	×	×	×	×			×

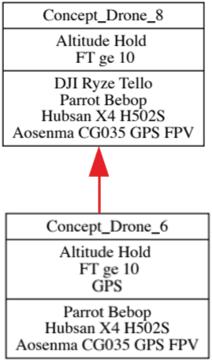
Concept

Concept_Drone_6
Altitude Hold FT ge 10 GPS
Parrot Bebop Hubsan X4 H502S Aosenma CG035 GPS FPV

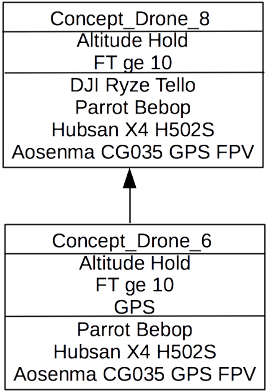
Drone	Gimbal	GPS	GLONASS	Avoidance	Headless	Altitude Hold	FT l 10	FT ge 10	FT ge 20
syma X4S Assault					×		×		
Syma X8G					×		×		
Parrot Bebop 		×				×		×	
DJI Ryze Tello						×		×	
Hubsan X4 H502S 		×			×	×		×	
Aosenma CG035 GPS FPV 	×	×			×	×		×	
DJI Mavic Air	×	×	×	×	×	×			×
Yuneec Typoon H Pro	×	×	×	×	×	×			×

Concept lattice: specialization

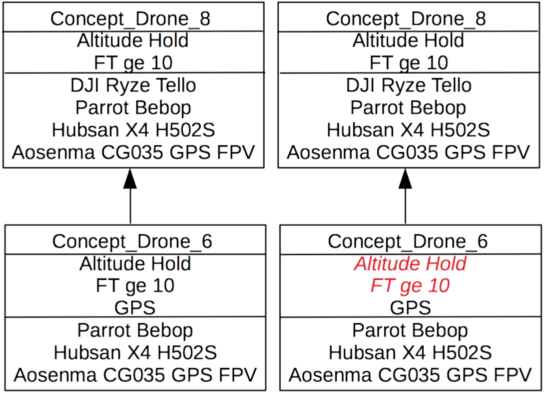




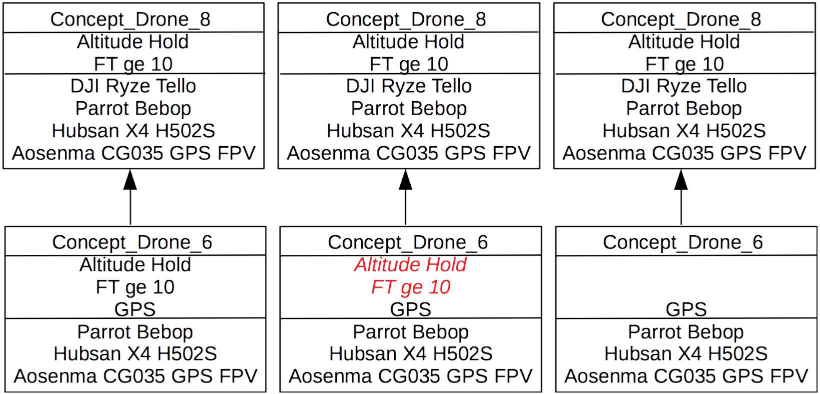
Concept lattice: top $\rightarrow$ bottom attribute inheritance



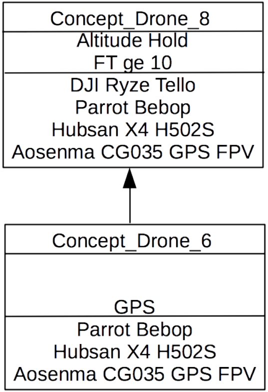
Concept lattice: top $\rightarrow$ bottom attribute inheritance



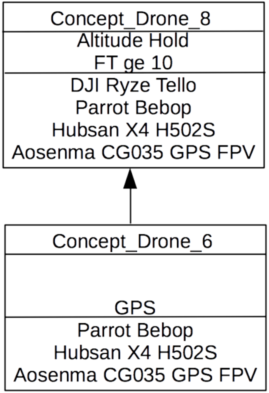
Concept lattice: top → bottom attribute inheritance

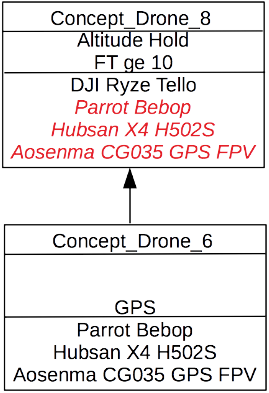


Concept lattice: bottom $\rightarrow$ top object inheritance

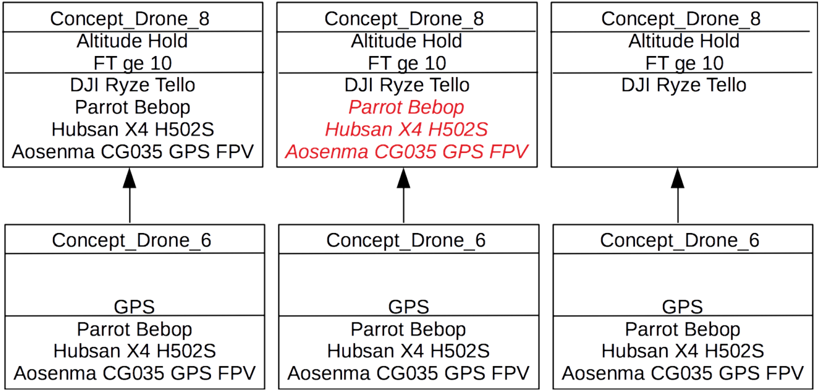


Concept lattice: bottom → top object inheritance

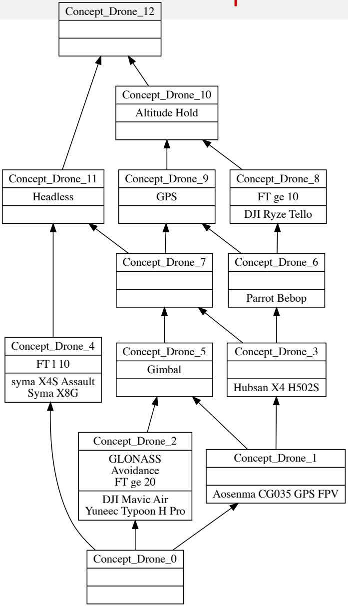




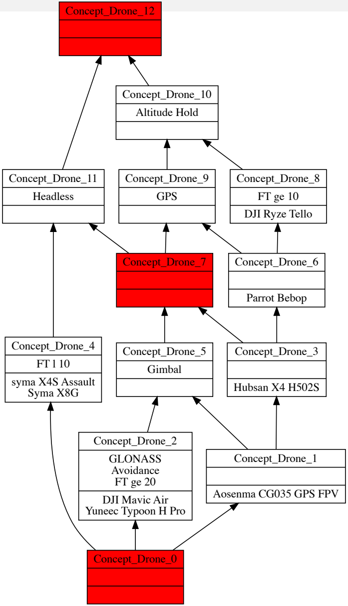
Concept lattice: bottom → top object inheritance



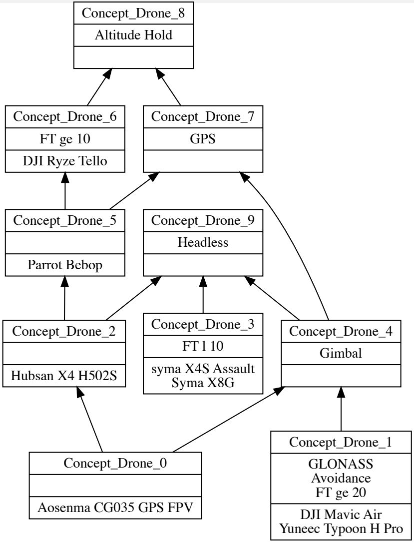
Concept lattice: simplified through inheritance



AOC-poset : restricted to Attributes and Objects introducers

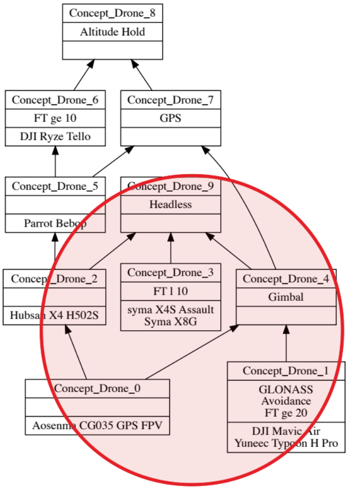


AOC-poset : restricted to Attributes and Objects introducers



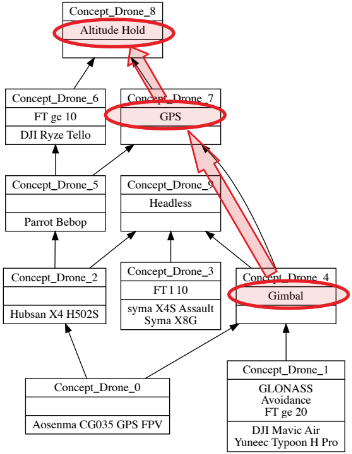
Sommaire

Knowledge engineering: Class discovery, concept extraction



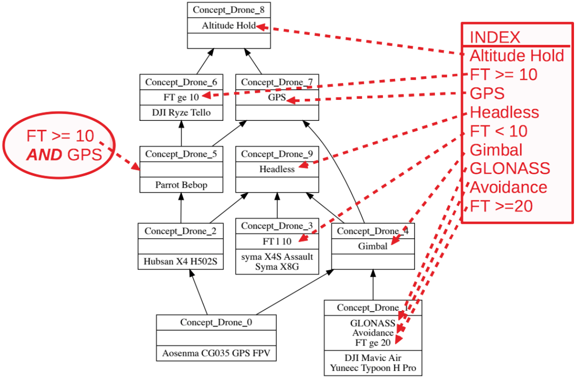
J. Poelmans, D. I. Ignatov, S. O. Kuznetsov, G. Dedene. Formal concept analysis in knowledge processing: A survey on applications. Expert Syst. Appl. 40(16) (2013)

Knowledge engineering: Implication rules extraction



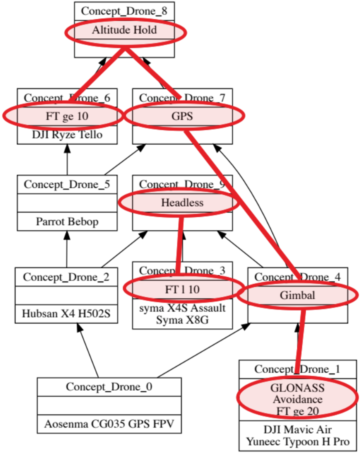
K. Bertet, C. Demko, J.-F. Viaud, C. Guérin. Lattices, closures systems and implication bases: A survey of structural aspects and algorithms. Theor. Comput. Sci. 743. (2018)

Index construction



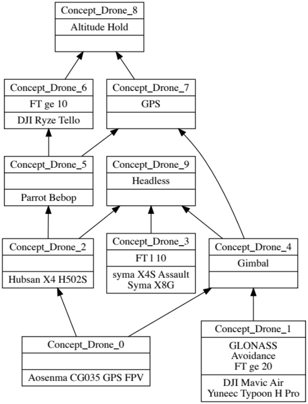
A. Gutierrez, M. Chein, M. Huchard, P. Pompidor. An Application of AOC-Posets: Indexing Large Corpuses for Text Generation Under Constraints. ISMIS 2017.

Decision tree



S. Guillas, K. Bertet, M. Visani, J-M. Ogier, and N. Girard. Some Links Between Decision Tree and Dichotomic Lattice. Concept Lattices and Applications. 2008

Ontology engineering



TBOX

AltitudeHold
GPS
(...)
Concept_Drone_7
PosDrone ≡ GPS ⊓ AltitudeHold

GLONASS ⊑ GPS
(...)

ABOX

ParrotBebop
DJIMavicAir
(...)

PosDrone(ParrotBebop)
GLONASS(DJIMavicAir)
(...)

F. Baader, B. Sertkaya. Applying Formal Concept Analysis to Description Logics. ICFCA 2004. LNAI 2961.

FCA for various data types

- Multi-valued attributes : integers, double, terms, structures, symbolic objects, etc. (Ganter et Wille, Polaillon, ...)
- Fuzzy (Belohlavek, Yahia et al., ...)
- Values taxonomies (Godin et al., Carpineto et Romano, ...)
- Logical description (Chaudron et al., Ferré et al., ...)
- Graphs (Ganter and Kuznetsov, Liquière, Prediger et Wille, Kötters et al., Ferré et al....)
- Relations (Priss, Rouane et al., ...)
- Polyadic (Sacarea, Tronca et al.)
- Pattern Structures (Kuznetsov, Napoli et al.)
- ....

Ressources

Books

- B. Ganter, R. Wille. Formal Concept Analysis, Mathematical Foundations, Springer-Verlag Berlin Heidelberg, 1999
- C. Carpineto, G. Romano. Concept Data Analysis: Theory and Applications John Wiley & Sons, 2004
- N. Caspard, B. Leclerc, B. Monjardet. Finite Ordered Sets: Concepts, Results and Uses. Cambridge University Press 2012.

Web page of Uta Priss

links to FCA software http://www.upriss.org.uk/fca/fca.html

Sommaire

A Flavor of Relational Concept Analysis

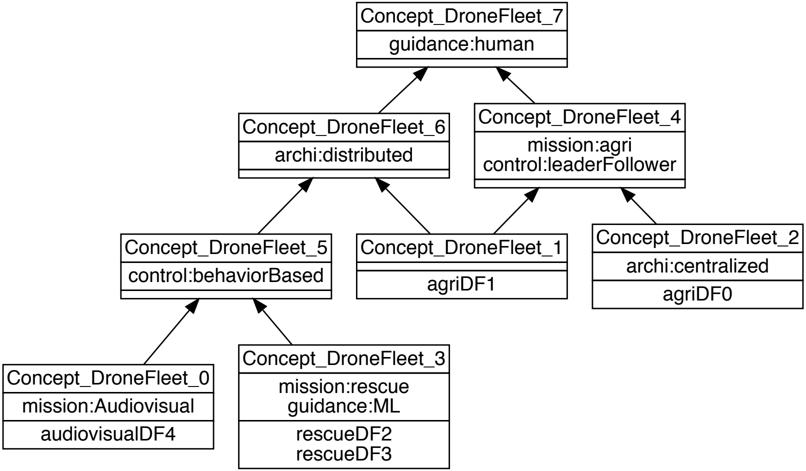
- Extends the purpose of FCA for taking into account object categories and links between objects
- Main principles:
  - a relational model based on an entity-relationship model
  - integrate relations between objects as *relational* attributes
  - iterative process
- RCA provides a set of interconnected 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

Drone fleet (Formal context)

DroneFleet	mission:agri	mission:rescue	mission:Audiovisual	archi:centralized	archi:distributed	guidance:human	guidance:ML	control:leaderFollower	control:behaviorBased
agriDF0	×			×		×		×	
agriDF1	×				×	×		×	
rescueDF2		×			×	×	×		×
rescueDF3		×			×	×	×		×
audiovisualDF4			×		×	×			×

Drone fleet (AOC-poset)



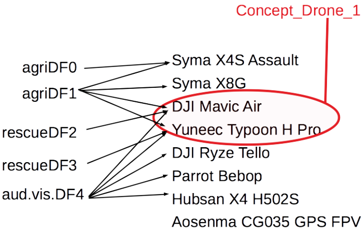
Drone fleet 2 Drone (Relational Context)

DroneFleet	Syma X4S Assault	Syma X8G	Parrot Bebop	DJI Ryze Tello	Hubsan X4 H502S	Aosenma CG035 GPS FPV	DJI Mavic Air	Yuneec Typoon H Pro
agriDF0	×							
agriDF1	×	×					×	×
rescueDF2							×	
rescueDF3								×
audiovisualDF4			×	×	×		×	×

Understanding scaling quantifiers

`rescueDF2` and `rescueDF3` do not share concrete drone types, but they share the fact that all their drones with `GLONASS`, `GPS`, `FT` ≥ 20, etc.

Relational attribute: $\exists\forall contains(Concept\_Drone\_1)$

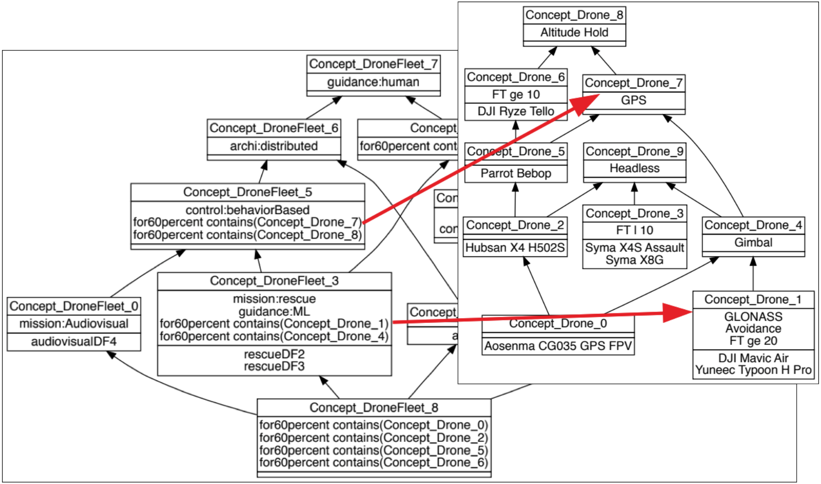


Drone Fleet	Examples of Relational Attributes
agriDF1, rescueDF2	∃ contains (Concept_Drone_1)
rescueDF2, rescueDF3	∃∀ contains (Concept_Drone_1)
agriDF1	∃∀≥50% contains (Concept_Drone_1)
aud.vis.DF4	∃∀≥40% contains (Concept_Drone_1)
agriDF1, aud.vis.DF4	∃⊇ contains (Concept_Drone_1)
rescueDF2, rescueDF3	∃⊇≥50% contains (Concept_Drone_1)

Drone fleet extended by relations to their drones

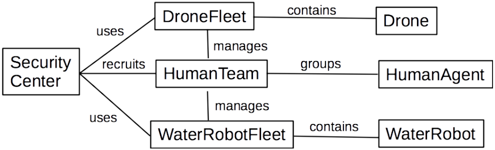
DroneFleet	mission:agri	mission:rescue	mission:Audiovisual	archi:centralized	archi:distributed	guidance:human	guidance:ML	control:leaderFollower	control:behaviorBased	∃≥60% contains(Concept_Drone_0)	∃≥60% contains(Concept_Drone_1)	∃≥60% contains(Concept_Drone_3)	∃≥60% contains(Concept_Drone_2)	∃≥60% contains(Concept_Drone_4)	∃≥60% contains(Concept_Drone_5)	∃≥60% contains(Concept_Drone_6)	∃≥60% contains(Concept_Drone_7)	∃≥60% contains(Concept_Drone_9)	∃≥60% contains(Concept_Drone_8)
agriDF0	×			×		×		×				×						×	
agriDF1	×				×	×		×										×	
rescueDF2		×			×	×	×		×		×			×			×	×	×
rescueDF3		×			×	×	×		×		×			×			×	×	×
audiovisualDF4			×		×	×			×								×		×

Drone fleets extended by relations to their drones (AOC-posets)



Rescue fleets have a majority of drones with GLONASS, Avoidance system and Flight Time $\geq$ 20mn

RCA in the general case



An iterative process

- Complex model with paths and cycles of any length
- Objects groups (concepts) are propagated along the paths and the cycles, step after step
- The process stops when no new concept appears

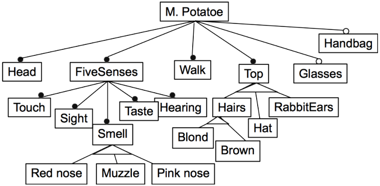
Tool

- http://dataqual.engees.unistra.fr/logiciels/rcaExplore

Sommaire

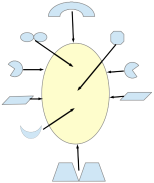
Focus 1: Product Line Engineering

Feature model



Assets

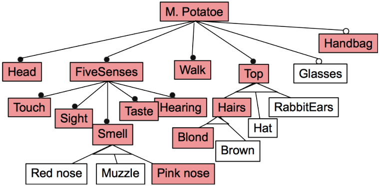
Variable Architecture



Derived Products

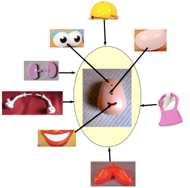
Building a product

Feature selection

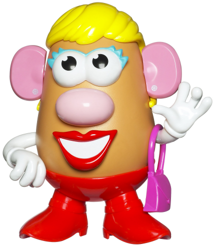


Selected Assets

Implemented Architecture



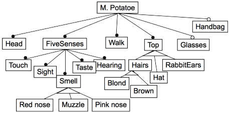
Product



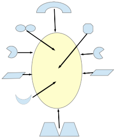
Product Line Reverse Engineering

Similar Products
developed in undisciplined manner

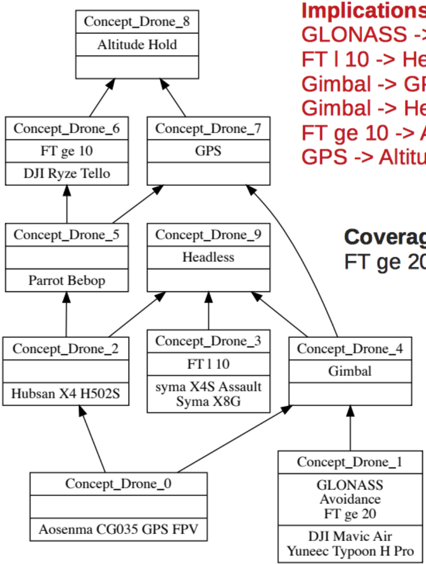
Which Feature model?



Which Assets?
Which Architecture?



Variability extraction with FCA and RCA



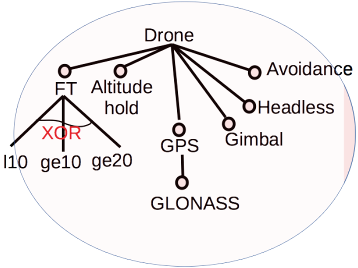
Implications
GLONASS -> Gimbal
FT l 10 -> Headless
Gimbal -> GPS
Gimbal -> Headless
FT ge 10 -> Altitude Hold
GPS -> Altitude Hold

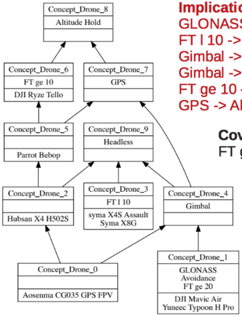
Coverage (candidate groups)
FT ge 20, FT ge 10, FT l 10

Co-occurrences
GLONASS <-> Avoidance
GLONASS <-> FT ge 20

Mutex
FT ge 10 → not T l 10
FT ge 10 → not GLONASS
Gimbal → not FT l 10
FT l 10 → not GLONASS

Theoretical variability vs. realized variability





Implications
GLONASS -> Gimbal
FT l 10 -> Headless
Gimbal -> GPS
Gimbal -> Headless
FT ge 10 -> Altitude Hold
GPS -> Altitude Hold

Coverage (candidate groups)
FT ge 20, FT ge 10, FT l 10

Co-occurrences
GLONASS <-> Avoidance
GLONASS <-> FT ge 20

Mutex
FT ge 10 → not T l 10
FT ge 10 → not GLONASS
Gimbal → not FT l 10
FT l 10 → not GLONASS

Feature Model
Theoretical variability

AOC-poset : Realized variability

Contributions to Product Lines

- Assist the construction of variability representations, e.g. Feature Models
- Assist the composition by union and intersection
- Exploring a product family
- Using RCA to represent interconnected product lines (like Drone fleets vs. Drones)

F. Loesch and E. Ploedereder. 2007. Restructuring Variability in Software Product Lines using Concept Analysis of Product Configurations. In Proc. of the 11th Europ. Conf. on Software Maintenance and Reengineering (CSMR'07)
S. She, U. Ryssel, N. Andersen, A. Wasowski, K. Czarnecki. Efficient synthesis of feature models. Information & Software Technology 56(9), 2014
J. Carbonnel, M. Huchard, C. Nebut. Modelling equivalence classes of feature models with concept lattices to assist their extraction from product descriptions. Journal of Systems & Software, Vol. 152, 2019

Sommaire

Focus 2: Services, Service Workflows

Approaches

- Select one Web service for a particular task
- Planning approaches to build a workflow satisfying a given input/output
- Instantiating an abstract workflow by mining Web services for each task

Issues

- Satisfy the expected functionality
- Satisfy QoS (Quality of Service) attributes
- Keep alternative choices (for backup concern)

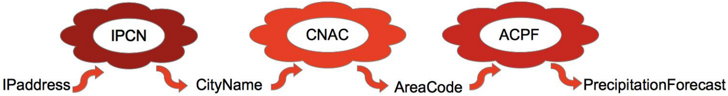
Hypotheses

Context

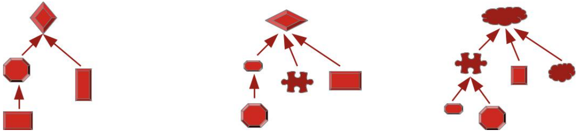
- Service is one functionality
  - complex WS can be considered as several services
- Semantic Web services: service input and output are described by concepts of an ontology
  - OWL-S, SAWSDL, WSMO, ...
- QoS: quality of service attributes are known
  - availability, response time, reputation, ...

Instantiating an abstract task workflow

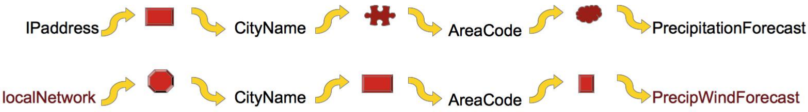
Abstract task workflow (credit Z. Azmeh)



Alternative concrete web services

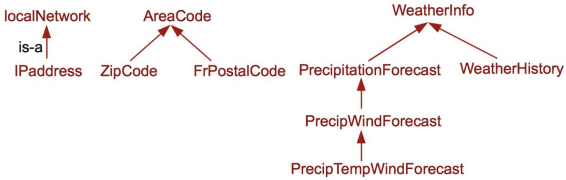


Alternative concrete workflows



Web Service Description

Ontology concepts for input/output

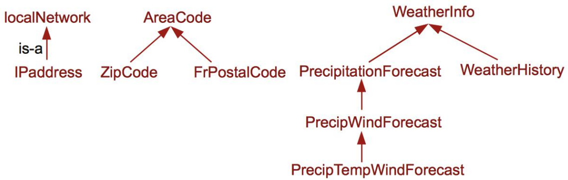


QoS attributes

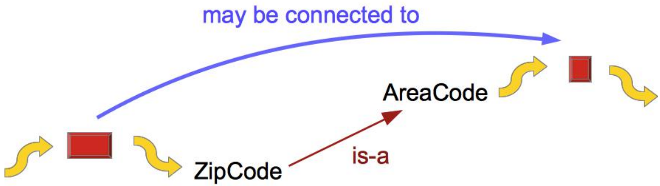
- availability: very low, low, medium, high, very high (best)
- response time: very low (best), low, medium, high, very high

Service connection

Ontology concepts

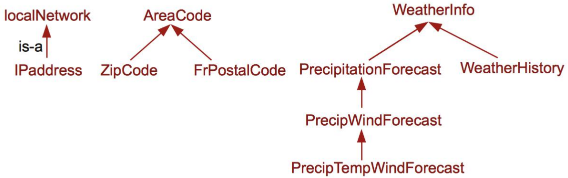


Connection principle

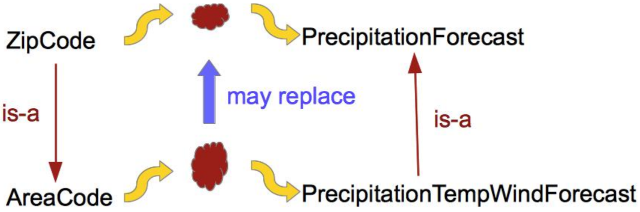


Service replacement

Ontology concepts



Replacement principle (require less, provide more)



WS Classification guided by Replacement
Formal context: IPCN

A query object `query_IPCN` is classified to identify easily the potential answers

IPCN	out:CN	in:IPadd	in:locNtw	vlA	lA	mA	hA	vhA	vlRT	lRT	mRT	hRT	vhRT
IP_to_CN_11	×	×		×							×	×	×
IP_to_CN_12	×	×		×	×							×	×
IP_to_CN_13	×	×		×	×	×	×	×		×	×	×	×
IP_to_CN_14	×	×		×	×	×	×	×	×	×	×	×	×
IP_to_CN_15	×	×	×	×	×	×	×	×	×	×	×	×	×
query_IPCN	×	×		×	×					×	×	×	×

WS Classification guided by Replacement
Formal context: CNAC

A query object `query_CNAC` is classified to identify easily the potential answers

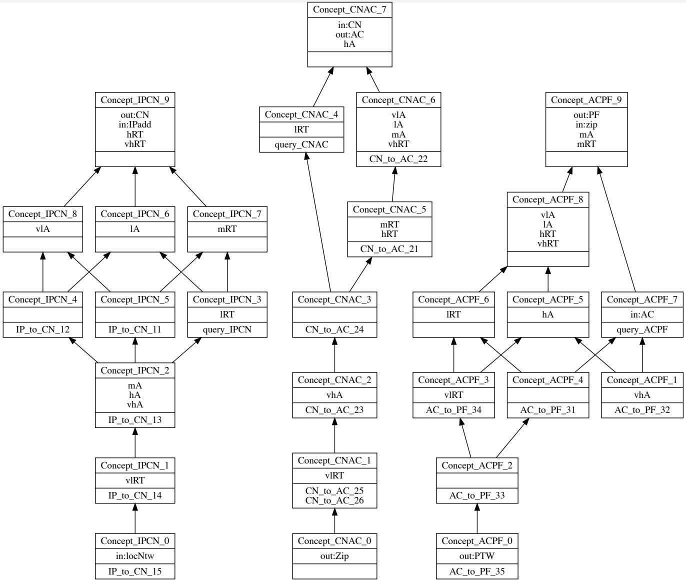
CNAC	in:CN	out:AC	out:Zip	vlA	lA	mA	hA	vhA	vlRT	lRT	mRT	hRT	vhRT
CN_to_AC_21	×	×		×	×	×	×				×	×	×
CN_to_AC_22	×	×		×	×	×	×						×
CN_to_AC_23	×	×		×	×	×	×	×		×	×	×	×
CN_to_AC_24	×	×		×	×	×	×			×	×	×	×
CN_to_AC_25	×	×		×	×	×	×	×	×	×	×	×	×
CN_to_AC_26	×	×		×	×	×	×	×	×	×	×	×	×
query_CNAC	×	×		×	×	×	×			×	×	×	×

WS Classification guided by Replacement
Formal context: ACPF

A query object `query_ACPF` is classified to identify easily the potential answers

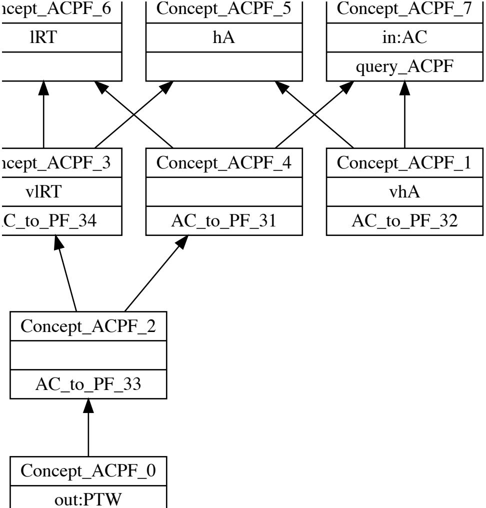
ACPF	out:PF	out:PTW	in:AC	in:zip	vlA	lA	mA	hA	vhA	vlRT	lRT	mRT	hRT	vhRT
AC_to_PF_31	×		×	×	×	×	×				×	×	×	×
AC_to_PF_32	×		×	×	×	×	×	×	×			×	×	×
AC_to_PF_33	×		×	×	×	×	×	×		×	×	×	×	×
AC_to_PF_34	×			×	×	×	×	×		×	×	×	×	×
AC_to_PF_35	×	×	×	×	×	×	×	×		×	×	×	×	×
query_ACPF	×		×	×	×	×	×					×	×	×

Concept Lattices of Web Services



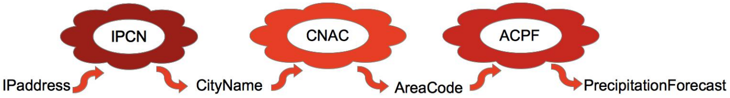
Concept Lattices of Web Services

Services introduced below `Concept_ACPF_7` are below `query_ACPF`, they satisfy the query

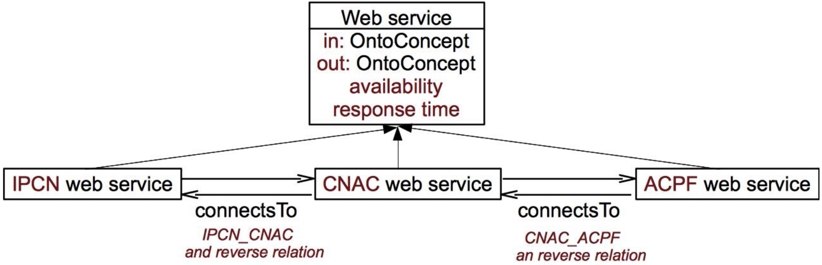


Using the classification for workflow instantiation

Abstract task workflow (credit Z. Azmeh)



Relational Data Model (scaling quantifier $\exists$)



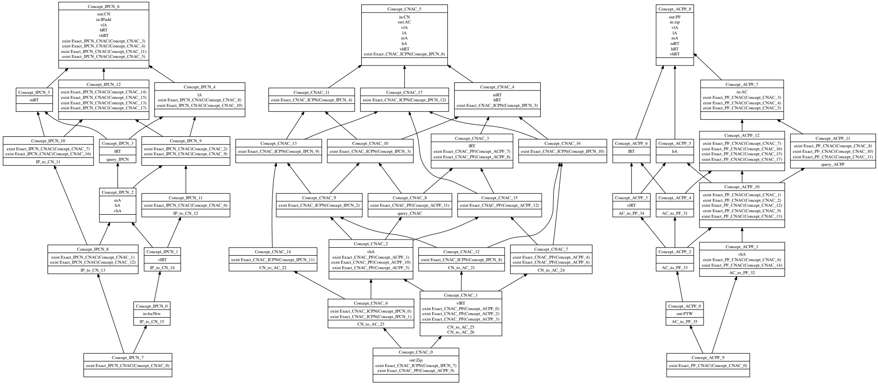
Relational contexts: connects to (reverse relation is not shown)

IPCN_CNAC	CN_to_AC_21	CN_to_AC_22	CN_to_AC_23	CN_to_AC_24	CN_to_AC_25	CN_to_AC_26	query_ACPF
IP_to_CN_11		×		×			
IP_to_CN_12			×				
IP_to_CN_13		×			×	×	
IP_to_CN_14			×				
IP_to_CN_15			×				
query_IPCN							×

Relational contexts: connects to (reverse relation is not shown)

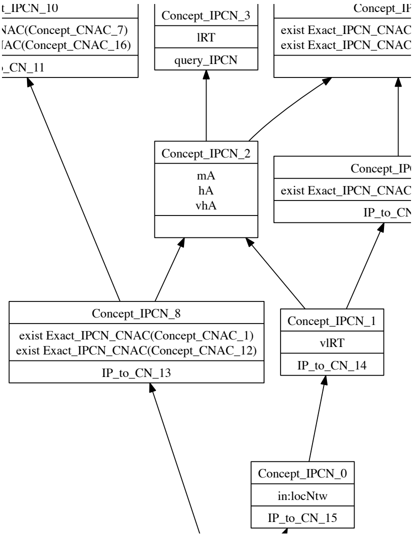
CNAC_ACPF	AC_to_PF_31	AC_to_PF_32	AC_to_PF_33	AC_to_PF_34	AC_to_PF_35	query_ACPF
CN_to_AC_21						
CN_to_AC_22						
CN_to_AC_23		×				
CN_to_AC_24	×					
CN_to_AC_25		×	×		×	
CN_to_AC_26		×	×		×	
query_CNAC						×

Lattice Family after RCA process

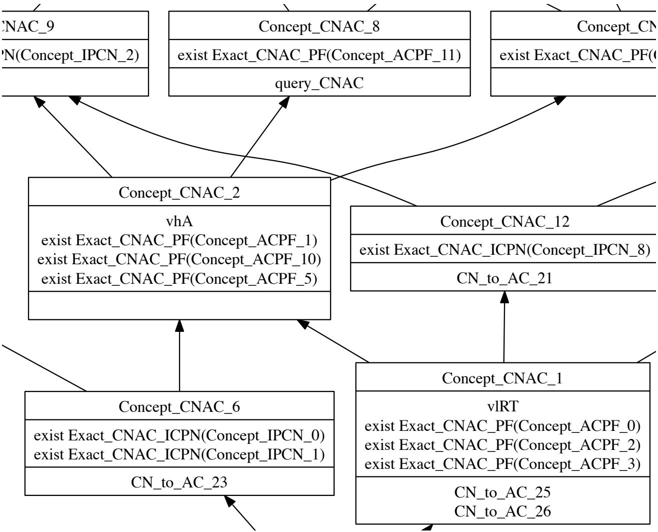


WS 13, 14, 15 satisfy query (with improved availability, response time or input);

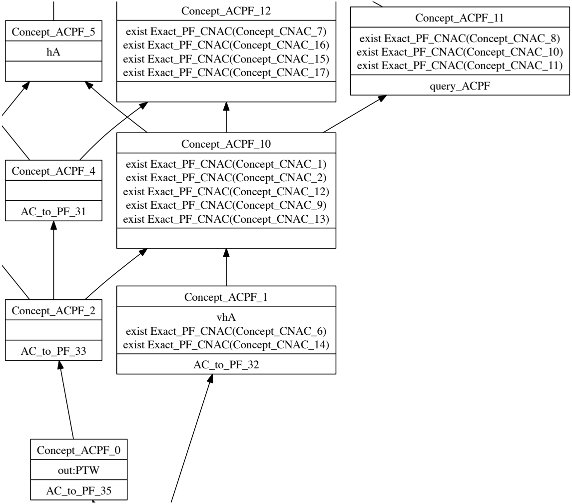
each connects to at least one service of Concept CNAC 8



WS 23, 25, 26 satisfy query (with improved availability and response time); 25, 26 almost equivalent; each connects to at least one service from Concept ACPF 11

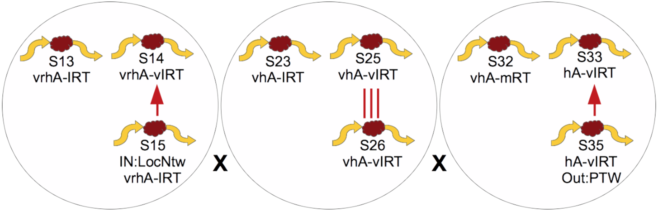


WS 32, 33, 35 satisfy query (with improved availability for 32); 35 more specific than 33 (gives more precision on weather)

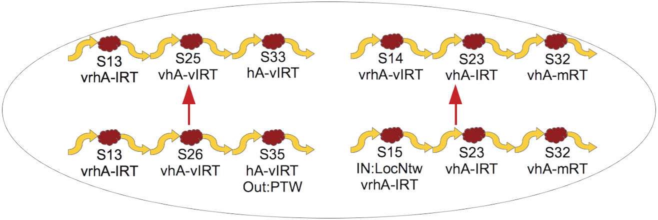


Candidate concrete workflows

Specialization (replacement and improvement) on services



Specialization (replacement and improvement) between candidate concrete workflows (excerpt)



Conclusion

The approach provides web services that:

- satisfy QoS properties ; can connect properly

Besides, several opportunities are:

- highlighted; classified along QoS properties and functionalities

Tracks for future research

- Encoding more information (e.g. various details on functionalities)
- Tools for visualization and navigation in and between lattices are needed
- Need for aligning ontologies for different web service sets
- Can be used to organize results of planning-based approaches

References

- Software component classification
  - N.Aboud, G. Arévalo, O. Bendavid, J.-R. Falleri, N. Haderer, M. Huchard, C. Tibermacine, C. Urtado, S. Vauttier. Building Hierarchical Component Directories JOT, 18(1)2:1-37, 2019
- Web service composition with backups
  - Zeina Azmeh, Marianne Huchard, Fady Hamoui, Naouel Moha: From Abstract to Executable BPEL Processes with Continuity Support. ICWS 2012: 368-375
- Querying RCA results
  - Zeina Azmeh, Marianne Huchard, Amedeo Napoli, Mohamed Rouane Hacene, Petko Valtchev: Querying Relational Concept Lattices. CLA 2011: 377-392
- Classifying semantic web services for workflow instantiation
  - Sara El Hassad, Master thesis, June 2013, Montpellier University (supervised by N. Moha, C. Tibermacine, M. Huchard)

Sommaire

Conclusion

FCA

- Objects + Attributes = Concept hierarchy
- Many extensions to take into account many data types

RCA

- Objects + Attributes + Relations = Interconnected Concept hierarchies
- Integrate information of complex entity-relationships
- Iterative process for concept propagation
- Tunable (algorithms, steps, quantifiers, excerpt of ER diagram)

Perspectives

Methodology

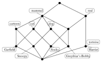
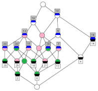
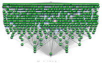
- On-demand algorithms
- Visualization
- Assistive tool for data exploration

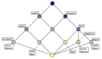
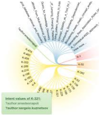
On-going applications

- Product Lines + service workflows
- Data exploration (e.g., environmental domain)

⇒ Exploring FCA and statistical ML

- as a pre-processing, as a post-processing
- in other combinations of methodological patterns, see inspiring lectures of Frank van Harmelen

Thank you!