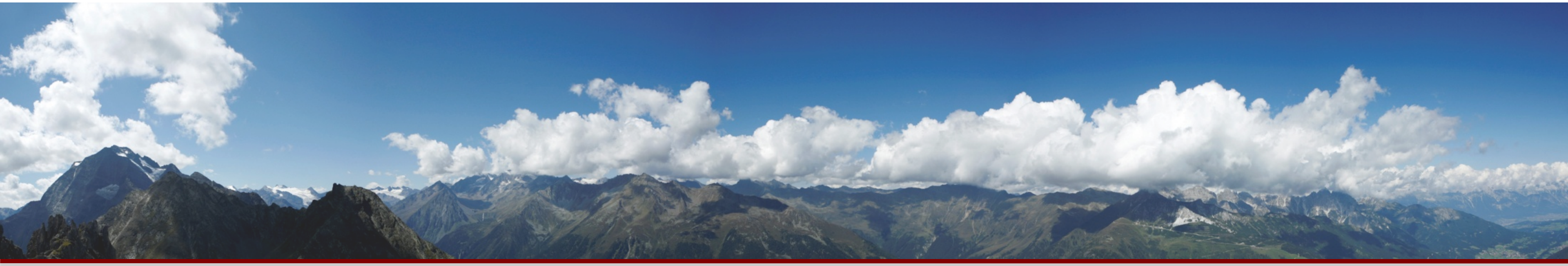
Artificial Intelligence
WS 2015/16

Formal Concept Analysis

Anna Fensel



Where are we?

STI · INNSBRUCK

#	Title
1	Introduction
2	Propositional Logic
3	Predicate Logic
4	Reasoning
5	Search Methods
6	CommonKADS
7	Problem-Solving Methods
8	Planning
9	Software Agents
10	Rule Learning
11	Inductive Logic Programming
12	**Formal Concept Analysis**
13	Neural Networks
14	Semantic Web and Services

Outline

- Motivation
- Technical Solution
  - Definitions
  - Concept Lattice
- Illustration by a Larger Example
- Extension
- Summary

From data to their conceptualization

MOTIVATION

Introduction

STI · INNSBRUCK

- In many applications we deal with huge amount of data
  - E.g. insurance company records
- Data by itself are not useful to support decisions
  - E.g. can I make an insurance contract to this new customer or it is too risky?
- Thus we need a set of methods to generate from a data set a “summary” that represent a conceptualization of the data set
  - E.g. what are similarities among different customers of an insurance company that divide them in different risk classes?
  - Age >25, City=Innsbruck => Low Risk
- This is a common task that is needed in several domains to support data analysis
  - Analysis of children suffering from diabetes
  - Marketing analysis of store departments or supermarkets
- Formal Concept Analysis is a technique that enables resolution of such problems

Formal Concept Analysis

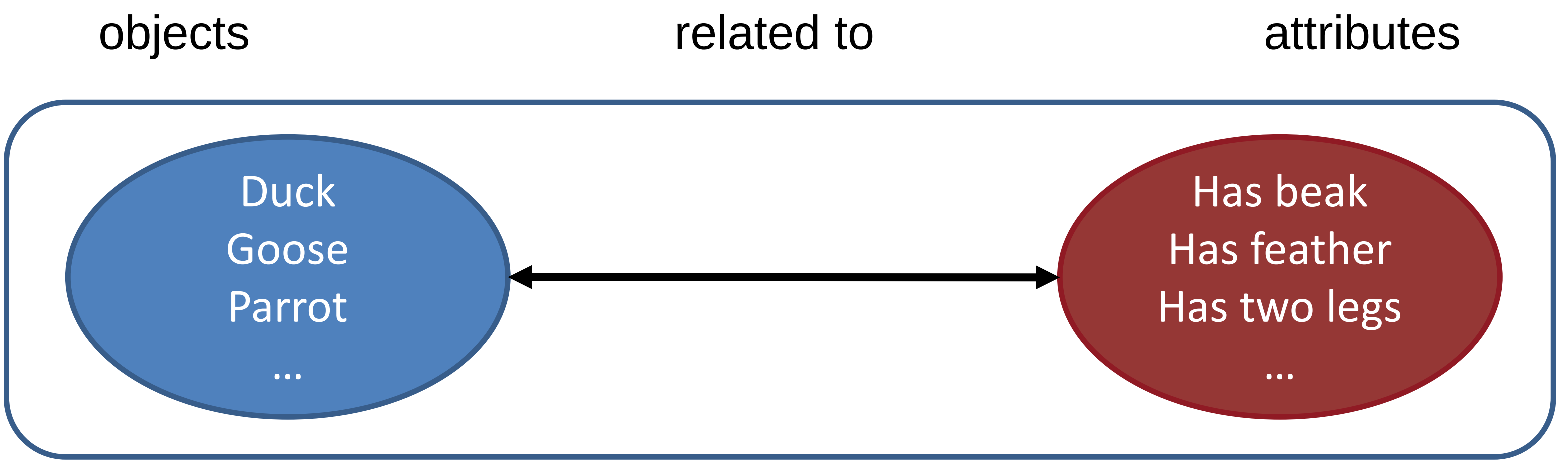
TECHNICAL SOLUTION

What is a Concept?

- What drives us to call an object a “bird”?
- Every object having certain attributes is called “bird”:
  - A bird has feathers
  - A bird has two legs
  - A bird has a beak
  - …
- All objects having these attributes are called “birds”:
  - Duck, goose, owl and parrot are birds
  - Penguins are birds, too
  - …

What is a Concept?

- This description of the concept "bird" is based on sets of



- Objects, attributes and a relation form a formal concept

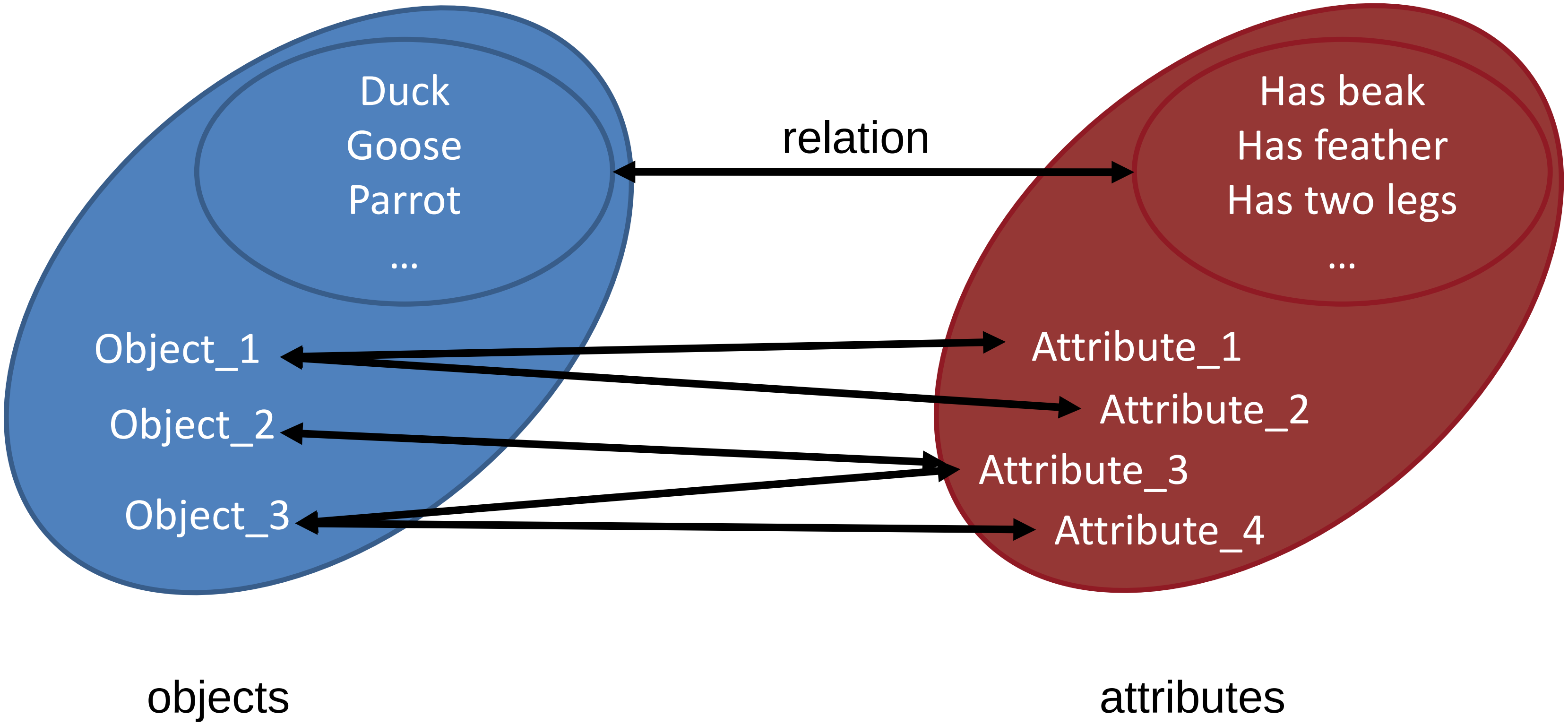
What is a Concept?

- So, a formal concept is constituted by two parts

- having a certain relation:
  - every object belonging to this concept has all the attributes in B
  - every attribute belonging to this concept is shared by all objects in A
- A is called the concept's extent
- B is called the concept's intent

The Universe of Discourse

- A repertoire of objects and attributes (which might or might not be related) constitutes the „context“ of our considerations



DEFINITIONS

- **Formal Concept Analysis** is a method used for investigating and processing explicitely given information, in order to allow for meaningful and comprehensive interpretation
  - An **analysis** of data
  - Structures of formal abstractions of **concepts** of human thought
  - **Formal** emphasizes that the concepts are mathematical objects, rather than concepts of mind

Formal Concept Analysis

- Formal Concept Analysis takes as input a matrix specifying a set of objects and the properties thereof, called attributes, and finds both all the “natural” clusters of attributes and all the “natural” clusters of objects in the input data, where
  - a “natural” object cluster is the set of all objects that share a common subset of attributes, and
  - a “natural” property cluster is the set of all attributes shared by one of the natural object clusters
- Natural property clusters correspond one-for-one with natural object clusters, and a concept is a pair containing both a natural property cluster and its corresponding natural object cluster
- The family of these concepts obeys the mathematical axioms defining a lattice, and is called a concept lattice

Definition: Formal Context

STI · INNSBRUCK

- **Context:** A triple (G, M, I) is a (formal) context if
  - G is a set of **objects** (Gegenstand)
  - M is a set of **attributes** (Merkmal)
  - I is a binary relation between G and M called **incidence**

- **Incidence** relation: $I \subseteq G \times M$
  - if $g \in G$, $m \in M$ in $(g,m) \in I$, then we know that "*object g has attribute m*" and we write gIm

- **Derivation** operators:
  - For $A \subseteq G$, $A' = \{m \in M \mid (g,m) \in I \text{ for all } g \in A\}$
  - For $B \subseteq M$, $B' = \{g \in G \mid (g,m) \in I \text{ for all } m \in B\}$

Definition: Formal Concept

STI · INNSBRUCK

- A pair (A,B) is a **formal concept** of (G,M,I) if and only if
  - $A \subseteq G$
  - $B \subseteq M$
  - A‘ = B, and A = B‘

- Note that at this point the **incidence relationship is closed**; i.e. all objects of the concept carry all its attributes and that there is no other object in G carrying all attributes of the concept

- A is called the **extent** (Umfang) of the concept (A,B)
- B is called the **intent** (Inhalt) of the concept (A,B)

Generating Formal Concepts

STI · INNSBRUCK

- Using the derivation operators we can derive formal concepts from our formal context with the following routine:
  1. Pick a set of objects A
  2. Derive the attributes A'
  3. Derive (A')'
  4. (A'',A') is a formal concept

- A dual approach can be taken starting with an attribute

Example

	small	medium	big	twolegs	fourlegs	feathers	hair	fly	hunt	run	swim	mane	hooves
dove	x	.	.	x	.	x	.	x	.	.	.	.	.
hen	x	.	.	x	.	x	.	.	.	.	.	.	.
duck	x	.	.	x	.	x	.	x	.	.	x	.	.
goose	x	.	.	x	.	x	.	x	.	.	x	.	.
owl	x	.	.	x	.	x	.	x	x	.	.	.	.
hawk	x	.	.	x	.	x	.	x	x	.	.	.	.
eagle	.	x	.	x	.	x	.	x	x	.	.	.	.
fox	.	x	.	.	x	.	x	.	x	x	.	.	.
dog	.	x	.	.	x	.	x	.	.	x	.	.	.
wolf	.	x	.	.	x	.	x	.	x	x	.	x	.
cat	x	.	.	.	x	.	x	.	x	x	.	.	.
tiger	.	.	x	.	x	.	x	.	x	x	.	.	.
lion	.	.	x	.	x	.	x	.	x	x	.	x	.
horse	.	.	x	.	x	.	x	.	.	x	.	x	x
zebra	.	.	x	.	x	.	x	.	.	x	.	x	x
cow	.	.	x	.	x	.	x	.	.	.	.	.	x

1. Pick any set of objects A, e.g. A={duck}.
2. Derive the attributes A'={small, two legs, feathers, fly, swim}
3. Derive (A')'={small, two legs, feathers, fly, swim}'={duck, goose}
4. (A'',A')=({duck, goose},{small, two legs, feathers, fly, swim}) is a *formal concept.*

[Bastian Wormuth and Peter Becker, Introduction to Formal Concept Analysis, 2nd International Conference of Formal Concept Analysis]

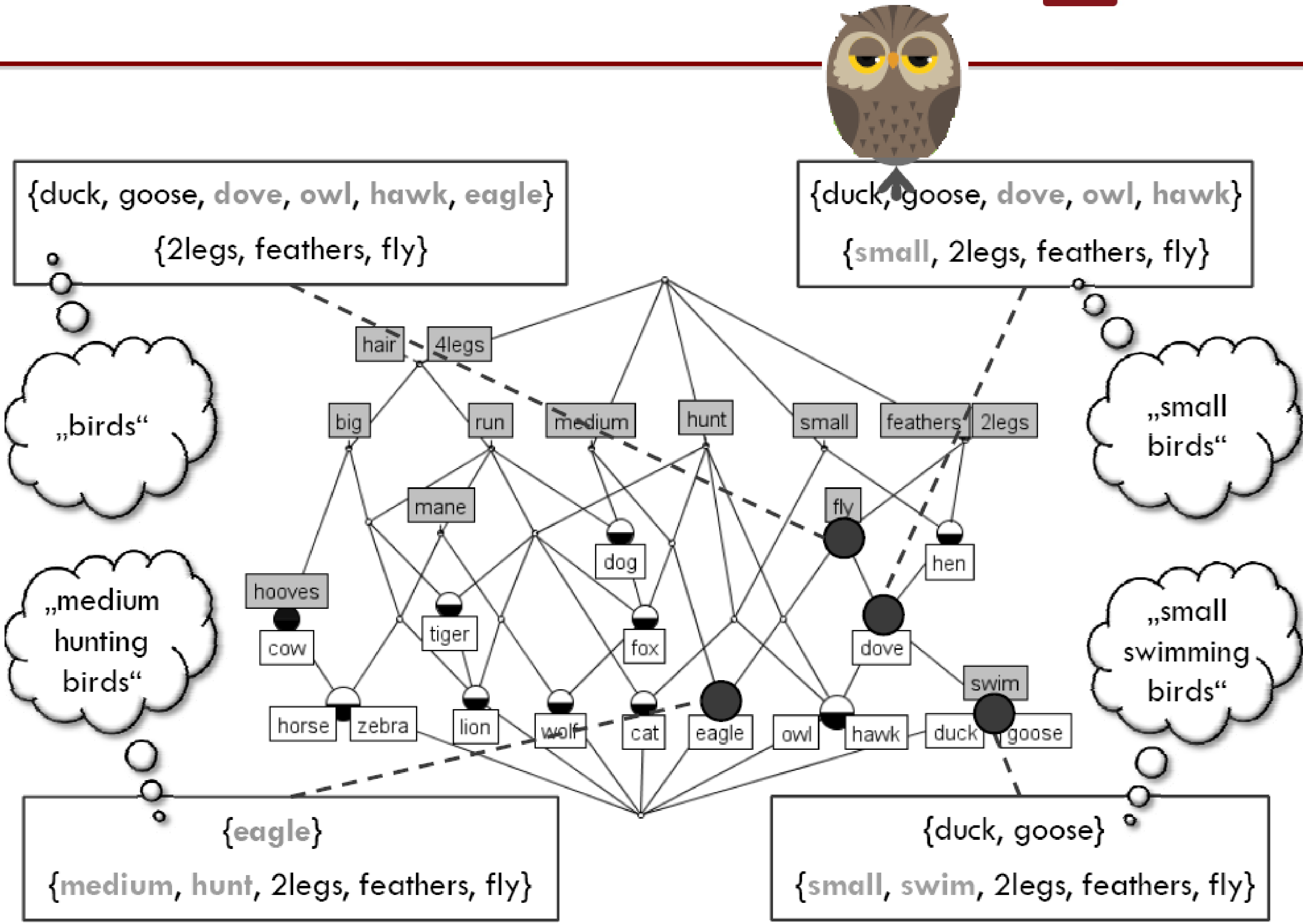
CONCEPT LATTICE

Definition: Concept Lattice

STI · INNSBRUCK

- The concepts of a given context are naturally ordered by a subconcept-superconcept relation:
  - $(A_1,B_1) \leq (A_2,B_2) :\Leftrightarrow A_1 \subseteq A_2$ $(\Leftrightarrow B_2 \subseteq B_1)$
- The ordered set of all formal concepts in (G,M,I) is denoted by $\mathfrak{B}$(G,M,I) and is called **concept lattice** (Begriffsverband)
- A concept lattice consists of the set of concepts of a formal context and the subconcept-superconcept relation between the concepts

Example


{duck, goose, dove, owl, hawk, eagle}
{2legs, feathers, fly}
„birds“
{duck, goose, dove, owl, hawk}
{small, 2legs, feathers, fly}
„small birds“
„medium hunting birds“
{eagle}
{medium, hunt, 2legs, feathers, fly}
„small swimming birds“
{duck, goose}
{small, swim, 2legs, feathers, fly}
hair
4legs
big
run
medium
hunt
small
feathers
2legs
mane
fly
dog
hen
hooves
cow
tiger
fox
dove
swim
horse
zebra
lion
wolf
cat
eagle
owl
hawk
duck
goose

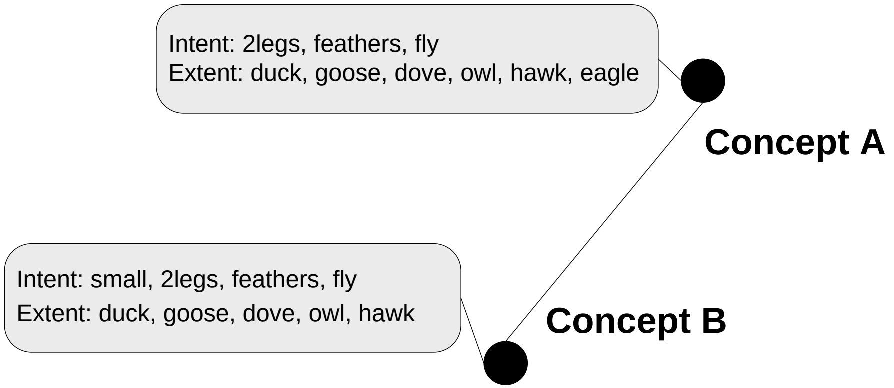
Extent and Intent in a Lattice

STI · INNSBRUCK

- The extent of a formal concept is given by all formal objects on the paths which lead ***down*** from the given concept node
  - The extent of an arbitrary concept is then found in the **principle ideal** generated by that concept

- The intent of a formal concept is given by all the formal attributes on the paths which lead ***up*** from the given concept node
  - The intent of an arbitrary concept is then found in the **principle filter** generated by that concept

Subconcepts in the Concept Lattice



- The Concept B is a subconcept of Concept A because
  - The extent of Concept B is a subset of the extent of Concept A
  - The intent of Concept B is a superset of the intent of Concept A
- All edges in the line diagram of a concept lattice represent this subconcept-superconcept relationship

Implication

STI · INNSBRUCK

- An **implication** A → B (between sets A,B∈M of attributes) holds in a formal context if and only if B⊆A“
  - i.e. if every object that has all attributes in A also has all attributes in B
  - e.g. if X has feather and has beak then is a bird
- The implication determines the concept lattice up to isomorphism and therefore offers an additional interpretation of the lattice structure
- Implications can be used for a step-wise construction of conceptual knowledge

Lexographic Ordering

STI · INNSBRUCK

- The **lexographic ordering** of sets of objects is given by:
  - for $A,B \subseteq G$, $i \in G = \{1, ..., n\}$ to order the objects
  - $A \prec_i B$ iff $i \in B - A$ and $A \cap \{1,...,i-1\} = B \cap \{1,...,i-1\}$
  - $A \prec B$ iff $\exists i$ such that $A \prec_i B$ (and there is only one i)

- The $\prec$ denotes a lexographic ordering and thus, every two distinct sets $A, B \subseteq G$ are comparable

- $B \subseteq G$ can also be represented in terms of a characteristics vector:
  - $G = \{1,2,3,4,5\}$, $B=\{1,2,5\}$
  - **characteristics vector** of B = 11001

Intuitive Algorithm for Creating a Concept Lattice

1. Start at top with all objects and no attributes
   - ({horse, zebra, lion, wolf, cat, eagle, owl, hawk, duck, goose},∅).

2. Make a concepts for the biggest set of attributes (i.e. hair and small)

3. Are there any objects that exactly match these attributes? No, so we label the concepts with only the attributes
   - ({fox, dog, wolf, cat, tiger, lion, horse, zebra, cow}, {hair})
   - ({dove, hen, duck, goose, owl, hawk, cat}, {small}).

Intuitive Algorithm for Creating a Concept Lattice

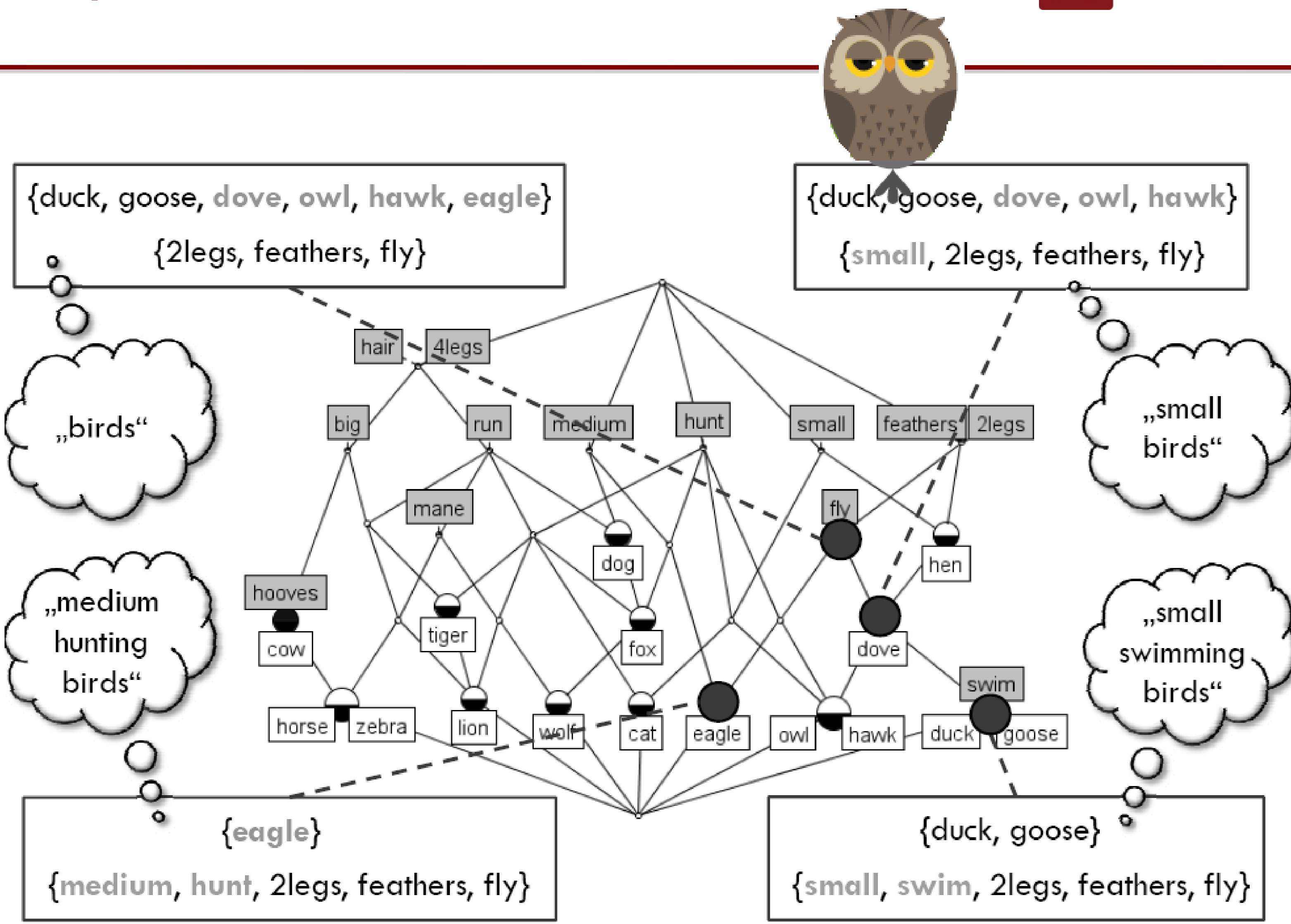
STI · INNSBRUCK

4. Add one attribute at a time to the attribute sets
   - fourlegs to small
   - fourlegs and mane to hair

5. Are there any objects that exactly match these attributes? Yes, so we label the concepts with objects. Giving:
   - ({cat},{small, fourlegs})
   - ({fox, dog, wolf, cat, tiger, lion, horse, zebra, cow}, {hair, fourlges})
   - ({horse, zebra, lion, wolf}, {hair, mane})

And so on… finally when all objects have been labelled, add the full collection of attributes to the bottom node

Example


{duck, goose, dove, owl, hawk, eagle}
{2legs, feathers, fly}
„birds“
{duck, goose, dove, owl, hawk}
{small, 2legs, feathers, fly}
„small birds“
„medium hunting birds“
„small swimming birds“
{eagle}
{medium, hunt, 2legs, feathers, fly}
{duck, goose}
{small, swim, 2legs, feathers, fly}
hair
4legs
big
run
medium
hunt
small
feathers
2legs
mane
fly
dog
hen
hooves
cow
tiger
fox
dove
swim
horse
zebra
lion
wolf
cat
eagle
owl
hawk
duck
goose

„Formal“ Construction of Concept Lattice

- This algorithm works on a finite context (G,M,I) with a lexographic ordering
- The lexographically smallest extent is Ø;
  - for i=1 we have {1,...,i-1} = Ø
- For an arbitrary $X \subseteq G$, one can find the lexographically next concept extent by checking all elements $y \in G - X$ (beginning with the lexographically largest) until $X \prec_i X \oplus i$ for the first time.
  - $\oplus$ denotes the operation of extending the object set
- $X \oplus i = ((X \cap \{1,...,i-1\}) \cup \{i\})$“
- $X \oplus i$ is the lexographically next extent

Example

STI · INNSBRUCK

Consider a context with G = {1,2,3} and M = {a,b,c} and incidence I = {(1,{a,b}), (2,{a}), (3,{b,c})}:

1. $X \oplus i = ((X \cap \{1,\ldots,i-1\}) \cup \{i\})''$
2. New extent $X \oplus i$ valid iff $X \prec_i X \oplus i$
3. $X \prec_i X \oplus i$ iff $i \in (X \oplus i) - X$ and $X \cap \{1,\ldots,i-1\} = (X \oplus i) \cap \{1,\ldots,i-1\}$

1. $\emptyset'' = \{a,b,c\}' = \emptyset$

   $\Rightarrow$ 1. extent: $\emptyset$

2. $\emptyset \oplus 3 = ((\emptyset \cap \{1,2\}) \cup \{3\})'' =$

   $= (\emptyset \cup \{3\})'' = \{3\}'' =$

   $= \{b,c\}' = \{3\}$

   Is $\emptyset \oplus 3$ valid?

   $\emptyset \prec_3 \{3\}$, as $3 \in \{3\} - \emptyset$ and $\emptyset \cap \{1,2\} = \{3\} \cap \{1,2\}$

   $\Rightarrow$ 2. extent: {3}

Example (2)

3. $\{3\} \oplus 2 = ((\{3\} \cap \{1\}) \cup \{2\})'' =$
   $= (\emptyset \cup \{2\})'' = \{2\}'' =$
   $= \{a\}' = \{1,2\}$

   Is $\{3\} \oplus 2$ valid?

   $\{3\} \nprec_2 \{1,2\}$ as $2 \in (\{1,2\} - \{3\} = \{1,2\})$
   but $(\{3\} \cap \{1\} = \emptyset) \neq (\{1,2\} \cap \{1\} = \{1\})$

4. $\{3\} \oplus 1 = ((\{3\} \cap \emptyset) \cup \{1\})'' =$
   $= (\emptyset \cup \{1\})'' = \{1\}'' =$
   $= \{a,b\}' = \{1\}$

   Is $\{3\} \oplus 1$ valid?

   $\{3\} \prec_1 \{1\}$, as $1 \in (\{1\} - \{3\} = \{1\})$ and $\{3\} \cap \{1\} = \{1\} \cap \emptyset = \emptyset$

   $\Rightarrow$ 3. extent: $\{1\}$

Example (3)

5. $\{1\} \oplus 3 = ((\{1\} \cap \{1,2\}) \cup \{3\})'' =$
   $= (\{1\} \cup \{3\})'' = \{1,3\}'' =$
   $= \{b\}' = \{1,3\}$

   Is $\{1\} \oplus 3$ valid?

   $\{1\} \prec_3 \{1,3\}$, as $3 \in (\{1,3\} - \{1\} = \{3\})$ and $\{1\} \cap \{1,2\} = \{1,3\} \cap \{1,2\} = \{1\}$

   $\Rightarrow$ 4. extent: {1,3}

6. $\{1,3\} \oplus 2 = ((\{1,3\} \cap \{1\}) \cup \{2\})'' =$
   $= (\{1\} \cup \{2\})'' = \{1,2\}'' =$
   $= \{a\}' = \{1,2\}$

   Is $\{1,3\} \oplus 2$ valid?

   $\{1,3\} \prec_2 \{1,2\}$, as $2 \in (\{1,2\} - \{1,3\} = \{2\})$ and $\{1,3\} \cap \{1\} = \{1,2\} \cap \{1\} = \{1\}$

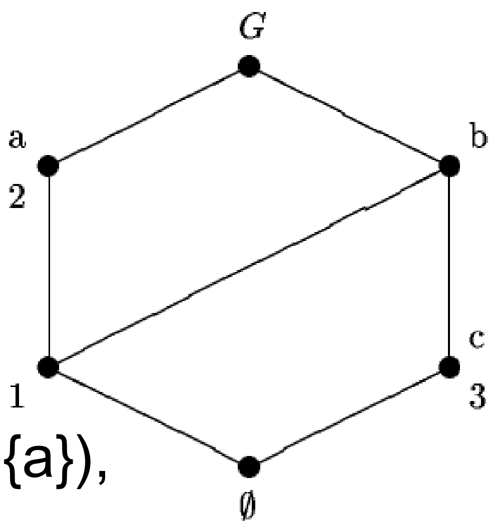
   $\Rightarrow$ 5. extent: {1,2}

Example (4)

7. $\{1,2\} \oplus 3 = ((\{1,2\} \cap \{1,2\}) \cup \{3\})'' =$
   $= (\{1,2\} \cup \{3\})'' = \{1,2,3\}'' =$
   $= \emptyset' = \{1,2,3\}$

   Is $\{1,2\} \oplus 3$ valid?
   $\{1,2\} \prec_3 \{1,2,3\}$, as $3 \in (\{1,2,3\} - \{1,2\} = \{3\})$ and $\{1,2\} \cap \{1,2\} = \{1,2,3\} \cap \{1,2\} = \{1,2\}$

   $\Rightarrow$ 6. extent: $\{1,2,3\}$



8. As $G = \{1,2,3\}$ there are no further extents

- $\mathfrak{B}(G,M,I) = \{(\emptyset,M),(\{1\},\{a,b\}),(\{3\},\{b,c\}),(\{1,2\},\{a\}),(\{1,3\},\{b\}),(G,\emptyset)\}$

ILLUSTRATION BY A LARGER EXAMPLE

From Data to Concept Representation

- Data are often represented in a table form

K0	sex	age
ADAM	M	21
BETTY	F	50
CHRIS	/	66
DORA	F	88
EVA	F	17
FRED	M	/
GEORGE	M	90
HARRY	M	50

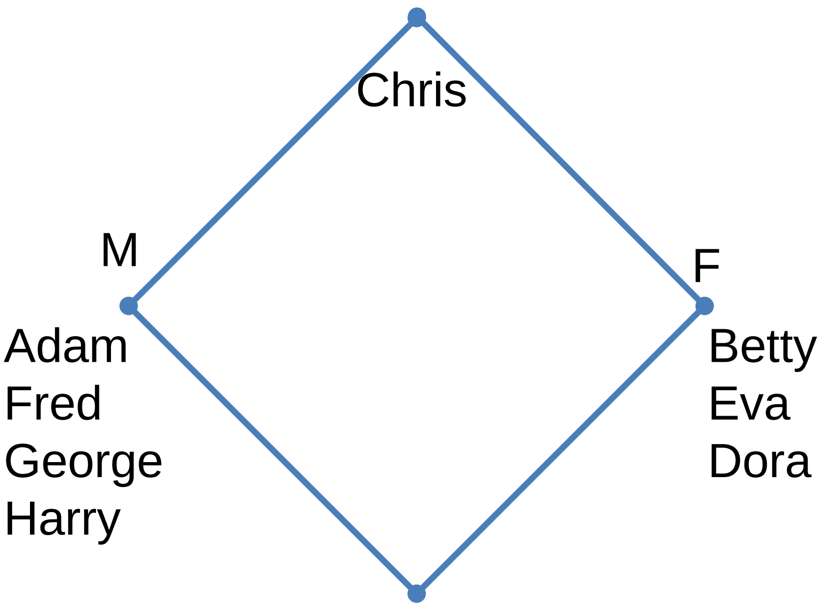
From Data to Concept Representation

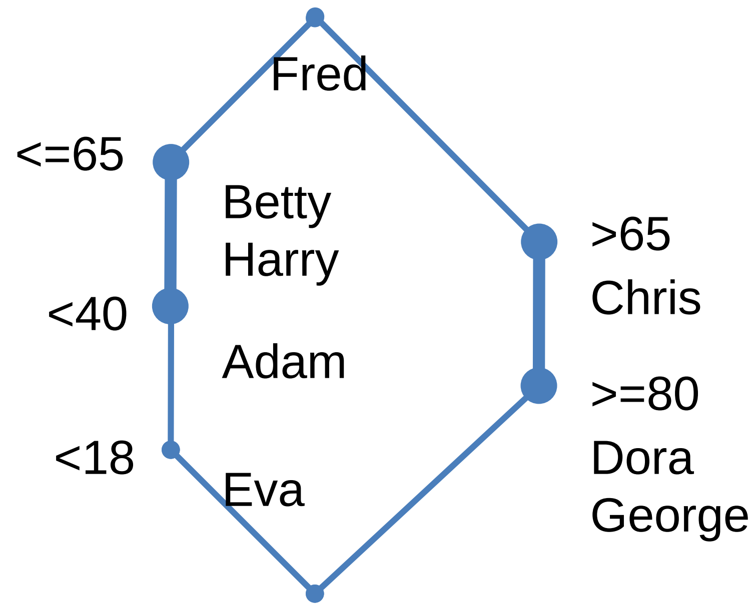
- The previous many-value context can be transformed to a formal context

K	sex		age				
	M	F	<18	<40	<=65	>65	>=80
ADAM	X			X	X		
BETTY		X			X		
CHRIS						X	
DORA		X				X	X
EVA		X	X	X	X		
FRED	X						
GEORGE	X					X	X
HARRY	X				X		

From Data to Concept Representation

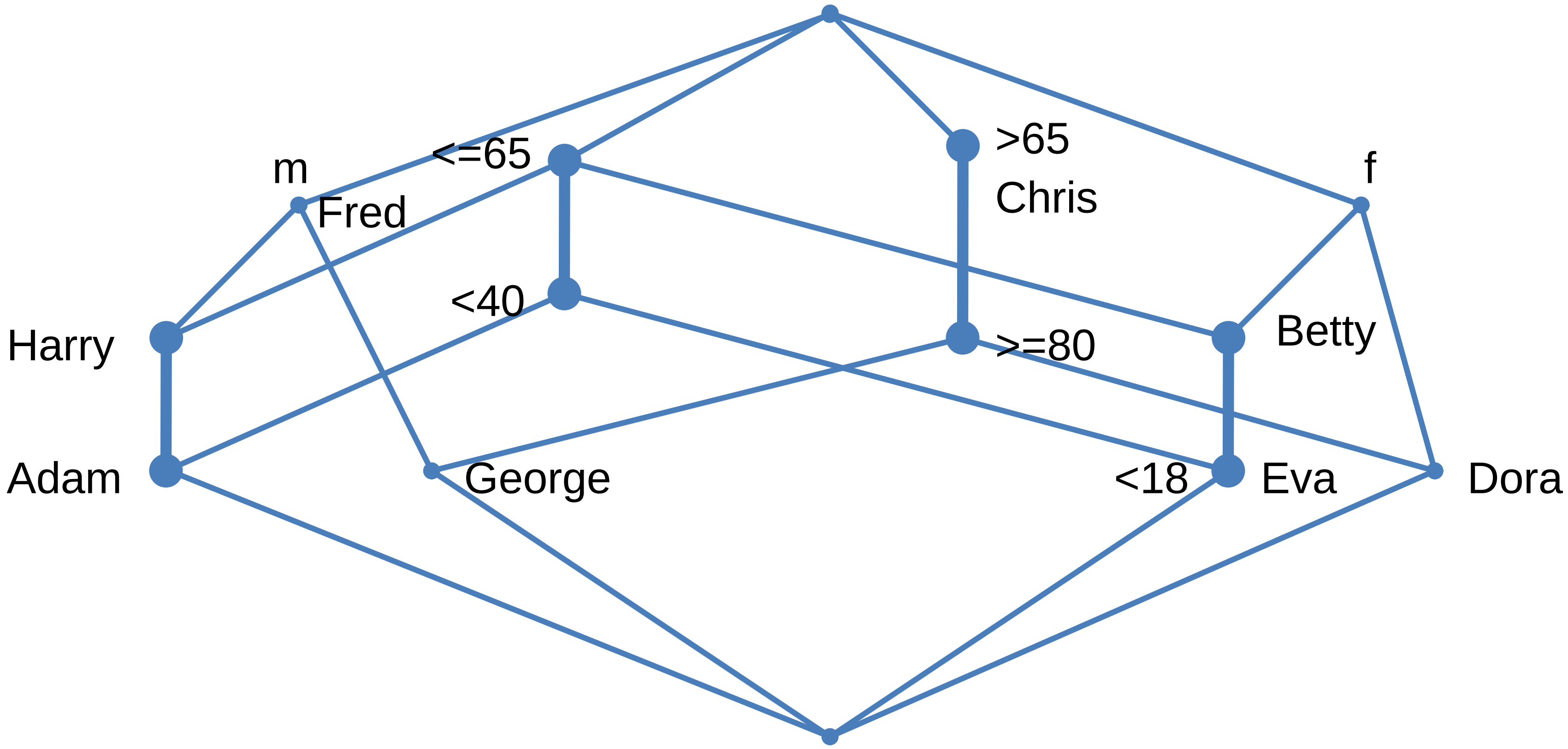
- By selecting a single view (i.e. an attribute space) we can create the following lattices





From Data to Concept Representation

- Diagrams can be combined to create a single lattice representing all the data space

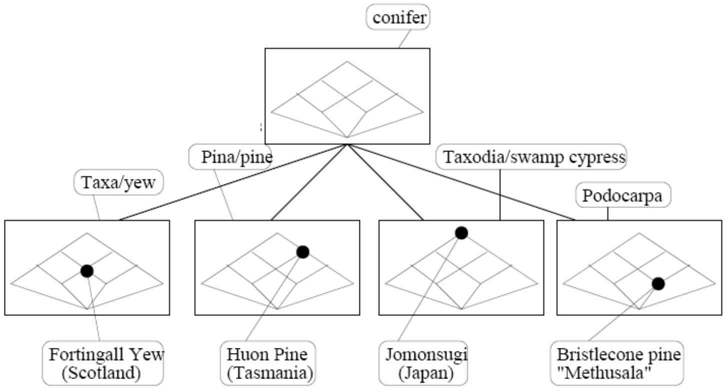


EXTENSIONS

Extensions

STI · INNSBRUCK

- Multi-valued Attributes
  - In common language settings, objects are not described by simply having an attribute or not
  - Attributes can have different values or notions
  - e.g. Fruit has color: red, green, brown...
  - An advantage of working with many-valued contexts is the resulting modularity
- Nested Conceptual Lattices
  - Often lattices tends to explode (many attributes, many objects)
  - It is possible to collide together more nodes to represent only most important concepts



SUMMARY

Summary

- Formal Concept Analysis is a method used for investigating and processing explicitely given information, in order to allow for meaningful and comprehensive interpretation

- A concept is given by a pair of objects and attributes within a formal context (G,M,I)

- The ordered set of all formal concepts in (G,M,I) is denoted by $\mathfrak{B}(G,M,I)$ and is called concept lattice

- Within a concept lattices it is possible to derive concept hierarchies, to determine super-concept or sub-concepts

REFERENCES

References

STI · INNSBRUCK

- Mandatory reading:
  – Bernhard Ganter, Rudolf Wille: Applied Lattice Theory: Formal Concept Analysis. In "General Lattice Theory", Birkhauser, 1998, ISBN 0-817-65239-6. http://www.math.tu-dresden.de/~ganter/psfiles/concept.ps

- Further reading:
  – Bernhard Ganter, Gerd Stumme, Rudolf Wille (Hg.): *Formal Concept Analysis: Foundations and Applications*. Springer, 2005, ISBN 3-540-27891-5.
  – Uta Priss: Formal Concept Analysis in Information Science. Annual Review of Information Science and Technology 40, 2006, pp. 521-543.
  – Rudolf Wille: *Introduction to Formal Concept Analysis*. TH Darmstadt (FB Mathematik), 1996.

- Wikipedia links:
  – http://en.wikipedia.org/wiki/Formal_concept_analysis
  – http://en.wikipedia.org/wiki/Lattice_%28order%29

Next lecture

#	Title
1	Introduction
2	Propositional Logic
3	Predicate Logic
4	Reasoning
5	Search Methods
6	CommonKADS
7	Problem-Solving Methods
8	Planning
9	Agents
10	Rule Learning
11	Inductive Logic Programming
12	Formal Concept Analysis
13	**Neural Networks**
14	**Semantic Web and Services**

Questions?

STI · INNSBRUCK