Algorithmic Knowledge Discovery with Concept Lattices

1. Relations, Orders, Diagrams and Lattices

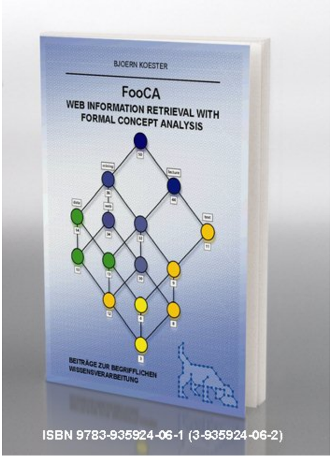
Sergei O. Kuznetsov

School of Applied Mathematics and Information Science
National Research University Higher School of Economics, Moscow, Russia

International Summer School "Information and Uncertainty"
Palacky University, Olomouc, June 4, 2012

Motivating applications. Metasearch with FooCA

- Metasearch for Web using concept lattices for representation, visualizing results and navigation
- http://www.bjoern-koester.de/
- Bjoern Koester



Construction of ontologies

[Cimiano et. al, 2003]

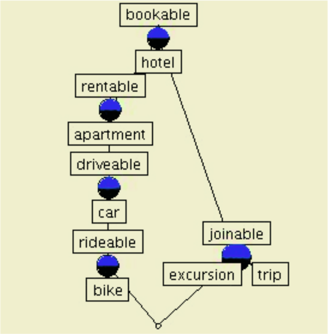
- Cimiano et. al, Automatic acquisition of taxonomies from text: FCA meets NLP, 2003
- Data on touristic business

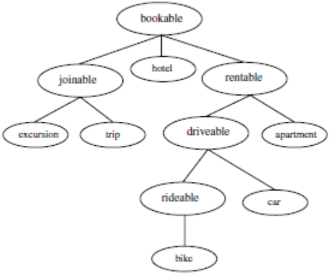
	bookable	rentable	driveable	rideable	joinable
hotel	x				
apartment	x	x			
car	x	x	x		
bike	x	x	x	x	
excursion	x				x
trip	x				x

Construction of ontologies

[Cimiano et. al, 2003]

- Concept lattice and partial order for tourist business

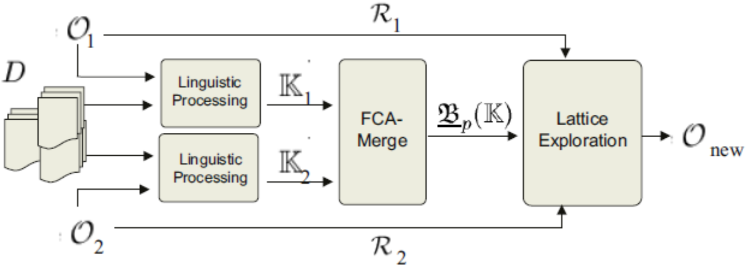




Merging of ontologies

[Stumme et. al, 2001]

- G. Stumme and A. Maedche. FCA–Merge: Bottom-up merging of ontologies, 2001

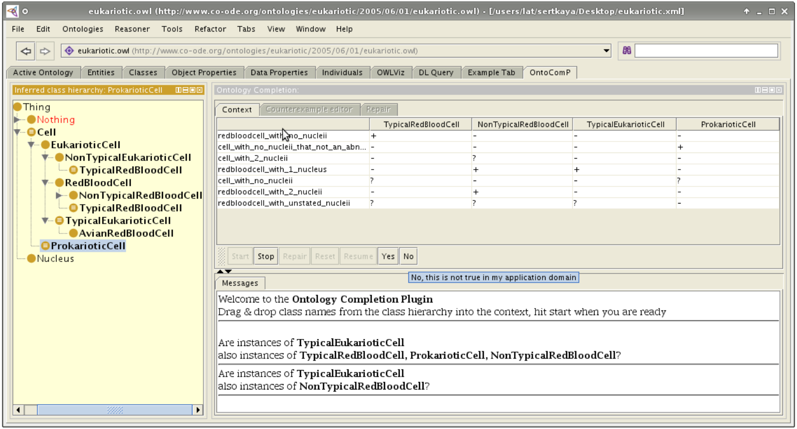


Controling quality of ontologies

[B. Sertkaya et. al, 2009]

Testing completeness of ontologies and updating OWL ontologies

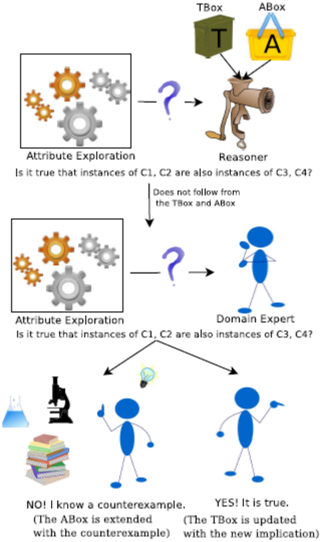
- B. Sertkaya, OntoComP: A Protégé Plugin for Completing OWL Ontologies, 2009



	TypicalRedBloodCell	NonTypicalRedBloodCell	TypicalEukarioticCell	ProkarioticCell
redbloodcell_with_no_nucleii	+	-	-	-
cell_with_no_nucleii_that_not_an_abn...	-	-	-	+
cell_with_2_nucleii	-	?	-	-
redbloodcell_with_1_nucleus	-	+	+	-
cell_with_no_nucleii	?	-	-	?
redbloodcell_with_2_nucleii	-	+	-	-
redbloodcell_with_unstated_nucleii	?	?	?	-

Controling quality of ontologies

[Bariş et. al, 2009]



Recommendation of advertising terms

Data

Data on purchases of advertising terms. Formal context $\mathbb{K}_{FT} = (F, T, I_{FT} \subseteq F \times T)$, F is the set of advertising companies, T is the set of terms, fIt means that company $f \in F$ bought term $t \in T$. The size of the context is 2000×3000.

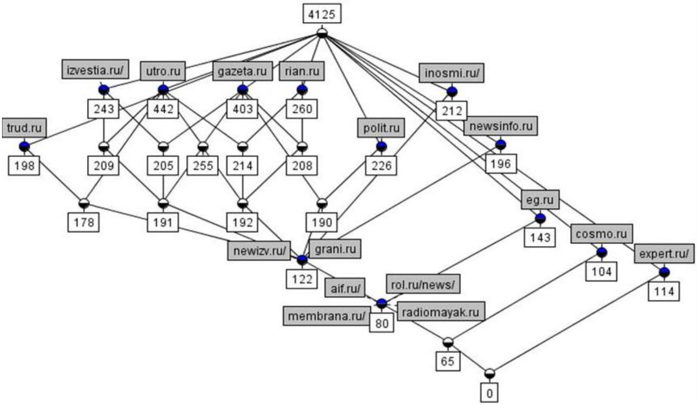
Problem statement

Detect markets of advertising terms for making bid recommendations

Solution tools

- FCA: constructing concepts and their generators
- constructing association rules
- association rules + morphology
- association rules + ontology

Taxonomy of internet audience



Modeling classes

- Robert Godin and Petko Valtchev, Formal Concept Analysis-Based Class Hierarchy Design in Object-Oriented Software Development, 2005

Analysis of software code

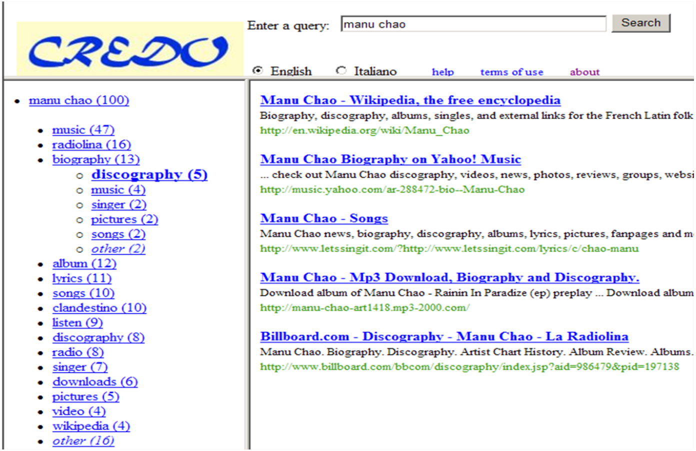
- Statical code analysis

  Gregor Snelting, Concept Lattices in Software Analysis, 2005
- Dynamical code analysis

  Gellier et. al, Concept Lattices in Software Analysis, 2005

Credo

- Metasearch system using concept lattices
- http://credo.fub.it
- Claudio Carpineto, Giovanni Romano. Concept Data Analysis: Theory and Applications



KVO

- International Research Group Knowledge, Visualisation and Ordering
- NLP, knowledge representation, information retrieval, data mining, usability knowledge models
- http://www.kvocentral.org/

Software

- Search Sleuth (metasearch system)
- Image Sleuth (search in collections of images)
- Mail Sleuth (plugin for e-mails)
- D-Sift (visualization of relational data in Web)
- ToscanaJ (data analysis)

Search Sleuth

- Processes results of search queries to Yahoo
- Passing to more general (more specific) categories by clicking -term (+term)

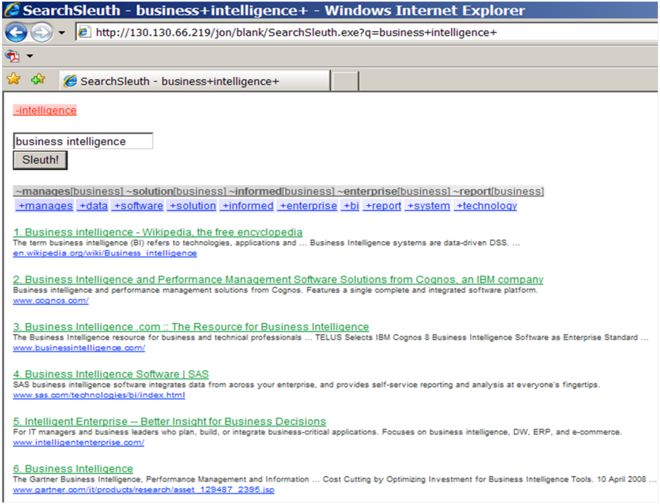
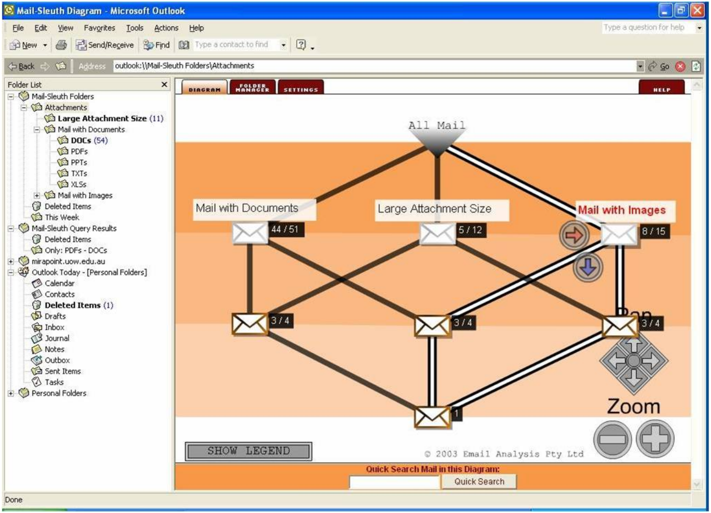


Image Sleuth

- FCA-based system for looking images, navigation and search in their collections

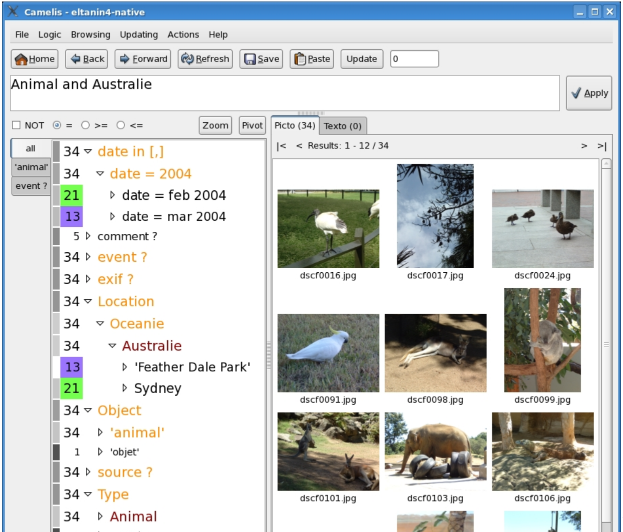
Mail Sleuth

- Plugin for Outlook, using concept lattices as a means of visualization and representing data from an e-mail account



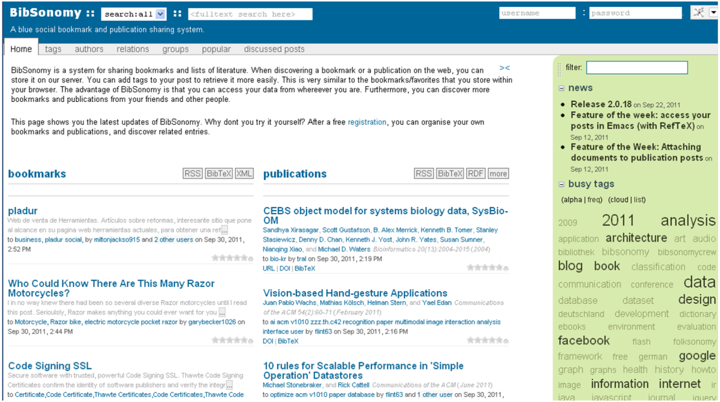
Camelis

- System of automatic indexing and navigation in data using concept lattices
- Sebastien Ferre



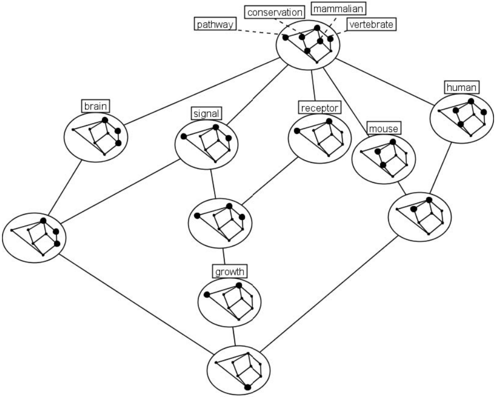
Bibsonomy

- http://bibsonomy.org/ a web-service of social bookmarks
- University of Kassel



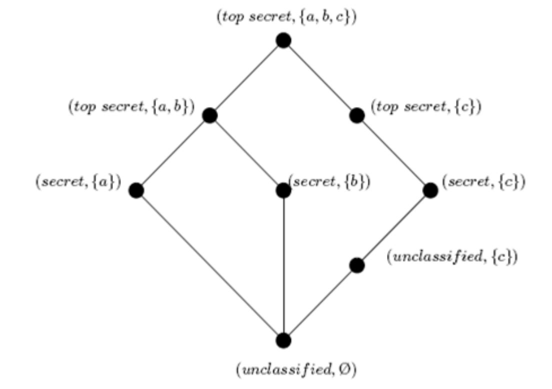
Epistemic communities

- S. Obiedkov, C. Roth (2006)
- studying "zebrafish" community



Access control models

- S. Obiedkov, 2006



Analysis of police reports

- Jonas Poelmans et al., KU Leuven, Belgium
- Paul Elzinga, Amsterdam-Amstelland Police, Netherlands
- Cordiet Project in cooperation with Higher School of Economics, Moscow

Outline

Main goal: Getting acquainted with applications of the lattice theory and Formal Concept Analysis in knowledge discovery and modern methods of data analysis.

1. Relations and ordered sets, diagrams, semilattices and lattices
2. Formal Concept Analysis (FCA)
3. FCA-based methods of knowledge discovery and data analysis
4. Algorithms and complexity of FCA-based knowledge discovery

Properties of binary relations

Let $R \subset A \times A$, then R is called

reflexive if $\forall a \in A \; aRa$
antireflexive if $\forall a \in A \quad \neg(aRa) \; (\Leftrightarrow aR^c a)$
symmetric if $\forall a, b \in A \quad aRb \Rightarrow bRa$
asymmetric if $\forall a, b \in A \quad aRb \Rightarrow \neg(bRa) \; (\Leftrightarrow bR^c a)$
antisymmetric if $\forall a, b \in A \quad aRb \;\&\; bRa \Rightarrow a = b$
transitive if $\forall a, b, c \in A \quad aRb \;\&\; bRc \Rightarrow aRc$
complete or linear if $\forall a, b \in A \quad a \neq b \Rightarrow aRb \vee bRa$.

Types of binary relations

- **Tolerance** is a reflexive and symmetric binary relation;
- **Equivalence** is a reflexive, symmetric, and transitive binary relation;
- **Quasi-order** or **preorder** is a reflexive and transitive binary relation;
- **Partial order** is a reflexive, transitive, and antisymmetric binary relation;
- **Strict order** is antireflexive and transitive binary relation.

Covering relation

Covering relation $\prec$ defined by order $\leq$ is defined as follows:

$$x \prec y := x \leq y,\ x \neq y,\ \nexists z \neq x, y \quad x \leq z \leq y$$

or, equivalently,

$$x \prec y := x < y,\ \nexists z \quad x < z < y.$$

Theorem. Let $a < b$ in a finite ordered set $(P, \leq)$. Then P contains a subset of elements $\{x_1, \ldots, x_l\}$ such that $a = x_1 \prec \ldots \prec x_l = b$.
Proof. By induction over the number of elements y with the property $a < y < b$.

Diagram of an order

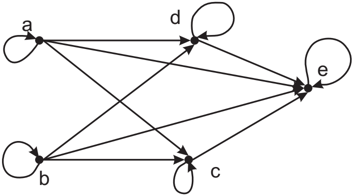
(Plane embedding) of a graph is an injection taking each vertex of a graph to a point in $R \times R$ and every edge of the graph to an interval joining endpoints.
(Hasse) diagram of an ordered set $(P, \leq)$ is a plane embedding of the graph of the covering relation $(P, \prec)$ with the following property:
$a \prec b \implies$ the point corresponding to vertex a has a less vertical coordinate than the point corresponding to vertex b.

Example. Order relation

	a	b	c	d	e
a	1	0	1	1	1
b	0	1	1	1	1
c	0	0	1	0	1
d	0	0	0	1	1
e	0	0	0	0	1

Example. Graph of an ordered set

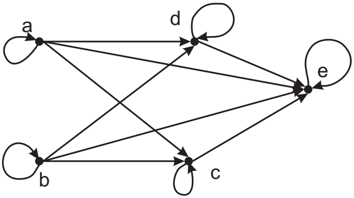
	a	b	c	d	e
a	1	0	1	1	1
b	0	1	1	1	1
c	0	0	1	0	1
d	0	0	0	1	1
e	0	0	0	0	1



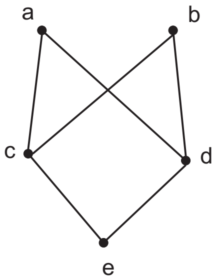
acyclic graph

Example. Diagram of an ordered set

	a	b	c	d	e
a	1	0	1	1	1
b	0	1	1	1	1
c	0	0	1	0	1
d	0	0	0	1	1
e	0	0	0	0	1



acyclic graph



order diagram

Bounds, supremums and infimums

Let $(P, \leq)$ be an ordered set and $A \subseteq P$.
Upper bound of subset $A \subseteq P$ is a set

$$\{b \in P \mid \forall a \in A \quad b \geq a\}.$$

Supremum of set $A \subseteq P$ is the least element b of the upper bound of A (if it exists):

1. $\forall a \in A \quad b \geq a$,
2. $\forall x \in P \ (\forall a \in A \quad x \geq a) \Rightarrow x \geq b$.

Supremum of A is denoted by **sup(A)** and also called **join** of A.

Dually, one defines **lower bound** of a subset $A \subseteq P$ and **infimum** of A or **inf(A)** as the largest element of the lower bound of A. Infimum of A is also called **meet** of A.

Semilattice

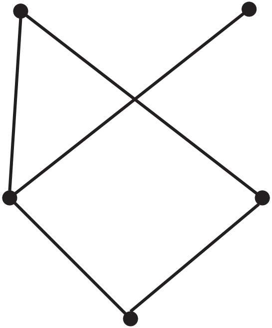
An ordered set $(SL, \leq)$ is called **upper semilattice** if any pair of its elements $\{x, y\} \subseteq SL$ has supremum (or join) $\sup\{x, y\}$.

Dually, a **lower semilattice** is defined wrt. infimum (or meet).

An ordered set $(SL, \leq)$ is called a **lower semilattice** if any pair of its elements $\{x, y\} \subseteq SL$ has infimum (or meet) $\inf\{x, y\}$.

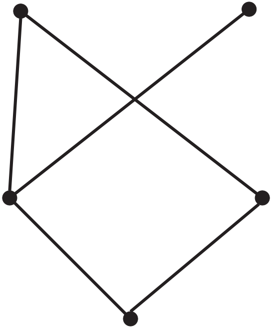
Semilattices. Examples

lower semilattice

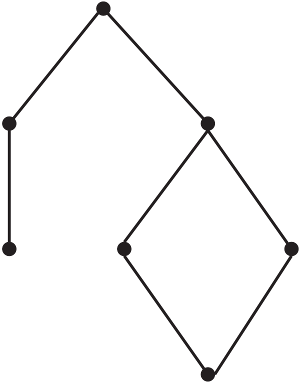


Semilattice. Examples

lower semilattice



upper semilattice

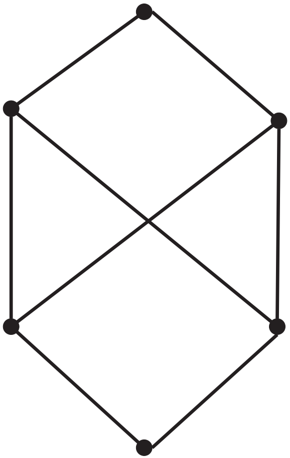


Lattices

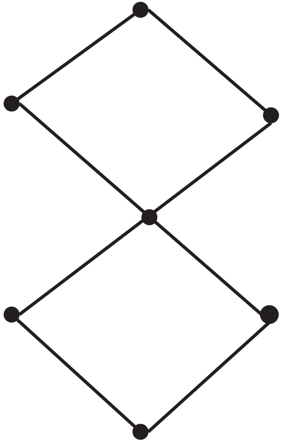
An ordered set $(L, \leq)$ is called a **lattice** if any pair of elements $\{x, y\} \subseteq L$ has supremum $\sup\{x, y\}$ and infimum $\inf\{x, y\}$.

Lattice. Examples.

the order is neither lower nor upper semilattice

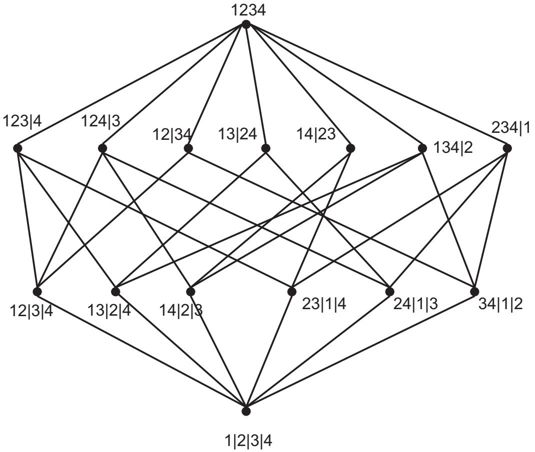


the order is a lattice



Lattice of partitions of a 4-element set

$A = \{1, 2, 3, 4\}$



Lattice. Another definition

Theorem. A set L is a lattice wrt. some partial order iff there are two operations $\vee$ and $\wedge$ on L, which satisfy the following properties for any $x, y, z \in L$:

- L1 $x \vee x = x, \quad x \wedge x = x$ (idempotence)
- L2 $x \vee y = y \vee x, \quad x \wedge y = y \wedge x$ (commutativity)
- L3 $x \vee (y \vee z) = (x \vee y) \vee z, \quad x \wedge (y \wedge z) = (x \wedge y) \wedge z$ (associativity)
- L4 $x = x \wedge (x \vee y) = x \vee (x \wedge y)$ (absorption)

Lattice. Another definition

Theorem. A set L is a lattice wrt. some partial order iff there are two operations $\vee$ and $\wedge$ on L, which satisfy the following properties for any $x, y, z \in L$:

L1 $x \vee x = x, \quad x \wedge x = x$ (idempotence)

L2 $x \vee y = y \vee x, \quad x \wedge y = y \wedge x$ (commutativity)

L3 $x \vee (y \vee z) = (x \vee y) \vee z, \quad x \wedge (y \wedge z) = (x \wedge y) \wedge z$ (associativity)

L4 $x = x \wedge (x \vee y) = x \vee (x \wedge y)$ (absorption)

The theorem allows one to consider the lattice as an algebra $(L, \vee, \wedge)$ with properties L1-L4. A **natural order** of the lattice given in this way is the relation "$\leq$" $\subseteq L \times L$ defined as $x \leq y \stackrel{\text{def}}{=} x \wedge y = x$ (or, equivalently, by $x \vee y = y$).

Complete lattices

A lattice is called **complete** if any its subset has infimum and supremum.

$$\bigvee \emptyset = \mathbf{0} \quad \bigwedge \emptyset = \mathbf{1}$$

All finite lattices are complete.

For an arbitrary subset of elements of a complete lattice one can write

$$\bigvee X, \quad \bigwedge X$$

due to associativity and commutativity of operations $\vee$ and $\wedge$.

Exercises

- Determine the properties of the relation $Q := \{(m, n) \mid m, n \in \mathbb{N} \ \& \ m = n^2\}$
- 'Prove that incomparability relation for an order is a tolerance relation
- Every subset of an ordered set is an ordered set (wrt. the restriction of the order relation)
- Is strict order antisymmetric?
- Construct a diagram of an order given by relation matrix
- Construct all nonisomorphic orders on a set of 4 (5) elements.
- Let there be a nonempty set A and P be the set of all orders on A. Let for $\rho, \sigma \in P$ one has $\rho \leq \sigma$ if $a\rho b$ implies $a\sigma b$. Prove that $(P, \leq)$ is an ordered set.
- Construct the diagram of partitions of a four-element set.

Literature on orders and lattices

G. Birkhoff, *Lattice Theory*, AMS, Providence, 1979. (Section 1.1)

G. Birkhoff and T.C. Bartee *Modern Applied Algebra*, McGraw-Hill, N.Y., 1970. (Chapter 2)

G. Grätzer, *Lattice Theory: Foundation*, Springer, 2011.

B. A. Davey and H. A. Priestley, *Introduction to Lattices and Order*, Cambridge University Press, 1990.