

Algorithmic Knowledge Discovery with Concept Lattices

1. Relations, Orders, Diagrams and Lattices

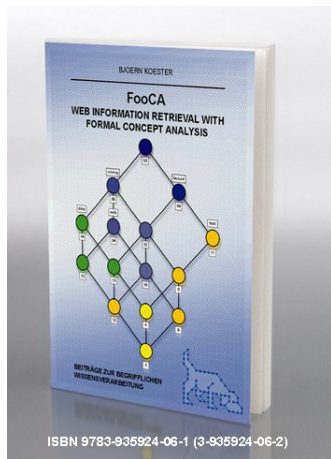
Sergei O. Kuznetsov

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]

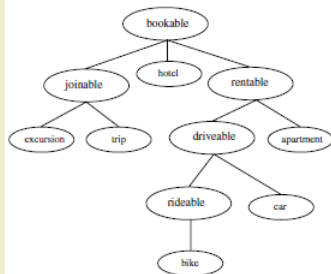
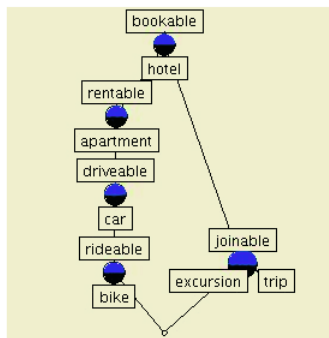
- ▶ Cimiano et. al, Automatic acquisition of taxonomies from text: FCA meets NLP, 2003
- ▶ Data on touristic business

	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]

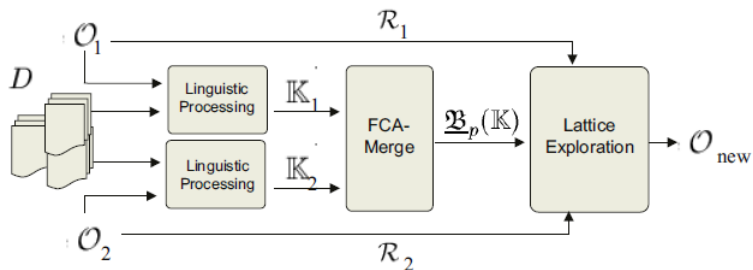
- ▶ Concept lattice and partial order for tourist business



Merging of ontologies

[Stumme et. al, 2001]

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



Controlling 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

The screenshot shows the Protégé interface with the OntoComP plugin active. On the left, the 'Inferred class hierarchy' for 'ProkarioticCell' is displayed, showing a tree structure of classes including Thing, Cell, EukarioticCell, NonTypicalEukarioticCell, TypicalRedBloodCell, RedBloodCell, NonTypicalRedBloodCell, TypicalRedBloodCell, TypicalEukarioticCell, AvianRedBloodCell, ProkarioticCell, and Nucleus.

The main window displays the 'Ontology Completion' dialog. It features a 'Context' tab and a 'Repair' button. Below these is a table with the following structure:

Context	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	?	?	?	-

Below the table are buttons for 'Start', 'Stop', 'Repair', 'Reset', 'Resume', 'Yes', and 'No'. A message box at the bottom contains the following text:

No, this is not true in my application domain

Messages

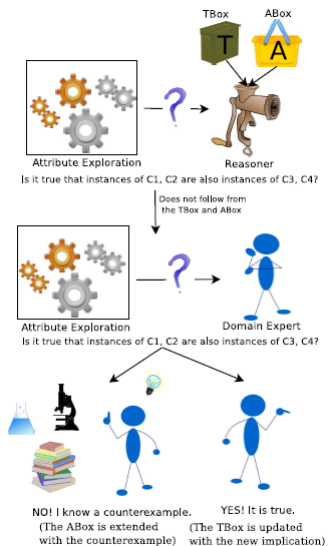
Welcome to the **Ontology Completion Plugin**
Drag & drop class names from the class hierarchy into the context, hit start when you are ready

Are instances of **TypicalEukarioticCell**
also instances of **TypicalRedBloodCell**, **ProkarioticCell**, **NonTypicalRedBloodCell**?

Are instances of **TypicalEukarioticCell**
also instances of **NonTypicalRedBloodCell**?

Controlling 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, flt means that company $f \in F$ bought term $t \in T$. The size of the context is 2000×3000 .

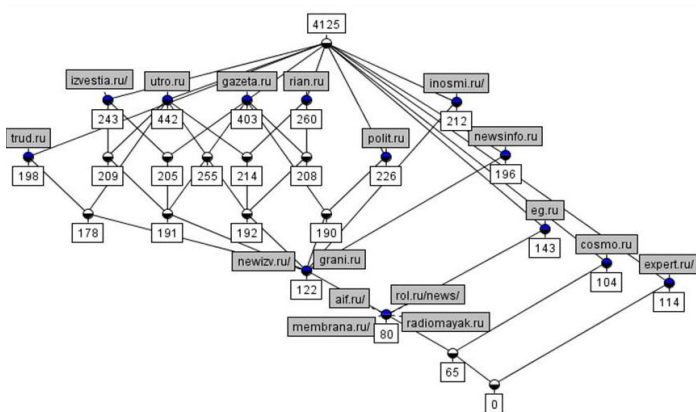
Problem statement

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

The screenshot shows the Credo metasearch engine interface. At the top left is the 'CREDO' logo in blue. To its right is a search bar with the text 'manu chao' and a 'Search' button. Below the search bar are navigation links for 'English', 'Italiano', 'help', 'terms of use', and 'about'. The main content area is divided into two columns. The left column contains a hierarchical list of search results for 'manu chao (100)', including categories like 'music (47)', 'radiolina (16)', 'biography (13)', 'discography (5)', 'album (12)', 'lyrics (11)', 'songs (10)', 'clandestino (10)', 'listen (9)', 'discography (8)', 'radio (8)', 'singer (7)', 'downloads (6)', 'pictures (5)', 'video (4)', 'wikipedia (4)', and 'other (16)'. The right column displays four search results with titles and brief descriptions: 'Manu Chao - Wikipedia, the free encyclopedia', 'Manu Chao Biography on Yahoo! Music', 'Manu Chao - Songs', and 'Manu Chao - Mp3 Download, Biography and Discography'. At the bottom of the page are navigation icons for back, forward, and search.

CREDO

Enter a query: Search

English Italiano help terms of use about

- [manu chao \(100\)](#)
 - [music \(47\)](#)
 - [radiolina \(16\)](#)
 - [biography \(13\)](#)
 - [discography \(5\)](#)
 - [music \(4\)](#)
 - [singer \(2\)](#)
 - [pictures \(2\)](#)
 - [songs \(2\)](#)
 - [other \(2\)](#)
 - [album \(12\)](#)
 - [lyrics \(11\)](#)
 - [songs \(10\)](#)
 - [clandestino \(10\)](#)
 - [listen \(9\)](#)
 - [discography \(8\)](#)
 - [radio \(8\)](#)
 - [singer \(7\)](#)
 - [downloads \(6\)](#)
 - [pictures \(5\)](#)
 - [video \(4\)](#)
 - [wikipedia \(4\)](#)
 - [other \(16\)](#)

[Manu Chao - Wikipedia, the free encyclopedia](#)
Biography, discography, albums, singles, and external links for the French Latin folk
http://en.wikipedia.org/wiki/Manu_Chao

[Manu Chao Biography on Yahoo! Music](#)
... check out Manu Chao discography, videos, news, photos, reviews, groups, websi
<http://music.yahoo.com/ar-288472-bio--Manu-Chao>

[Manu Chao - Songs](#)
Manu Chao news, biography, discography, albums, lyrics, pictures, fanpages and m
<http://www.letsingit.com/?http://www.letsingit.com/lyrics/c/chao-manu>

[Manu Chao - Mp3 Download, Biography and Discography.](#)
Download album of Manu Chao - Rainin In Paradize (ep) preplay ... Download album
<http://manu-chao-art1418.mp3-2000.com/>

[Billboard.com - Discography - Manu Chao - La Radiolina](#)
Manu Chao. Biography. Discography. Artist Chart History. Album Review. Albums.
<http://www.billboard.com/bbcom/discography/index.jsp?aid=986479&pid=197138>



- ▶ 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)

SearchSleuth - business+intelligence+ - Windows Internet Explorer

http://130.130.66.219/jon/blank/SearchSleuth.exe?q=business+intelligence+

SearchSleuth - business+intelligence+

[-intelligence](#)

business intelligence

Sleuth!

[-manages](#) [business|](#) [-solution](#) [business|](#) [-informed](#) [business|](#) [-enterprise](#) [business|](#) [-report](#) [business|](#)
[+manages](#) [+data](#) [+software](#) [+solution](#) [+informed](#) [+enterprise](#) [+bj](#) [+report](#) [+system](#) [+technology](#)

[1. Business intelligence - Wikipedia, the free encyclopedia](#)
The term business intelligence (BI) refers to technologies, applications and ... Business Intelligence systems are data-driven DSS. ...
en.wikipedia.org/wiki/Business_Intelligence

[2. Business Intelligence and Performance Management Software Solutions from Cognos, an IBM company](#)
Business intelligence and performance management solutions from Cognos. Features a single complete and integrated software platform.
www.cognos.com/

[3. Business Intelligence.com :: The Resource for Business Intelligence](#)
The Business Intelligence resource for business and technical professionals ... TELUS Selects IBM Cognos 8 Business Intelligence Software as Enterprise Standard ...
www.businessintelligence.com/

[4. Business Intelligence Software | SAS](#)
SAS business intelligence software integrates data from across your enterprise, and provides self-service reporting and analysis at everyone's fingertips.
www.sas.com/technologies/bi/index.html

[5. Intelligent Enterprise -- Better Insight for Business Decisions](#)
For IT managers and business leaders who plan, build, or integrate business-critical applications. Focuses on business intelligence, DW, ERP, and e-commerce.
www.intelligententerprise.com/

[6. Business Intelligence](#)
The Gartner Business Intelligence, Performance Management and Information ... Cost Cutting by Optimizing Investment for Business Intelligence Tools. 10 April 2008 ...
www.gartner.com/it/products/research/asset_129487_2395.jsp

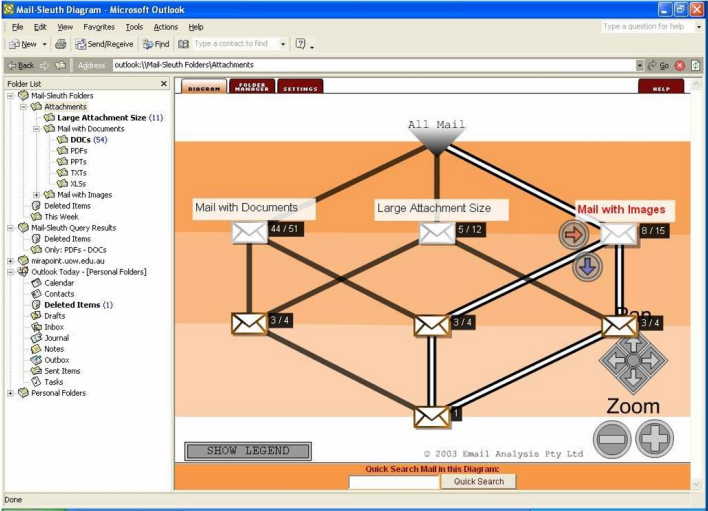
Image Sleuth

- 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

The screenshot displays the Camelis software interface. The window title is "Camelis - eltanin4-native". The menu bar includes "File", "Logic", "Browsing", "Updating", "Actions", and "Help". The toolbar contains icons for Home, Back, Forward, Refresh, Save, Paste, and Update, along with an "Update" button and a numeric input field set to "0".

The main search area contains the text "Animal and Australie" and an "Apply" button. Below this, there are navigation controls including "Zoom" and "Pivot" buttons, and a status bar showing "Picto (34)" and "Texto (0)".

The left sidebar shows a concept lattice structure with the following nodes and counts:

- all: 34
- date in [,]: 34
- date = 2004: 34
- date = feb 2004: 21
- date = mar 2004: 13
- comment?: 5
- event?: 34
- exif?: 34
- Location: 34
- Oceanie: 34
- Australie: 34
- 'Feather Dale Park': 13
- Sydney: 21
- Object: 34
- 'animal': 34
- 'objet': 1
- source?: 34
- Type: 34
- Animal: 34

The right pane displays a grid of 12 images, each with a caption below it:

- dscr0016.jpg: A white bird on a fence.
- dscr0017.jpg: A blue sky with clouds.
- dscr0024.jpg: Two ducks on a paved area.
- dscr0091.jpg: A white bird in a field.
- dscr0098.jpg: A dog lying on the ground.
- dscr0099.jpg: A grey animal on a tree branch.
- dscr0101.jpg: A small animal in a field.
- dscr0103.jpg: An elephant in a field.
- dscr0106.jpg: A small animal in a field.

Bibsonomy

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

BibSonomy :: search:all :: [Fulltext search here] [username] [password]

A blue social bookmark and publication sharing system.

Home tags authors relations groups popular discussed posts

BibSonomy is a system for sharing bookmarks and lists of literature. When discovering a bookmark or a publication on the web, you can store it on our server. You can add tags to your post to retrieve it more easily. This is very similar to the bookmarks/favorites that you store within your browser. The advantage of BibSonomy is that you can access your data from wherever you are. Furthermore, you can discover more bookmarks and publications from your friends and other people.

This page shows you the latest updates of BibSonomy. Why dont you try it yourself? After a free registration, you can organise your own bookmarks and publications, and discover related entries.

bookmarks [RSS] [BibTeX] [XML] **publications** [RSS] [BibTeX] [RDF] [more]

pladur
Web de visita de Herramientas. Artículos sobre reformas, interesante sitio que pone al alcance en su página web herramientas actuales, para obtener una ref...
to business, pladur social, by mitorjacks0915 and 2 other users on Sep 30, 2011, 2:52 PM ★★★★★

Who Could Know There Are This Many Razor Motorcycles?
[It is] how many there had been so several diverse Razor motorcycles until I read this post. Seriously, Razor makes anything you could ever want for you...
to Motorcycle, Razor bike, electric motorcycle pocket razor by garybecker1026 on Sep 30, 2011, 2:44 PM ★★★★★

Code Signing SSL
Secure software with trusted, powerful Code Signing SSL. Thawte Code Signing Certificates confirm the identity of software publishers and verify the integ...
to Certificate, Code Certificate, Thawte Certificates, Code Certificates, Thawte Code

CEBS object model for systems biology data, SysBio-OM
Sandhya Krasagar, Scott Gustafson, B. Alex Merrick, Kenneth B. Toner, Stanley Stasiewicz, Denny D. Chan, Kenneth J. Yost, John R. Yates, Susan Sumner, Nanqing Xiao, and Michael D. Winters. *Bioinformatics* 20(13):2004-2015 (2004)
to bio-kr by tral on Sep 30, 2011, 2:19 PM ★★★★★
URL | DOI | BibTeX

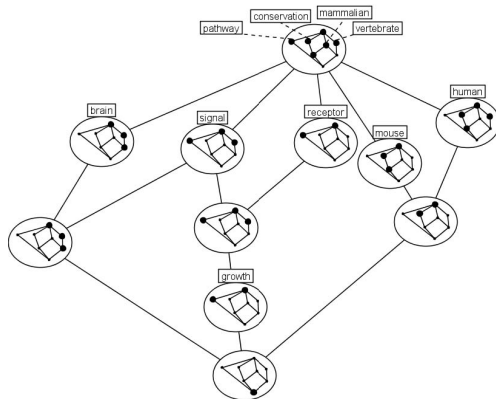
Vision-based Hand-gesture Applications
Juan Pablo Viecha, Melissa Kölsch, Heeman Stern, and Yeat Eidan. *Communications of the ACM* 54(2):60-71 (February 2011)
to ai acm v1010 zzz.th.c42 recognition paper multimodal image interaction analysis interface user by flrt63 on Sep 30, 2011, 2:16 PM ★★★★★
DOI | BibTeX

10 rules for Scalable Performance in 'Simple Operation' Databases
Michael Stonebraker, and Rick Cattell. *Communications of the ACM* (June 2011)
to optimize acm v1010 paper database by flrt63 and 1 other user on Sep 30, 2011,

filter: [input]
= news
• Release 2.0.18 on Sep 22, 2011
• Feature of the week: access your posts in Emacs (with RefTeX) on Sep 12, 2011
• Feature of the Week: Attaching documents to publication posts on Sep 12, 2011
= busy tags
(alpha | fre) (cloud | list)
2009 2011 analysis
application architecture art audio
bibliothek bibsonomy bibsonomycrew
blog book classification code
communication conference data
database dataset design
deutschland development dictionary
ebooks environment evaluation
facebook flash folksonomy
framework free german google
graph graphs health history howto
image information internet ir
java javascrint journal journal

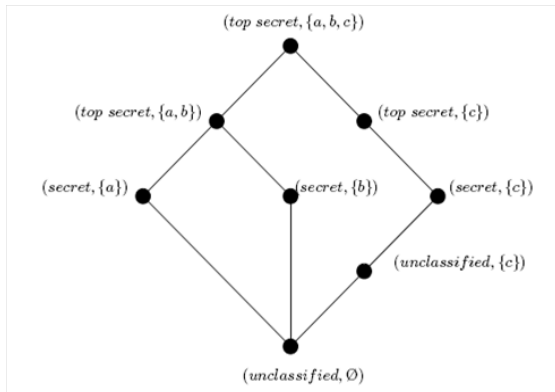
Epistemic communities

- ▶ 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 $\forall a \in A \ \neg(aRa) \ (\Leftrightarrow aR^c a)$

symmetric if $\forall a, b \in A \ aRb \Rightarrow bRa$

asymmetric if $\forall a, b \in A \ aRb \Rightarrow \neg(bRa) \ (\Leftrightarrow bR^c a)$

antisymmetric if $\forall a, b \in A \ aRb \ \& \ bRa \Rightarrow a = b$

transitive if $\forall a, b, c \in A \ aRb \ \& \ bRc \Rightarrow aRc$

complete or linear if $\forall a, b \in A \ a \neq b \Rightarrow aRb \vee bRa$.

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 $\{x_1, \dots, x_l\}$ such that $a = x_1 \prec \dots \prec x_l = b$.

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:

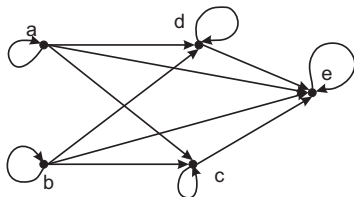
$a \prec b \implies$ the point corresponding to vertex a has a less vertical coordinate than the point corresponding to vertex b .

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

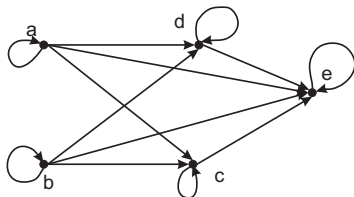
	a	b	c	d	e
a	1	0	1	1	1
b	0	1	1	1	1
c	0	0	1	0	1
d	0	0	0	1	1
e	0	0	0	0	1



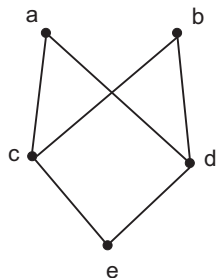
acyclic graph

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

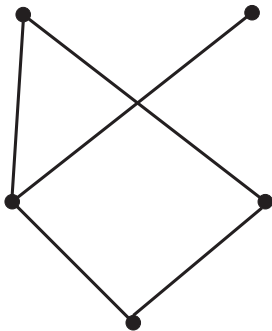
An ordered set (SL, \leq) is called **upper semilattice** if any pair of its elements $\{x, y\} \subseteq SL$ has supremum (or join) $\sup\{x, y\}$.

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) $\inf\{x, y\}$.

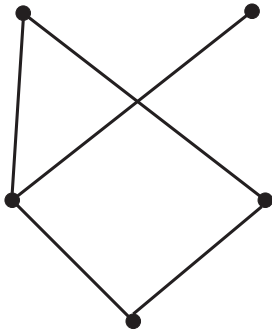
Semilattices. Examples

lower semilattice

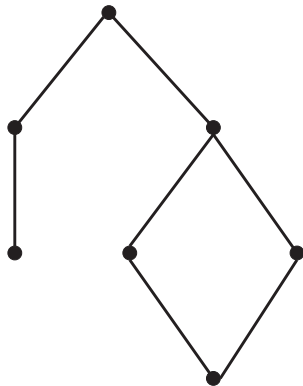


Semilattice. Examples

lower semilattice



upper semilattice

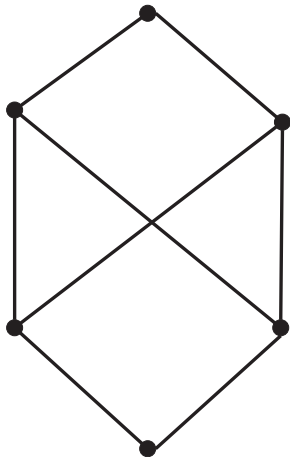


Lattices

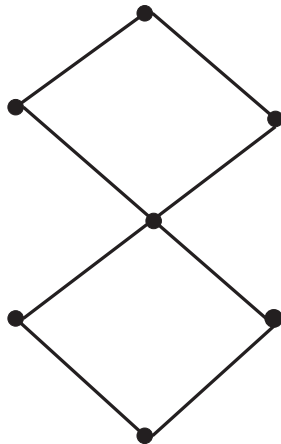
An ordered set (L, \leq) is called a **lattice** if any pair of elements $\{x, y\} \subseteq L$ has supremum $\sup\{x, y\}$ and infimum $\inf\{x, y\}$.

Lattice. Examples.

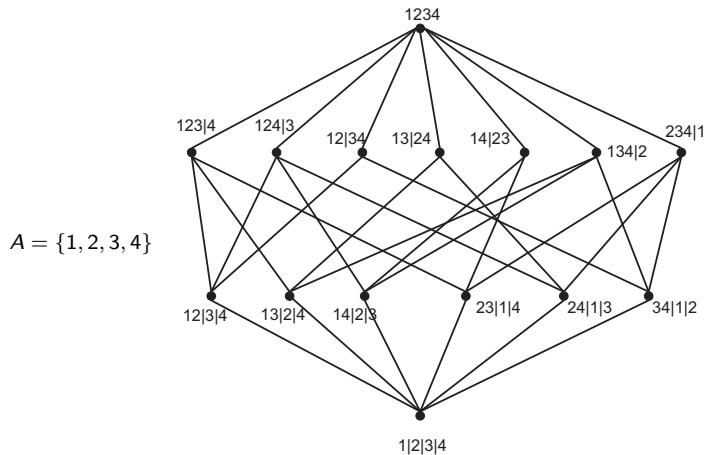
the order is neither lower nor upper semilattice



the order is a lattice



Lattice of partitions of a 4-element set



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.