Formal Concept Analysis

II Closure Systems and Implications

Sebastian Rudolph

Computational Logic Group
Technische Universität Dresden

slides based on a lecture by Prof. Gerd Stumme

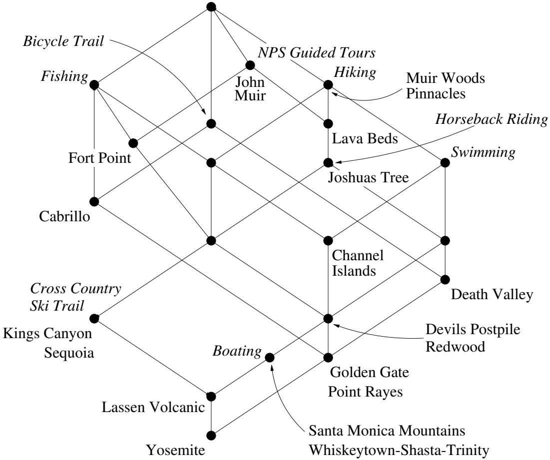
Agenda

4 Implications

- Implications
- Attribute Logic
- Concept Intents and Implications
- Implications and Closure Systems
- Pseudo-Intents and the Stem Base
- Computing the Stem Base With NEXT CLOSURE
- Bases of Association Rules

Implications

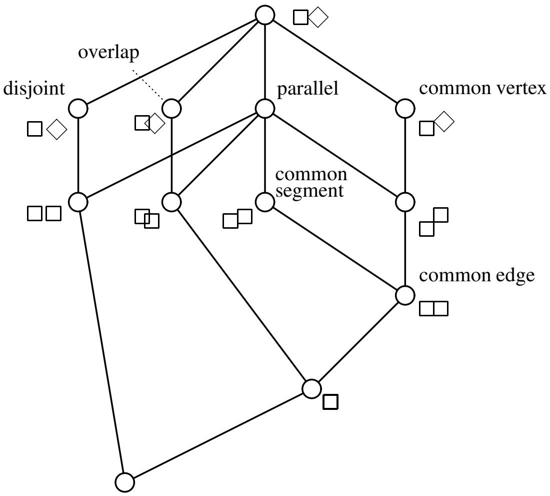
Def.: An *implication* $X \to Y$ *holds* in a context, if every object that has all attributes from X also has all attributes from Y.



Examples:

- $\{$*Swimming*$\} \to \{$*Hiking*$\}$
- $\{$*Boating*$\} \to \{$*Swimming, Hiking, NPS Guided Tours, Fishing, Horseback Riding*$\}$
- $\{$*Bicycle Trail, NPS Guided Tours*$\} \to \{$*Swimming, Hiking, Horseback Riding*$\}$

Attribute Logic



We are dealing with implications over an possibly *infinite* set of objects!

Concept Intents and Implications

Def.: A subset $T \subseteq M$ *respects* an implication $A \to B$,
if $A \nsubseteq T$ or $B \subseteq T$ holds.

(We then also say that T is a *model* of $A \to B$.)

T *respects a set* $\mathcal{L}$ of implications, if T respects every implication in $\mathcal{L}$.

Lemma: An implication $A \to B$ holds in a context, iff $B \subseteq A''$ ($\Leftrightarrow A' \subseteq B'$). It is then respected by all concept intents.

Implications and Closure Systems

Lemma: If $\mathcal{L}$ is a set of implications in M, then

$$\mathrm{Mod}(\mathcal{L}) := \{X \subseteq M \mid X \text{ respects } \mathcal{L}\}$$

is a closure system on M.
The respective closure operator $X \mapsto \mathcal{L}(X)$ is constructed in the following way: For a set $X \subseteq M$, let

$$X^{\mathcal{L}} := X \cup \bigcup \{B \mid A \to B \in \mathcal{L}, A \subseteq X\}.$$

We form the sets $X^{\mathcal{L}}, X^{\mathcal{L}\mathcal{L}}, X^{\mathcal{L}\mathcal{L}\mathcal{L}}, \ldots$ until a set

$$\mathcal{L}(X) := X^{\mathcal{L}\ldots\mathcal{L}}$$

is obtained with $\mathcal{L}(X)^{\mathcal{L}} = \mathcal{L}(X)$ (i.e., a fixpoint).[1] $\mathcal{L}(X)$ is then the closure of X for the closure system $\mathrm{Mod}(\mathcal{L})$.

[1] If M is infinite, this may require infinitely many iterations.

Implications and Closure Systems

Def.: An implication $A \to B$ *follows (semantically)* from a set $\mathcal{L}$ of implications in M if each subset of M respecting $\mathcal{L}$ also respects $A \to B$. A family of implications is called *closed* if every implication following from $\mathcal{L}$ is already contained in $\mathcal{L}$.

Lemma: A set $\mathcal{L}$ of implications in M is closed, iff the following conditions (*Armstrong Rules*) are satisfied for all $W, X, Y, Z \subseteq M$:

1. $X \to X \in \mathcal{L}$,
2. If $X \to Y \in \mathcal{L}$, then $X \cup Z \to Y \in \mathcal{L}$,
3. If $X \to Y \in \mathcal{L}$ and $Y \cup Z \to W \in \mathcal{L}$, then $X \cup Z \to W \in \mathcal{L}$.

Remark: You should know these rules from the database lecture!

Pseudo-Intents and the Stem Base

Def.: A set $\mathcal{L}$ of implications of a context (G, M, I) is called *complete*, if every implication that holds in (G, M, I) follows from $\mathcal{L}$.
A set $\mathcal{L}$ of implications is called *non-redundant* if no implication in $\mathcal{L}$ follows from other implications in $\mathcal{L}$.

Def.: $P \subseteq M$ is called *pseudo intent* of (G, M, I), if

- $P \neq P''$, and
- if $Q \subsetneq P$ is a pseudo intent, then $Q'' \subseteq P$.

Theorem: The set of implications

$$\mathcal{L} := \{P \to P'' \mid P \text{ is pseudo intent}\}$$

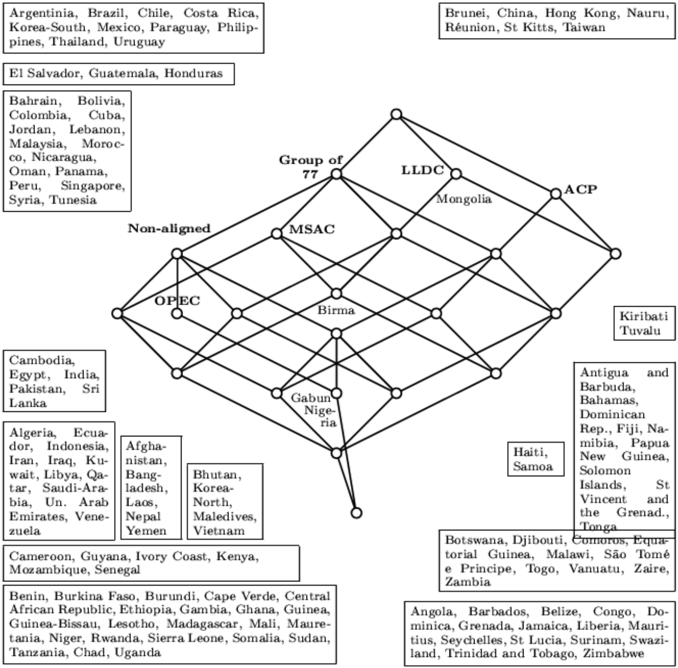
is non-redundant and complete. We call $\mathcal{L}$ the *stem base*.

Pseudo-Intents and the Stem Base

Example: membership of developing countries in supranational groups
(Source: Lexikon Dritte Welt. Rowohlt-Verlag, Reinbek 1993)

	Group of 77	Non-aligned	LLDC	MSAC	OPEC	ACP
Afghanistan	×	×	×	×		
Algeria	×	×			×	
Angola	×	×				×
Antigua and Barbuda	×					×
Argentina	×					
Bahamas	×					×
Bahrain	×	×				
Bangladesh	×	×	×	×		
Barbados	×	×				×
Belize	×	×				×
Benin	×	×	×	×		×
Bhutan	×	×	×			
Bolivia	×	×				
Botswana	×	×	×			×
Brazil	×					
Brunei						
Burkina Faso	×	×	×	×		×
Burundi	×	×	×	×		×
Cambodia	×	×		×		
Cameroon	×	×		×		×
Cape Verde	×	×	×	×		×
Central African Rep.	×	×	×	×		×
Chad	×	×	×	×		×
Chile	×					
China						
Colombia	×	×				
Comoros	×	×	×			×
Congo	×	×				×
Costa Rica	×					
Cuba	×	×				
Djibouti	×	×	×			×
Dominica	×	×				×
Dominican Rep.	×					×
Ecuador	×	×			×	
Egypt	×	×		×		
El Salvador	×			×		
Equatorial Guinea	×	×	×			×
Ethiopia	×	×	×	×		×
Fiji	×					×
Gabon	×	×			×	×
Gambia	×	×	×	×		×
Ghana	×	×	×	×		×
Grenada	×	×				×
Guatemala	×			×		
Guinea	×	×	×	×		×
Guinea-Bissau	×	×	×	×		×
Guyana	×	×		×		×
Haiti	×		×	×		×
Honduras	×			×		
Hong Kong						
India	×	×		×		
Indonesia	×	×			×	
Iran	×	×			×	
Iraq	×	×			×	
Ivory Coast	×	×		×		×
Jamaica	×	×				×
Jordan	×	×				
Kenya	×	×		×		×
Kiribati			×			×
Korea-North	×	×	×			
Korea-South	×					
Kuwait	×	×			×	
Laos	×	×	×	×		
Lebanon	×	×				
Lesotho	×	×	×	×		×
Liberia	×	×				×
Libya	×	×			×	
Madagascar	×	×	×	×		×
Malawi	×	×	×			×
Malaysia	×	×				
Maledives	×	×	×			
Mali	×	×	×	×		×
Mauretania	×	×	×	×		×
Mauritius	×	×				×
Mexico	×					
Mongolia			×			
Morocco	×	×				
Mozambique	×	×		×		×
Myanmar	×		×	×		
Namibia	×					×
Nauru						
Nepal	×	×	×	×		
Nicaragua	×	×				
Niger	×	×	×	×		×
Nigeria	×	×			×	×
Oman	×	×				
Pakistan	×	×		×		
Panama	×	×				
Papua New Guinea	×					×
Paraguay	×					
Peru	×	×				
Philippines	×					
Qatar	×	×			×	
Réunion						
Rwanda	×	×	×	×		×
Samoa	×		×	×		×
São Tomé e Principe	×	×	×			×
Saudi Arabia	×	×			×	
Senegal	×	×		×		×
Seychelles	×	×				×
Sierra Leone	×	×	×	×		×
Singapore	×	×				
Solomon Islands	×					×
Somalia	×	×	×	×		×
Sri Lanka	×	×		×		
St Kitts						
St Lucia	×	×				×
St Vincent& Grenad.	×					×
Sudan	×	×	×	×		×
Surinam	×	×				×
Swaziland	×	×				×
Syria	×	×				
Taiwan						
Tanzania	×	×	×	×		×
Thailand	×					
Togo	×	×	×			×
Tonga	×					×
Trinidad and Tobago	×	×				×
Tunisia	×	×				
Tuvalu			×			×
Uganda	×	×	×	×		×
United Arab Emirates	×	×			×	
Uruguay	×					
Vanuatu	×	×	×			×
Venezuela	×	×			×	
Vietnam	×	×	×			
Yemen	×	×	×	×		
Zaire	×	×	×			×
Zambia	×	×	×			×
Zimbabwe	×	×				×

The abbreviations stand for: LLDC := Least Developed Countries, MSAC := Most Seriously Affected Countries, OPEC := Organization of Petrol Exporting Countries, ACP := African, Caribbean and Pacific Countries.


Argentinia, Brazil, Chile, Costa Rica, Korea-South, Mexico, Paraguay, Philippines, Thailand, Uruguay
Brunei, China, Hong Kong, Nauru, Réunion, St Kitts, Taiwan
El Salvador, Guatemala, Honduras
Bahrain, Bolivia, Colombia, Cuba, Jordan, Lebanon, Malaysia, Morocco, Nicaragua, Oman, Panama, Peru, Singapore, Syria, Tunesia
Group of 77
LLDC
Mongolia
ACP
Non-aligned
MSAC
OPEC
Birma
Kiribati Tuvalu
Cambodia, Egypt, India, Pakistan, Sri Lanka
Gabun Nigeria
Antigua and Barbuda, Bahamas, Dominican Rep., Fiji, Namibia, Papua New Guinea, Solomon Islands, St Vincent and the Grenad., Tonga
Algeria, Ecuador, Indonesia, Iran, Iraq, Kuwait, Libya, Qatar, Saudi-Arabia, Un. Arab Emirates, Venezuela
Afghanistan, Bangladesh, Laos, Nepal Yemen
Bhutan, Korea-North, Maledives, Vietnam
Haiti, Samoa
Botswana, Djibouti, Comoros, Equatorial Guinea, Malawi, São Tomé e Principe, Togo, Vanuatu, Zaire, Zambia
Cameroon, Guyana, Ivory Coast, Kenya, Mozambique, Senegal
Benin, Burkina Faso, Burundi, Cape Verde, Central African Republic, Ethiopia, Gambia, Ghana, Guinea, Guinea-Bissau, Lesotho, Madagascar, Mali, Mauretania, Niger, Rwanda, Sierra Leone, Somalia, Sudan, Tanzania, Chad, Uganda
Angola, Barbados, Belize, Congo, Dominica, Grenada, Jamaica, Liberia, Mauritius, Seychelles, St Lucia, Surinam, Swaziland, Trinidad and Tobago, Zimbabwe

Pseudo-Intents and the Stem Base

stem base of the developing countries context:

$$\{\text{OPEC}\} \to \{\text{Group of 77, Non-Alligned}\}$$
$$\{\text{MSAC}\} \to \{\text{Group of 77}\}$$
$$\{\text{Non-Alligned}\} \to \{\text{Group of 77}\}$$
$$\{\text{Group of 77, Non-Alligned, MSAC, OPEC}\} \to \{\text{LLDC, AKP}\}$$
$$\{\text{Group of 77, Non-Alligned, LLDC, OPEC}\} \to \{\text{MSAC, AKP}\}$$

Computing the Stem Base With Next Closure

The algorithm Next Closure to compute all concept intents and the stem base:

1. The set $\mathcal{L}$ of all implications is initialized to $\mathcal{L} = \emptyset$.
2. The lectically first concept intent or pseudo-intent is $\emptyset$.
3. If A is an intent or a pseudo-intent, the lectically next intent/pseudo-intent is computed by checking all $i \in M \backslash A$ in descending order, until $A <_i \mathcal{L}(A + i)$ holds.
   Then $\mathcal{L}(A + i)$ is the next intent or pseudo-intent.
4. If $\mathcal{L}(A + i) = (\mathcal{L}(A + i))''$ holds, then $\mathcal{L}(A + i)$ is a concept intent, otherwise it is a pseudo-intent and the implication $\mathcal{L}(A + i) \to (\mathcal{L}(A + i))''$ is added to $\mathcal{L}$.
5. If $\mathcal{L}(A + i) = M$, finish. Else, set $A \leftarrow \mathcal{L}(A + i)$ and continue with Step 3.

Computing the Stem Base With Next Closure

Example:

	a	b	c	e
1	×		×	
2		×		×
3		×	×	×

A	i	$A+i$	$\mathcal{L}(A+i)$	$A <_i \mathcal{L}(A+i)$?	$(\mathcal{L}(A+i))''$	$\mathcal{L}$	new intent

Agenda

4 Implications
- Implications
- Attribute Logic
- Concept Intents and Implications
- Implications and Closure Systems
- Pseudo-Intents and the Stem Base
- Computing the Stem Base With NEXT CLOSURE
- Bases of Association Rules

Bases of Association Rules

{veil color: white, gill spacing: close} $\rightarrow$ {gill attachment: free}
support: 78.52% confidence: 99.60%

The input data to compute association rules can be represented as a formal context (G, M, I):

- M is a set of *items* (things, products of a market basket),
- G contains the *transaction ids*,
- and the relation I the *list of transactions*.

Bases of Association Rules

{veil color: white, gill spacing: close} → {gill attachment: free}
support: 78.52% confidence: 99.60%

The *support* of an implication is the fraction of all objects that have all attributes from the premise and the conclusion.
(repetition: the support of an attribute set $X \subseteq M$ is $\text{supp}(X) := \frac{|X'|}{|G|}$.)
Def.: The *support of a rule* $X \to Y$ is given by

$$\text{supp}(X \to Y) := \text{supp}(X \cup Y)$$

The *confidence* is the fraction of all objects that fulfill both the premise and the conclusion among those objects that fulfill the premise.
Def.: The *confidence of a rule* $X \to Y$ is given by

$$\text{conf}(X \to Y) := \frac{\text{supp}(X \cup Y)}{\text{supp}(X)}$$

Bases of Association Rules

{veil color: white, gill spacing: close} → {gill attachment: free}
support: 78.52% confidence: 99.60%

Classical data mining task: Find for given $minsupp, minconf \in [0, 1]$ all rules with a support and confidence above these bounds.

Our task: finding a *base* of rules, i.e., a minimal set of rules from which all other rules follow.

Bases of Association Rules

From $B' = B'''$ follows

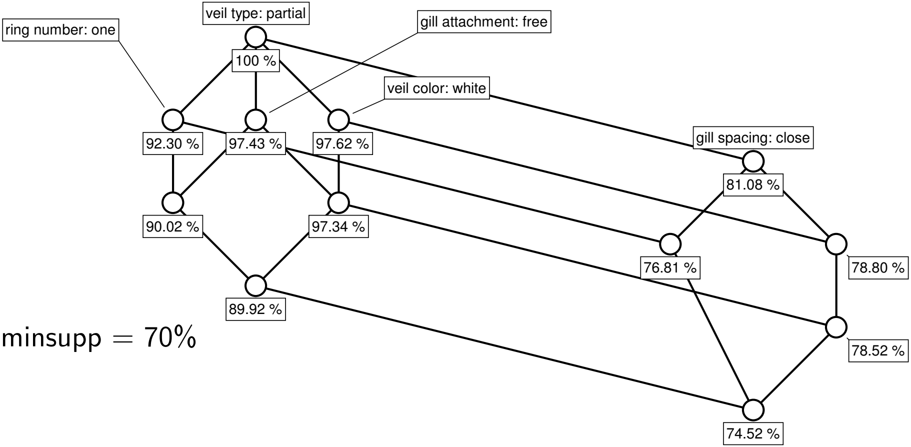
$$\text{supp}(B) = \frac{|B'|}{|G|} = \frac{|B'''|}{|G|} = \text{supp}(B'')$$

Theorem: $X \to Y$ and $X'' \to Y''$ have the same support and the same confidence.

To compute *all* association rules it is thus sufficient to compute the support of all frequent sets with $B = B''$ (i.e., the intents of the iceberg concept lattice).

Bases of Association Rules

The Benefit of Iceberg Concept Lattices (Compared to Frequent Itemsets)



minsupp = 70%

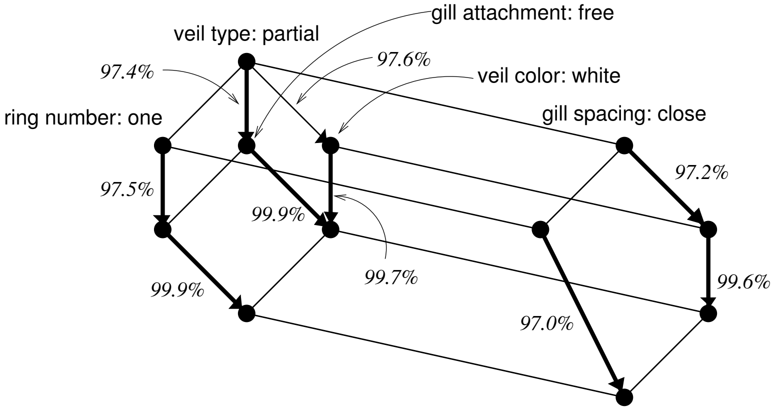
32 frequent itemsets are represented by 12 frequent concept intents

→ *more efficient computation (e.g., TITANIC)*

→ *fewer rules (without loss of information!)*

Bases of Association Rules

The Benefit of Iceberg Concept Lattices (Compared to Frequent Itemsets)



Association rules can be visualized in the (iceberg) concept lattice:

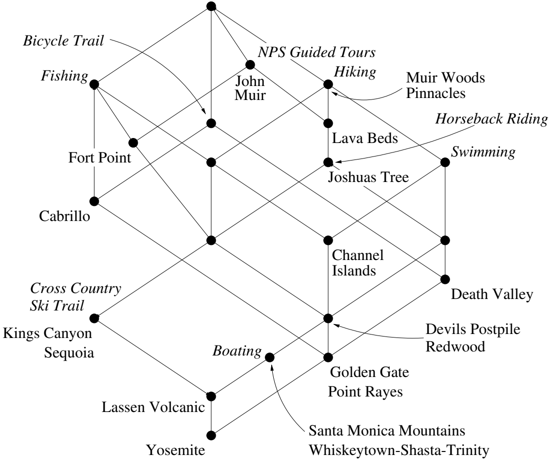
exact association rules (implications): $conf = 100\%$

(approximate) association rules: $conf < 100\%$

Bases of Association Rules: Exact Association Rules

...can be read off from the stem base. In concept lattices we can read them directly off from the diagram:

Lemma: An implication $X \to Y$ holds, iff the largest concept that is below the concepts that are generated by the attributes of X is below all concepts that are generated by the attributes in Y.



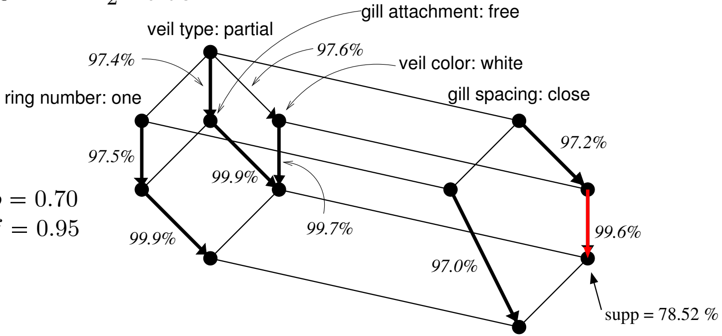
Examples:

- $\{$*Swimming*$\} \to \{$*Hiking*$\}$ ($supp = 10/19 \approx 52.6\%$, $conf = 100\%$)
- $\{$*Boating*$\} \to \{$*Swimming, Hiking, NPS Guided Tours, Fishing, Horseback Riding*$\}$ ($supp = 4/19 \approx 21.0\%$, $conf = 100\%$)
- $\{$*Bicycle Trail, NPS Guided Tours*$\} \to \{$*Swimming, Hiking, Horseback Riding*$\}$ ($supp = 4/19 \approx 21.0\%$, $conf = 100\%$)

Bases of Association Rules

Def.: The *Luxenburger basis* contains all valid approximate association rules $X \to Y$, such that concepts (A_1, B_1) and (A_2, B_2) exist, with (A_1, B_1) being a direct upper neighbor of (A_2, B_2), such that $X = B_1$ and $X \cup Y = B_2$ holds.

$minsupp = 0.70$
$minconf = 0.95$



Every arrow shows a rule of the basis. E.g., the right arrow stands for {veil type: partial, gill spacing: close, veil color: white} → {gill attachment: free} ($conf = 99.6\%$, $supp = 78.52\%$)

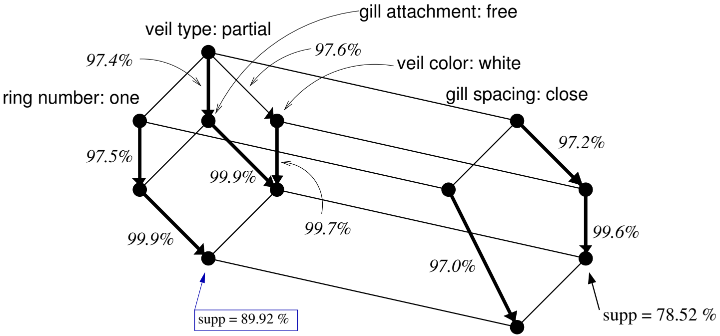
Bases of Association Rules

Theorem: From the Luxenburger basis all approximate association rules (incl. support and confidence) can be derived by the following rules:

- $\phi(X \to Y) = \phi(X \to Y \setminus Z)$, for $\phi \in \{\text{conf}, \text{supp}\}, Z \subseteq X$
- $\phi(X'' \to Y'') = \phi(X \to Y)$
- $\text{conf}(X \to X) = 1$
- $\text{conf}(X \to Y) = p, \text{conf}(Y \to Z) = q \Rightarrow \text{conf}(X \to Z) = pq$ for all frequent concept intents $X \subset Y \subset Z$.
- $\text{supp}(X \to Z) = \text{supp}(Y \to Z)$ for all $X, Y \subseteq Z$

The basis is minimal with respect to this property.

Bases of Association Rules



example

{ring number: one} → {veil color: white}

- has a support of 89.92% (the support of the largest concept which contains both attributes in its intent)
- and confidence $97.5\% \cdot 99.9\% \approx 97.4\%$.

Some experimental results

Dataset (Minsupp)	Exact rules	stem basis	Minconf	asssociation rules	Luxenburger basis
T10I4D100K (0.5%)	0	0	90%	16,269	3,511
			70%	20,419	4,004
			50%	21,686	4,191
			30%	22,952	4,519
MUSHROOMS (30%)	7,476	69	90%	12,911	563
			70%	37,671	968
			50%	56,703	1,169
			30%	71,412	1,260
C20D10K (50%)	2,277	11	90%	36,012	1,379
			70%	89,601	1,948
			50%	116,791	1,948
			30%	116,791	1,948
C73D10K (90%)	52,035	15	95%	1,606,726	4,052
			90%	2,053,896	4,089
			85%	2,053,936	4,089
			80%	2,053,936	4,089