The University Of Sheffield.

Über die Entwicklung von Programmen zur Begriffsanalyse

Robert Jäschke

Information School, University of Sheffield

Ernst-Schröder-Seminar, Darmstadt, 3.12.2016

Ausgangslage 2006

- ToscanaJ, Siena, Elba, Galicia, ConExp, ...
- NEXT CLOSURE, TITANIC, ...
- 1. Conceptual Structures Tools Interoperability Workshop

Ausgangslage 2006

- ToscanaJ, Siena, Elba, Galicia, ConExp, ...
- NEXT CLOSURE, TITANIC, ...
- 1. Conceptual Structures Tools Interoperability Workshop

→ nichts passendes für triadische FBA

Agenda

Triadische FBA mit TRIAS

Merkmalsexploration im Web

ConExp-NG

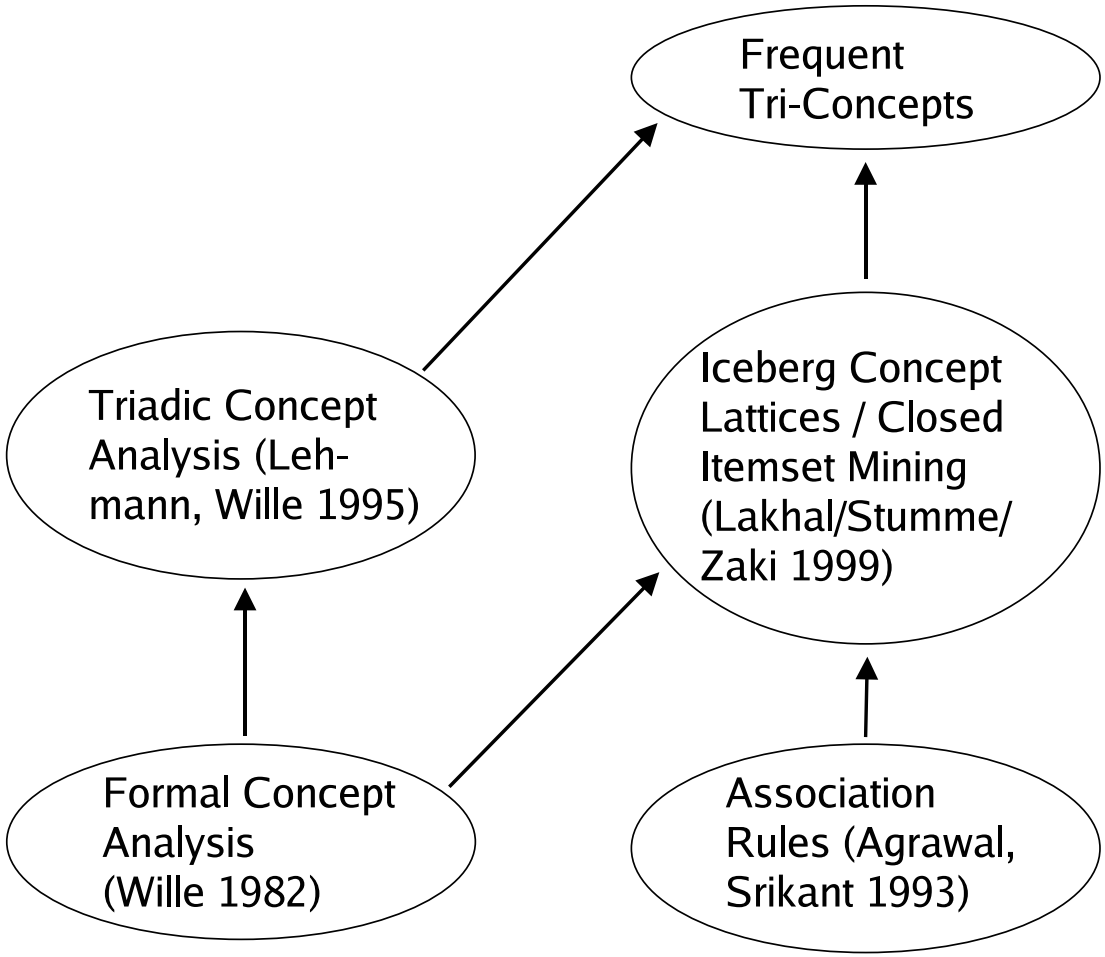
Agenda

Triadische FBA mit TRIAS

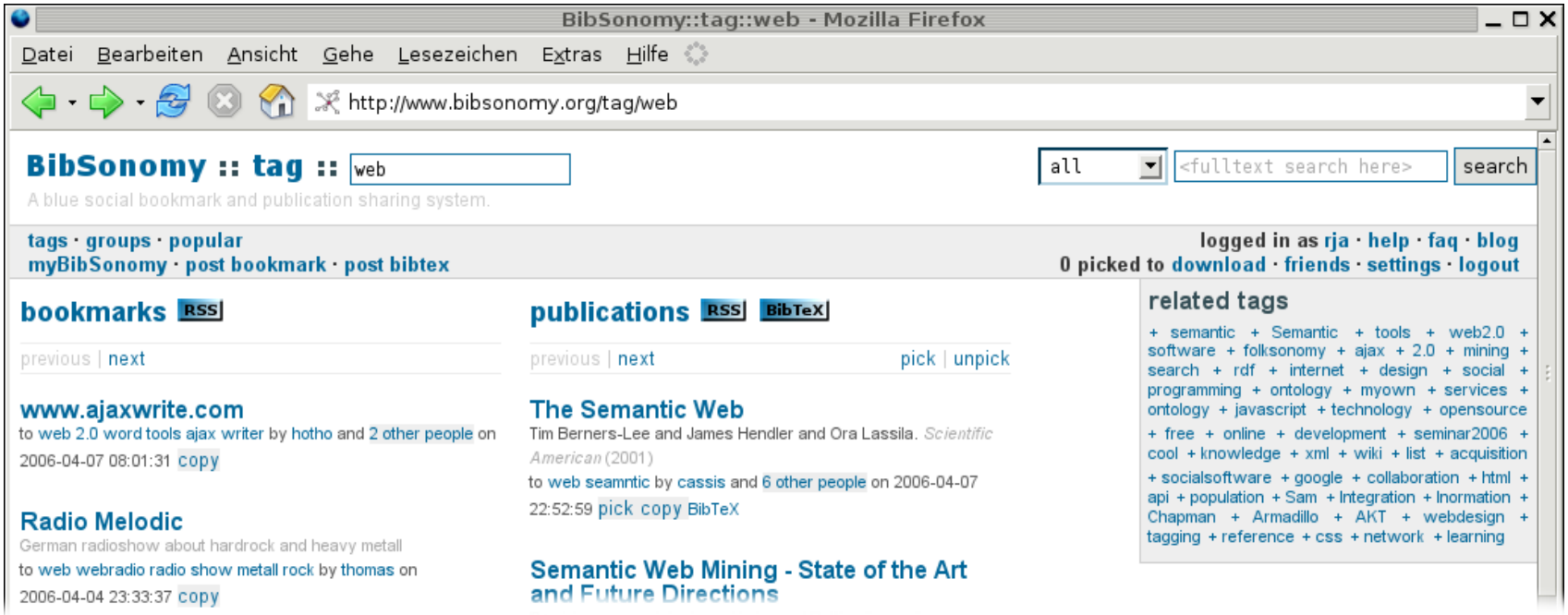
Merkmalsexploration im Web

ConExp-NG

Geschichte der Eisberg-Tri-Verbände



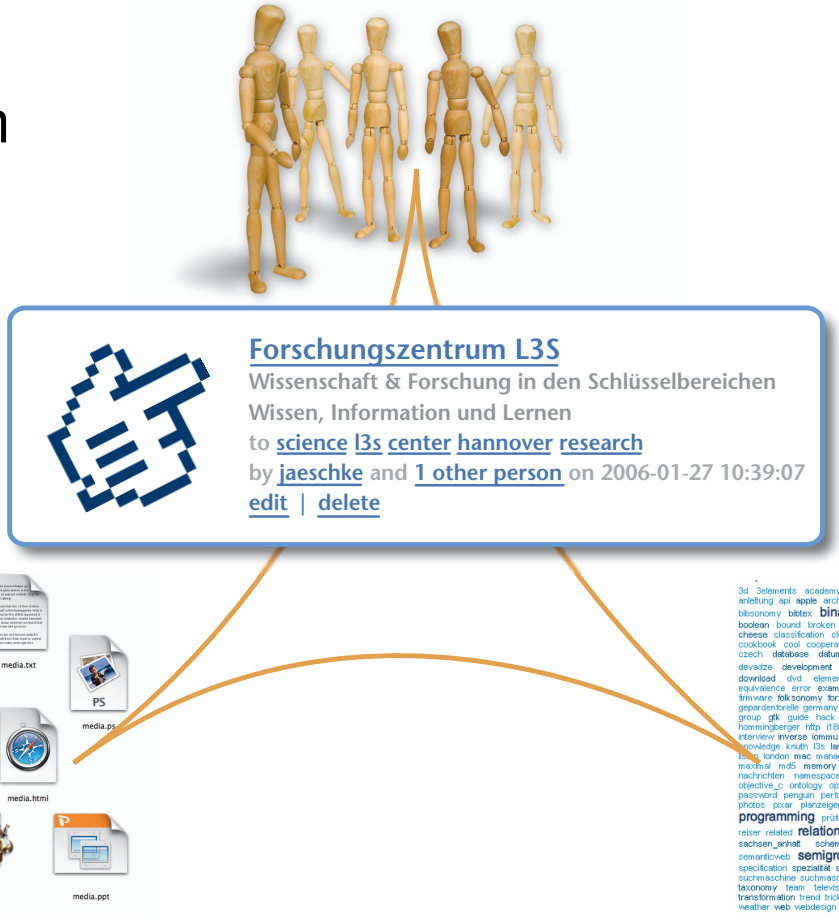
Motivation



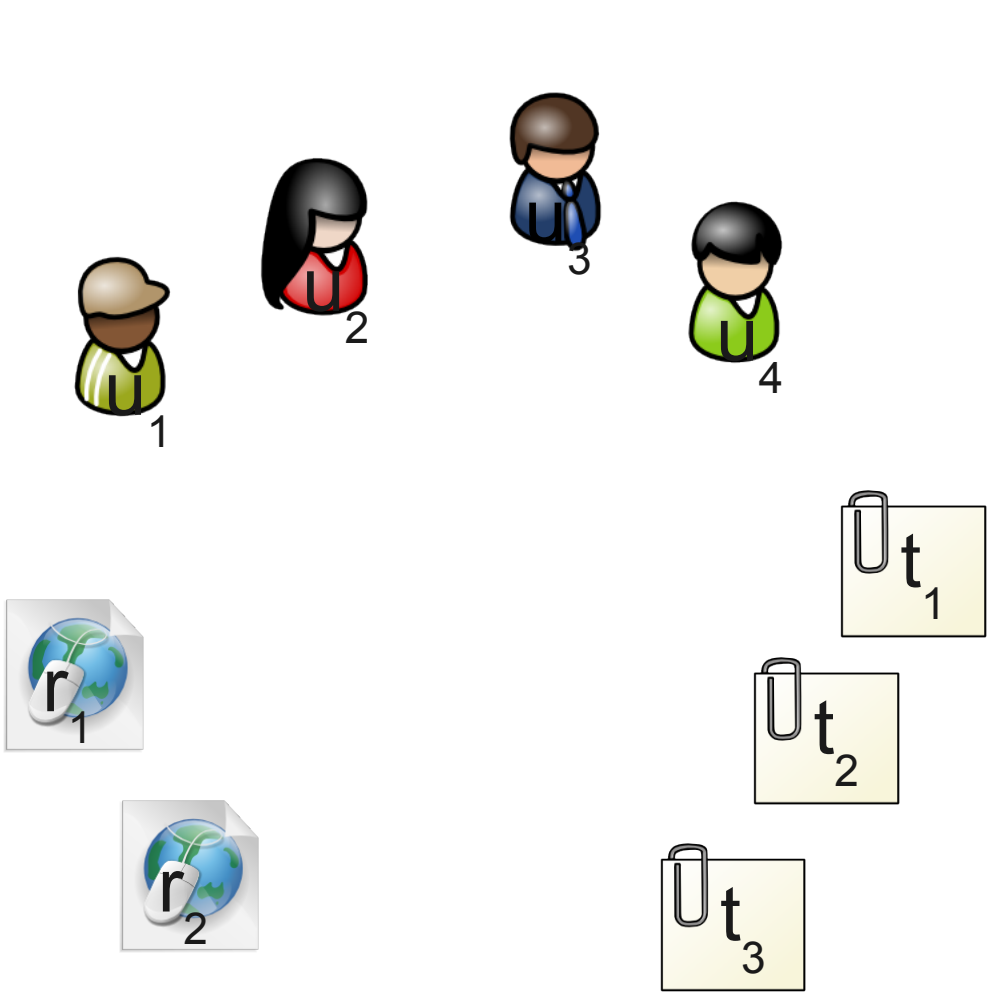
- Social Tagging: Delicious, Flickr, 43Things, BibSonomy, ...
- *Benutzer* annotieren *Ressourcen* mit *Schlagwörtern (Tags)*
- Beispiel: *jaeschke*, {*fca*, *tools* }, *https://github.com/fcatools*

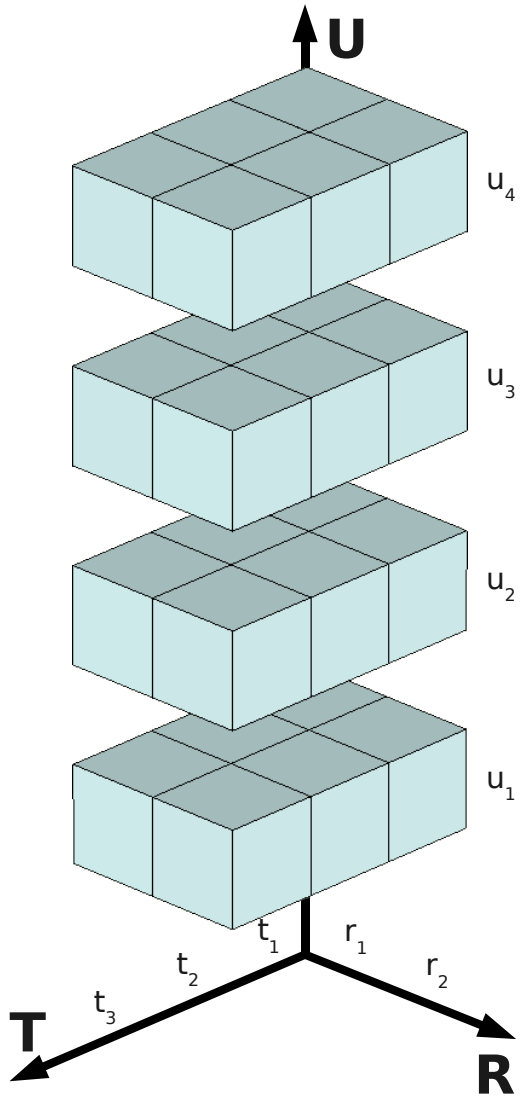
Folksonomien

- Struktur von Social-Tagging-Systemen
- Verknüpft Benutzer, Tags und Ressourcen
- Begriffliche Struktur, durch Beitrag vieler Menschen erschaffen

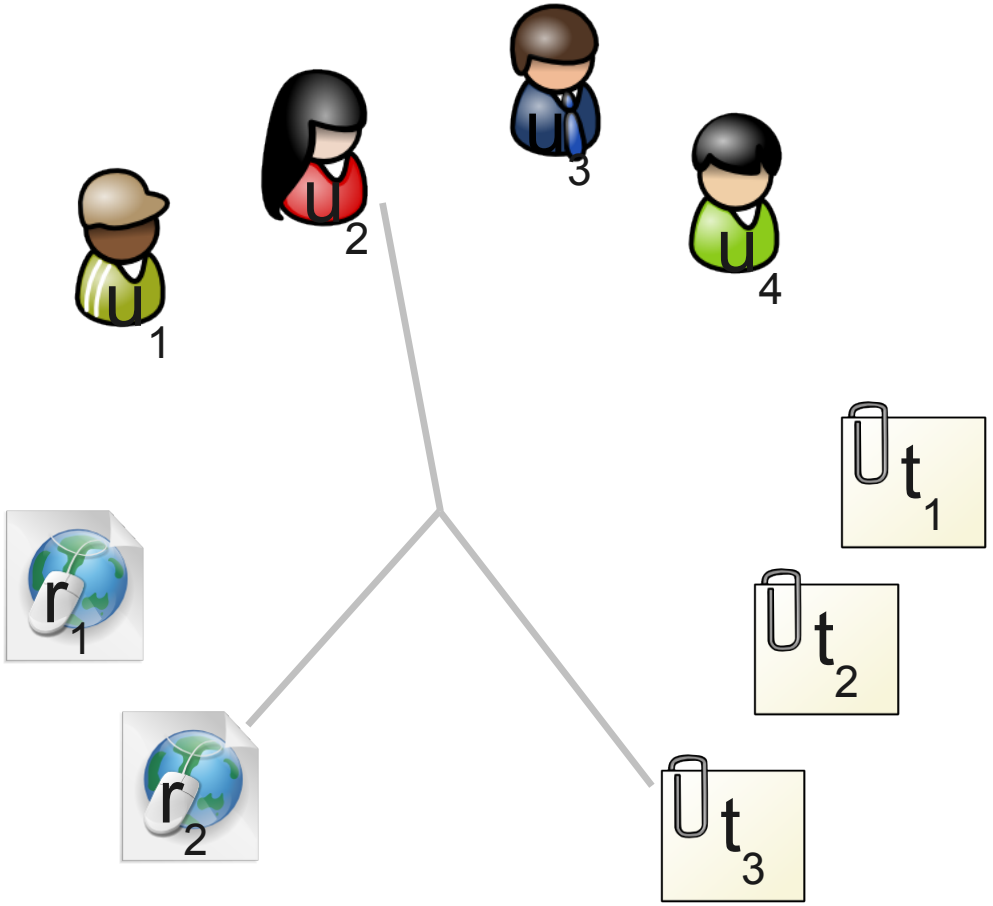


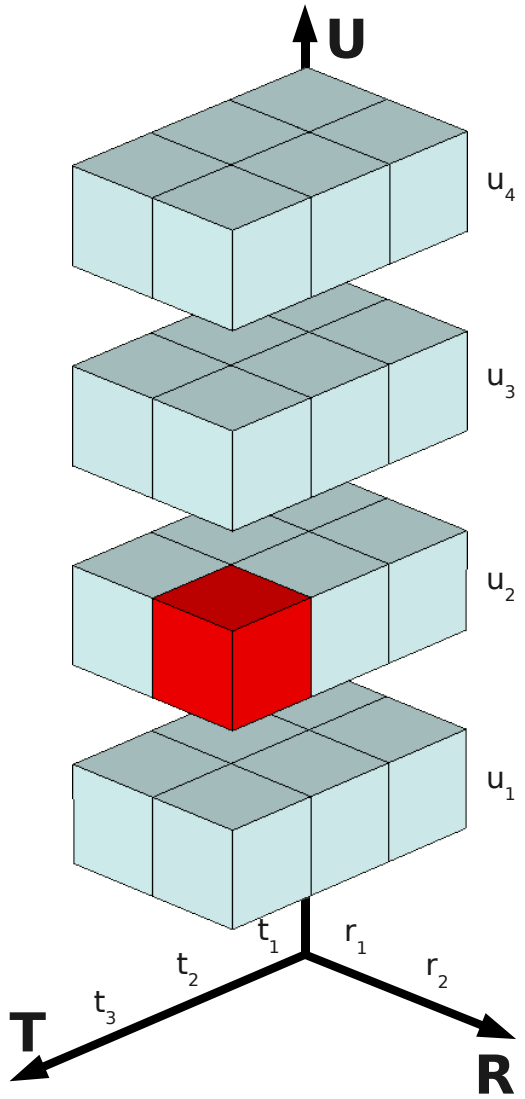
Folksonomien



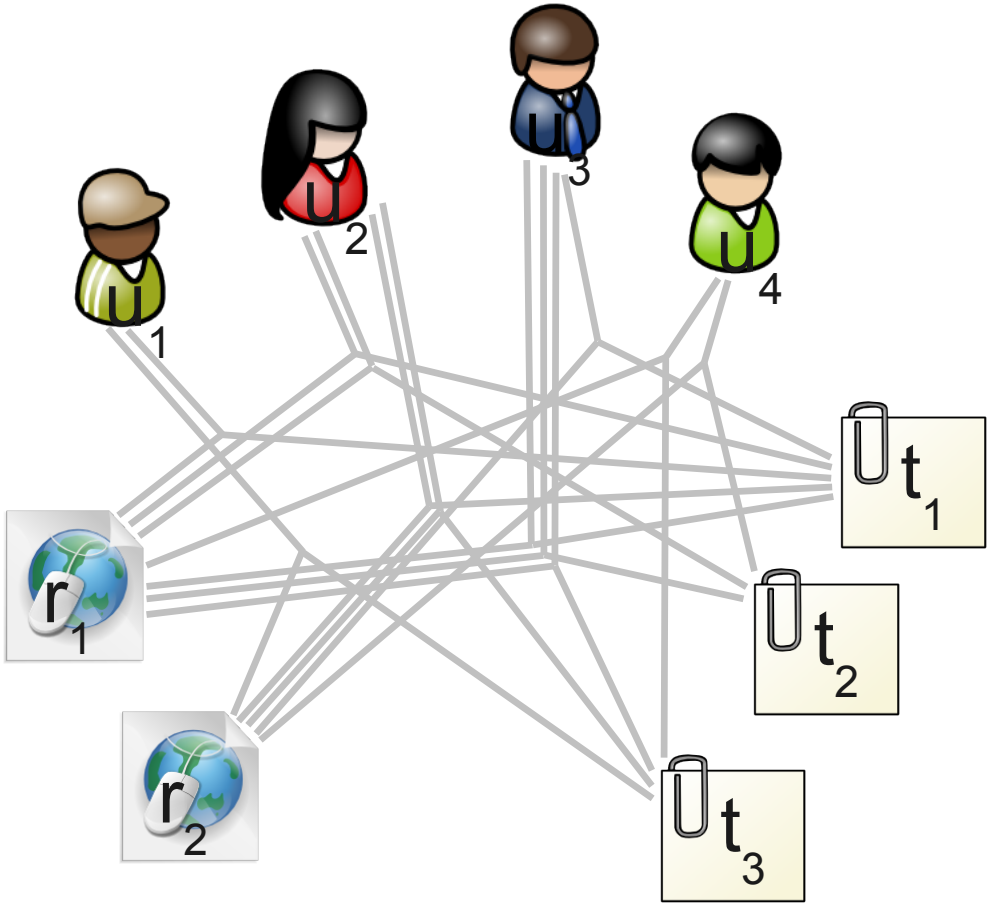


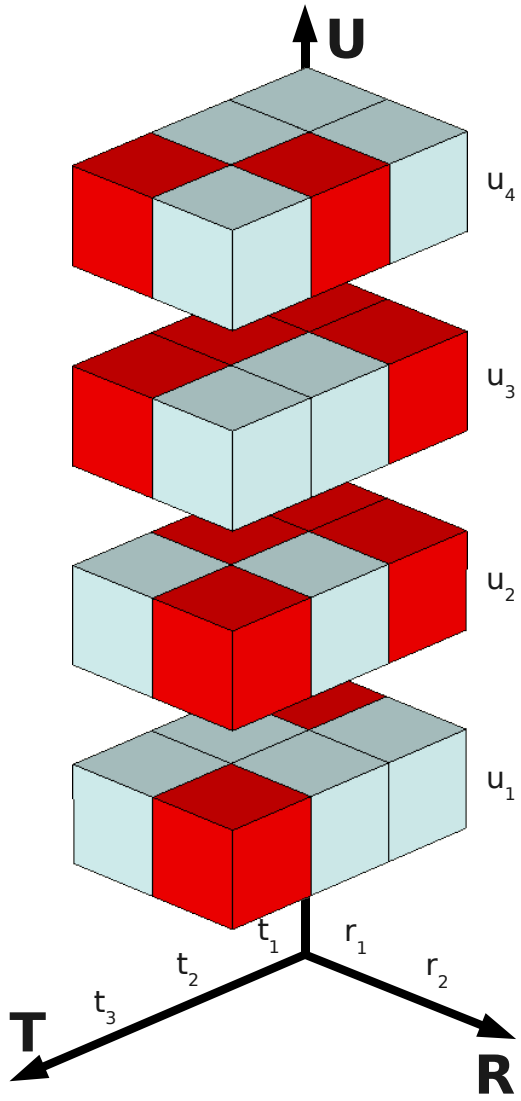
Folksonomien





Folksonomien





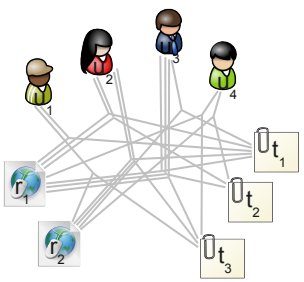
Folksonomien

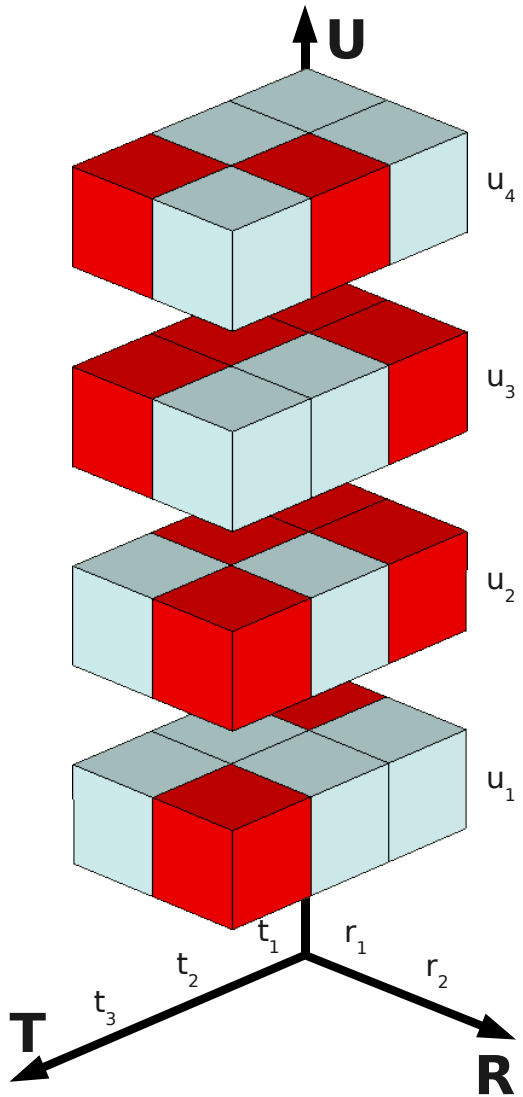
Definition (Folksonomie)

Ein Tupel $\mathbb{F} := (U, T, R, Y)$ mit

- U, T, R endliche Mengen von Benutzern, Tags und Ressourcen
- $Y \subseteq U \times T \times R$ ternäre Relation

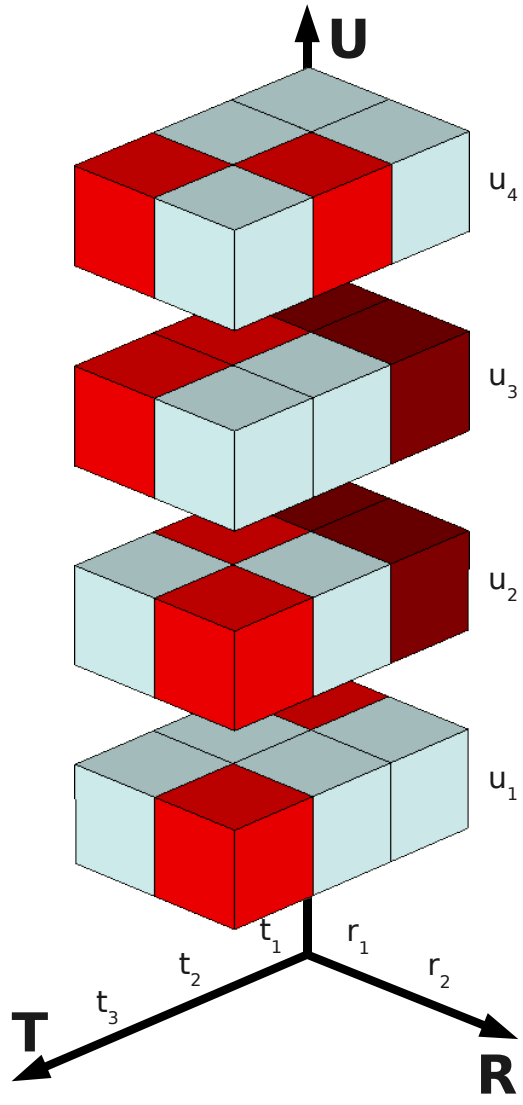
- Tripartiter Hypergraph
- Boolescher 3-dimensionaler Tensor
- Triadischer Formaler Kontext





Ziele

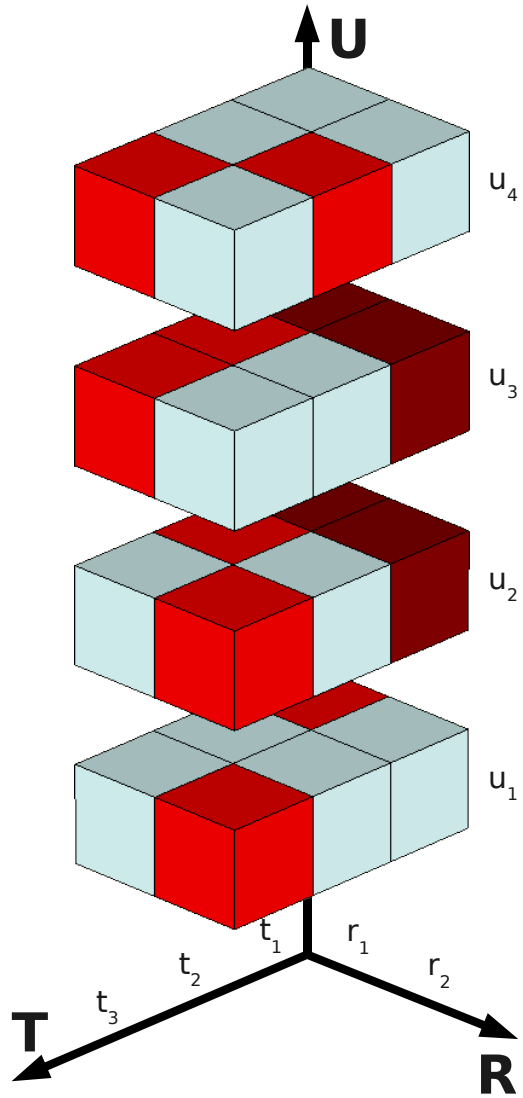
- begriffliches Clustern von Folksonomien
  - interessante Begriffe/Cluster finden
  - Browsing, Community-Entdeckung, Empfehlungen unterstützen
  - Einblick in die Struktur der Folksonomie gewinnen



Ziele

- begriffliches Clustern von Folksonomien
  - interessante Begriffe/Cluster finden
  - Browsing, Community-Entdeckung, Empfehlungen unterstützen
  - Einblick in die Struktur der Folksonomie gewinnen
- Tri-Begriff (A, B, C): maximaler Quader, in dem jeder Nutzer aus A jede Ressource aus C mit allen Tags aus B versehen hat

  → geteilte Konzeptualisierung



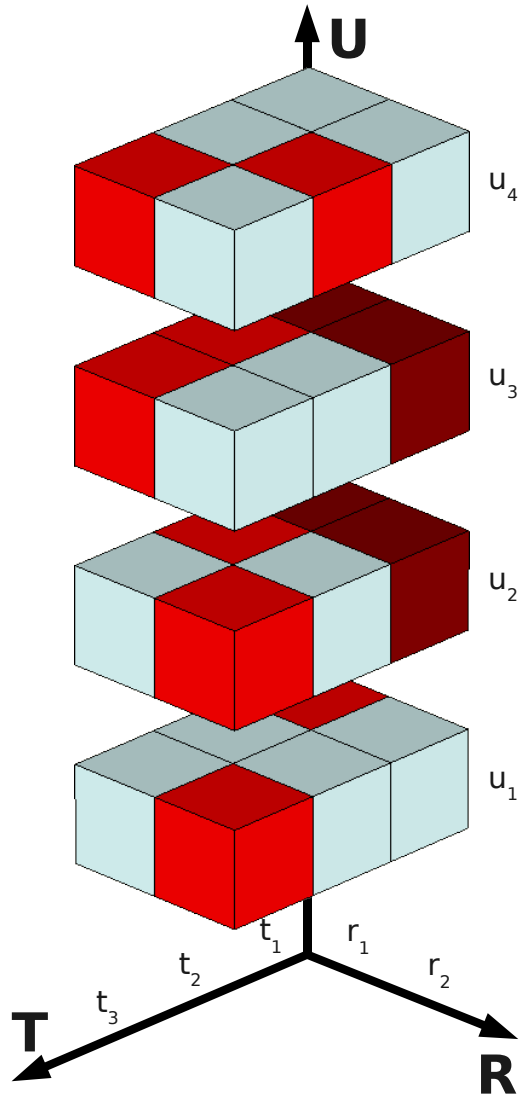
Tri-Begriff

Wir betrachten $\mathbb{F} = (U, T, R, Y)$ als triadischen Formalen Kontext.

Definition (Tri-Begriff)

Tripel (A, B, C) mit $A \subseteq U$, $B \subseteq T$, $C \subseteq R$ und $A \times B \times C \subseteq Y$, so dass keine der drei Komponenten vergrößert werden kann, ohne die Bedingung $A \times B \times C \subseteq Y$ zu verletzen.

→ Natürliche Erweiterung Formaler Begriffe



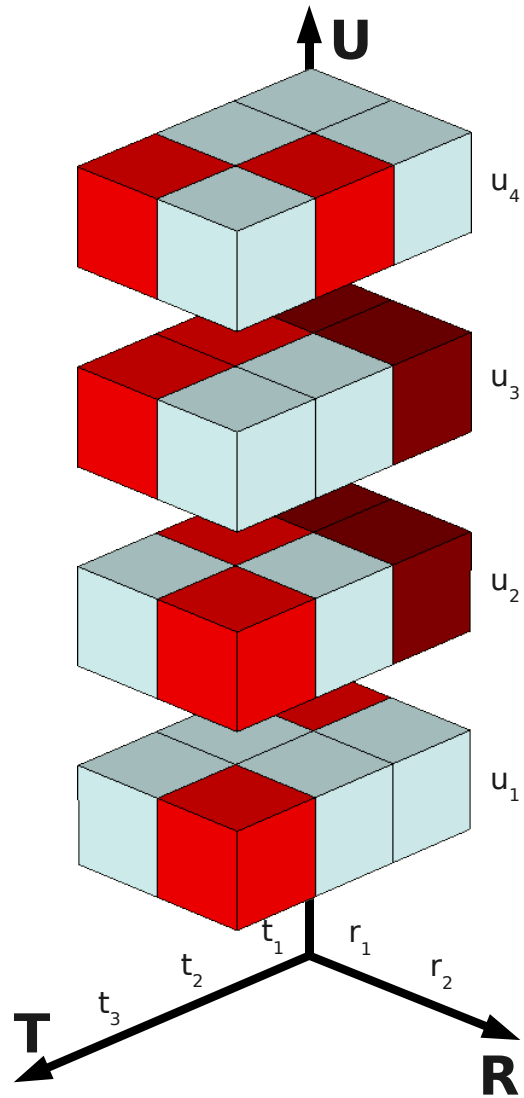
Tri-FBA

- Drei Quasi-Ordnungen $\lesssim_1, \lesssim_2, \lesssim_3$:
  $(A_1, A_2, A_3) \lesssim_i (B_1, B_2, B_3)$
  $:\Leftrightarrow\ A_i \subseteq B_i$, für $i = 1, 2, 3$.

  $\rightarrow$ *Begriffs-Tri-Verband* $\underline{\mathfrak{B}}(\mathbb{K})$ des triadischen Kontexts $\mathbb{K}$

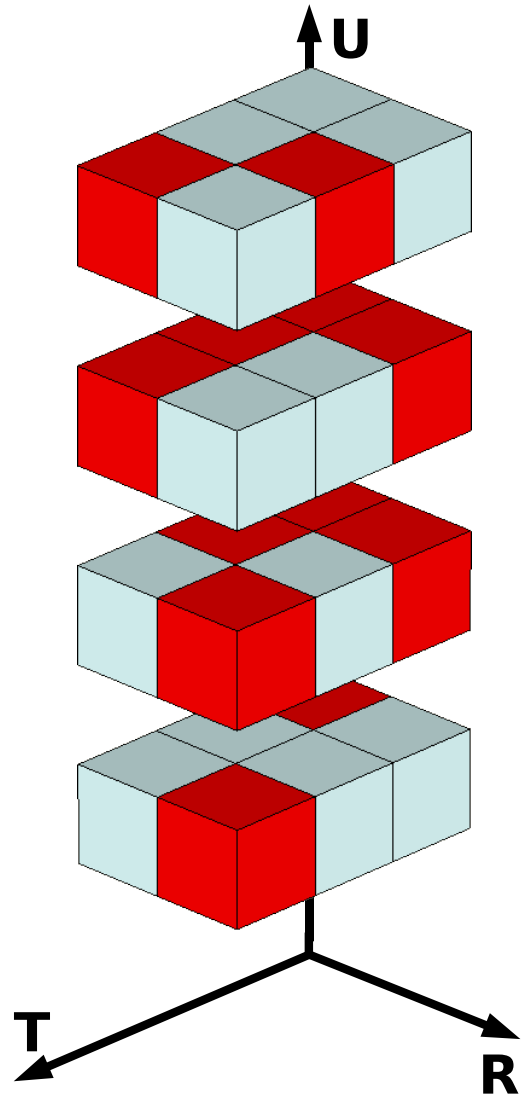
- Gegeben Support-Schranken τ_u, τ_t, τ_r :
  Tri-Begriff (A, B, C) *häufig*
  $:\Leftrightarrow\ |A| \geq \tau_u$, $|B| \geq \tau_t$, und $|C| \geq \tau_r$

  $\rightarrow$ *Eisberg-Tri-Verband*



Der TRIAS-Algorithmus

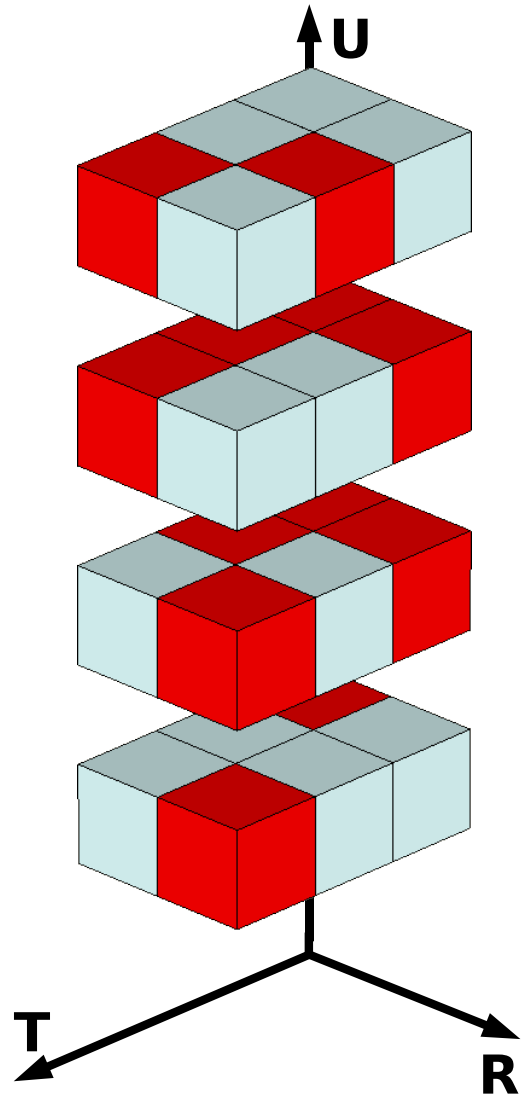
Berechnet den Eisberg-Tri-Verband eines triadischen Formalen Kontexts



Der TRIAS-Algorithmus

Berechnet den Eisberg-Tri-Verband eines triadischen Formalen Kontexts

Ablauf

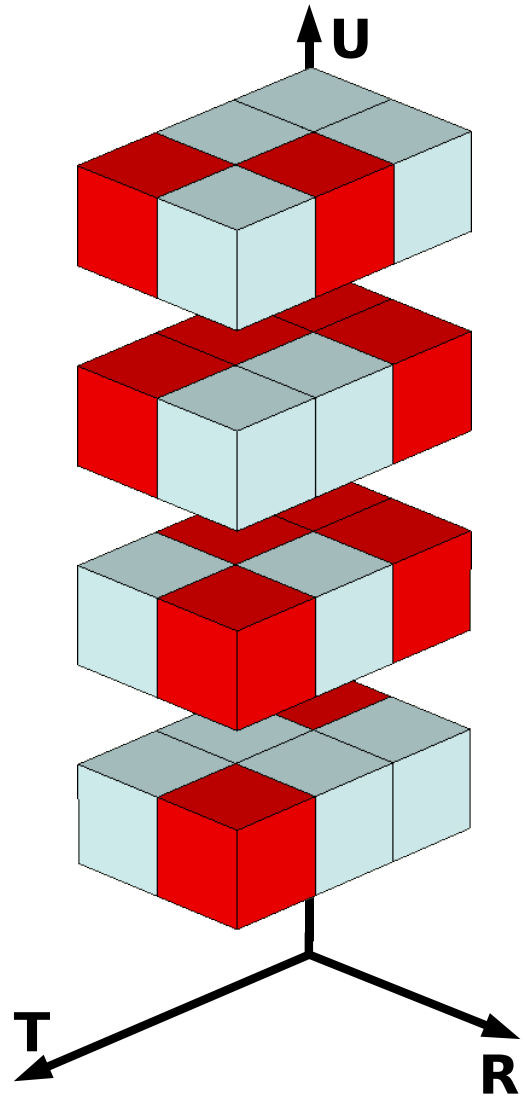


Der TRIAS-Algorithmus

Berechnet den Eisberg-Tri-Verband eines triadischen Formalen Kontexts

Ablauf

- Sei $\tilde{Y} := \{(u, (t, r)) \mid (u, t, r) \in Y\}$

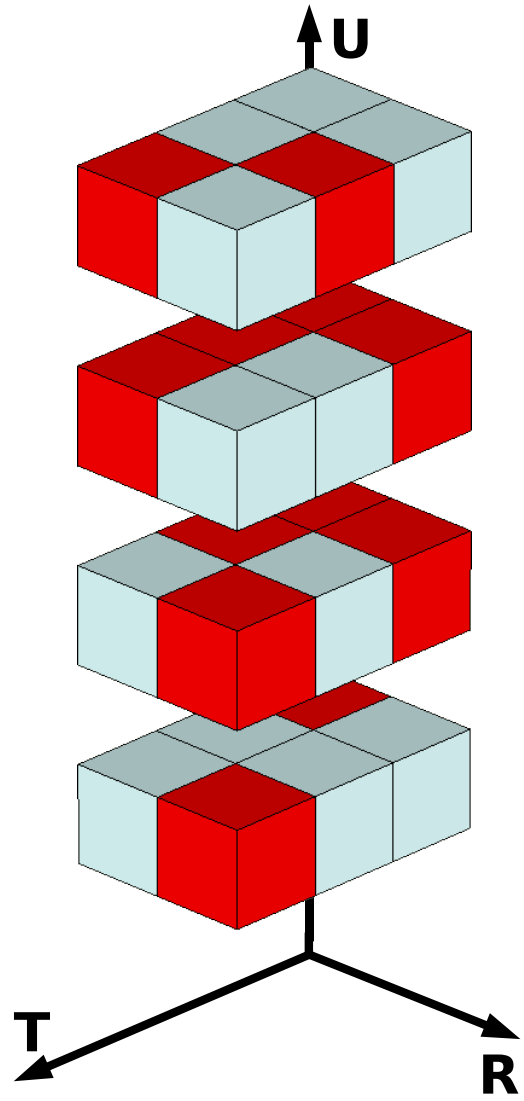


Der TRIAS-Algorithmus

Berechnet den Eisberg-Tri-Verband eines triadischen Formalen Kontexts

Ablauf

- Sei $\tilde{Y} := \{(u, (t, r)) \mid (u, t, r) \in Y\}$
- Schleife: Finde (häufige) Begriffe (A, I) in $(U, T \times R, \tilde{Y})$



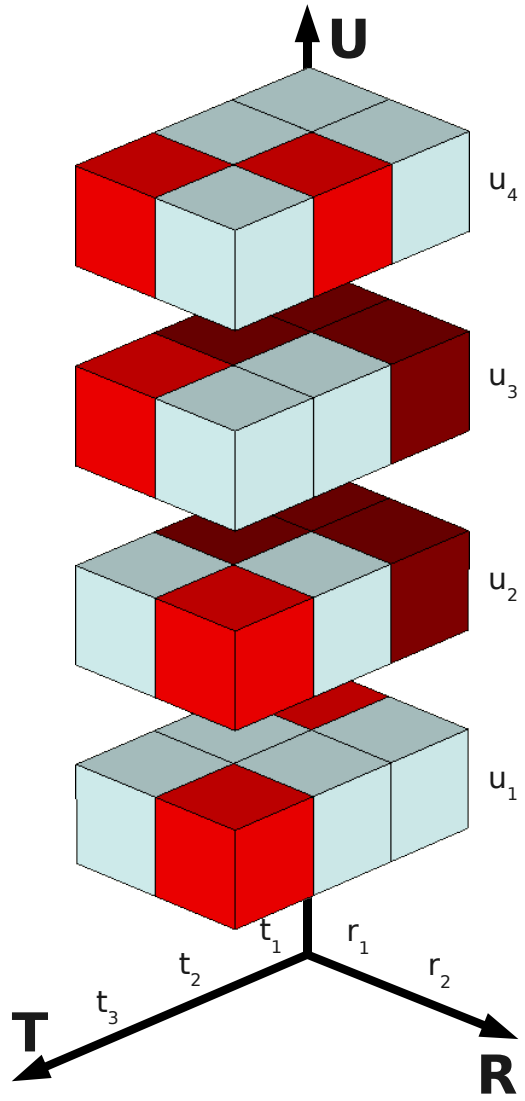
Der TRIAS-Algorithmus

Berechnet den Eisberg-Tri-Verband eines triadischen Formalen Kontexts

Ablauf

- Sei $\tilde{Y} := \{(u, (t, r)) \mid (u, t, r) \in Y\}$
- Schleife: Finde (häufige) Begriffe (A, I) in $(U, T \times R, \tilde{Y})$

Im Beispiel:
$(A, I) = \big(\{u_2, u_3\},\ \{(t_1, r_1), (t_1, r_2), (t_2, r_1)\}\big)$



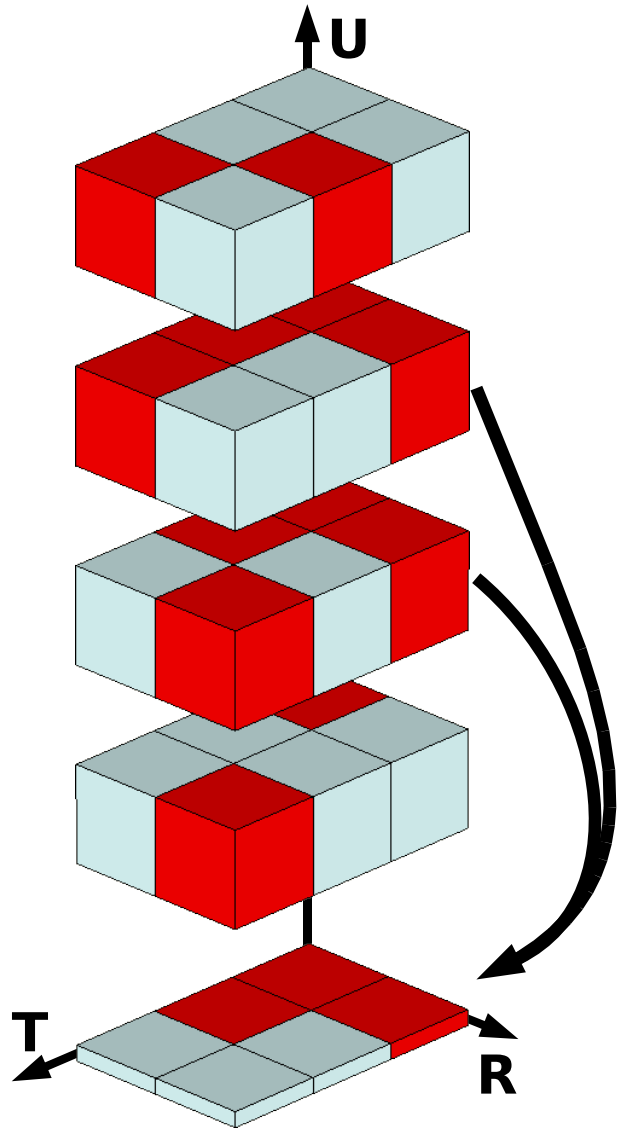
Der TRIAS-Algorithmus

Berechnet den Eisberg-Tri-Verband eines triadischen Formalen Kontexts

Ablauf

- Sei $\tilde{Y} := \{(u, (t, r)) \mid (u, t, r) \in Y\}$
- Schleife: Finde (häufige) Begriffe (A, I) in $(U, T \times R, \tilde{Y})$
  - Schleife: Finde (häufige) Begriffe (B, C) in (T, R, I)

Im Beispiel:
$(T, R, I) = (T, R, \{(t_1, r_1), (t_1, r_2), (t_2, r_1)\})$



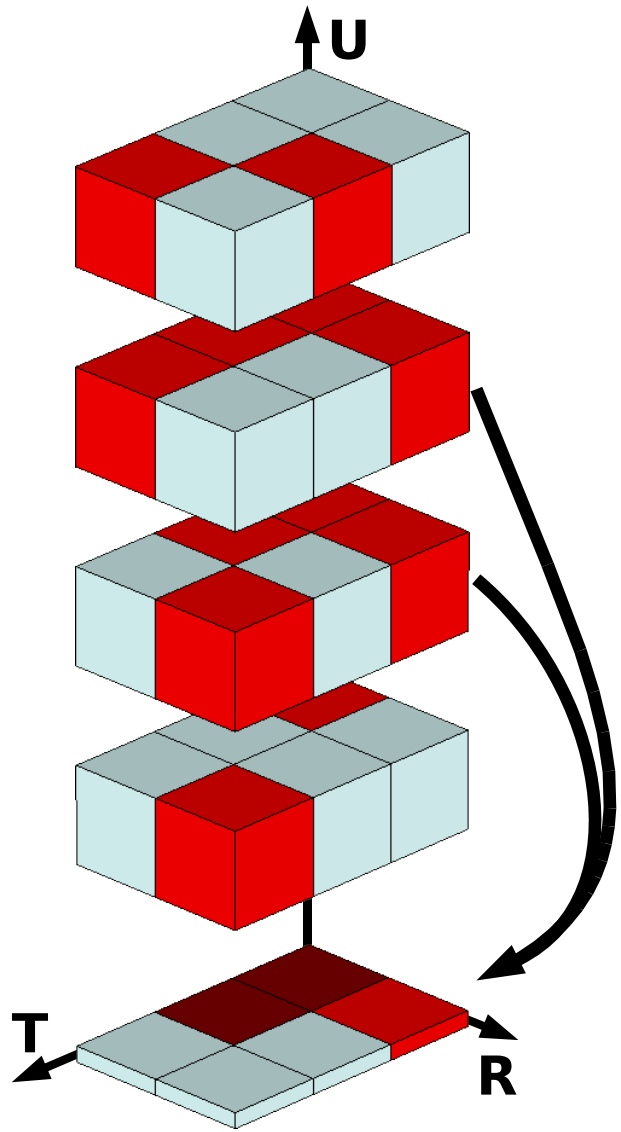
Der TRIAS-Algorithmus

Berechnet den Eisberg-Tri-Verband eines triadischen Formalen Kontexts

Ablauf

- Sei $\tilde{Y} := \{(u, (t, r)) \mid (u, t, r) \in Y\}$
- Schleife: Finde (häufige) Begriffe (A, I) in $(U, T \times R, \tilde{Y})$
  - Schleife: Finde (häufige) Begriffe (B, C) in (T, R, I)

Im Beispiel:
$(B, C) = (\{t_1\}, \{r_1, r_2\})$



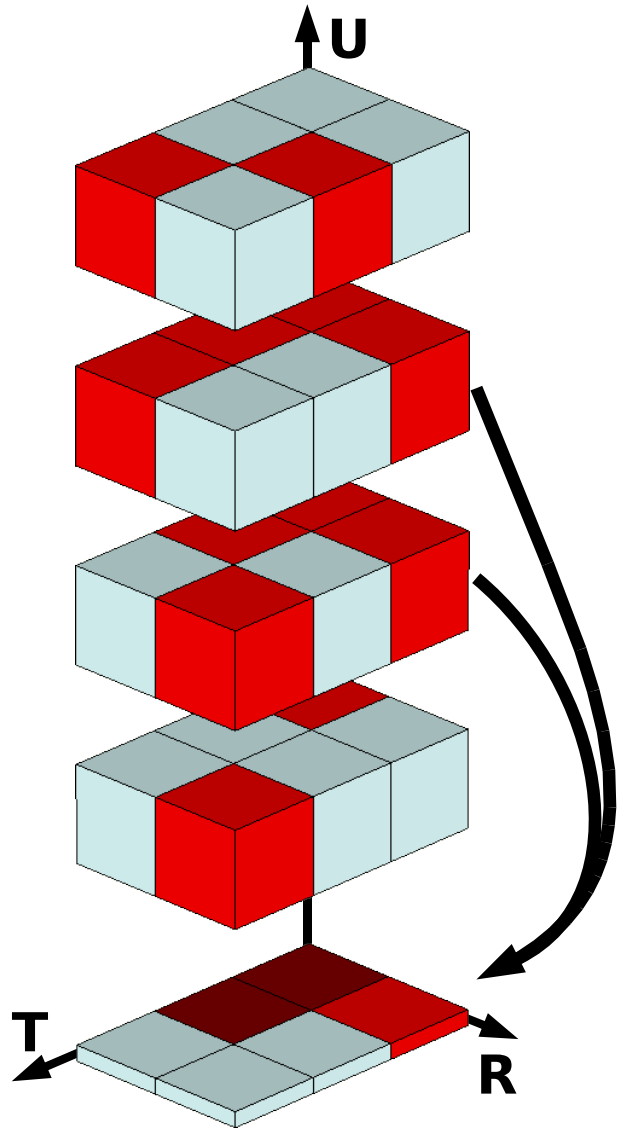
Der TRIAS-Algorithmus

Berechnet den Eisberg-Tri-Verband eines triadischen Formalen Kontexts

Ablauf

- Sei $\tilde{Y} := \{(u, (t, r)) \mid (u, t, r) \in Y\}$
- Schleife: Finde (häufige) Begriffe (A, I) in $(U, T \times R, \tilde{Y})$
  - Schleife: Finde (häufige) Begriffe (B, C) in (T, R, I)
    - Falls $A = (B \times C)^{\tilde{Y}}$, dann gib (A, B, C) aus

Im Beispiel:
$(B \times C)^{\tilde{Y}} = (\{t_1\} \times \{r_1, r_2\})^{\tilde{Y}}$



Der TRIAS-Algorithmus

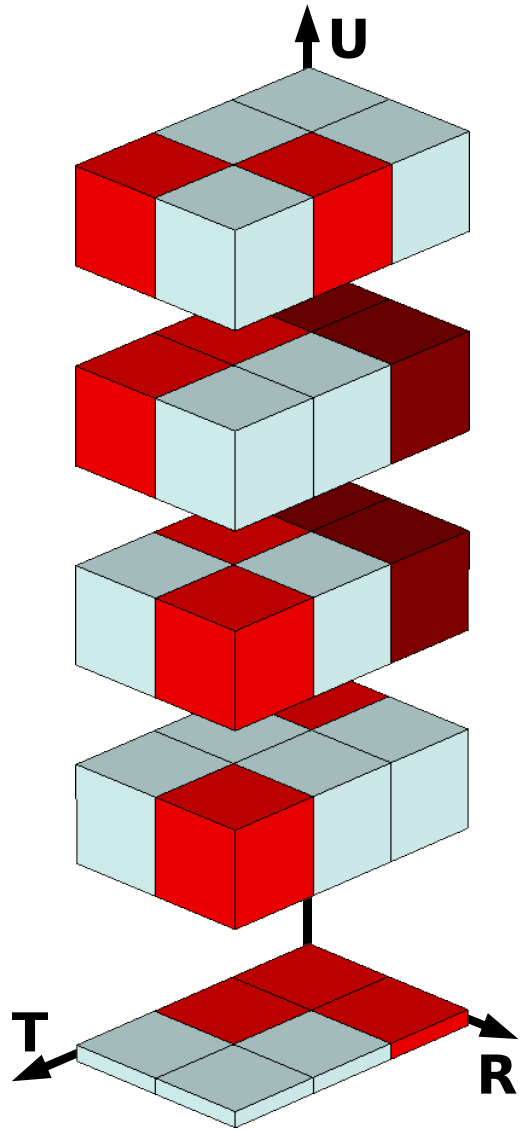
Berechnet den Eisberg-Tri-Verband eines triadischen Formalen Kontexts

Ablauf

- Sei $\tilde{Y} := \{(u, (t, r)) \mid (u, t, r) \in Y\}$
- Schleife: Finde (häufige) Begriffe (A, I) in $(U, T \times R, \tilde{Y})$
  - Schleife: Finde (häufige) Begriffe (B, C) in (T, R, I)
    - Falls $A = (B \times C)^{\tilde{Y}}$, dann gib (A, B, C) aus

Im Beispiel:

$$(B \times C)^{\tilde{Y}} = (\{t_1\} \times \{r_1, r_2\})^{\tilde{Y}} = \{u_2, u_3\} = A$$



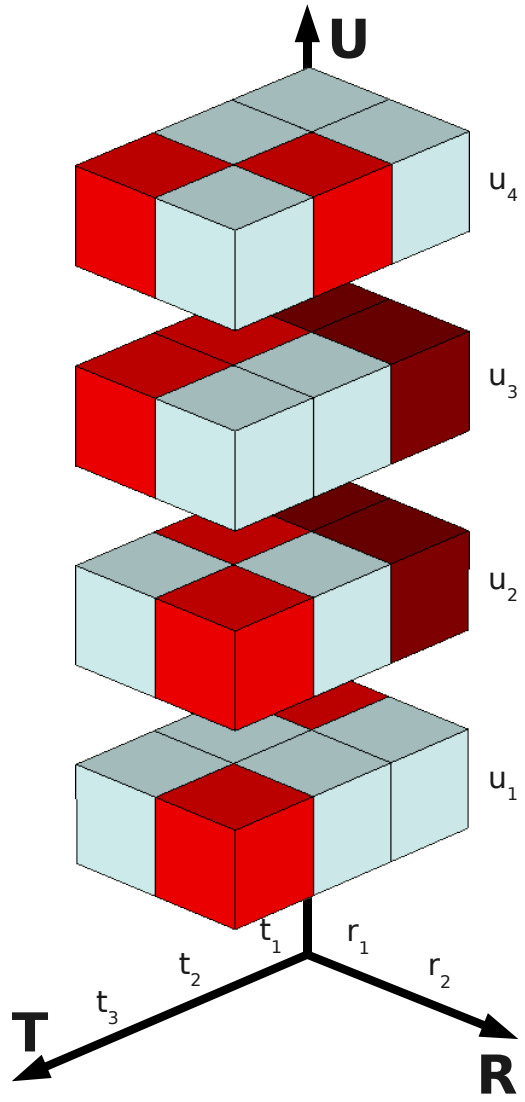
Der TRIAS-Algorithmus

Berechnet den Eisberg-Tri-Verband eines triadischen Formalen Kontexts

Ablauf

- Sei $\tilde{Y} := \{(u, (t, r)) \mid (u, t, r) \in Y\}$
- Schleife: Finde (häufige) Begriffe (A, I) in $(U, T \times R, \tilde{Y})$
  - Schleife: Finde (häufige) Begriffe (B, C) in (T, R, I)
    - Falls $A = (B \times C)^{\tilde{Y}}$, dann gib (A, B, C) aus

Im Beispiel:
$(A, B, C) = (\{u_2, u_3\}, \{t_1\}, \{r_1, r_2\})$



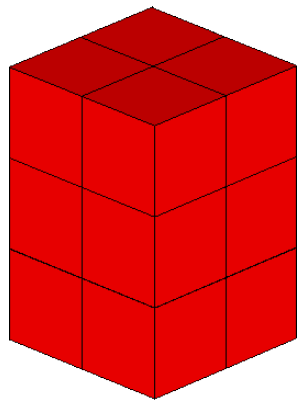
Qualitative Evaluation

BibSonomy-Datensatz:

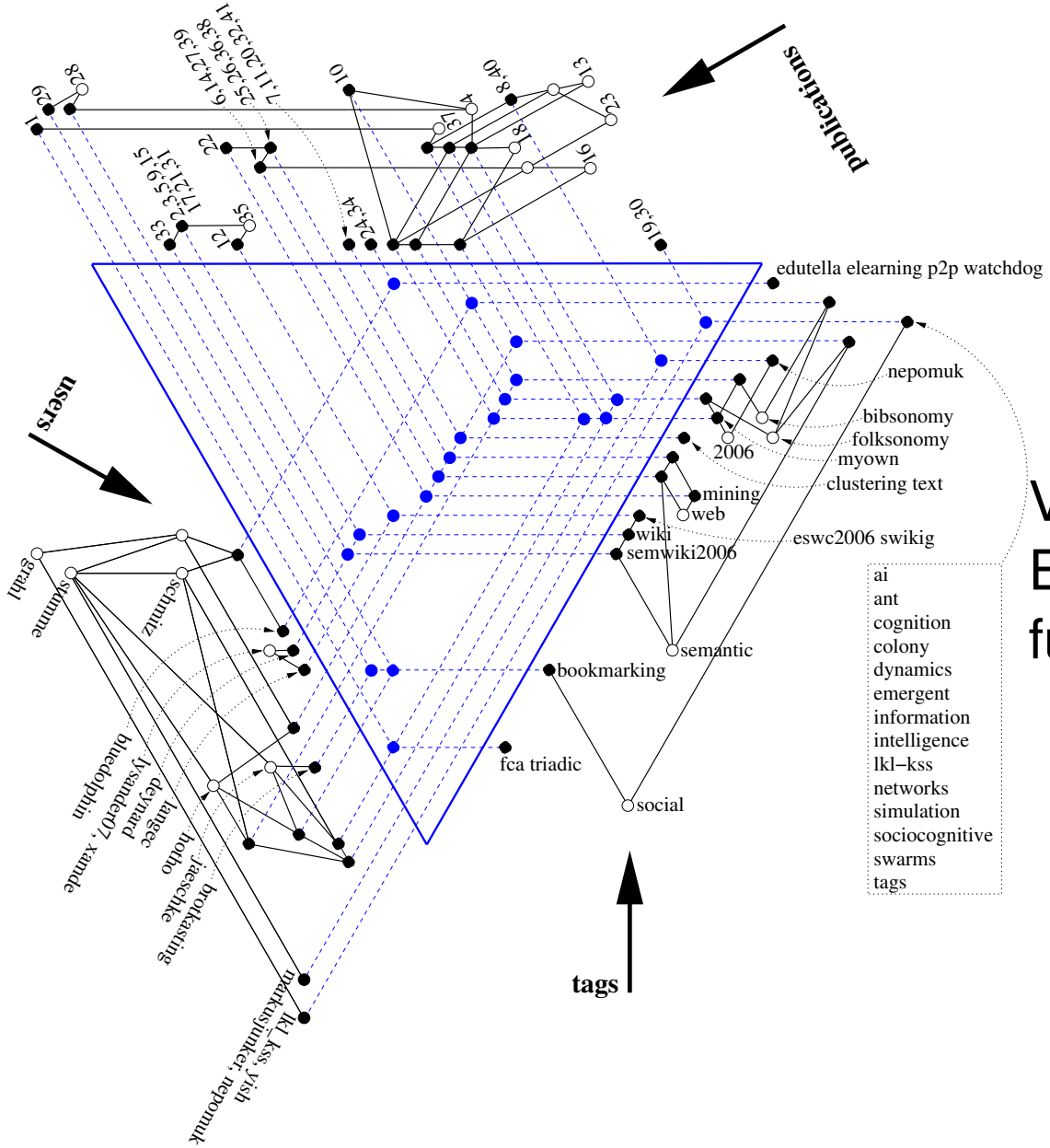
- Alle Publikationsreferenzen bis zum 23.11.2006
- Entfernt: DBLP, Posts mit dem Tag „imported“
- $|U| = 262$, $|T| = 5\,954$, $|R| = 11\,101$, $|Y| = 44\,944$

Ergebnis:

- 13 992 Tri-Begriffe (75 Minuten auf 2 GHz PC)
- Mit Support-Schranken $\tau_u = 3$, $\tau_t = 2$, $\tau_r = 2$:
  - 21 Tri-Begriffe
  - Enthalten 41 Publikationen, 15 Benutzer und 36 Tags

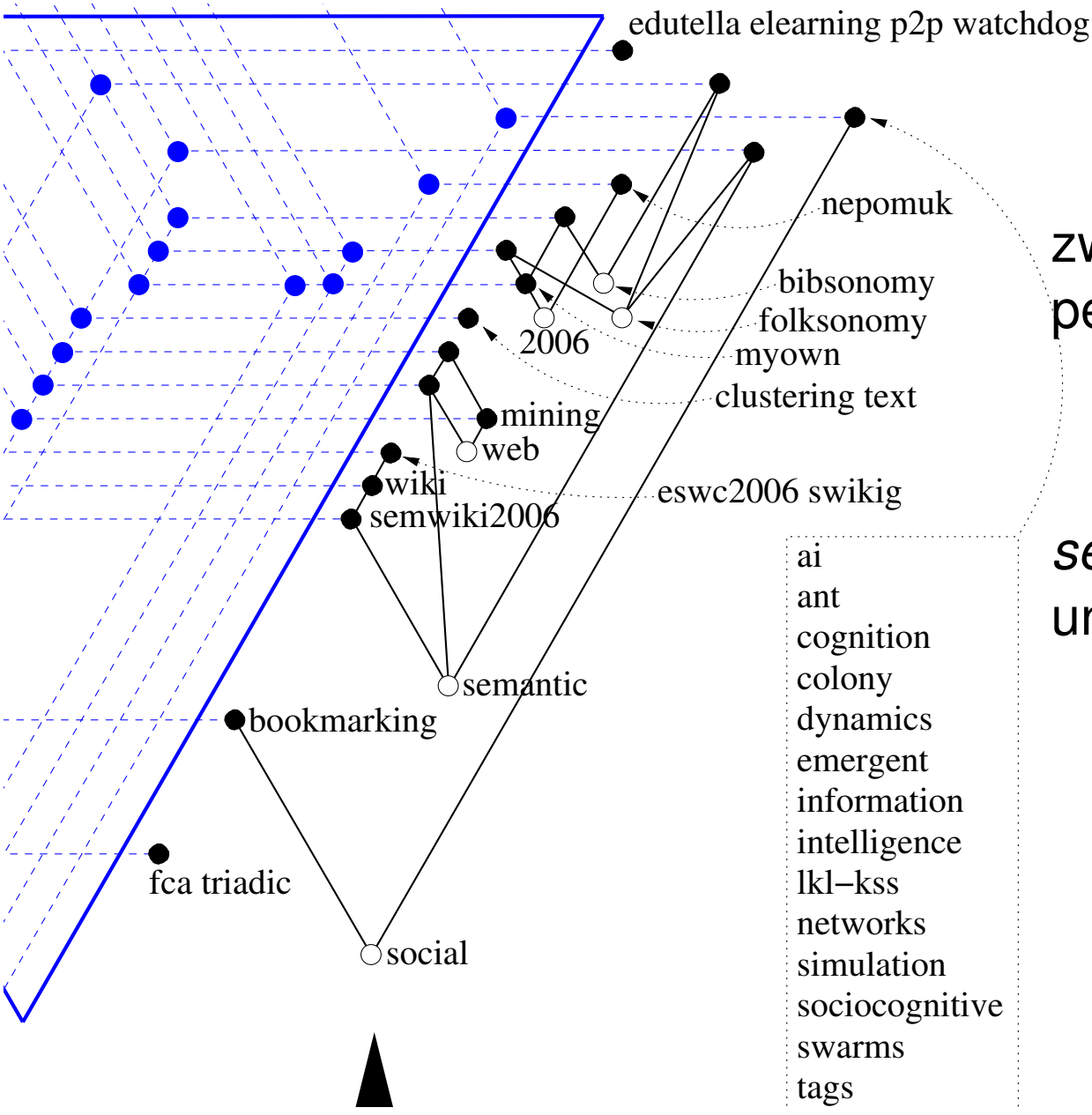


Qualitative Evaluation



Visualisierung des Eisberg-Tri-Verbandes für $\tau_u = 3$, $\tau_t = 2$, $\tau_r = 2$

Qualitative Evaluation



zwei Themengruppen:

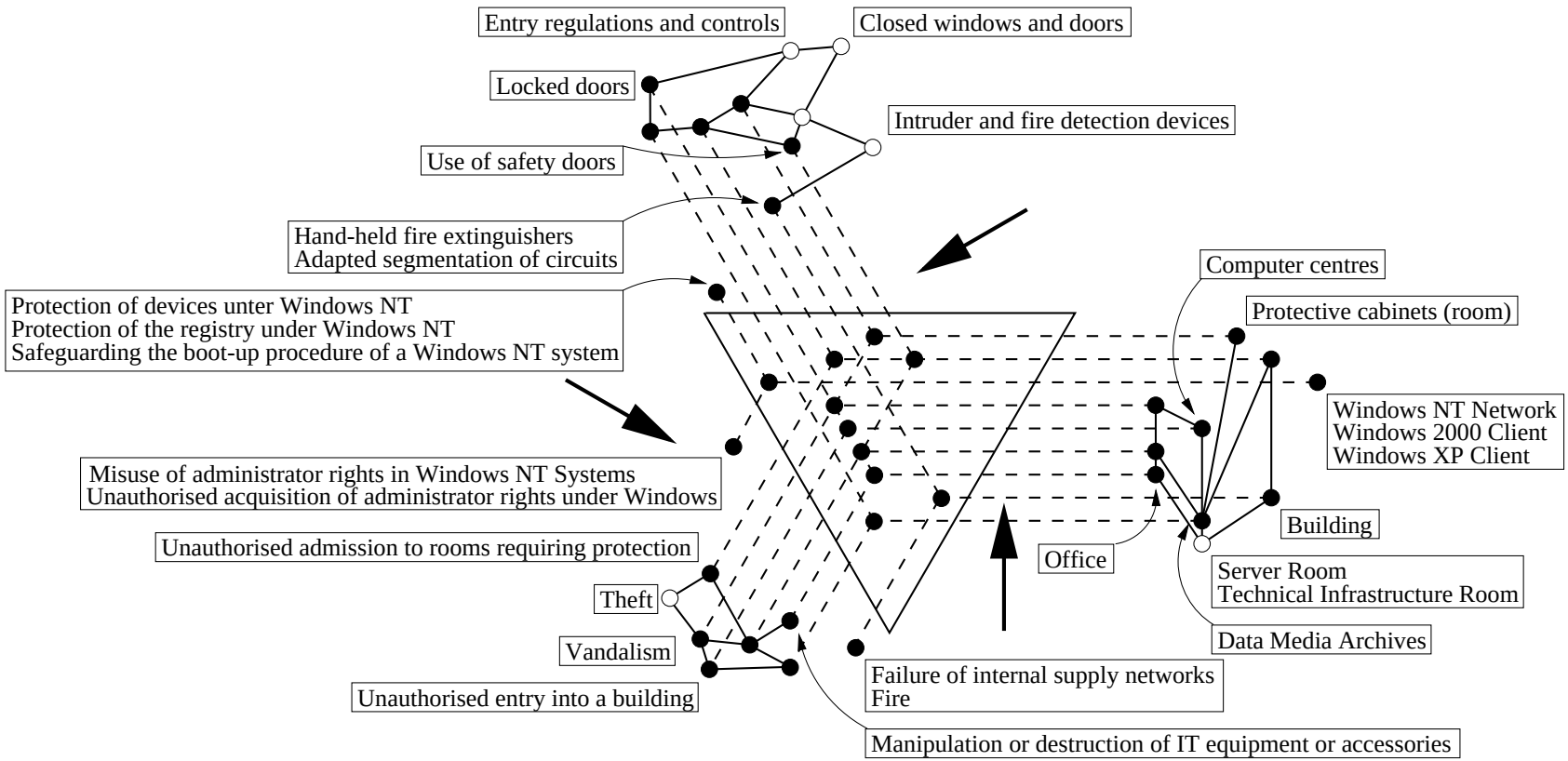
- semantic
- social

semantic weiter unterteilt:

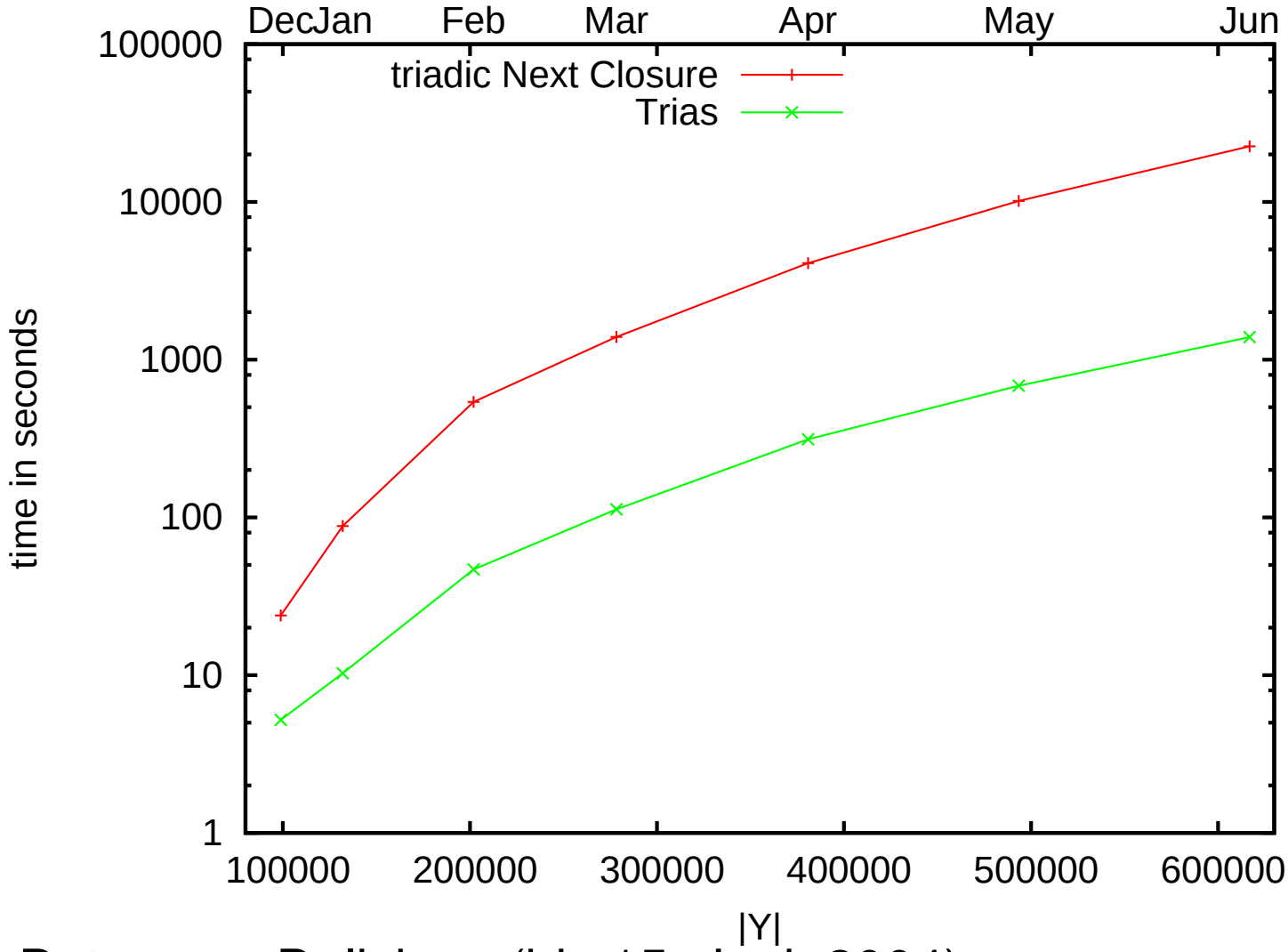
- wiki
- web
- folksonomy

BSI-Grundschutzhandbuch 2003

- 66 IT-Komponenten, 377 Bedrohungen, 912 Sicherheitsmaßnahmen, 3751 Tri-Begriffe
- häufige Tri-Begriffe für $\tau_u = 3, \tau_t = 2, \tau_r = 3$:



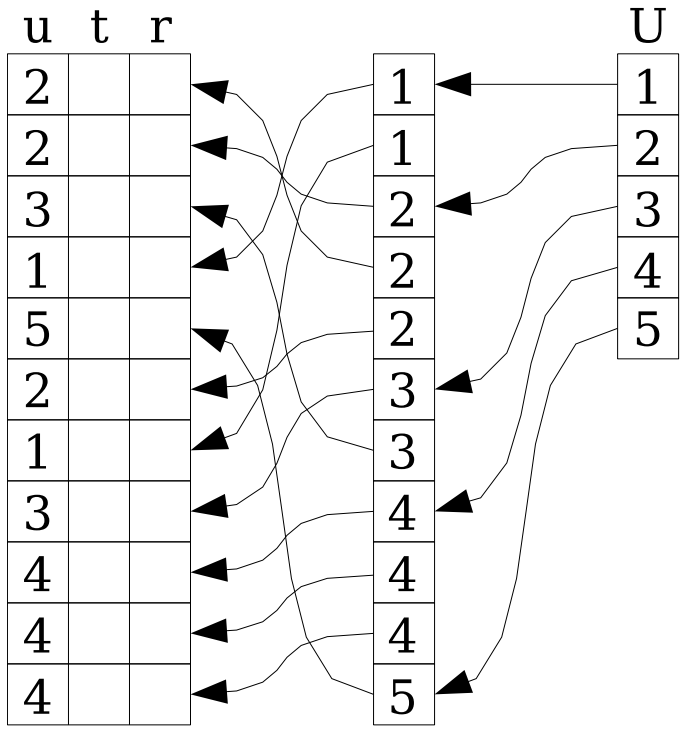
Evaluation der Laufzeit



- Daten von Delicious (bis 15. Juni, 2004)
- $|U| = 3,301$, $|T| = 30,416$, $|R| = 22,036$, $|Y| = 616,819$
- $\tau_u := \tau_t := \tau_r := 2$

Implementierungsdetails

- Java
- Ein-/Ausgabe: Datei, JDBC, RDF
- Konfiguration: Java, Kommandozeile, Properties-Dateien
- JUnit-Tests
- `http://github.com/fcatools/trias-algorithm`



Implementierungsdetails

- SOAP-Webinterface
- `http://github.com/fcatools/trias-webapp`
- Signatur:

```
public TriConcept<String>[] trias(
    final Context<String> context,
    final int minSuppObjects,
    final int minSuppAttributes,
    final int minSuppConditions)
```

Zusammenfassung

- Algorithmus zum Clustern triadischer Daten
- Zusammenführung von triadischer Begriffsanalyse und Frequent-Closed-Itemset-Mining
- neue Impulse zur Analyse multi-dimensionaler Daten durch Folksonomien
- fehlende Visualisierungsmöglichkeit
- weitere Referenzen:
  `http://www.bibsonomy.org/tag/fca`

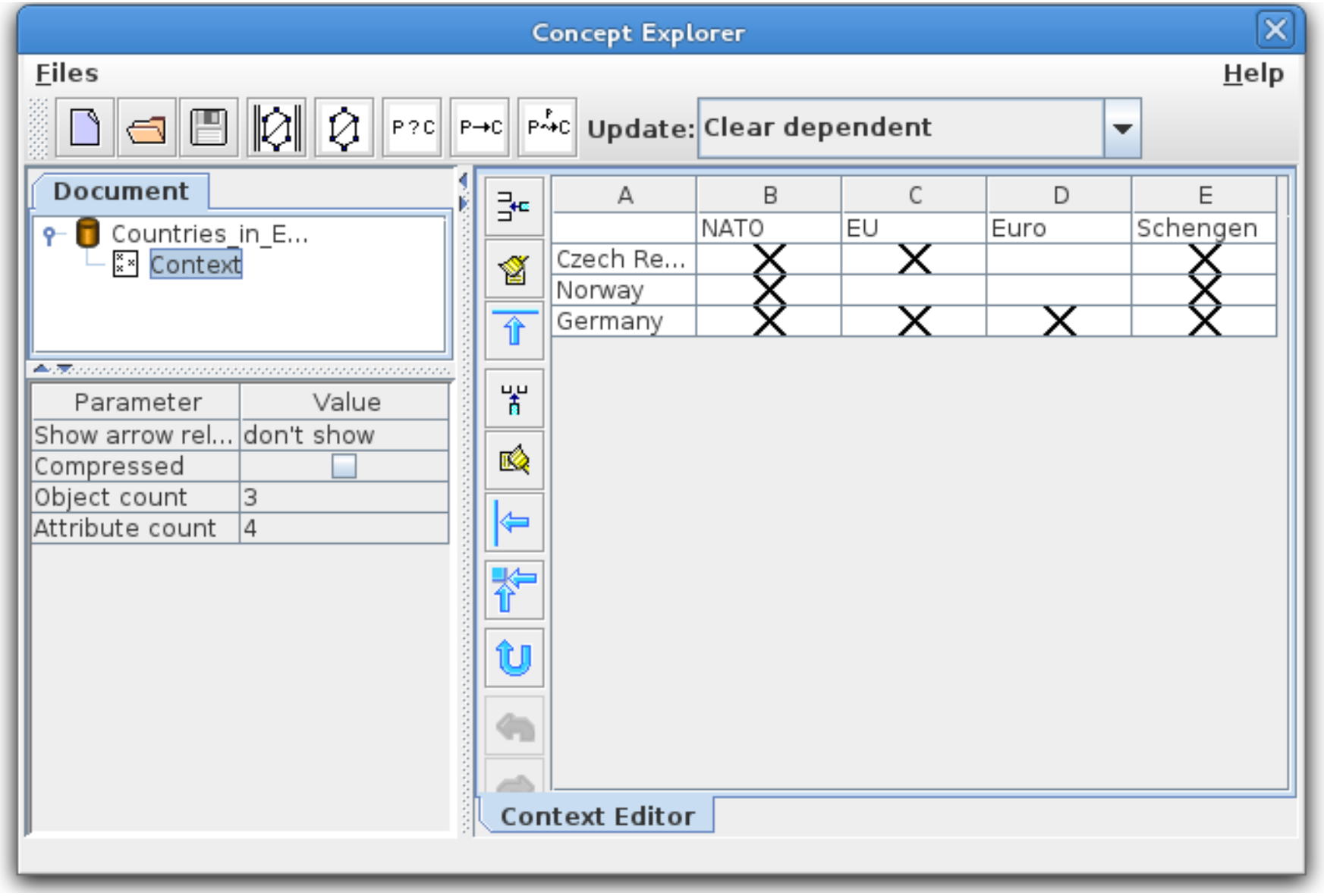
Agenda

Triadische FBA mit TRIAS

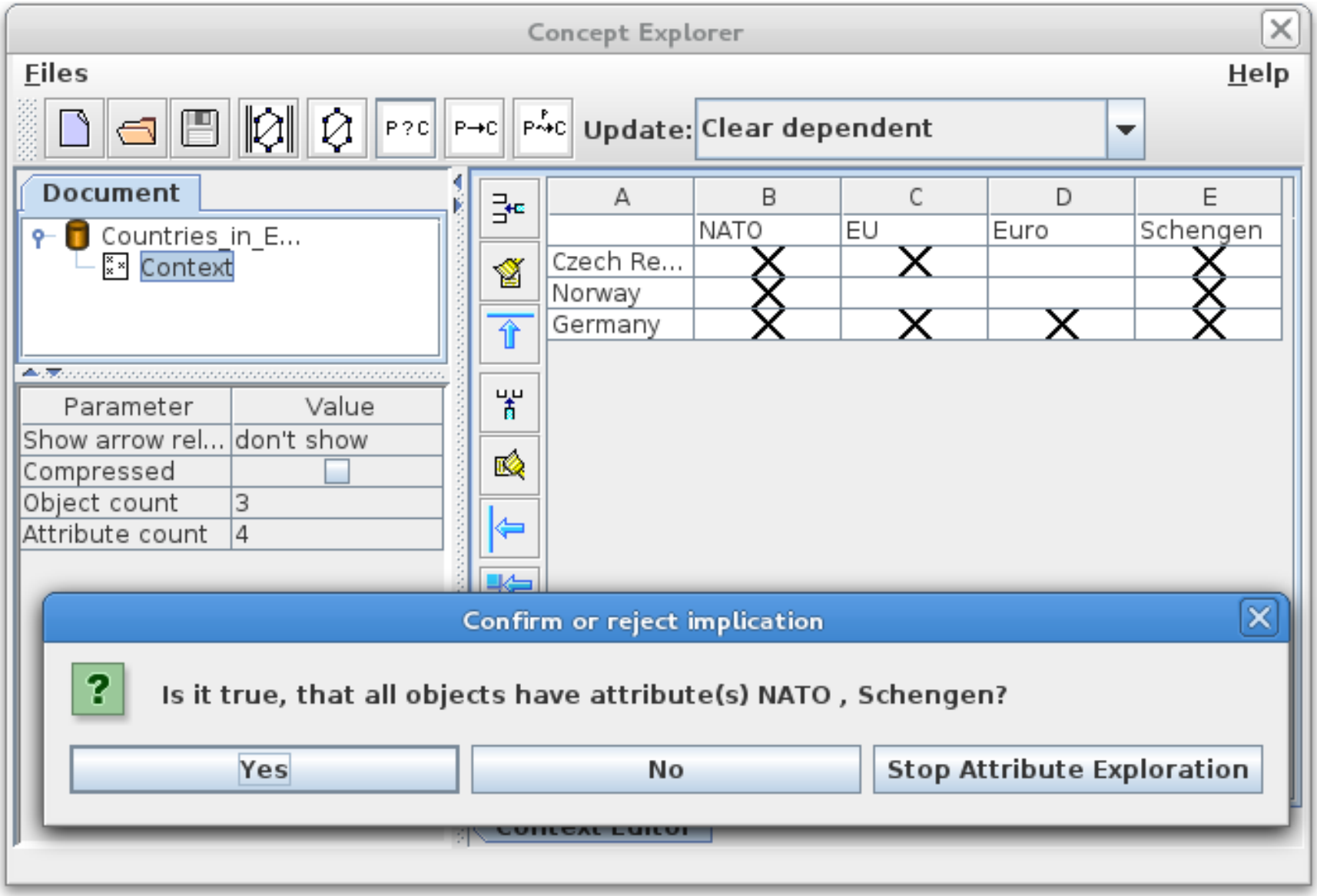
Merkmalsexploration im Web

ConExp-NG

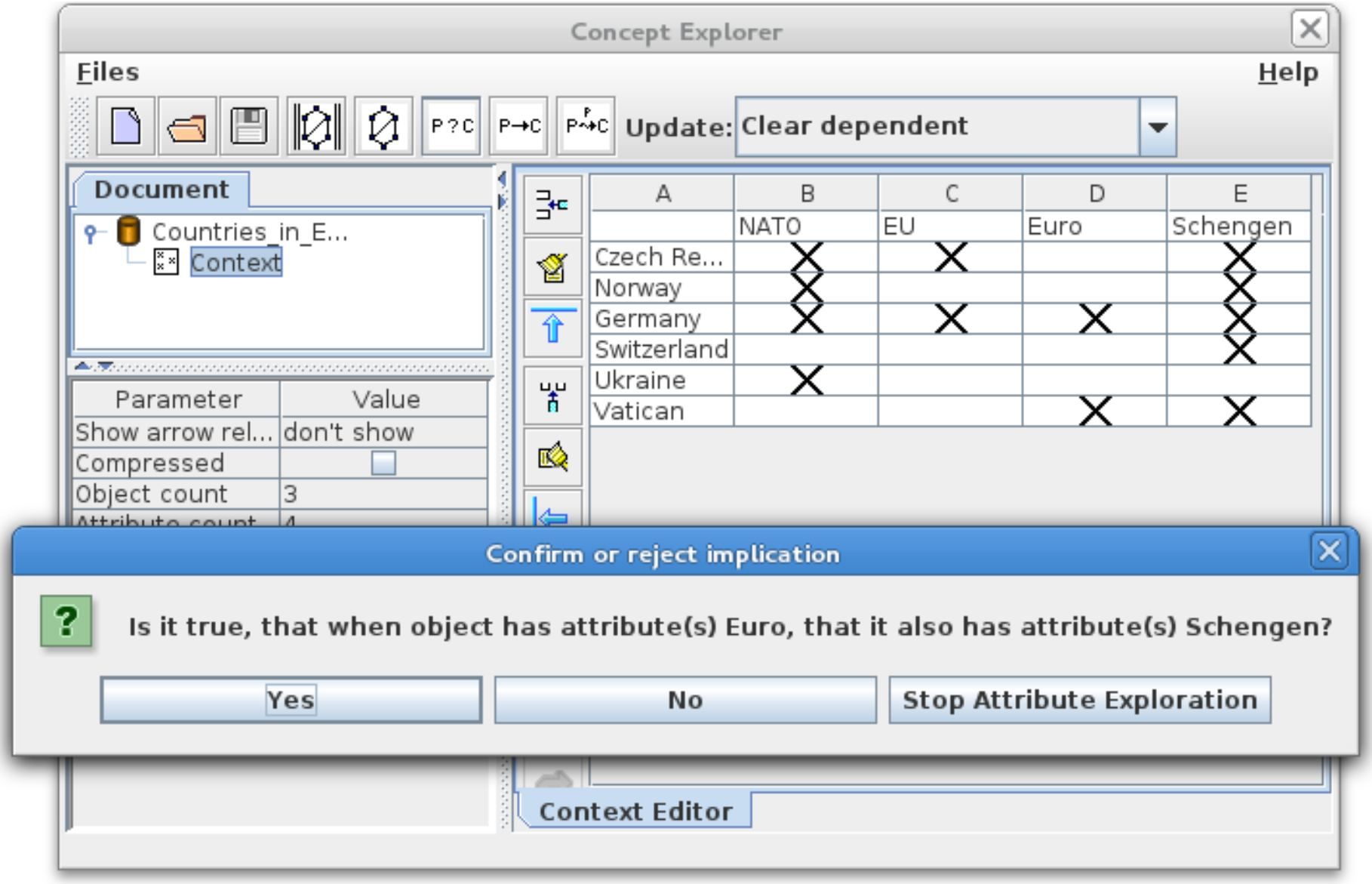
Motivation


Concept Explorer
Files
Help
P ? C
P→C
Update:
Clear dependent
Document
Countries_in_E...
Context
Parameter
Value
Show arrow rel...
don't show
Compressed
Object count
3
Attribute count
4
A
B
C
D
E
NATO
EU
Euro
Schengen
Czech Re...
Norway
Germany
Context Editor

Motivation


Concept Explorer
Files
Help
Update:
Clear dependent
Document
Countries_in_E...
Context
Parameter
Value
Show arrow rel...
don't show
Compressed
Object count
3
Attribute count
4
A
B
C
D
E
NATO
EU
Euro
Schengen
Czech Re...
Norway
Germany
Confirm or reject implication
Is it true, that all objects have attribute(s) NATO , Schengen?
Yes
No
Stop Attribute Exploration

Motivation


Concept Explorer
Files
Help
Update:
Clear dependent
Document
Countries_in_E...
Context
Parameter
Value
Show arrow rel...
don't show
Compressed
Object count
3
A
B
C
D
E
NATO
EU
Euro
Schengen
Czech Re...
Norway
Germany
Switzerland
Ukraine
Vatican
Confirm or reject implication
Is it true, that when object has attribute(s) Euro, that it also has attribute(s) Schengen?
Yes
No
Stop Attribute Exploration
Context Editor

Motivation

Wie können wir Merkmalsexploration im Web unterstützen?

- Merkmalsexploration typischerweise durch Experten
- große Menge an Information im Web verfügbar
- (und auch viele Experten)
- Wie können wir diese Ressourcen zur Unterstützung der Experten nutzen?
  - beim Finden von Gegenbeispielen
  - beim Bereitstellen von Kontext

Merkmalsexploration

Grundbergriffe:

- Formaler Kontext $\mathbb{K} := (G, M, I)$
- eine *Implikation* $L \to R$ (für $L, R \subseteq M$) *gilt* in $\mathbb{K}$, gdw. jeder Gegenstand mit allen Merkmalen aus L auch alle Merkmale aus R hat
- ein Gegenstand c *widerlegt* eine Implikation $L \to R$, gdw. $L \subseteq c'$ aber $R \not\subseteq c'$ (d.h., es gibt ein $m \in R$ mit $m \notin c'$)

Merkmalsexploration ist dann:

- Berechung aller geltenden Implikationen
- für jede neue Implikation $L \to R$, Experten fragen, ob
  $L \to R$ gilt (dann $L \to R$ hinzufügen), oder
  $L \to R$ durch Gegenbeispiel c widerlegt wird (dann $\mathbb{K}$ um c erweitern)

Ansatz

1. Implikationen in Anfragen bzw. Anfragemengen umformen
2. Anfragen an System stellen und Experten Ergebnis zeigen
3. Experten entscheiden lassen, ob Implikation gilt oder nicht

Das Web ist groß, es gibt viele Möglichkeiten für Anfragen.
→ Wir fokussieren auf: Social Question Answering, Crowdsourcing, Linked Open Data, Web-Suche.

Wie die Anfragen stellen?

Abstrakte Anfrage-Strategie

Annahme: die meisten Informationen im Web sind *faktisch*, d.h., Informationen über Gegenstände und ihre Merkmale

aber: Implikationen sind *terminologisches* Wissen, welches deutlich seltener und schwerer zu finden ist

Ansatz: Suche nach *Gegenbeispielen*, d.h., faktischem Wissen

Abstrakte Anfragestrategie

- $\varphi[x]$: Instanz-Anfrage in einer spezifischen *Anfragesprache*
- $Ans(\varphi[x])$: Menge der Gegenstände d für die $\varphi[d]$ wahr ist
- Anfrage $q[x]$ für ein Gegenbeispiel von $L \rightarrow R$:

$$q[x] := \bigwedge_{l \in L} \varphi_l[x] \ \wedge \bigvee_{r \in R} \neg \varphi_r[x]$$

- alternativ: $Q = \{q_r[x] \mid r \in R\}$ mit

$$q_r[x] := \bigwedge_{l \in L} \varphi_l[x] \ \wedge \neg \varphi_r[x]$$

dann

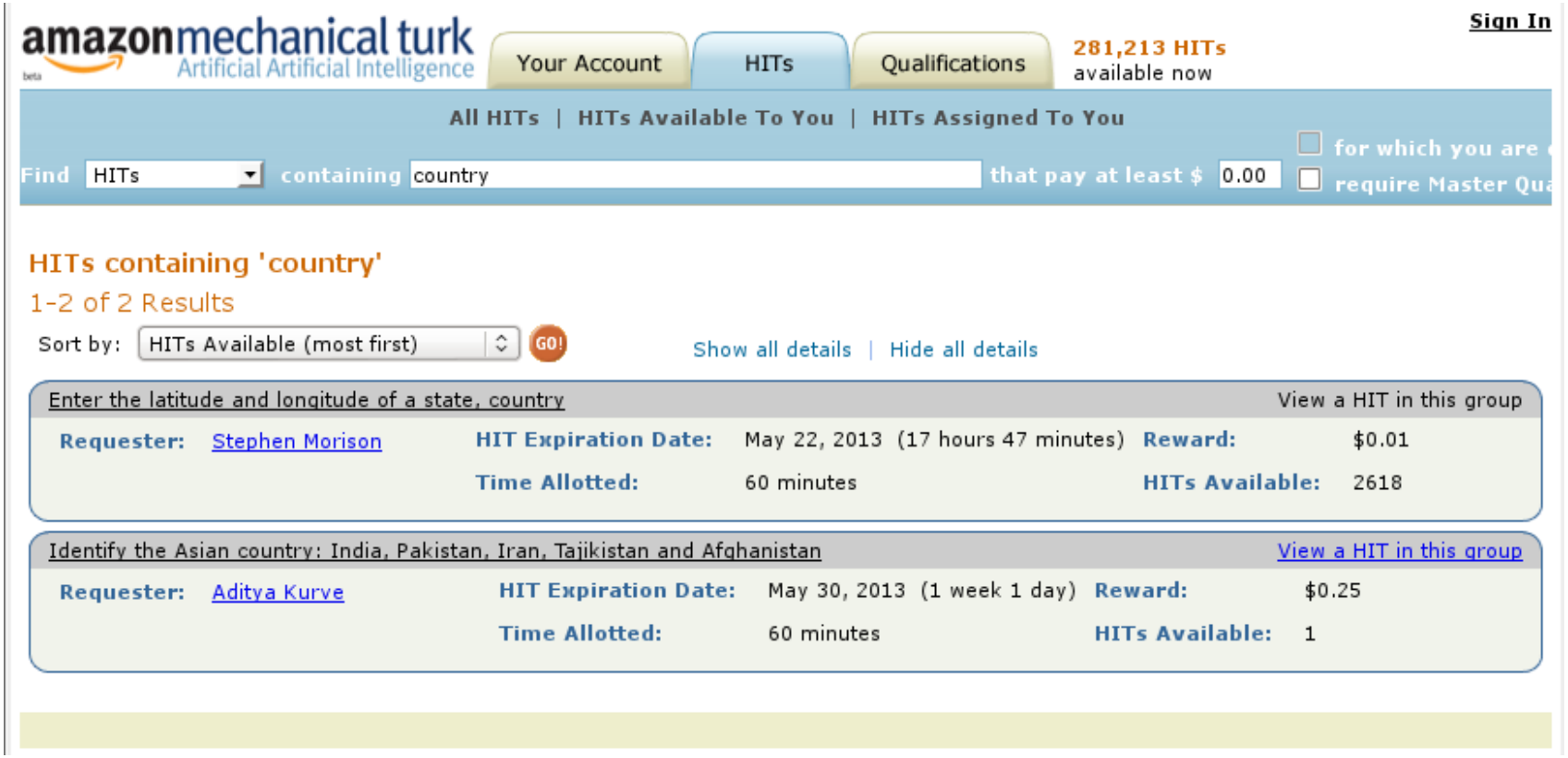
$$Ans(q[x]) = \bigcup_{q_r \in Q} Ans(q_r[x])$$

Social Question Answering/Crowdsourcing

Idee: Anfrage an andere Experten weiterleiten

- Social Question Answering (z.B. StackExchange, Yahoo! Answers)

Social Question Answering/Crowdsourcing



Idee: Anfrage an andere Experten weiterleiten

- Crowdsourcing (z.B. Amazon Mechanical Turk)

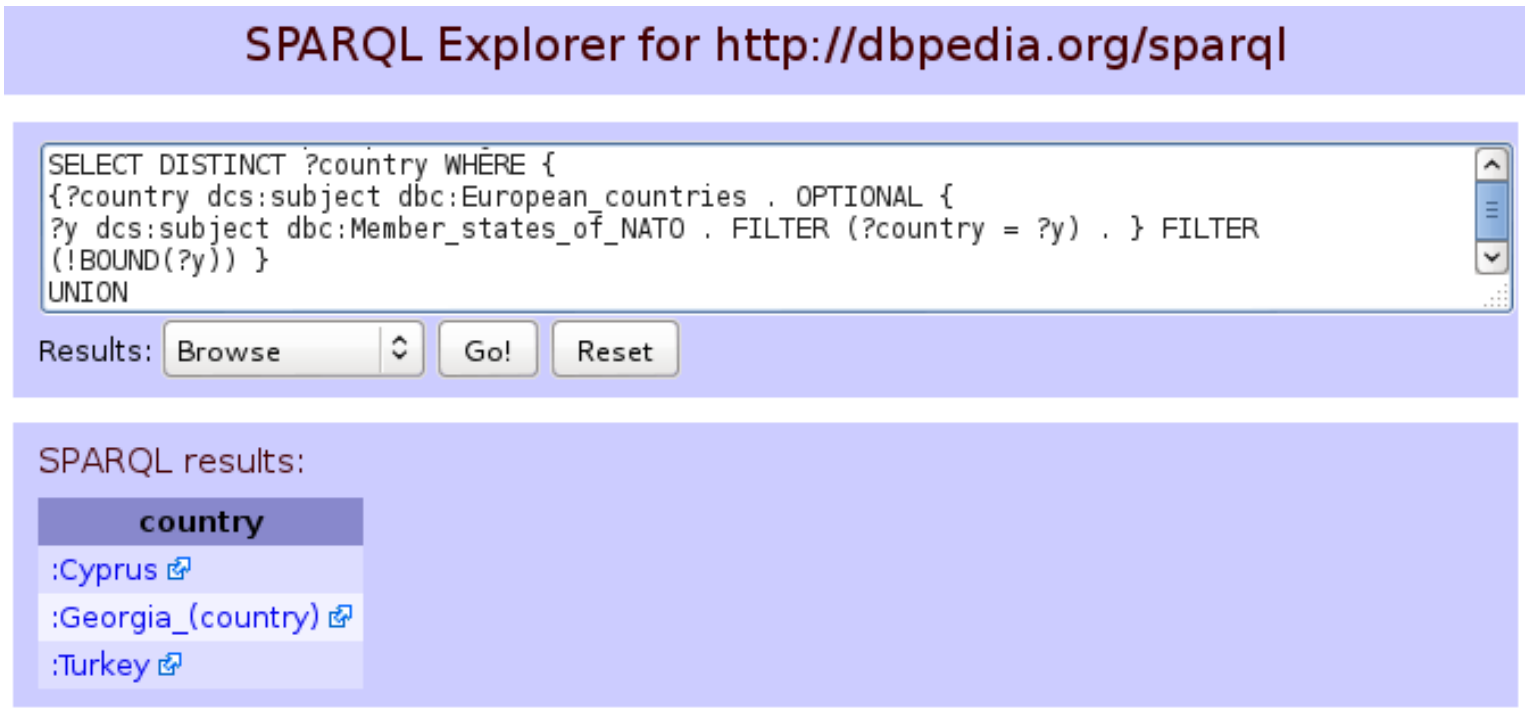
Social Question Answering/Crowdsourcing

Idee: Anfrage an andere Experten weiterleiten

- andere Ansätze (z.B. E-Mail, (Micro-)Blog)

Linked Open Data Cloud

- Teil des *Semantic Web*
- große Wissensbasen wie Wikidata, YAGO oder DBpedia
- typischerweise Möglichkeit für *SPARQL*-Anfragen
- *Einschränkung:* noch vglw. begrenztes Wissen



Web-Suchmaschinen

Annahmen:

1. Viele Webseiten beschreiben einzelne Gegenstände.
2. Für jedes Merkmal gibt es einen Suchbegriff, dessen Vorkommen in einer Webseite ein guter Indikator dafür ist, dass der beschriebene Gegenstand das Merkmal hat.
3. Ebenso ist die Abwesenheit eines Suchbegriffes ein guter Indikator dafür, dass der Gegenstand das Merkmal nicht hat.

Web-Suchmaschinen

Für eine Implikation $L \to R$ stellen wir für jedes $r \in R$ die Anfrage

$$q_r[x] := +l_1 \,␣\, +l_2 \,␣\, \ldots \,␣\, +l_{|L|} \,␣\, -r$$

und zeigen den Experten die Ergebnisliste

zusätzlich Einschränkungen möglich auf:

- Gegenstandsdomäne d: $q_r[x] := +d \,␣\, +l_1 \,␣\, +l_2 \,␣\, \ldots \,␣\, +l_{|L|} \,␣\, -r$
- Website s: $q_r[x] := \texttt{+site:}s \,␣\, +l_1 \,␣\, +l_2 \,␣\, \ldots \,␣\, +l_{|L|} \,␣\, -r$

Implementierung

- Java-basierter Prototyp (Web-Anwendung)
- nutzt die FCAlib-Implementierung von NEXT CLOSURE
- Anfragen an Microsofts Bing-API
- `http://github.com/fcatools/web-attribute-exploration`

Beispiel

	NATO	EU	Euro	Schengen
Czech Republik	×	×		×
Norway	×			×
Germany	×	×	×	×

Erste Implikation ist $\emptyset \rightarrow \{\text{NATO}, \text{Schengen}\}$

- Anfragen:

  `-NATO`

  `-Schengen`
- aber: Suchmaschinen unterstützen reine Negation nicht
- daher: Hinzufügen der Domäne “Countries in Europe”:

  `+'Countries in Europe' -NATO`

  `+'Countries in Europe' -Schengen`

Beispiel

1. Countries of Europe - Geography Home Page - Geography at About.com
   http://geography.about.com/library/maps/blreurope.htm
   The countries of Europe.: city geography europe maps serbia and montenegro bosnia and herzegovina countries of europe
2. European Countries
   http://www.aneki.com/europe.html
   List of Countries in Europe with maps, statistics, and country comparisons of all the European Nations
3. Countries in Europe - YouTube
   http://www.youtube.com/watch?v=k2AMN0L-rfk
   This video includes all the countries in Europe except Vaticana,Monaco,San Marino,Andora,Liechtenstein,Malta .I hope it will be helpfull. The song is aria ...
4. ESL - English Exercises: Countries in Europe
   http://www.englishexercises.org/exercise.asp?id=8524
   Fullscreen: Countries in Europe by robirimini. Date: 30 - Sep - 2012 Level: elementary Age: 8-11. Description: Three exercices about Italy, Germany, France, Spain ...
5. List of countries in EUROPE, list of countries of EUROPE, European ...
   http://www.countries-ofthe-world.com/countries-of-europe.html
   Europe is unique continent which is not surrounded by water from all directions, and has overland border with the neighbouring continent Asia. Definition of correct ...
6. Pages Europe, Europe Yellow Pages, Europe Business Directory ...
   http://pageseurope.com/
   ... Europe Hotels, All Inclusive Europe, Europe Real Estate, Travel to Europe, European Vacations, Jobs in Europe, Map of Europe, European Football, Countries in Europe
7. Countries in EUROPE - Cost of Living City to City
   http://www.earthcosts.com/countries-europe/
   New posts: Hot thread with new posts: No new posts: Hot thread with no new posts: Thread is closed
8. The Largest Countries in Europe | Infoplease.com
   http://www.infoplease.com/askeds/largest-countries-europe.html
   The Question: According to your info Sweden is the fourth largest country in Europe. Which are the three bigger ones?
9. How Many Countries are there in Europe? - Buzzle
   http://www.buzzle.com/articles/how-many-countries-in-europe.html
   Europe is one of the most beautiful continents in the world, with a variety of countries and cultures in its fold. Wondering how many countries are there ...
10. Richest Countries in Europe
    http://www.aneki.com/europe_richest.html
    List of the richest countries in Europe ... Plutocrats: The Rise of the New Global Super-Rich and the Fall of Everyone Else

→ Hinzufügen der Einschränkung auf en.wikipedia.org:

```
+‘Countries in Europe’ +site:en.wikipedia.org -NATO
+‘Countries in Europe’ +site:en.wikipedia.org
-Schengen
```

Beispiel

Web-Based Attribute Exploration

Formal Context

Countries in Europe	NATO	EU	Euro	Schengen
Czech Republic	x	x		x
Norway	x			x
Germany	x	x	x	x

change context

Attribute Exploration

The current implication is: **[] ⇒ [Schengen, NATO]**.

You can either accept it or provide a counterexample:

	NATO	EU	Euro	Schengen
	☐	☐	☐	☐

Can you find a find a counterexample within the following web search results?

1. **+"Countries in Europe" -"Schengen" site:en.wikipedia.org**
2. +"Countries in Europe" -"NATO" site:en.wikipedia.org

+"Countries in Europe" -"Schengen" site:en.wikipedia.org

custom search

1. Former countries in Europe after 1815 - Wikipedia, the free ...
   http://en.wikipedia.org/wiki/Former_countries_in_Europe_after_1815
   This article gives a detailed listing of all the countries, (including puppet states), that have existed in Europe since the Congress of Vienna in 1815 to the present ...
2. Category:Former countries in Europe - Wikipedia, the free encyclopedia
   http://en.wikipedia.org/wiki/Category:Former_countries_in_Europe
   Former countries in Europe after 1815; List of early East Slavic states; List of East Slavic duchies; List of historic states of Germany; List of historic states of Italy
3. Category:Countries in Europe - Wikipedia, the free encyclopedia
   http://en.wikipedia.org/wiki/Category:Countries_in_Europe
   Subcategories. This category has the following 58 subcategories, out of 58 total.
4. List of national capitals of countries in Europe by area ...
   http://en.wikipedia.org/wiki/List_of_national_capitals_of_countries_in_Europe_by_area
   This list includes the capitals of European countries by area. The chart is below. Rank City Country Area (km 2) 1. Ankara Turkey 7003251600000000000 2,516 ...
5. List of sovereign states and dependent territories by population ...
   http://en.wikipedia.org/wiki/Largest_countries_in_Europe

Beispiel

Web-Based Attribute Exploration

Formal Context

Countries in Europe	NATO	EU	Euro	Schengen
Czech Republic	x	x		x
Norway	x			x
Germany	x	x	x	x

change context

Attribute Exploration

The current implication is: **[] ⇒ [Schengen, NATO]**.

You can either accept it or provide a counterexample:

	NATO	EU	Euro	Schengen
	☐	☐	☐	☐

Can you find a find a counterexample within the following web search results?

1. **+"Countries in Europe" -"Schengen" site:en.wikipedia.org**
2. +"Countries in Europe" -"NATO" site:en.wikipedia.org

+"Countries in Europe" -"Schengen" site:en.wikipedia.org

custom search

Beispiel

Can you find a find a counterexample within the following web search results?

1. **+"Countries in Europe" -"Schengen" site:en.wikipedia.org**
2. +"Countries in Europe" -"NATO" site:en.wikipedia.org

+"Countries in Europe" -"Schengen" site:en.wikipedia.org

custom search

1. Former countries in Europe after 1815 - Wikipedia, the free ...
   http://en.wikipedia.org/wiki/Former_countries_in_Europe_after_1815
   This article gives a detailed listing of all the countries, (including puppet states), that have existed in Europe since the Congress of Vienna in 1815 to the present ...
2. Category:Former countries in Europe - Wikipedia, the free encyclopedia
   http://en.wikipedia.org/wiki/Category:Former_countries_in_Europe
   Former countries in Europe after 1815; List of early East Slavic states; List of East Slavic duchies; List of historic states of Germany; List of historic states of Italy
3. Category:Countries in Europe - Wikipedia, the free encyclopedia
   http://en.wikipedia.org/wiki/Category:Countries_in_Europe
   Subcategories. This category has the following 58 subcategories, out of 58 total.
4. List of national capitals of countries in Europe by area ...
   http://en.wikipedia.org/wiki/List_of_national_capitals_of_countries_in_Europe_by_area
   This list includes the capitals of European countries by area. The chart is below. Rank City Country Area (km 2) 1. Ankara Turkey 7003251600000000000 2,516 ...
5. List of sovereign states and dependent territories by population ...
   http://en.wikipedia.org/wiki/Largest_countries_in_Europe

Beispiel

Warum sind keine Länder unter den ersten Ergebnissen?

- Abgleich nur mit dem *Text* der Webseiten
  → Einschränkung auf Domäne wirkungslos
- Seiten mit Suchbegriffen im Titel werden höher bewertet
- *Bewertung durch PageRank* bevorzugt Seiten mit vielen eingehenden Links

Warum fehlen Länder, die Gegenbeispiele wären?
Warum werden Länder zurückgegeben, die keine Gegenbeispiele sind??
→ beschränkte Gültigkeit der Annahmen 2 und 3:

- Webseiten *enthalten nicht* immer die Wörter entsprechend den Merkmalen der Gegenstände, die sie beschreiben
- Webseiten *enthalten* manchmal Wörter entsprechend Merkmalen die die Gegenstände, die sie beschreiben, *nicht haben*

Schlussfolgerung

- *Gegenbeispiele zu finden ist schwierig*, oft auch für menschliche Experten
- vorgeschlagener Ansatz
  - ist *eingeschränkt anwendbar* für Web-Suche,
  - könnte nützlich sein zur *Exploration von Linked Open Data*,
  - oder zur Fehlersuche
- Generelles Problem der *Open World Assumption*

Schlussfolgerung

Einige der Probleme könnten gelöst werden durch

- mehr Informationen zu Gegenständen + Merkmalen
- Kombination der Quellen (1. Linked Open Data, 2. Web-Suche, 3. Crowdsourcing/Social Question Answering)
- weitergehende Analyse der gefundenen Seiten
- Warten auf bessere Wissensbasen ;-)

Agenda

Triadische FBA mit TRIAS

Merkmalsexploration im Web

ConExp-NG

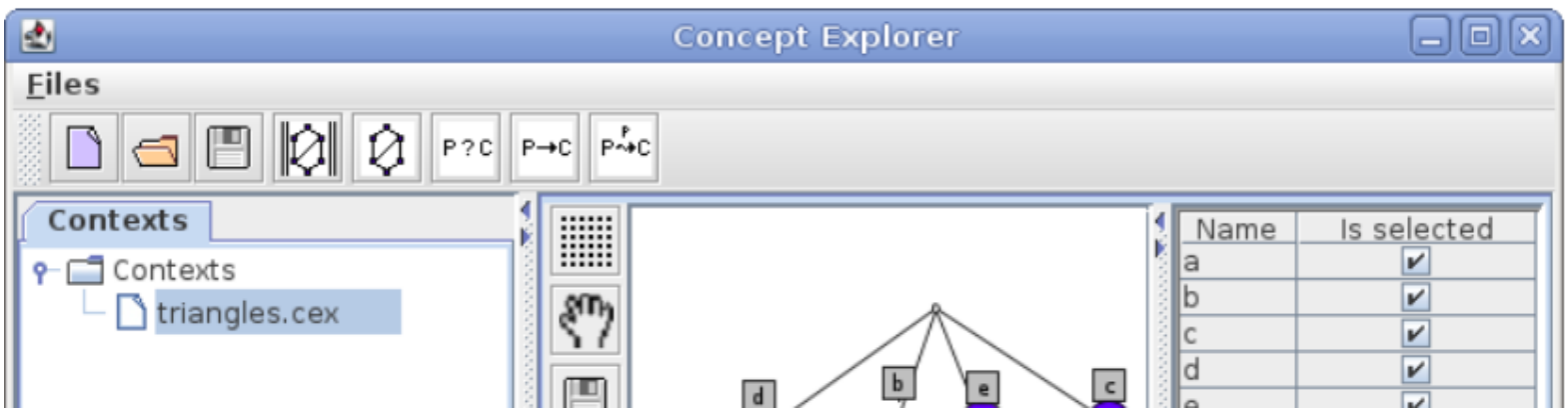
Vektorgraphik für ConExp

Mittels Formaler Begriffsanalyse kann man Begriffe und Begriffshierarchien aus Daten gewinnen. Ein dazu oft verwendetes Werkzeug ist der Concept Explorer (ConExp) - ein Java-Programm zum Eingeben von Kontexten sowie zum Analysieren und Visualisieren von Begriffsverbänden.

Ziel dieser Arbeit ist es, ConExp um einen Vektorgraphik-Export zu erweitern, so dass die damit gezeichneten Begriffsverbände weiterverwendet werden können. Dabei kommen insbesondere die Formate PostScript, PDF und LaTeX in Frage. Des weiteren soll ein XML-basiertes Export-Format entwickelt und implementiert sowie ggf. weitere notwendige Anpassungen vorgenommen werden.

Vorwissen über Formale Begriffsanalyse ist empfehlenswert aber nicht notwendig. Gute Java-Kenntnisse, insbesondere in der GUI-Programmierung sind von Vorteil.

Kontakt: Robert Jäschke

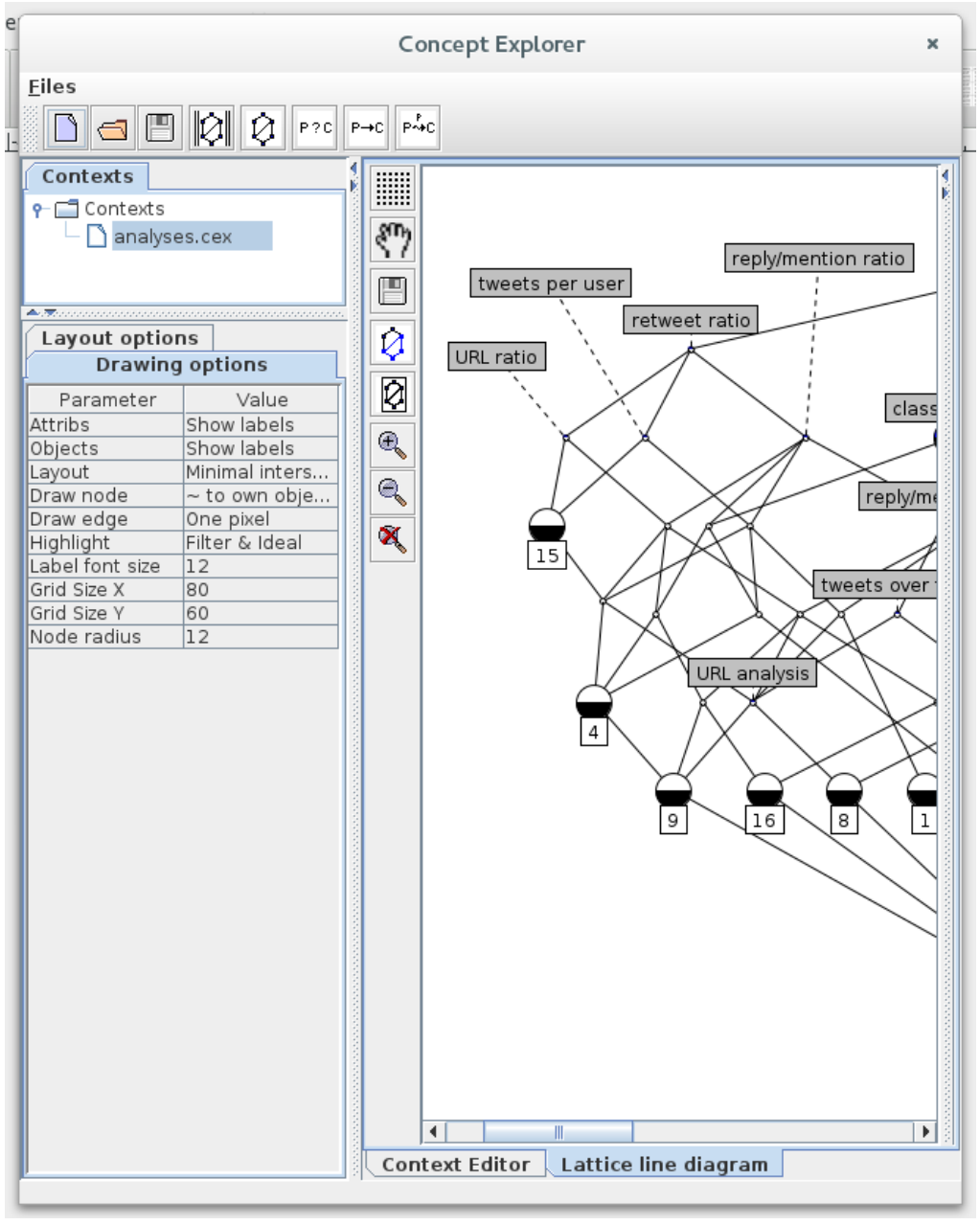


http://www.kbs.uni-hannover.de/~jaeschke/2013/02/19/Vektorgraphik_fuer_ConExp.html

Motivation

ConExp sehr beliebt, aber

- fehlender Vektorgrafikexport
- nicht modular
- schwierig zu erweitern



Motivation

vorhanden: FCAlib von Barış Sertkaya

- Java-Implementierung einer Bibliothek für FBA
- Klassenstruktur für Kontexte, Gegenstände, Begriffe, Implikationen, Hüllenoperatoren, Merkmalsexploration, Experten, etc.
- Implementierung von Next Closure für Merkmalsexploration
- modular, erweiterbar, gut wartbar
- `http://github.com/fcatools/fcalib`

hinzu kamen:

- Daniel Borchmann: conexp-clj benötigt Visualisierung
- drei sehr motivierte Studierende

Entwicklungsphasen

04/2013–08/2013 Studentenprojekt

- modularer Aufbau
- Re-Implementierung der vorhandenen Features
- Reverse-Engineering/Neuimplementierung Visualisierung

10/2013–02/2014 Studentenprojekt

- deutliche Verbesserungen und Bereinigungen

03/2014–03/2015 studentische Hilfskraft

- weitere Verbesserungen

Architektur

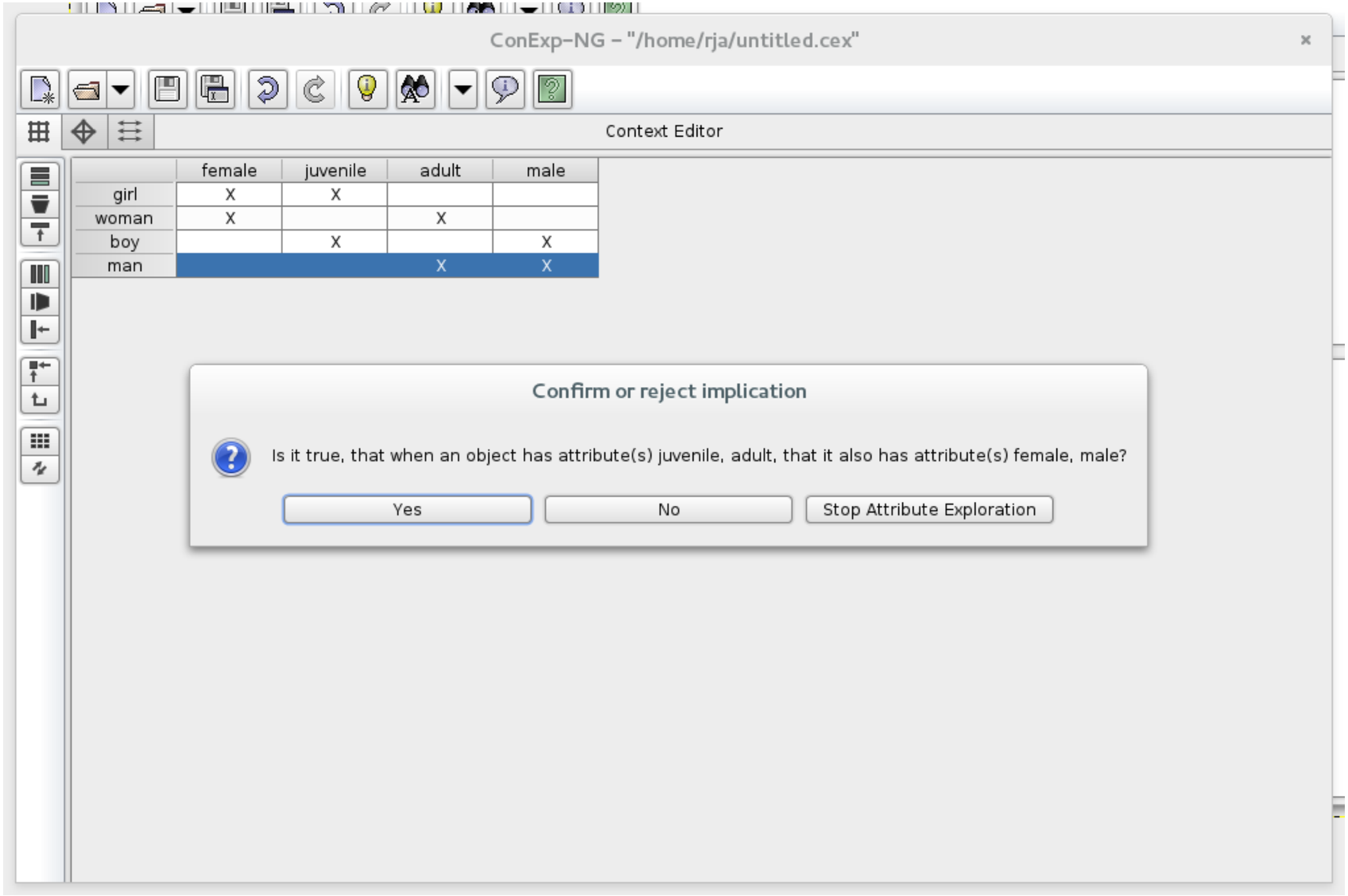
Wiederverwendung vorhandener Bibliotheken:

- FCAlib: Datenstrukturen, Next Closure
- Batik: Vektorgraphik (SVG, EPS, PDF)
- WebLaF: GUI

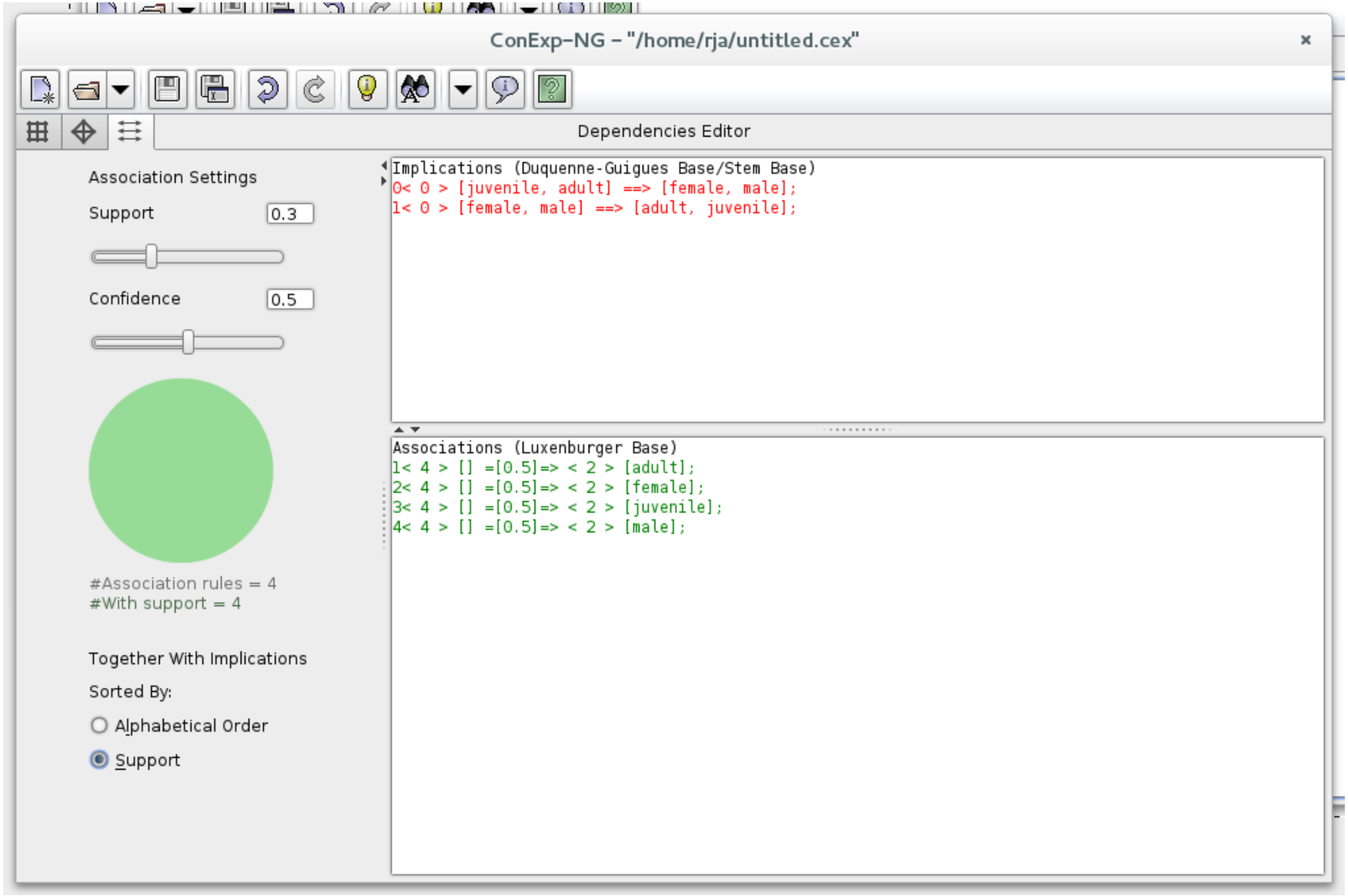
moderne Softwarentwicklung:

- Maven
- GitHub (`https://github.com/fcatools/conexp-ng`)

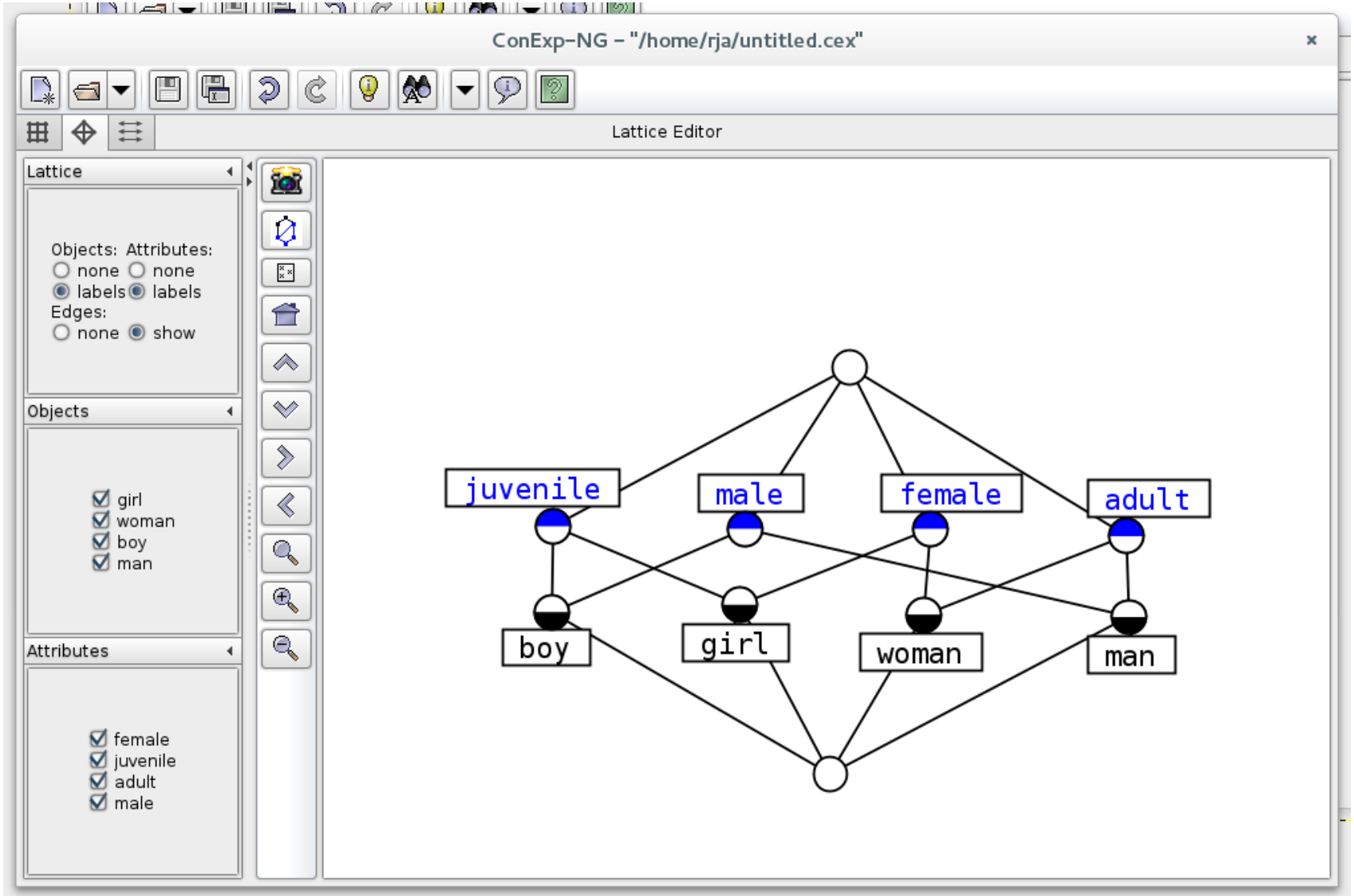
Features


ConExp-NG – "/home/rja/untitled.cex"
Context Editor
female
juvenile
adult
male
girl
woman
boy
man
Confirm or reject implication
Is it true, that when an object has attribute(s) juvenile, adult, that it also has attribute(s) female, male?
Yes
No
Stop Attribute Exploration

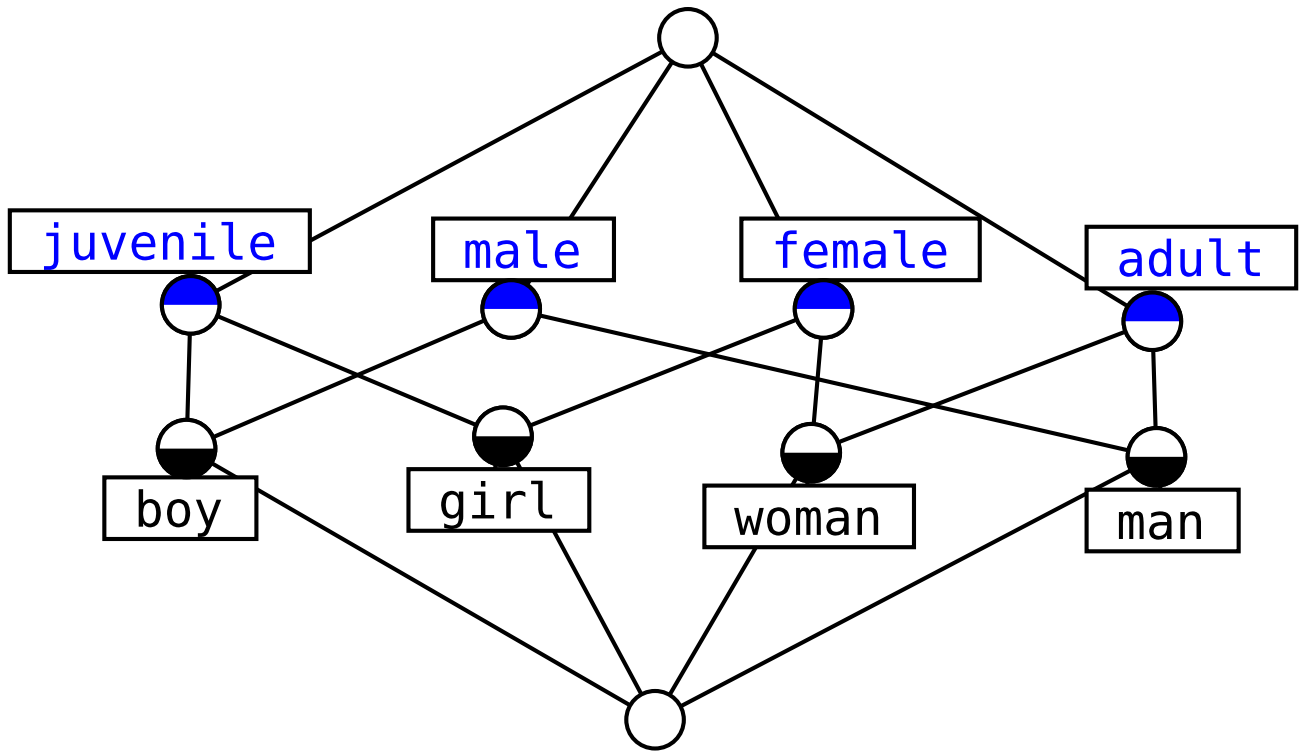
Features


ConExp-NG – "/home/rja/untitled.cex"
Dependencies Editor
Association Settings
Support
0.3
Confidence
0.5
#Association rules = 4
#With support = 4
Together With Implications
Sorted By:
Alphabetical Order
Support
Implications (Duquenne-Guigues Base/Stem Base)
0< 0 > [juvenile, adult] ==> [female, male];
1< 0 > [female, male] ==> [adult, juvenile];
Associations (Luxenburger Base)
1< 4 > [] =[0.5]=> < 2 > [adult];
2< 4 > [] =[0.5]=> < 2 > [female];
3< 4 > [] =[0.5]=> < 2 > [juvenile];
4< 4 > [] =[0.5]=> < 2 > [male];

Features


ConExp-NG – "/home/rja/untitled.cex"
Lattice Editor
Lattice
Objects: Attributes:
none none
labels labels
Edges:
none show
Objects
girl
woman
boy
man
Attributes
female
juvenile
adult
male
juvenile
male
female
adult
boy
girl
woman
man

Features



Status und Ausblick

- grundlegende Funktionalität vorhanden
- gute Ausgangsbasis für Weiterentwicklung
- notwendige Verbesserungen
  - Optionen für Verbandsvisualisierung
  - Layoutalgorithmen
  - Dateiformate
  - Separation in Bibliotheken
- eigene Ideen hinzufügen:
  `http://github.com/fcatools/conexp-ng/issues`

Stand 2016

`https://github.com/fcatools`:

- FCAlib
- ConExp-FX
- conexp-clj
- ConExp-NG
- Trias

`http://www.upriss.org.uk/fca/fcasoftware.html`

Dank an





- Gerd Stumme, Andreas Hotho, Bernhard Ganter, Sebastian Rudolph & Christoph Schmitz
- das Team von ConExp-NG: Torsten Casselt, Eugen Kiss, Jan Kassel & David Bormann
- Daniel Borchmann & Johannes Wollbold