Discovering Shared Conceptualizations in Folksonomies

Robert Jäschke [a,b] , Andreas Hotho [a] , Christoph Schmitz [a] , Bernhard Ganter [c] , Gerd Stumme [a,b]

[a]*Knowledge & Data Engineering Group, University of Kassel*
Wilhelmshöher Allee 73, 34121 Kassel, Germany
http://www.kde.cs.uni-kassel.de

[b]*Research Center L3S, Appelstr. 9a, 30167 Hannover, Germany*
http://www.l3s.de

[c]*Institute for Algebra, Dresden University of Technology*
Zellescher Weg 12 – 14, 01062 Dresden, Germany
http://www.math.tu-dresden.de/∼ganter/

Abstract

Social bookmark tools are rapidly emerging on the Web. In such systems users are setting up lightweight conceptual structures called folksonomies. Unlike ontologies, shared conceptualisations are not formalised, but rather implicit. We present a new data mining task, the *mining of all frequent tri-concepts*, together with an efficient algorithm, for discovering these implicit shared conceptualisations. Our approach extends the data mining task of discovering all closed itemsets to three-dimensional data structures to allow for mining folksonomies. We provide a formal definition of the problem, and present an efficient algorithm for its solution. Finally, we show the applicability of our approach on three large real-world examples.

Key words: Folksonomies, Tagging, Formal Concept Analysis

1. Introduction

Social resource sharing systems on the web, such as the shared photo gallery Flickr[1] or the bookmarking system del.icio.us,[2] have acquired large numbers of users within a few years. Flickr is known to have more than 1.5 million users,[3] while del.icio.us has celebrated crossing the 1 million users threshold in 2006.[4] The reason for their immediate success is the fact that no specific skills are needed for participating, and that these tools yield immediate benefit for each individual user (e.g. organizing ones bookmarks in a browser-independent, persistent fashion) without too much overhead.

The core data structure of a social resource sharing system is a *folksonomy*. It consists of assignments of arbitrary keywords – called 'tags' – to resources by users. Folksonomies are thus a lightweight knowledge representation for sharing knowledge on the web.

1.1. *Discovering shared conceptualisations*

Unlike ontologies, folksonomies do not suffer from the knowledge acquisition bottleneck, as the significant provision of content by many people shows. On the other hand, folksonomies – unlike ontologies [29] – do not explicitly state shared conceptualisations, nor do they force users to use the same tags. However, the us-

Email addresses: jaeschke@cs.uni-kassel.de (Robert Jäschke), hotho@cs.uni-kassel.de (Andreas Hotho), schmitz@cs.uni-kassel.de (Christoph Schmitz), bernhard.ganter@tu-dresden.de (Bernhard Ganter), stumme@cs.uni-kassel.de (Gerd Stumme).

[1] http://www.flickr.com

[2] http://del.icio.us

[3] http://money.cnn.com/magazines/business2/business2_archive/2005/12/01/8364623/

[4] http://blog.del.icio.us/blog/2006/09/million.html

age of tags of users with similar interests tends to converge to a shared vocabulary. Our intention is to discover these shared conceptualisations that are hidden in a folksonomy. To this end, we present in this paper an algorithm, TRIAS, for discovering subsets of folksonomy users who implicitly agree (on subsets of resources) on a common conceptualization.

Our algorithm will return a tri-ordered [5] set of triples, where each triple (A, B, C) consists of a set A of users, a set B of tags, and a set C of resources. These triples – called *tri-concepts* in the sequel – have the property that each user in A has tagged each resource in C with all tags from B, and that none of these sets can be extended without shrinking one of the other two dimensions. Each retrieved triple indicates thus a set A of users who (implicitly) share a conceptualisation, where the set B of tags is the intension of the concept, and the set C of resources is its extension. We can additionally impose minimum support constraints on each of the three dimensions 'users', 'tags', and 'resources', to retrieve the most significant shared concepts only.

1.2. *The problem of closed itemset mining in triadic data*

From a data mining perspective, the discovery of shared conceptualisations opens a new research field which may prove interesting also outside the folksonomy domain: 'Closed itemset mining in triadic data', which is located on the confluence of the research areas of association rule mining and Formal Concept Analysis.

Formal Concept Analysis (FCA) [74,25] is a mathematical theory that formalizes the concept of 'concept', and allows for computing concept hierarchies out of data tables. At the end of last century, one discovered that it also provides an elegant framework for significantly reducing the effort of mining association rules [50,78,64]. A new research area emerged which became known as *closed itemset mining* in the data mining community and as *iceberg concept lattices* [68] in FCA.

Independent of this development, Formal Concept Analysis has been extended about ten years ago to deal with three-dimensional data [40]. This line of *Triadic Concept Analysis* did not receive a broad attention up to now. With the rise of *folksonomies* as core data structure of social resource sharing systems, however, the interest in Triadic Concept Analysis increased again.

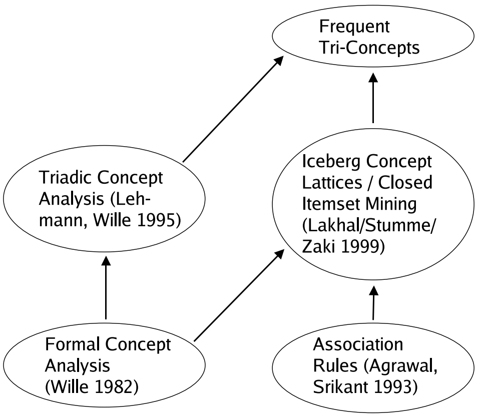

Frequent Tri-Concepts
Triadic Concept Analysis (Lehmann, Wille 1995)
Iceberg Concept Lattices / Closed Itemset Mining (Lakhal/Stumme/Zaki 1999)
Formal Concept Analysis (Wille 1982)
Association Rules (Agrawal, Srikant 1993)

Fig. 1. History of iceberg tri-lattices

With this paper, we initiate the confluence of both lines of research, Triadic Concept Analysis and closed itemset mining (see Figure 1). In particular, we give a formal definition of the *problem of mining all frequent tri-concepts* (in other terms: the three-dimensional version of mining all frequent closed itemsets), and present our algorithm TRIAS for mining all frequent tri-concepts of a given dataset.

With its sets of users, tags, and resources, folksonomies have one additional dimension compared to typical basket analysis datasets (which consist of the two dimensions 'items' and 'transactions'). Informally spoken, the task of mining all frequent *tri-sets* is to discover all triples of sets of users, tags, and resources, resp., such that, for each triple of sets, all users in the first set have assigned all tags in the second set to all resources in the third set, and that the cardinalities of the three sets are above predefined minimum support thresholds. [6]

As in the classical case, the resulting set of all frequent tri-sets is usually too large, and can be condensed without any loss of information. To this end, we adapt the notion of iceberg concept lattices (aka closed itemsets) to the three-dimensional nature of folksonomies. With our TRIAS algorithm, we provide an efficient method for computing all frequent tri-concepts.

1.3. *Contribution and organisation of the paper*

In this paper, we present the following contributions:

- a formal definition of the problem of mining frequent tri-concepts,
- TRIAS, an efficient algorithm for solving the problem,

[5] See Section 2.4 for details.

[6] In classical association rule mining, the thresholds equal the minimum support and minimal length thresholds.

– and a conceptual analysis of two social bookmarking systems and an IT security manual by means of this algorithm.

The paper is organized as follows. In the next section, we introduce folksonomies and social resource sharing systems in more detail and motivate the need of a conceptual clustering approach for this kind of data. In Section 2, we discuss the state of the art and related work in the research areas of folksonomies, Ontology Learning, Formal Concept Analysis, and closed itemset mining. In Section 3.1, we provide the formal definition of the problem of mining all frequent tri-concepts; in Section 3.2, we introduce our TRIAS algorithm; and in Section 3.3, we evaluate its performance. In Section 4, we apply our approach on three large-scale real-world applications: the folksonomy of the popular bookmark sharing system del.icio.us, the collection of publications in our social reference management system BibSonomy, and a manual for protecting IT infrastructure. Section 5 concludes with an outlook on future work. Parts of this article have been presented as a short paper at the Intl. Conf. on Data Mining 2006 [35] and at the Intl. Conf. on Conceptual Structures 2007 [36].

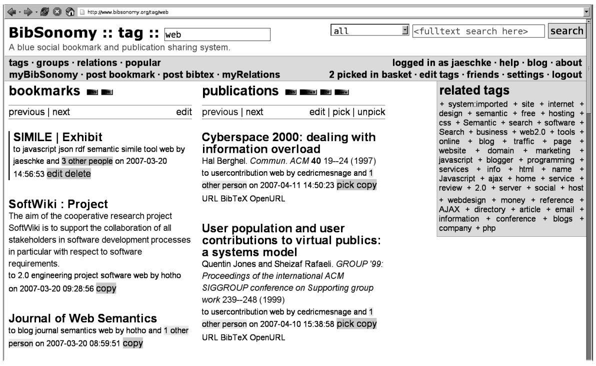
2. Basic Notions and State of the Art

In this section, we recall the basic notions and discuss the state of the art of the research areas relevant to this article: Folksonomies, Ontology Learning, Formal Concept Analysis and its triadic version, and the mining of closed itemsets.

2.1. *Social Resource Sharing Systems and Folksonomies*

Social resource sharing systems are web-based systems that allow users to upload their resources, and to label them with arbitrary words, so-called *tags*. Each system has a specific type of resources it supports. Flickr, for instance, enables the sharing of photos, del.icio.us the sharing of bookmarks, CiteULike [7] and Connotea [8] the sharing of bibliographic references, and 43Things [9] even the sharing of goals in private life. Our own system, *BibSonomy* [10] ([33], see Figure 2), allows the sharing of bookmarks and BibTeX entries simultaneously.

In their core, these systems are all very similar. Once a user is logged in, he can add a resource to the system, and assign arbitrary tags to it. The collection of all his assignments is his *personomy*, the collection of all personomies constitutes the *folksonomy*. The user can explore his personomy, as well as the personomies of the other users, in all dimensions: for a given user one can see all resources he has uploaded, together with the tags he has assigned to them (see Figure 2); when clicking on a resource one sees which other users have uploaded this resource and how they tagged it; and when clicking on a tag one sees who assigned it to which resources.


http://www.bibsonomy.org/tag/web
BibSonomy :: tag :: web
all
<fulltext search here>
search
tags · groups · relations · popular
myBibSonomy · post bookmark · post bibtex · myRelations
logged in as jaeschke · help · blog · about
2 picked in basket · edit tags · friends · settings · logout
bookmarks
publications
related tags
previous | next
edit
previous | next
edit | pick | unpick

A blue social bookmark and publication sharing system.

SIMILE | Exhibit
to javascript json rdf semantic simile tool web by jaeschke and 3 other people on 2007-03-20 14:56:53 edit delete

SoftWiki : Project
The aim of the cooperative research project SoftWiki is to support the collaboration of all stakeholders in software development processes in particular with respect to software requirements.
to 2.0 engineering project software web by hotho on 2007-03-20 09:28:56 copy

Journal of Web Semantics
to blog journal semantics web by hotho and 1 other person on 2007-03-20 08:59:51 copy

Cyberspace 2000: dealing with information overload
Hal Berghel. *Commun. ACM* **40** 19--24 (1997)
to usercontribution web by cedricmesnage and 1 other person on 2007-04-11 14:50:23 pick copy
URL BibTeX OpenURL

User population and user contributions to virtual publics: a systems model
Quentin Jones and Sheizaf Rafaeli. *GROUP '99: Proceedings of the international ACM SIGGROUP conference on Supporting group work* 239--248 (1999)
to usercontribution web by cedricmesnage and 1 other person on 2007-04-10 15:38:58 pick copy
URL BibTeX OpenURL

+ system:imported + site + internet + design + semantic + free + hosting + css + Semantic + search + software + Search + business + web2.0 + tools + online + blog + traffic + page + website + domain + marketing + javascript + blogger + programming + services + info + html + name + Javascript + ajax + home + service + review + 2.0 + server + social + host + webdesign + money + reference + AJAX + directory + article + email + information + conference + blogs + company + php

Fig. 2. Bibsonomy displays bookmarks and (BibTeX-based) bibliographic references simultaneously.

The word “folksonomy” is a blend of the words “taxonomy” and “folk”, and stands for conceptual structures created by the people [73]. Folksonomies are thus a bottom-up complement to more formalized Semantic Web technologies, as they rely on *emergent semantics* [61,62] which result from the converging use of the same vocabulary. The main difference to “classical” ontology engineering approaches is their aim to respect to the largest possible extent the request of non-expert users not to be bothered with any formal modeling overhead. Intelligent techniques may well be inside the system, but should be hidden from the user.

A folksonomy describes the users, resources, and tags, and the user-based assignment of tags to resources. We recall here our formal definition of folksonomies [34], which is also underlying our BibSonomy system.

Definition 1 *A* folksonomy *is a tuple* $\mathbb{F} := (U, T, R, Y, \prec)$ *where*

- U*,* T*, and* R *are finite sets, whose elements are called* users, tags*, and* resources*, resp.,*
- Y *is a ternary relation between them, i. e.,* $Y \subseteq U \times T \times R$*, whose elements are called tag assignments (*tas *for short), and*
- $\prec$ *is a user-specific subtag/supertag-relation, i. e.,* $\prec\ \subseteq U \times T \times T$*, called* is-a relation.

The personomy $\mathbb{P}_u$ *of a given user* $u \in U$ *is the restric-*

[7] http://www.citeulike.org

[8] http://www.connotea.org

[9] http://www.43things.com

[10] http://www.bibsonomy.org

tion of $\mathbb{F}$ to u, i. e., $\mathbb{P}_u := (T_u, R_u, I_u, \prec_u)$ with $I_u := \{(t,r) \in T \times R \mid (u,t,r) \in Y\}$, $T_u := \pi_1(I_u)$, $R_u := \pi_2(I_u)$, and $\prec_u := \{(t_1,t_2) \in T \times T \mid (u,t_1,t_2) \in \prec\}$, where π_i denotes the projection on the ith dimension.

Users are typically described by their user ID, and tags may be arbitrary strings. What is considered as a resource depends on the type of system. For instance, in del.icio.us, the resources are URLs, in flickr, the resources are pictures, and in BibSonomy they are either URLs or publication entries.

As the is-a relation $\prec$ was only implemented in a rudimentary way (so-called 'bundles' in del.icio.us) in one of the systems considered in our paper at the time of writing,[11] we will ignore it for the purpose of this paper. Therefore, we will consider a folksonomy as a four-tuple $\mathbb{F} := (U, T, R, Y)$, without the $\prec$ relation.

Related Work. While the scientific community has only begun to explore folksonomies as a knowledge representation mechanism as well as a source of data which can be mined for different purposes, there is a growing number of publications concerned with the various aspects of this new phenomenon. Overviews of social bookmarking tools with special emphasis on folksonomies are provided by [31] and [43], as well as [46] and [60] who discuss strengths and limitations of folksonomies. Recent papers include [28] and [21] which focus on analyzing and visualizing the structure of folksonomies. The knowledge discovery, information retrieval, and knowledge engineering communities are currently becoming involved in this development, e. g., by enhancing recommendations given by the systems, to improving search and ranking, and structuring the knowledge in a systematic way.

Cattuto et al. [17] investigate statistical properties of tagging systems and introduce a stochastic model of user behaviour; [30] analyses the dynamics and semantics of tagging systems, and [39] introduces further techniques to structure the tripartite network of folksonomies. Recently, work on more specialized topics such as structure mining on folksonomies – e. g. to visualize trends [21] has been presented.

In [34], we presented FolkRank, a differential version of the PageRank algorithm [11] for computing topic-specific rankings of users, tags, and resources in a folksonomy. In [57], we computed association rules on del.ico.us data.

2.2. *Ontology Learning*

The term *ontology learning* was first introduced by Mädche and Staab in [44]. It stands for the task of (semi-)automatically constructing an ontology or a domain model. Usually machine learning or data mining algorithms are applied mostly on textual data to extract the hidden conceptualization from the data and to make it explicit. Revealing the hidden conceptualization of an author partially written in a text document can be seen as a kind of reverse engineering task (cf. [18]). All ontology learning approaches try to support the knowledge engineer by setting up the ontology. Recent advances in ontology learning are described in [12].

In this paper, we describe one step for learning ontologies from folksonomies. Other approaches are discussed in the next paragraph.

Related Work. Approaches trying to analyze the weakly structured information of folksonomies and use this to learn conceptualization or ontologies are still rare. Among them is the work of Mika [47], who defines a model of semantic-social networks for extracting lightweight ontologies from del.icio.us. Besides calculating measures like the clustering coefficient, (local) betweenness centrality or the network constraint on the extracted one-mode network, Mika uses co-occurrence techniques for clustering the folksonomy.

Paul Heymann and Hector Garcia-Molina [32] propose a new clustering algorithm to construct a tag hierarchy. Schmitz proposes in [58] the construction of a subsumption tree consisting of Flickr tags based on the tag co-occurrence network of tags. Both approaches are showing ways to construct an ontology, but both are using only parts of the information of an folksonomy as they are based on an aggregated graph rather than the full folksonomy.

2.3. *Formal Concept Analysis*

Formal Concept Analysis (FCA) is a conceptual clustering technique that formalizes the concept of 'concept' as established in the international standard ISO 704: a concept is considered as a unit of thought constituted of two parts: its extension and its intension [74,25]. This understanding of 'concept' is first mentioned explicitly in the Logic of Port Royal [4]. To allow a formal description of extensions and intensions, FCA starts with a *(formal) context*:

Definition 2 ([74]) *A formal context is a triple $\mathbb{K} := (G, M, I)$ which consists of a set G of objects [German:*

[11] BibSonomy now provides the $\prec$ hierarchy as 'relations'.

Gegenstände], a set M of attributes [Merkmale], and a binary relation $I \subseteq G \times M$. $(g,m) \in I$ is read as "object g has attribute m".

This data structure equals the set of transactions used for association rule mining, if we consider M as the set of items and G as the set of transactions.

Definition 3 ([74]) *For $A \subseteq G$, let*

$$A^I := \{m \in M \mid \forall g \in A\colon (g,m) \in I\} \;;$$

and dually, for $B \subseteq M$, let

$$B^I := \{g \in G \mid \forall m \in B\colon (g,m) \in I\} \;.$$

Now, a formal concept *is a pair (A,B) with $A \subseteq G$, $B \subseteq M$, $A^I = B$ and $B^I = A$. A is called* extent *and B is called* intent *of the concept.*

This is equivalent to saying that $A \times B \subseteq I$ such that neither A nor B be can be enlarged without violating this condition.

Definition 4 ([74]) *The set $\mathfrak{B}(\mathbb{K})$ of all concepts of a formal context $\mathbb{K}$ together with the partial order $(A_1,B_1) \leq (A_2,B_2) :\Leftrightarrow A_1 \subseteq A_2$ (which is equivalent to $B_1 \supseteq B_2$) is a complete lattice, called the* concept lattice *of $\mathbb{K}$.*

The concept lattice is a hierarchical conceptual clustering of the data which can be visualised by a Hasse diagram. This visualisation technique has been used in many applications for qualitative data analysis [24]. An example of a Hasse diagram is given in Figure 6 and described in more detail in Section 4.1.

Related Work. FCA has grown over the years to a powerful theory for data analysis, information retrieval, and knowledge discovery [65]. In Artificial Intelligence (AI), FCA is used as a knowledge representation mechanism [66] and as conceptual clustering method [63,15,48]. In database theory, FCA has been extensively used for class hierarchy design and management [49,77,20,72,56,27].

The amount of publications on Formal Concept Analysis is abundant. A good starting point for the lecture are the textbooks [25,16,24], the collection of FCA publications in BibSonomy, [12] and the proceedings of the Intl. Conference on Formal Concept Analysis [13] and the Intl. Conference on Conceptual Structures [14] series.

2.4. *Triadic Concept Analysis*

Inspired by the pragmatic philosophy of Charles S. Peirce with its three universal categories [54], Rudolf Wille and Fritz Lehmann extended Formal Concept Analysis in 1995 with a third category:

Definition 5 ([40]) *A* triadic formal context *is a quadruple $\mathbb{F} := (G,M,B,Y)$ where G, M, and B are sets, and Y is a ternary relation between G, M, and B, i. e., $Y \subseteq G \times M \times B$. The elements of G, M, and B are called* (formal) objects, attributes, *and* conditions, *resp, and $(g,m,b) \in Y$ is read "object g has attribute m under condition b".*

A triadic formal context models exactly the structure of a folksonomy $\mathbb{F} := (U,T,R,Y)$ without tag hierarchy $\prec$.

Definition 6 ([40]) *A* triadic concept *of $\mathbb{F}$ is a triple (A_1,A_2,A_3) with $A_1 \subseteq G$, $A_2 \subseteq M$, and $A_3 \subseteq B$ with $A_1 \times A_2 \times A_3 \subseteq Y$ such that none of its three components can be enlarged without violating this condition.*

From each of the three dimensions one obtains a quasi-order $\lesssim_1$, $\lesssim_2$, and $\lesssim_3$, resp., on the set of all tri-concepts: For $i = 1,2,3$, let $(A_1,A_2,A_3) \lesssim_i (B_1,B_2,B_3)$ iff $A_i \subseteq B_i$.

The definition of a triadic concept is the natural extension of the definition of a formal concept to the triadic case. Alternatively the definition can be described with $\cdot^I$ operators similar to the dyadic case, but as there are now three dimensions involved, the notation (which we omit here, cf. [40]) becomes more complex.

Lemma 1 ([40]) *For two tri-concepts $\mathfrak{a}$ and $\mathfrak{b}$, and for $i \neq j \neq k \neq i$, $\mathfrak{a} \lesssim_i \mathfrak{b}$ and $\mathfrak{a} \lesssim_j \mathfrak{b}$ implies $\mathfrak{b} \lesssim_k \mathfrak{a}$.*

This implication is the triadic version of the dyadic proposition that for two dyadic concepts (A_1,A_2) and (B_1,B_2) holds $A_1 \subseteq B_1$ iff $B_2 \subseteq A_1$. In the dyadic case, the two orders induced by the concept extents and the concept intents, resp. are thus dually isomorphic. This allows for visualising the concept lattice in just one diagram and is at the same time the justification for the famous support pruning strategy in the Apriori algorithm. In the triadic case, the relationship between the three quasi-orders is unfortunately weaker (as seen above), which makes both the mining (see Section 3.2) and the visualisation (see Section 4.2) more complex. Figures 7 – 9 show examples of diagrams of triadic concept lattices; they are discussed in detail in Section 4.

Lehmann and Wille present in [40] an extension of the theory of ordered sets and (concept) lattices to the triadic case, and discuss structural properties. This approach initiated research on the theory of *concept trilattices*.

[12] http://www.bibsonomy.org/tag/fca

[13] http://www.informatik.uni-trier.de/~ley/db/conf/icfca/

[14] http://www.informatik.uni-trier.de/~ley/db/conf/iccs/

Whereas there have been some significant publications on the mathematical properties of trilattices (see below), this approach had no large impact on real-world applications up to now. This is mainly due to its abovementioned resistance to scalable visualisations. With the rise of social resource sharing systems on the web, triadic data move again in the focus of many researchers. In this setting, one needs – beside a more scalable visualisation paradigm – knowledge discovery and information retrieval methods and algorithms that are able to handle very large datasets.

Related Work. Following the initial paper [40] by Lehmann and Wille, several researchers started to analyse the mathematical properties of trilattices, e. g., [7–9,19,23,75,76]. [40] and [19] present several ways to project a triadic context to a dyadic one. [67] presents a model for navigating a triadic context by visualising concept lattices of such projections. In [57], we discussed how to compute association rules from a triadic context, based on these (and other) projections. A first step towards truly 'triadic association rules' has been done in [23].

2.5. *Closed Itemset Mining*

In terms of Formal Concept Analysis, the task of mining frequent itemsets [1] can be described as follows: Given a formal context $\mathbb{K} = (G, M, I)$ and a threshold minsupp $\in [0, 1]$, determine all subsets B of M where the *support* $\operatorname{supp}(B) := \frac{\operatorname{card}(B^I)}{\operatorname{card}(G)}$ (with B^I as defined above) is larger than the threshold minsupp. In warehouse basket analysis, M is the set of items and G is the set of transactions.

The set of these so-called *frequent itemsets* itself is usually not considered as a final result of the mining process, but rather an intermediate step. Its most prominent use are association rules [1]. Association rules are for instance used in warehouse basket analysis, where the warehouse management is interested in learning about products that are frequently bought together.

Since determining the frequent itemsets is the computationally most expensive part, most research has focused on this aspect. Most algorithms follow the way of the well-known Apriori algorithm [2], which is traversing iteratively the set of all itemsets in a levelwise manner. Algorithms based on this approach have to extract the supports of *all* frequent itemsets from the database. However, this is by no means necessary.

It turned out that FCA can significantly improve both the efficiency and the effectiveness of frequent itemset mining. [50,78,64] discovered independently that it is sufficient to consider the intents of those concepts where the cardinality of their extent is above the minimum support threshold. These frequent concept intents are called *closed itemsets* in association rule mining, because the set of all concept intents is a closure system (i. e., it is closed under set intersection). The corresponding closure operator is the consecutive application of the two $\cdot^I$ operators defined in the previous subsection. I. e., for an itemset B, the set B^{II} is the smallest concept intent containing B. This closure operator will be used in the TRIAS algorithm in Section 3.2.

In FCA, the equivalent notion is that of an *iceberg concept lattice* [68], which is the $\bigvee$–semi-lattice $\{(A, B) \in \underline{\mathfrak{B}}(\mathbb{K}) \mid \frac{\text{card(A)}}{\text{card(G)}} \geq \text{minsupp}\}$ with the order defined in Section 2.3. The iceberg concept lattice visualises the most frequent concepts of a dataset [68], and allows for an efficient visualisation of a basis (condensed set) of association rules [69,52]. These bases allow to reduce the number of rules significantly without losing any information.

Related Work. The problem of mining frequent itemsets arose first as a sub-problem of mining association rules [1], but it then turned out to be present in a variety of problems: mining sequential patterns [3], episodes [45], association rules [2], correlations [59], multi-dimensional patterns [37,41], maximal itemsets [6,79,42], closed itemsets [71,50,51,53].

The first algorithm based on the combination of association rule mining with FCA was Close [50], followed by A-Close [51], ChARM [78], Pascal [5], Closet [53], and Titanic [68], each having its own way to exploit the closure operator which is hidden in the data. Many algorithms can be found at the Frequent Itemset Mining Implementations Repository. [15]

Beside closed itemsets, other condensed representations have been studied: key sets [5]/free sets [10], δ-free sets [10], non-derivable itemsets [14], disjunction free sets [13], and k-free sets [55]. Closed itemsets and other condensed representations can be used for defining bases of association rules [69,52].

3. Mining all Frequent Tri-Concepts of a Folksonomy

In this section we formalize the problem of mining all frequent tri-concepts of a folksonomy, present the

[15] http://fimi.cs.helsinki.fi/

TRIAS algorithm for its efficient solution, and discuss its performance.

3.1. *The Problem of Mining all Frequent Tri-Concepts*

We will now formalize the problem of mining all frequent tri-concepts. We start with an adaptation of the notion of 'frequent itemsets' to the triadic case.

Definition 7 *Let $\mathbb{F} := (U, T, R, Y)$ be a folksonomy/triadic context. A* tri-set *of $\mathbb{F}$ is a triple (A, B, C) with $A \subseteq U$, $B \subseteq T$, $C \subseteq R$ such that $A \times B \times C \subseteq Y$.*

As folksonomies have three dimensions which are completely symmetric, one can establish minimum support thresholds on all of them. The general problem of mining frequent tri-sets is then the following:

Problem 1 (Mining all frequent tri-sets) *Let $\mathbb{F} := (U, T, R, Y)$ be a folksonomy/triadic context, and let u-minsupp, t-minsupp, r-minsupp $\in [0, 1]$. The task of mining all frequent tri-sets consists in determining all tri-sets (A, B, C) of $\mathbb{F}$ with $\frac{|A|}{|U|} \geq u$-minsupp, $\frac{|B|}{|T|} \geq t$-minsupp, and $\frac{|C|}{|R|} \geq r$-minsupp.*

This is actually a harder problem than the direct adaptation of frequency to one more dimension: In classical frequent itemset mining, one has a constraint – the frequency – only on one dimension (the number of transactions). Thus the equivalent triadic version of the problem would need two minimum support thresholds only (say u-minsupp and t-minsupp). However, this seems not natural as it breaks the symmetry of the problem. Hence we decided to go for the harder problem directly (which equals in the dyadic case the addition of a minimal length constraint on the itemsets). The lighter version with only two constraints is then just a special case (e. g., by letting r-minsupp := 0).

As in the dyadic case, our thresholds are monotonic/antimonotonic constraints: If (A_1, B_1, C_1) with A_1 being maximal for $A_1 \times B_1 \times C_1 \subseteq Y$[16] is not u-frequent, then all (A_2, B_2, C_2) with $B_1 \subseteq B_2$ and $C_1 \subseteq C_2$ are not u-frequent either. The same holds symmetrically for the other two dimensions.

With the step from two to three dimensions, however, the direct symmetry between monotonicity and anti-monotonicity (which results in the dyadic case from the dual order isomorphism between the set of concept extents and the set of concept intents) breaks. All we have in the triadic case is the following lemma which results (via the three quasi-orders defined in Section 2.4) from the triadic Galois connection [8] induced by a triadic context.

Lemma 2 (cf. [40]) *Let both (A_1, B_1, C_1) and (A_2, B_2, C_2) be tri-sets with A_i being maximal for $A_i \times B_i \times C_i \subseteq Y$, for $i = 1, 2$.*[17] *If $B_1 \subseteq B_2$ and $C_1 \subseteq C_2$ then $A_2 \subseteq A_1$. The same holds symmetrically for the other two directions.*

As the set of all frequent tri-sets is highly redundant, we will in particular consider a specific condensed representation, i. e., a subset which contains the same information, namely the set of all frequent tri-concepts.

Definition 8 *A tri-set is a frequent tri-concept if it is both a tri-concept and a frequent tri-set.*

Problem 2 (Mining all frequent tri-concepts) *Let $\mathbb{F} := (U, T, R, Y)$ be a folksonomy/triadic context, and let u-minsupp, t-minsupp, r-minsupp $\in [0, 1]$. The task of mining all frequent tri-concepts consists in determining all tri-concepts (A, B, C) of $\mathbb{F}$ with $\frac{|A|}{|U|} \geq u$-minsupp, $\frac{|B|}{|T|} \geq t$-minsupp, and $\frac{|C|}{|R|} \geq r$-minsupp.*

Sometimes it is more convenient to use absolute rather than relative thresholds. For this case we let $\tau_u := |U| \cdot u$-minsupp, $\tau_t := |T| \cdot t$-minsupp, and $\tau_r := |R| \cdot r$-minsupp.

Once Problem 2 is solved, we obtain the answer to Problem 1 in a straightforward enumeration as $\{(A, B, C) \mid \exists$ frequent tri-concept $(\hat{A}, \hat{B}, \hat{C})$: $A \subseteq \hat{A}, B \subseteq \hat{B}, C \subseteq \hat{C}, |A| \geq \tau_u, |B| \geq \tau_t, |C| \geq \tau_r\}$.

3.2. *The* TRIAS *Algorithm for Mining all Frequent Tri-Concepts*

Our algorithm for mining all frequent tri-concepts of a folksonomy $\mathbb{F} := (U, T, R, Y)$ is listed as Algorithm 3.1. A prior version was used for analysing psychological studies [38]. That application varied from TRIAS as it aimed at an iterative pruning of the data set. Furthermore, it did not take into account any frequency constraints.

We let $\tilde{Y} := \{(u, (t, r)) \mid (u, t, r) \in Y\}$, and we identify the elements of U, T, and R with natural numbers, i. e. $U = \{1, \ldots, |U|\}$ (and symmetrically for T, R). In both its outer and its inner loop, TRIAS calls the pairs of subroutines *FirstFrequentConcept*$((G, M, I), \tau)$ and *NextFrequentConcept*$((A, B), (G, M, I), \tau)$. These two routines provide an enumeration of all frequent dyadic concepts (A, B) of the formal (dyadic) context (G, M, I). The context is passed over as input parameter. *FirstFrequentConcept*

[16] In the dyadic case this condition is implicitly covered by the use of B^I in the definition of the support since, for any given $B \subseteq M$, the set B^I is always maximal with $B^I \times B \subseteq I$.

[17] This holds in particular if the tri-sets are tri-concepts, see Lemma 1.

```
TRIAS($U, T, R, Y, \tau_u, \tau_t, \tau_r$)
1.  begin
2.    $\tilde{Y} := \{(u,(t,r)) \mid (u,t,r) \in Y\}$
3.    $(A,I) :=$ FirstFrequentConcept($(U, T \times R, \tilde{Y}), \tau_u$)
4     repeat
5.      if $|I| \geq \tau_t \cdot \tau_r$ then begin
6.        $(B,C) :=$ FirstFrequentConcept($(T,R,I), \tau_t$)
7.        repeat
8.          if $|C| \geq \tau_r$ then
9.            if $A = (B \times C)^{\tilde{Y}}$ then output($A,B,C$)
10.       until not NextFrequentConcept($(B,C), (T,R,I), \tau_t$)
11.     endif
12.   until not NextFrequentConcept($(A,I), (U, T \times R, \tilde{Y}), \tau_u$)
13.end
```

Algorithm 3.1: The TRIAS algorithm for mining all frequent tri-concepts

```
FirstFrequentConcept($(G,M,I), \tau$)
1.  begin
2.    $A := \emptyset^I$
3.    $B := A^I$
4.    if $|A| < \tau$ then
5.      NextFrequentConcept($(A,B), (G,M,I), \tau$)
6.    endif
7.    return $(A,B)$
8.  end
```

Algorithm 3.2: The *FirstFreqentConcept* function of the TRIAS algorithm

```
NextFreqentConcept($(A,B), (G,M,I), \tau$)
1.  begin
2.    while defined($i$) begin
3.      $A := (B \oplus i)^I$
4.      if $|A| \geq \tau$ then
5.        $D := A^I$
6.        if $B <_i D$ then
7.          $B := D$
8.          return true
9.        endif
10.     endif
11.     $i :=$ max($M \setminus B \cap \{1, \ldots, i-1\}$
12.   end
13.   return false
14.end
```

Algorithm 3.3: The *NextFreqentConcept* function of the TRIAS algorithm

returns in (A,B) the first concept of the enumeration. *NextFrequentConcept* takes the current concept (A,B) and modifies it to the next concept of the enumeration. This way, we compute all frequent maximal cuboids in the relation Y by consecutively computing maximal rectangles in the binary relations $\tilde{Y}$ and I, resp, where the condition in line 9 of Algorithm 3.1 checks if the rectangle layers form a maximal cuboid. Note that $A \subseteq (B \times C)^{\tilde{Y}}$ trivially holds, because of $A = I^{\tilde{Y}}$ and $(B \times C) \subseteq I$. Hence only "$\supseteq$" has to be checked.

For computing all (frequent) maximal rectangles in a binary relation, one can resort to any algorithm for computing (iceberg) concept lattices. The enumeration can be done in any convenient way. For the inner and the outer loop, one could use different algorithms for that task.

In our implementation we equipped the NEXTCLOSURE algorithm [22,25] of the fourth author with frequency pruning for implementing the *FirstFrequentConcept* and *NextFrequentConcept* routines (see Algorithms 3.2 and 3.3, resp.) for both the outer and the inner loop. This algorithm has the advantage that it needs almost no space in main memory.

NEXTCLOSURE computes the concepts of a dyadic formal context (G,M,I) in a particular order, starting with the concept $(\emptyset^I, \emptyset^{II})$. For a given concept (A,B), NEXTCLOSURE computes the concept (C,D) whose intent D is the next set after B in the so-called *lectic* order. The lectic order on sets is a total order and is equivalent to the lexicographic order of bit vectors representing those sets.

To find the next concept we define, for $B \subseteq M$ and $i \in M$,

$$B \oplus i := (B \cap \{1, \ldots, i-1\}) \cup \{i\}.$$

By applying the closure operator $X \mapsto X^{II}$ to $B \oplus i$, the algorithm computes, for a given B, the set $D := (B \oplus i)^{II}$. This is the lectically next intent, if $B <_i D$ holds, meaning that i is the smallest element in which B and D differ, and $i \in D$.

The method *NextFrequentConcept* adopts this idea and additionally checks if the computed extent $A := (B \oplus i)^I$ fulfills the minimal support criterion before computing the intent $D := A^I$. This is done in line 4 of Algorithm 3.3 by considering the extent A only if it is large enough.

Taking a closer look on the function $\cdot^I$ revealed that it demands the computation of several set intersections at a time. Since profiling showed that this is the main bottleneck of the algorithm, we optimized this by first ordering the sets to be intersected by size (with the smallest set first). Then the algorithm recursively intersects them with a procedure used for merge-sort. This is possible, since every itemset of the binary context can be accessed as ordered list in the data structure described in the following.

Because two sortings of Y are needed, instead of storing both, we just store the permutations for every order and an additional offset table which allows constant time access to the triples of a given tag, user, or resource. The chosen approach is exemplified in Figure 3. The table on the left contains the unsorted triples Y of which only the values from U are shown here. The table in the middle describes the permutation which allows to access the triples in lexicographic order. Finally, the right table contains, for every element $u \in U$, an offset

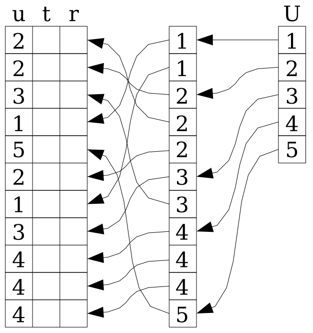


Fig. 3. Accessing triples in sorted order

which points to the position in the second table, which points to the first triple of that user in the Y list. Together, all this allows constant time access to the sorted tag-resource set of every user.

3.3. *Performance of the* TRIAS *Algorithm*

As in the dyadic case, the number of (frequent) tri-concepts may grow exponentially in the worst case. Biedermann has shown in [9] that the concept tri-lattice of the triadic context of size $n \times n \times n$ where only the main diagonal is empty has size 3^n. In typical applications, however, one is far from this theoretical boundary. Therefore we focus on empirical evaluations on a large scale real-world dataset.

For measuring the runtime and the number of frequent concepts we have evaluated the performance of TRIAS on a snapshot of the del.icio.us system (which is described in more detail in Section 4.1). It consists of all users, tags, resources and tag assignments we could download that were entered to the system on or before June 15, 2004. From this base set we created monthly snapshots as follows. $\mathbb{F}_0$ contains all tag assignments performed on or before Dec 15, 2003, together with the involved users, tags, and resources; $\mathbb{F}_1$ all tag assignments performed on or before Jan 15, 2004, together with the involved users, tags, and resources; and so on until $\mathbb{F}_6$ which contains all tag assignments performed on or before June 15, 2004, together with the involved tags, users, and resources. This represents seven monotonously growing contexts describing the del.icio.us folksonomy at different points in time. For mining frequent tri-sets and frequent tri-concepts we used minimum support values of $\tau_u := \tau_t := \tau_r := 2$ and measured the run-time of our Java implementations on a dual-core Opteron system with 2 GHz and 8 GB RAM.

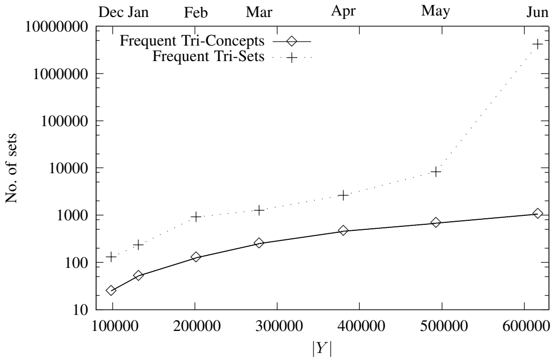


Fig. 4. Number of frequent tri-sets vs. number of frequent tri-concepts

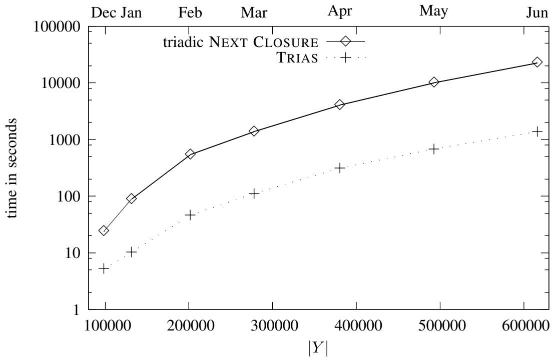


Fig. 5. Runtime of triadic NEXT CLOSURE and TRIAS algorithm on del.icio.us datasets

Figure 4 shows the number of frequent tri-concepts versus the number of frequent tri-sets on the logarithmically scaled y-axis, whereas the x-axis depicts the number of triples in Y – which grows from 98,870 triples in Dec 2003 to 616,819 in June 2004. It shows a massive increase of frequent tri-sets in June 2004 with only a modest growth of the number of frequent tri-concepts. This difference results from the fact that more and more users appear and start to agree on a common vocabulary, which leads to more frequent tri-concepts with larger volumes from June 2004 on. Such large concepts (like those shown in Table 1) contain combinatorially many frequent tri-sets.

One can observe that the number of frequent tri-sets of every snapshot is always at least one magnitude of size larger than the number of frequent tri-concepts. Consequently, computing frequent tri-sets is much more demanding than computing frequent tri-concepts – without providing any additional information.

A comparison of the speed improvement gained from not computing all tri-concepts with an algorithm like NEXT CLOSURE and afterwards pruning the non-frequent concepts but using the TRIAS algorithm for directly mining frequent tri-concepts is shown in Figure 5. The logarithmically scaled y-axis depicts the runtime of the algorithms in seconds while the x-axis shows again the size of the Y relation. One can see that computing all tri-concepts is more than one magnitude more expensive than mining only the frequent tri-concepts one is interested in.

With these observations we conclude that the TRIAS algorithm provides an efficient method to mine frequent tri-concepts in large scale conceptual structures.

4. Applications

We have applied the algorithm on three real-world data sets: the social bookmarking system del.icio.us, the IT Baseline Security Manual of the German Federal Office for Information Security, and the collection of publications in our social reference management system BibSonomy.

4.1. *The Social Bookmarking System del.icio.us*

First, we have analyzed the popular social bookmarking sytem del.icio.us with our approach. Del.icio.us is a server-based system with a simple-to-use interface that allows users to organize and share bookmarks on the internet. It is able to store for each URL, in addition to the tags assigned to it, a description and a note.

For detecting communities of users which have the same tagging behaviour (an thus share their conceptualisations), we ran the TRIAS algorithm on a del.icio.us snapshot consisting of all users, resources, tags and tag assignments we could download that were entered to the system on or before June 15, 2004 [34]. The resulting folksonomy consists of $|U| = 3,301$ users, $|T| = 30,416$ different tags, $|R| = 220,366$ resources (URLs), which are linked by $|Y| = 616,819$ triples.

As a first step, we ran TRIAS on the dataset without restricting the minimum supports (i. e., $\tau_u := \tau_t := \tau_r := 0$). The resulting concept tri-lattice consists of $246,167$ tri-concepts. We then investigated the concepts which contain two or more users, tags and resources, i. e., with $\tau_u := \tau_t := \tau_r := 2$. There were $1,062$ such tri-concepts. [18]

Figure 1 shows three examples. The first of them shows that the two users *bibi* and *poppy* have assigned the three tags *women*, *cinema*, and *film* to all the ten listed web pages, which are all about women in movies or women in the movie industry.

Table 1
Examples of frequent tri-concepts of del.icio.us

A	bibi poppy
B	women cinema film
C	http://www.reelwomen.org/ http://www.people.virginia.edu/~pm9k/libsci/womFilm.html http://www.lib.berkeley.edu/MRC/womenbib.html http://www.beaconcinema.com/womfest/ http://www.widc.org/ http://www.wftv.org.uk/home.asp http://www.feminist.com/resources/artspeech/media/femfilm.htm http://www.duke.edu/web/film/pioneers/ http://www.womenfilmnet.org/index.htm#top http://208.55.250.228/

A	fischer gnat
B	css design web
C	http://www.quirksmode.org/ http://webhost.bridgew.edu/etribou/layouts/ http://www.picment.com/articles/css/funwithforms/ http://www.alistapart.com/articles/sprites/

A	angusf carlomazza
B	css design web
C	http://www.positioniseverything.net/index.php http://www.fu2k.org/alex/css/layouts/3Col_NN4_FMFM.mhtml http://glish.com/css/home.asp http://www.maxdesign.com.au/presentation/process/index.cfm http://unraveled.com/projects/css_tabs/

The two lower tri-concepts show that different tri-concepts with the same extent can co-exist. [19] The first of them shows that the two users *fischer* and *gnat* agree (implicitly) in their assignments of the tags *css*, *web*, and *design* to the four listed URLs, while the users *angusf* and *carlomazza* agree in assigning the same tags to five completely different URLs. When inspecting the corresponding web pages, one finds out that the content of all resources is indeed very much related. These two related tri-concepts may be exploited further for extracting relations between tags or for recommending to all of the four users to study the posts of the other three.

Next, we wanted to study in more detail shared conceptualisations around the tags *css*, *web*, and *design*. To this end, we computed the concept lattice that is shown in Figure 6. Its formal context (G, M, I) was constructed as follows. Its set G of objects was ex-

[18] Larger thresholds did not provide any results any more. This comes from the fact that we took a rather early snapshot of del.icio.us, where the numbers of users, tags, and resources were still rather small. See also Section 3.3.

[19] This is in contrast to the situation in the dyadic case, where equality in one dimension implies equality in the other one.

tracted from the set of all resources by selecting all those resources which were tagged with at least one of these three tags by at least $k_1 \in \mathbb{N}$ users. The set M contains all tags. A tag $t \in M$ is defined to be related to a resource $r \in G$ (i. e., $(r,t) \in I$) iff $\frac{|\{u \in U | (u,t,r) \in Y\}|}{|\{u \in U | \exists r' \in R : (u,t,r') \in Y\}|} \geq k_2$, for a given $k_2 \in [0,1]$.

In this analysis, we have set $k_1 = 5$. This means that a resource was considered only if at least five users assigned it to at least one of the tags *css*, *web*, and *design*. This resulted in 575 resources. The second pruning parameter was set to $k_2 = 0.5$, i. e., at least half of the users who considered a resource had to use a particular tag, otherwise the tag was not assigned to the resource. This resulted in a relatively sparse assignment which reflects only rather strong shared conceptualisations. This way, only 22 tags were assigned to at least one resource; and only 297 out of the 575 resources received at least one tag.

The resulting concept lattice is displayed in Figure 6. Because of space restrictions, we pruned from it the tags *rest, cms, wiki, xml, fonts, wordpress, google, search, color, art*, and *music*. These tags formed singletons (i. e., separate nodes that were connected only to the top and to the bottom element of the lattice) with one or two resources each.

Each node in the diagram is a formal concept according to the definition in Section 2.3, i. e., a pair (A,B) where A is its extent (all resources belonging to it), and B is its intent (all tags belonging to it). In the diagram, the extent of a concept consists of all resources attached to the concepts or to any of its sub-concepts; and the intent consists of all tags that are attached to the concept or to any of its super-concepts. The leftmost concept, for instance, has the two URLs starting with "www.fiftyfoureleven..." as extent, and the set $\{$*php, css*$\}$ of tags as intent. The top node represents the concept (G, G^I), and the bottom node the concept (M^I, M).

The diagram shows that most agreement exists for the usage of the tag *css*, as it was assigned (according to our majority vote with the k_2 threshold) to 235 resources, while *web* was assigned to only 14 resources, and *design* to 31 resources. Apparently, the latter are too general or polysemous terms to reach a large agreement about their usage.

The resulting concept lattice could now be used for building a concept hierarchy. It suggests to the ontology engineer, e. g., to model *architecture* as a sub-concept of *design*. Another use of the concept lattice is a collaborative filtering approach to web search. When a user is for instance searching for "*web design*", the system could recommend him the web pages http://www.alistapart.com/articles/elastic and http://9rules.com/version2/.

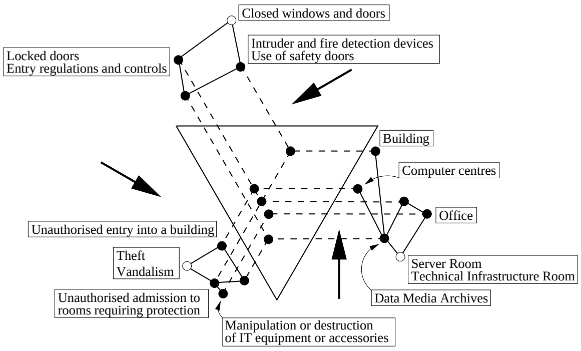

Closed windows and doors
Locked doors
Entry regulations and controls
Intruder and fire detection devices
Use of safety doors
Building
Computer centres
Office
Unauthorised entry into a building
Theft
Vandalism
Server Room
Technical Infrastructure Room
Unauthorised admission to
rooms requiring protection
Data Media Archives
Manipulation or destruction
of IT equipment or accessories

Fig. 7. All frequent tri-concepts of the IT Baseline Security Manual for $\tau_u = \tau_t = \tau_r = 3$.

4.2. *IT Baseline Protection Manual*

To illustrate another use of iceberg tri-lattices, we focus now on a non-folksonomy application. The IT Baseline Security Manual [26] of the German Federal Office for Information Security provides a description of a threat scenario and standard security measures for typical IT systems, and detailed descriptions of safeguards to assist with their implementation. [20]

Unlike a folksonomy, this manual has not been set up by an open group of users, but by a closed group of experts of the federal office. The manual has thus carefully been designed by domain specialists, and can be considered as an ontology (a formal specification of the shared conceptualisation of the experts of the federal office) – structured in form of a triadic context. Here, we use our knowledge discovery approach not for discovering a shared conceptualisation, but for analysing it. Even though the manual is smaller than a typical folksonomy resulting from a social bookmarking system, it is still by far too large to be analysed without technical support.

The core data of the manual forms a triadic context (U,T,R,Y). We consider as objects U the 66 IT components, as attributes T the 377 listed threats, and as conditions R the 912 safeguards. They are related by $5,680$ triples. [21]

From this dataset, we have computed the iceberg concept lattice for $\tau_u = \tau_t = \tau_r = 3$. Its visualisation in Figure 7 follows the conventions introduced in [40]. The five nodes in the middle are the five resulting fre-

[20] The online version of the manual is available at http://www.bsi.de/gshb/.

[21] See [19,70,76] for other analyses of this dataset.

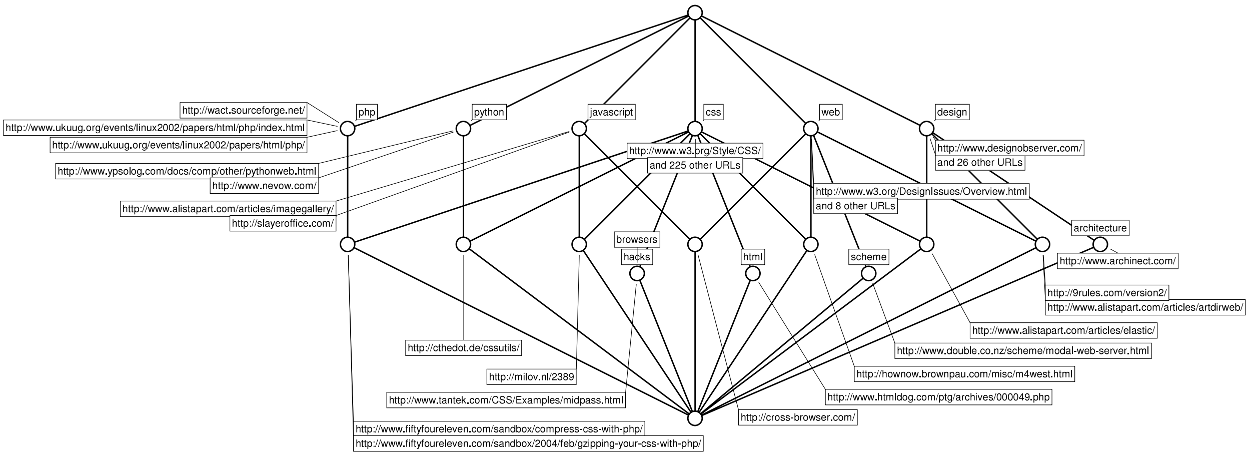

http://wact.sourceforge.net/
http://www.ukuug.org/events/linux2002/papers/html/php/index.html
http://www.ukuug.org/events/linux2002/papers/html/php/
php
python
javascript
css
web
design
http://www.ypsolog.com/docs/comp/other/pythonweb.html
http://www.nevow.com/
http://www.alistapart.com/articles/imagegallery/
http://slayeroffice.com/
http://www.w3.org/Style/CSS/
and 225 other URLs
http://www.designobserver.com/
and 26 other URLs
http://www.w3.org/DesignIssues/Overview.html
and 8 other URLs
architecture
browsers
hacks
html
scheme
http://www.archinect.com/
http://9rules.com/version2/
http://www.alistapart.com/articles/artdirweb/
http://www.alistapart.com/articles/elastic/
http://www.double.co.nz/scheme/modal-web-server.html
http://hownow.brownpau.com/misc/m4west.html
http://www.htmldog.com/ptg/archives/000049.php
http://cthedot.de/cssutils/
http://milov.nl/2389
http://www.tantek.com/CSS/Examples/midpass.html
http://cross-browser.com/
http://www.fiftyfoureleven.com/sandbox/compress-css-with-php/
http://www.fiftyfoureleven.com/sandbox/2004/feb/gzipping-your-css-with-php/

Fig. 6. Most relevant tags and resources related to *css*, *web*, and *design*.

quent tri-concepts. The sets of users, tags, and resources composing a tri-concept can be read off the three sides of the triangle. There, three Hasse diagrams display the three quasi-orders $\lesssim_1$, $\lesssim_2$, and $\lesssim_3$ as introduced in Section 2.4. The arrows guide the reader to the larger elements of each quasi-order. Each node in a hierarchy represents the set containing the labels attached to it plus all labels below. The empty nodes are not part of the quasi-order. They are just used to be able to place each label once only. In the IT components hierarchy on the right, for instance, the leftmost node represents the set $\{$*Computer Centres, Data Media Archives, Server Room, Technical Infrastructure Room*$\}$.

A node in the middle of the diagram represents then the tri-concept consisting of the three components it projects to. The left-most tri-concept, for instance, is the tri-concept ($\{$*Computer Centres, Server Room, Data Media Archives, Technical Infrastructure Room*$\}$, $\{$*Unauthorised entry into a building, Theft, Vandalism*$\}$, $\{$*Locked doors, Entry regulations and controls, Closed windows and doors*$\}$).

The three corners of the inner triangle are not realised (as there are no nodes on them). They stand for the trisets $(\emptyset, T, R)$, $(U, \emptyset, R)$, and $(U, T, \emptyset)$, resp., and are only realised if the first, second, or third threshold is set to zero.

The manual distinguishes seven classes of IT components, like *Networked Systems* and *Telecommunications*. The fact that all components that occur in the most frequent tri-concepts (i. e., the six components in the right-most hierarchy) are of the *Infrastructure* class indicates that this class was modeled with the highest level of detail. Surprisingly it surpasses more typical IT classes like the two mentioned above.

For having a closer look, we decrease the minimum thresholds, e. g., to $\tau_u = 3, \tau_t = \tau_r = 2$. The resulting tri-lattice is shown in Figure 8. It contains the previous five tri-concepts plus five new ones. We see that again the major contribution comes from the *Infrastructure* class, which is now extended by *Protective cabinets*. Additionally some more of the combinations of these components became frequent, indicated by the additional nodes in the right hierarchy.

With the decreasing thresholds, the lower left hierarchy grew as well. It contains now additionally four threats in two separated nodes. These nodes are not comparable (in terms of set inclusion) with the already existing nodes. The threats in the lower one of them – *Failure of internal supply networks, Fire* – are extending the list of threats against the *Infrastructure* class via the IT component *Building*. The upper hierarchy shows the safeguards against these new threats: *Hand-held fire extinguishers* and *Adapted segmentation of circuits*.

The threats in the uppermost isolated node of the lower left hierarchy – *Misuse of administrator rights [. . .]* and *Unauthorised acquisition [. . .]* – belong to a new class of IT components, as they are related to the new isolated node with three Windows operating systems in the right diagram. The safeguards against these threats are listed at the isolated node in the upper diagram. The IT components that seem to be endangered secondmost are thus – after IT infrastructure rooms – Windows operating systems. At least they are modeled with greater detail as other operating systems that show up when decreasing the thresholds further.

If we decrease the minimum thresholds further, we can discover this way more and more details, until we finally reach with $\tau_u = \tau_t = \tau_r = 0$ all $3,751$ tri-concepts of this dataset.

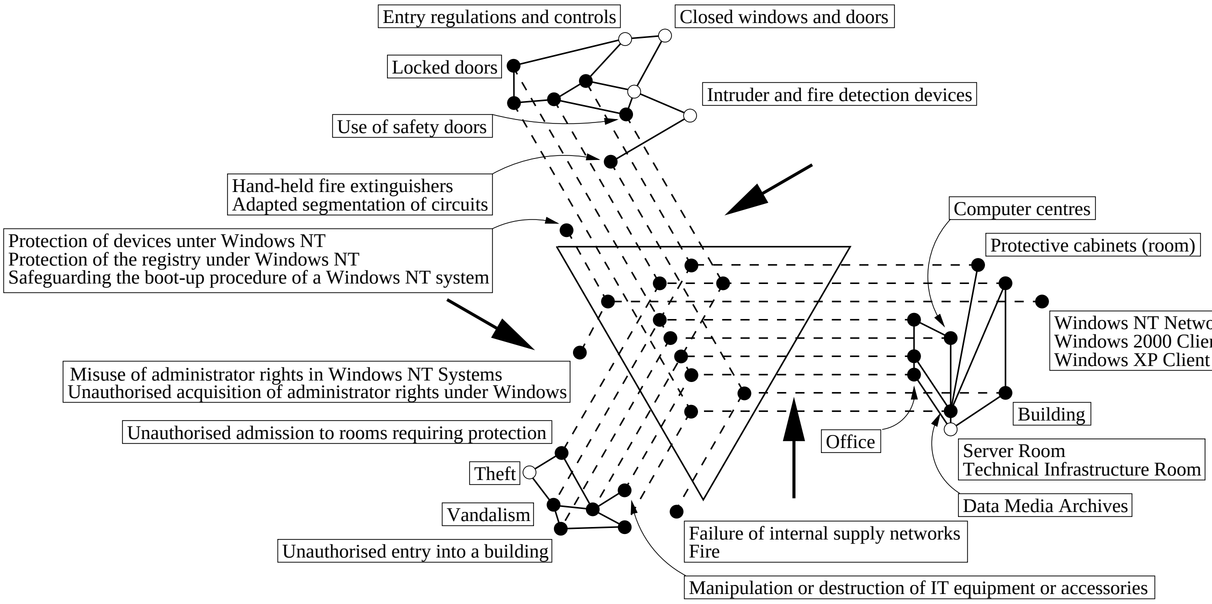

Entry regulations and controls
Closed windows and doors
Locked doors
Intruder and fire detection devices
Use of safety doors
Hand-held fire extinguishers
Adapted segmentation of circuits
Protection of devices unter Windows NT
Protection of the registry under Windows NT
Safeguarding the boot-up procedure of a Windows NT system
Computer centres
Protective cabinets (room)
Windows NT Network
Windows 2000 Client
Windows XP Client
Misuse of administrator rights in Windows NT Systems
Unauthorised acquisition of administrator rights under Windows
Unauthorised admission to rooms requiring protection
Building
Office
Server Room
Technical Infrastructure Room
Theft
Vandalism
Data Media Archives
Unauthorised entry into a building
Failure of internal supply networks
Fire
Manipulation or destruction of IT equipment or accessories

Fig. 8. All frequent tri-concepts of the IT Baseline Security Manual for $\tau_u = 3, \tau_t = \tau_r = 2$.

4.3. *Conceptual Analysis of the BibSonomy Publication Data*

We conclude the list of applications with another social resource sharing system. BibSonomy[22] is a social bookmark and publication management system that is run by the Knowledge & Data Engineering Group at the University of Kassel. Beside sharing bookmarks, BibSonomy enables the sharing of publication lists. It provides several output formats, including BIBTEX, formatted HTML, RTF, EndNote, XML, RDF, and RSS-Feeds. BibSonomy can thus be used for generating reference lists for scientific publications and annual reports, as well as for personal, group, and project homepages – supporting researchers in their everyday business. As a folksonomy offers the possibility to add more than one tag to a resource, documents can be found following different search paths, unlike books in a library which can only be placed in one physical location.

For our analysis we focused on the publication management part of BibSonomy. We first made a snapshot of BibSonomy's publication entries, including all publication posts made until November 23, 2006 at 13:30 CET. From the snapshot we excluded the publication posts from the DBLP computer science bibliography[23] since they are automatically inserted and all owned by one user and all tagged with the same tag (*dblp*). Therefore they do not provide meaningful information about shared conceptualisations. Similarly we excluded all tag assignments with the tag *imported* and all publication posts which exclusively have this tag, because it is automatically assigned to all posts which were added by one of the import functions. The resulting snapshot contains $|Y| = 44,944$ tag assignments built by $|U| = 262$ users, containing $|R| = 11,101$ publication references tagged with $|T| = 5,954$ distinct tags.[24]

The TRIAS algorithm needed 75 minutes on a 2 GHz AMD Opteron machine to compute all 13,992 tri-concepts of this dataset. Among those there are 12,659 tri-concepts which contain only one user, representing the individual conceptualisations of the users. (These could be used to present personal concept hierarchies by means of dyadic Hasse diagrams.) The remaining 1,333 tri-concepts thus all contain at least two users and therefore represent shared concepts. To further analyse these concepts, we next take a closer look on the tri-concepts which contain at least three users, two tags and two publication entries (i. e., with minimal support values $\tau_u = 3$, $\tau_t = 2$, $\tau_r = 2$). Each of these 21 tri-concepts expresses the fact that all of its users tagged all its publications with all its tags.

The diagram in Figure 9 on the following page shows the triadic concept lattice of all these 21 tri-concepts. The titles of the publications in the figure are substi-

[22] http://www.bibsonomy.org

[23] http://www.informatik.uni-trier.de/~ley/db/

[24] BibSonomy benchmark datasets are available for scientific purposes, see http://www.bibsonomy.org/faq.

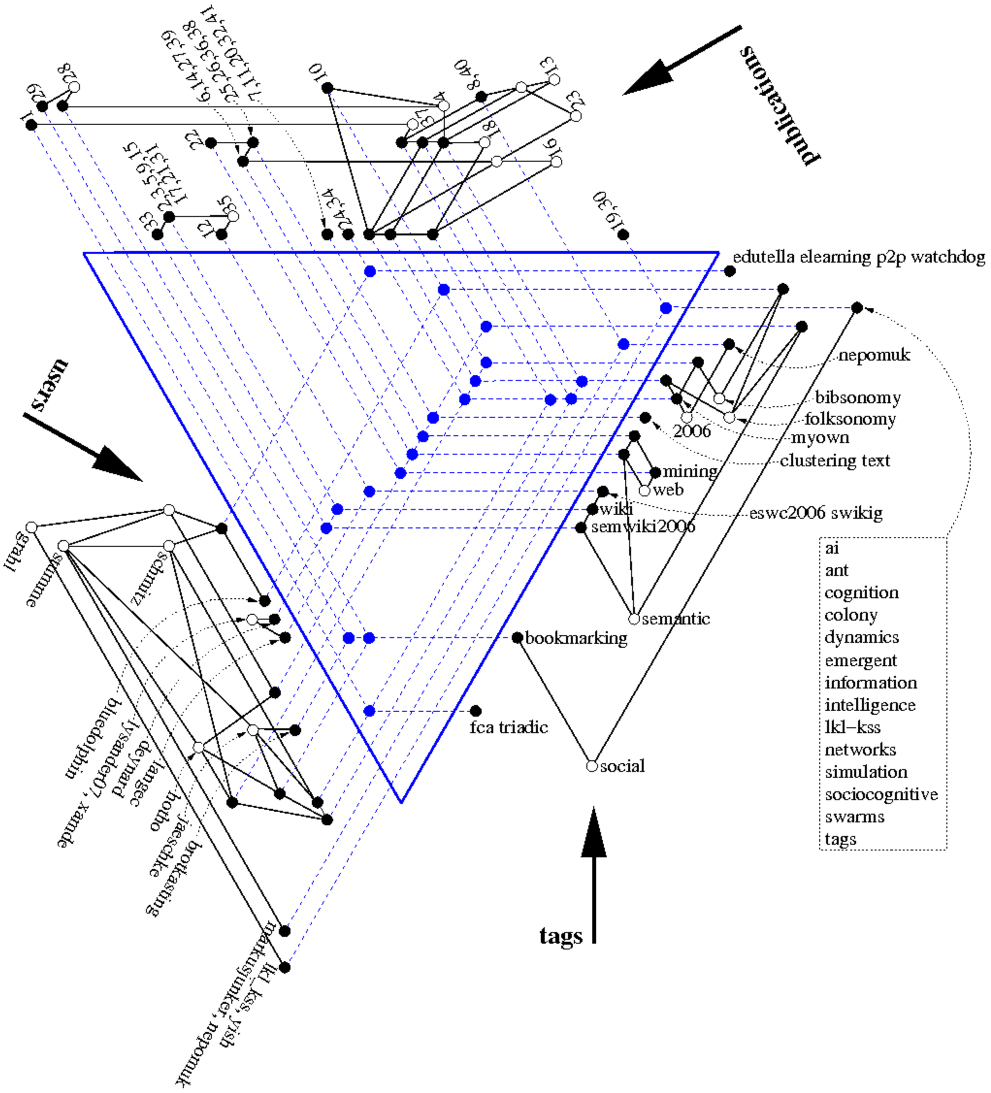


Fig. 9. All frequent tri-concepts of the BibSonomy publications for $\tau_u = 3$, $\tau_t = 2$, $\tau_r = 2$.

tuted by numbers for space reasons. The corresponding titles can be found in Table 2, the full bibliographic information was tagged in BibSonomy (after the evaluation) with the tag *trias_example*.[25] As in Figures 7 and 8, the 21 nodes in the center of the triangle represent the 21 frequent tri-concepts. The sets of users, tags, and resources composing a tri-concept can be read off the three sides of the triangle.

For instance, the lower-most node in the triangle represents the tri-concept consisting of the set $\{$*jaeschke, schmitz, stumme*$\}$ of users, the set $\{$*fca, triadic*$\}$ of tags, and the set $\{$*1, 37*$\}$ of resources. Similarly, the node in the user hierarchy labelled *brotkasting* represents not only the user *brotkasting* but also all users in nodes laying below this node. Therefore the users *jaeschke* and – since it is located below both *brotkasting* and *jaeschke* – *stumme* also belong to this node. Note that it fulfills thus the minimal support constraint $\tau_u = 3$ for the users.

A closer look on the tag hierarchy reveals the content of the most central publications in the system. The tag *social* co-occurs with most of the tags. On the level of generality defined by the τ thresholds, this tag is (together with the tags *ai* (meaning Artificial Intelligence), ..., *tags*) assigned by the users *lkl_kss* and *yish* to the publications *19* and *30*, (together with the tag *bookmarking*) by the users *hotho, jaeschke, stumme* to the publications *4* and *28*, and (again together with the tag *bookmarking*) by the users *brotkasting, jaeschke, stumme* to

[25] http://www.bibsonomy.org/group/kde/trias_example?items=50

Table 2
The mapping of publication IDs to publication titles.

ID	Publication Title
1	A Finite-State Model for On-Line Analytical Processing in Triadic Contexts
2	Annotation and Navigation in Semantic Wikis
3	A Semantic Wiki for Mathematical Knowledge Management
4	BibSonomy: A Social Bookmark and Publication Sharing System
5	Bringing the "Wiki-Way" to the Semantic Web with Rhizome
6	Building and Using the Semantic Web
7	Conceptual Clustering of Text Clusters
8	Content Aggregation on Knowledge Bases using Graph Clustering
9	Creating and using Semantic Web information with Makna
10	Emergent Semantics in BibSonomy
11	Explaining Text Clustering Results using Semantic Structures
12	Harvesting Wiki Consensus - Using Wikipedia Entries as Ontology Elements
13	Information Retrieval in Folksonomies: Search and Ranking
14	KAON – Towards a Large Scale Semantic Web
15	Kaukolu: Hub of the Semantic Corporate Intranet
16	Kollaboratives Wissensmanagement
17	Learning with Semantic Wikis
18	Mining Association Rules in Folksonomies
19	On Self-Regulated Swarms, Societal Memory, Speed and Dynamics
20	Ontologies improve text document clustering
21	Proceedings of the First Workshop on Semantic Wikis – From Wiki To Semantics
22	Proc. of the European Web Mining Forum 2005
23	Semantic Network Analysis of Ontologies
24	Semantic Resource Management for the Web: An ELearning Application.
25	Semantic Web Mining
26	Semantic Web Mining and the Representation, Analysis, and Evolution of Web Space
27	Semantic Web Mining for Building Information Portals (Position Paper)
28	Social Bookmarking Tools (I): A General Review
29	Social Bookmarking Tools (II). A Case Study – Connotea
30	Social Cognitive Maps, Swarm Collective Perception and Distributed Search on Dynamic Landscapes
31	SweetWiki : Semantic Web Enabled Technologies in Wiki
32	Text Clustering Based on Background Knowledge
33	The ABCDE Format Enabling Semantic Conference Proceedings
34	The Courseware Watchdog: an Ontology-based tool for Finding and Organizing Learning Material
35	Towards a Wiki Interchange Format (WIF) – Opening Semantic Wiki Content and Metadata
36	Towards Semantic Web Mining
37	TRIAS - An Algorithm for Mining Iceberg Tri-Lattices
38	Usage Mining for and on the Semantic Web (Book)
39	Usage Mining for and on the Semantic Web (Workshop)
40	Wege zur Entdeckung von Communities in Folksonomies
41	WordNet improves text document clustering

the publications *28* and *29*. The tags as well as the corresponding publication titles indicate that the two sets of users {*lkl_kss, yish*} and {*brotkasting, hotho, jaeschke, stumme*} form two sub-communities which both work on social phenomena in the Web 2.0, but from different perspectives.

A second topical group is spanned by the tag *semantic*, which occurs in three different contexts. The first is on semantic wikis, which correlates with the isolated group {2, …, 31, 12, 33, 35} of publications, and the – equally isolated – group {*lysander07, xamde, deynard, langec*} of users. The second context in which the tag *semantic* occurs is on Semantic Web Mining, being connected by the users {*grahl, hotho, stumme*} with different combinations of the additional tags *web* and *mining* to the publications *6, 14, 22, 25, 26, 27, 36, 38*, and *39*. These assignments are witnessed by the three tri-concepts in the very middle of the diagram. On the same line are two more tri-concepts, which indicate that these users are also interested in *text clustering* and in *nepomuk* (the acronym of a European project). The third context in which the tag *semantic* occurs is in combination with *folksonomy*. This provides a link to the group {*2006, myown, nepomuk, bibsonomy, folksonomy*} of tags which are used by the authors of this paper and by other researchers from the European project Nepomuk [26] to describe their own publications.

Two more topical groups can be found at the top and bottom of the tags quasi-order. One is related to a Peer-to-Peer eLearning application, and the other to triadic Formal Concept Analysis.

Since the diagram shows the frequent tri-concepts only, we cannot deduce from the absence of a relationship that two objects are not related at all. When the thresholds are lowered, links between the topical islands discussed above will show up.

Concluding we see that iceberg tri-concept lattices provide a means for exploring the flat structure of folksonomies – just as iceberg concept lattices in the dyadic case. One may be surprised by the relatively small numbers of frequent tri-concepts. This shows – just as in the dyadic case – that the closeness condition provides a strong criterion for pruning the result set without loss of information.

5. Conclusion and Outlook

In this paper, we have presented a formal definition of the problem of mining all frequent tri-concepts, and

[26] http://nepomuk.semanticdesktop.org/

have presented an efficient algorithm for its solution. We have empirically studied the performance of the algorithm, and have presented two real-world applications.

This work opens a series of challenging tasks for future research. (i) An important issue for the presentation of the results is the development of a visualisation metaphor to display small, medium, and large (frequent) concept tri-lattices, and to provide efficient means for navigating and browsing them. (ii) Continuing the research on association rules, a natural next step would be the development of 'triadic association rules', combining thus the developments in triadic FCA and association rule mining. (iii) The natural next step after discovering shared conceptualisations would be to formalize them in an ontology. We plan thus to extend our approach to an ontology learning application. (iv) These steps together lead to a development which is currently undertaken in the European project 'Nepomuk – The Social Semantic Desktop': the exploitation of TRIAS for discovering and managing communities in a peer to peer network of semantic desktops.

Acknowledgement. Part of this research was funded by the EU in the Nepomuk project (FP6-027705).

References

[1] R. Agrawal, T. Imielinski, A. Swami, Mining association rules between sets of items in large databases, in: Proc. of SIGMOD 1993, ACM Press, 1993.

[2] R. Agrawal, R. Srikant, Fast algorithms for mining association rules in large databases, in: Proceedings of the 20th international conference on Very Large Data Bases (VLDB'94), Morgan Kaufmann, 1994.

[3] R. Agrawal, R. Srikant, Mining sequential patterns, in: Proceedings of the 11th International Conference on Data Engineering (ICDE'95), IEEE Computer Society Press, 1995.

[4] A. Arnauld, P. Nicole, La logique ou l'art de penser — contenant, outre les règles communes, plusieurs observations nouvelles, propres à former le jugement, Ch. Saveux, 1668.

[5] Y. Bastide, R. Taouil, N. Pasquier, G. Stumme, L. Lakhal, Mining frequent patterns with counting inference., SIGKDD Explorations, Special Issue on Scalable Algorithms 2 (2) (2000) 71–80.

[6] R. J. Bayardo, Efficiently mining long patterns from databases, in: Proceedings of the 1998 ACM SIGMOD international conference on Management of Data (SIGMOD'98), ACM Press, 1998.

[7] K. Biedermann, How triadic diagrams represent conceptual structures, in: D. Lukose, H. S. Delugach, M. Keeler, L. Searle, J. F. Sowa (eds.), Conceptual Structures: Fulfilling Peirce's Dream, No. 1257 in LNAI, Springer, Heidelberg, 1997.

[8] K. Biedermann, Triadic Galois connections., in: K. Denecke, O. Lüders (eds.), General algebra and applications in discrete mathematics, Shaker Verlag, Aachen, 1997.

[9] K. Biedermann, Powerset trilattices, in: M. Mugnier, M. Chein (eds.), Conceptual Structures: Theory, Tools and Applications, vol. 1453 of Lecture Notes in Computer Science, Springer, 1998.

[10] J.-F. Boulicaut, A. Bykowski, C. Rigotti, Approximation of frequency queris by means of free-sets, in: Principles of Data Mining and Knowledge Discovery, 2000.
URL http://citeseer.ist.psu.edu/boulicaut00approximation.html

[11] S. Brin, L. Page, The Anatomy of a Large-Scale Hypertextual Web Search Engine, Computer Networks and ISDN Systems 30 (1-7) (1998) 107–117.

[12] P. Buitelaar, P. Cimiano, B. Magnini (eds.), Ontology Learning from Text: Methods, Evaluation and Applications, vol. 123 of Frontiers in Artificial Intelligence, IOS Press, 2005.

[13] A. Bykowski, C. Rigotti, A condensed representation to find frequent patterns., in: PODS, 2001.
URL http://dblp.uni-trier.de/db/conf/pods/pods2001.html#BykowskiR01

[14] T. Calders, B. Goethals, Mining all non-derivable frequent itemsets., in: PKDD, 2002.
URL http://dblp.uni-trier.de/db/conf/pkdd/pkdd2002.html#CaldersG02

[15] C. Carpineto, G. Romano, GALOIS: An order-theoretic approach to conceptual clustering., in: Machine Learning *Proc. ICML 1993*, Morgan Kaufmann Publications, 1993.

[16] C. Carpineto, G. Romano, Concept Data Analysis, Wiley, 2004.

[17] C. Cattuto, V. Loreto, L. Pietronero, Collaborative tagging and semiotic dynamics, arXiv:cs.CY/0605015 (May 2006).
URL http://arxiv.org/abs/cs/0605015

[18] P. Cimiano, Ontology Learning and Population from Text: Algorithms, Evaluation and Applications, Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2006.
URL http://portal.acm.org/citation.cfm?id=1177318

[19] F. Dau, R. Wille, On the modal unterstanding of triadic contexts., in: R. Decker, W. Gaul (eds.), Classification and Information Processing at the Turn of the Millenium, Proc. Gesellschaft für Klassifikation, 2001.

[20] H. Dicky, C. Dony, M. Huchard, T. Libourel, On automatic class insertion with overloading., in: OOPSLA 1996, 1996.

[21] M. Dubinko, R. Kumar, J. Magnani, J. Novak, P. Raghavan, A. Tomkins, Visualizing tags over time, in: Proc. of the 15th International WWW Conference, 2006.

[22] B. Ganter, Algorithmen zur formalen Begriffsanalyse, in: B. Ganter, R. Wille, K. E. Wolff (eds.), Beiträge zur Begriffsanalyse, B.I.–Wissenschaftsverlag, Mannheim, 1987, pp. 241–254.

[23] B. Ganter, S. A. Obiedkov, Implications in triadic contexts, in: Conceptual Structures at Work: 12th International Conference on Conceptual Structures, vol. 3127 of Lecture Notes in Computer Science, Springer, 2004.

[24] B. Ganter, G. Stumme, R. Wille (eds.), Formal Concept Analysis – Foundations and Applications, vol. 3626 of LNAI, Springer, Heidelberg, 2005 (2005).

[25] B. Ganter, R. Wille, Formal Concept Analysis: Mathematical Foundations, Springer, 1999.

[26] German Federal Office for Information Security, IT Baseline Protection Manual (October 2003).
URL http://www.bsi.de/gshb/

[27] R. Godin, H. Mili, G. Mineau, R. Missaoui, A. Arfi, T. Chau, Design of class hierarchies based on concept (galois) lattices., TAPOS 4 (2) (1998) 117–134.

[28] S. Golder, B. A. Huberman, The structure of collaborative tagging systems, Tech. rep., Information Dynamics Lab, HP

Labs (Aug 2005).
URL http://arxiv.org/abs/cs.DL/0508082

[29] T. R. Gruber, Towards principles for the design of ontologies used for knowledge sharing, in: N. Guarino, R. Poli (eds.), Formal Ontology in Conceptual Analysis and Knowledge Representation, Kluwer Academic Publishers, Deventer, The Netherlands, 1993.

[30] H. Halpin, V. Robu, H. Shepard, The dynamics and semantics of collaborative tagging, in: Proceedings of the 1st Semantic Authoring and Annotation Workshop (SAAW'06), 2006.
URL http://sunsite.informatik.rwth-aachen.de/Publications/CEUR-WS/Vol-209/saaw06-full01-halpin.pdf

[31] T. Hammond, T. Hannay, B. Lund, J. Scott, Social Bookmarking Tools (I): A General Review, D-Lib Magazine 11 (4).

[32] P. Heymann, H. Garcia-Molina, Collaborative creation of communal hierarchical taxonomies in social tagging systems, Tech. Rep. 2006-10, Computer Science Department (April 2006).
URL http://dbpubs.stanford.edu:8090/pub/2006-10

[33] A. Hotho, R. Jäschke, C. Schmitz, G. Stumme, BibSonomy: A social bookmark and publication sharing system, in: Proceedings of the Conceptual Structures Tool Interoperability Workshop at the 14th International Conference on Conceptual Structures, 2006.

[34] A. Hotho, R. Jäschke, C. Schmitz, G. Stumme, Information retrieval in folksonomies: Search and ranking, in: Y. Sure, J. Domingue (eds.), The Semantic Web: Research and Applications, vol. 4011 of LNAI, Springer, Heidelberg, 2006.

[35] R. Jäschke, A. Hotho, C. Schmitz, B. Ganter, G. Stumme, Trias - an algorithm for mining iceberg tri-lattices, in: Proceedings of the 6th IEEE International Conference on Data Mining (ICDM 06), IEEE Computer Society, Hong Kong, 2006.

[36] R. Jäschke, A. Hotho, C. Schmitz, G. Stumme, Analysis of the publication sharing behaviour in BibSonomy, in: U. Priss, S. Polovina, R. Hill (eds.), Proceedings of the 15th International Conference on Conceptual Structures (ICCS 2007), vol. 4604 of Lecture Notes in Artificial Intelligence, Springer-Verlag, Berlin, Heidelberg, 2007.

[37] M. Kamber, J. Han, Y. Chiang, Metarule-guided mining of multi-dimensional association rules using data cubes., in: Proc. of the 3rd KDD Int'l Conf., 1997.

[38] S. Krolak-Schwerdt, P. Orlik, B. Ganter, TRIPAT: a model for analyzing three-mode binary data, in: H. H. Bock, W. Lenski, M. M. Richter (eds.), Studies in Classification, Data Analysis, and Knowledge Organization, vol. 4 of Information systems and data analysis, Springer, Berlin, 1994, pp. 298–307.

[39] R. Lambiotte, M. Ausloos, Collaborative tagging as a tripartite network, arXiv:cs.DS/0512090 (Dec 2005).
URL http://arxiv.org/abs/cs.DS/0512090

[40] F. Lehmann, R. Wille, A triadic approach to formal concept analysis, in: G. Ellis, R. Levinson, W. Rich, J. F. Sowa (eds.), Conceptual structures: applications, implementation and theory, vol. 954 of Lecture Notes in Artificial Intelligence, Springer Verlag, 1995.

[41] B. Lent, R. Agrawal, R. Srikant, Discovering trends in text databases, in: Proceedings of the 3rd international conference on Knowledge Discovery and Data mining (KDD'97), AAAI Press, 1997.

[42] D. Lin, M. Kedem, A new algorithm for discovering the maximum frequent set., in: Proceedings of the 6th Int'l Conf.on Extending Database Technology (EDBT), 1998.

[43] B. Lund, T. Hammond, M. Flack, T. Hannay, Social Bookmarking Tools (II): A Case Study - Connotea, D-Lib Magazine 11 (4).

[44] A. Maedche, S. Staab, Ontology learning for the semantic web, IEEE Intelligent Systems 16 (2) (2001) 72–79.

[45] H. Mannila, Methods and problems in data mining, in: Proceedings of the 6th biennial International Conference on Database Theory (ICDT'97), Lecture Notes in Computer Science, Vol. 1186, Springer-Verlag, 1997.

[46] A. Mathes, Folksonomies – Cooperative Classification and Communication Through Shared Metadata (December 2004).
URL http://www.adammathes.com/academic/computer-mediated-communication/folksonomies.html

[47] P. Mika, Ontologies Are Us: A Unified Model of Social Networks and Semantics, in: Y. Gil, E. Motta, V. R. Benjamins, M. A. Musen (eds.), ISWC 2005, vol. 3729 of LNCS, Springer-Verlag, 2005.

[48] G. Mineau, G., R. Godin, Automatic structuring of knowledge bases by conceptual clustering., IEEE Transactions on Knowledge and Data Engineering 7 (5) (1985) 824–829.

[49] M. Missikoff, M. Scholl, An algorithm for insertion into a lattice: application to type classification., in: Proc. 3rd Intl. Conf. FODO 1989, vol. 367 of LNCS, Springer, Heidelberg, 1989.

[50] N. Pasquier, Y. Bastide, R. Taouil, L. Lakhal, Closed set based discovery of small covers for association rules, in: Actes des 15èmes journées Bases de Données Avancées (BDA'99), 1999.

[51] N. Pasquier, Y. Bastide, R. Taouil, L. Lakhal, Discovering frequent closed itemsets for association rules, in: Proceedings of the 7th biennial International Conference on Database Theory (ICDT'99), Lecture Notes in Computer Science, Vol. 1540, Springer-Verlag, 1999.

[52] N. Pasquier, R. Taouil, Y. Bastide, G. Stumme, L. Lakhal, Generating a condensed representation for association rules, J. Intelligent Information Systems (JIIS) 24 (1) (2005) 29–60.

[53] J. Pei, J. Han, R. Mao, Closet: An efficient algorithm for mining frequent closed itemsets., in: ACM SIGMOD Workshop on Research Issues in Data Mining and Knowledge Discovery, 2000.

[54] C. S. Peirce, Collected Papers, Harvard Universit Press, Cambridge, 1931–1935.

[55] F. Rioult, Extraction de connaissances dans les bases de donnees comportant des valeurs manquantes ou un grand nombre d'attributs, Ph.D. thesis, Université de Caen Basse-Normandie (2005).

[56] I. Schmitt, G. Saake, Merging inheritance hierarchies for database integration., in: Proc. 3rd IFCIS Intl. Conf. on Cooperative Information Systems, New York City, Nework USA, 1996.

[57] C. Schmitz, A. Hotho, R. Jäschke, G. Stumme, Mining association rules in folksonomies, in: V. Batagelj, H.-H. Bock, A. Ferligoj, A. Žiberna (eds.), Data Science and Classification: Proc. of the 10th IFCS Conf., Studies in Classification, Data Analysis, and Knowledge Organization, Springer, Berlin, Heidelberg, 2006.

[58] P. Schmitz, Inducing ontology from flickr tags., in: Collaborative Web Tagging Workshop at WWW2006, Edinburgh, Scotland, 2006.
URL http://www.ibiblio.org/www_tagging/2006/22.pdf

[59] C. Silverstein, S. Brin, R. Motwani, Beyond market baskets : Generalizing association rules to dependence rules, Data Mining and Knowledge Discovery 2 (1) (1998) 39–68.

[60] E. Speller, Library student journal: Collaborative tagging, folksonomies, distributed classification or ethnoclassification: a literature review. (February 2007).
URL http://informatics.buffalo.edu/org/lsj/articles/speller_2007_2_collaborative.php

[61] S. Staab, S. Santini, F. Nack, L. Steels, A. Maedche, Emergent semantics, Intelligent Systems, IEEE [see also IEEE Expert] 17 (1) (2002) 78–86.
URL
http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=988491

[62] L. Steels, The origins of ontologies and communication conventions in multi-agent systems, Autonomous Agents and Multi-Agent Systems 1 (2) (1998) 169–194.
URL http://www.isrl.uiuc.edu/~amag/langev/paper/steels98theOrigins.html

[63] S. Strahringer, R. Wille, Conceptual clustering via convex-ordinal structures., in: O. Opitz, B. Lausen, R. Klar (eds.), *Information and Classification*, Springer, Berlin–Heidelberg, 1993.

[64] G. Stumme, Conceptual knowledge discovery with frequent concept lattices, FB4-Preprint 2043, TU Darmstadt (1999).
URL http://www.kde.cs.uni-kassel.de/stumme/papers/1999/P2043.pdf

[65] G. Stumme, Begriffliche Wissensverarbeitung–Methoden und Anwendungen, Springer, Heidelberg, 2000.

[66] G. Stumme, Off to new shores – conceptual knowledge discovery and processing, Intl. J. Human-Comuter Studies (IJHCS) 59 (3) (2003) 287–325.

[67] G. Stumme, A finite state model for on-line analytical processing in triadic contexts., in: B. Ganter, R. Godin (eds.), Proceedings of the 3rd International Conference on Formal Concept Analysis, vol. 3403 of Lecture Notes in Computer Science, Springer, 2005.

[68] G. Stumme, R. Taouil, Y. Bastide, N. Pasqier, L. Lakhal, Computing iceberg concept lattices with titanic., J. on Knowledge and Data Engineering 42 (2) (2002) 189–222.

[69] G. Stumme, R. Taouil, Y. Bastide, N. Pasquier, L. Lakhal, Intelligent structuring and reducing of association rules with formal concept analysis., in: F. Baader, G. Brewker, T. Eiter (eds.), KI 2001: Advances in Artificial Intelligence, vol. 2174 of LNAI, Springer, Heidelberg, 2001.

[70] H. Söll, Begriffliche Analyse triadischer Daten: Das IT-Grundschutzhandbuch des Bundesamts für Sicherheit in der Informationstechnik, Diploma thesis, FB Mathematik, TU Darmstadt, Darmstadt (April 1998).

[71] R. Taouil, Algorithmique du treillis des fermés : application à l'analyse formelle de concepts et aux bases de données, Ph.D. thesis, Université de Clermont-Ferrand II (2000).

[72] K. Waiyamai, R. Taouil, L. Lakhal, Towards an object database approach for managing concept lattices., in: Proc. 16th Intl. Conf. on Conceptual Modeling, vol. 1331 of LNCS, Springer, Heidelberg, 1997.

[73] T. V. Wal, Folksonomy (2007).
URL http://vanderwal.net/folksonomy.html

[74] R. Wille, Restructuring lattice theory: An approach based on hierarchies of concepts, in: I. Rival (ed.), Ordered Sets, Reidel, Dordrecht-Boston, 1982.

[75] R. Wille, The basic theorem of triadic concept analysis., Order 12 (1995) 149–158.

[76] R. Wille, M. Zickwolff, Grundlagen einer triadischen Begriffsanalyse, in: G. Stumme, R. Wille (eds.), Begriffliche Wissensverarbeitung. Methoden und Anwendungen, Springer-Verlag, Berlin-Heidelberg, 2000.

[77] A. Yahia, L. Lakhal, J. P. Bordat, R. Cicchetti, io2: An algorithmic method for building inheritance graphs in object database design., in: Proc. 15th Intl. Conf. on Conceptual Modeling, vol. 1157 of LNCS, Springer, Heidelberg, 1996.

[78] M. J. Zaki, C.-J. Hsiao, Charm: An efficient algorithm for closed association rule mining. technical report 99–10, Tech. rep., Computer Science Dept., Rensselaer Polytechnic (October 1999).

[79] M. J. Zaki, S. Parthasarathy, M. Ogihara, W. Li, New algorithms for fast discovery of association rules, in: Proceedings of the 3rd international conference on Knowledge Discovery and Data mining (KDD'97), AAAI Press, 1997.