

Visualizing and Interacting with Concept Hierarchies

Michel Crampes*, Michel Plantié**

*Laboratoire LGI2P, Ecole des Mines d'Ales,
site EERIE, Parc George Besse, 30000 Nîmes, France*

Abstract

Concept Hierarchies and Formal Concept Analysis are theoretically well grounded and largely experimented methods. They rely on line diagrams called Galois lattices for visualizing and analysing object-attribute sets. Galois lattices are visually seducing and conceptually rich for experts. However they present important drawbacks due to their concept oriented overall structure: analysing what they show is difficult for non experts, navigation is cumbersome, interaction is poor, and scalability is a deep bottleneck for visual interpretation even for experts.

In this paper we introduce semantic probes as a means to overcome many of these problems and extend usability and application possibilities of traditional FCA visualization methods. Semantic probes are visual user centred objects which extract and organize reduced Galois sub-hierarchies. They are simpler, clearer, and they provide a better navigation support through a rich set of interaction possibilities. Since probe driven sub-hierarchies are limited to users' focus, scalability is under control and interpretation is facilitated. After some successful experiments, several applications are being developed with the remaining problem of finding a compromise between simplicity and conceptual expressivity.

Keywords: Information visualization, Formal Concept Analysis, Galois sub-hierarchy

1. Introduction

Visualization and interaction are two major supports for searching and analysing object sets. A lot of methods and tools have been proposed to organize, represent and display objects for providing users with immediate access to subsets of objects according to users' intention or some internal logic. But when the size

*Principal Corresponding author

**Corresponding author

Email addresses: michel.crampes@mines-ales.fr (Michel Crampes),
michel.plantie@mines-ales.fr (Michel Plantié)

URL: <http://www.lgi2p.ema.fr> (Michel Crampes)

of sets is important, most of visual displays become difficult to interpret and interaction turns into complex manipulation. Scalability is a serious bottleneck. This is a paradox because visualization loses efficiency with complex sets where it is expected to provide solutions for managing complexity (Chen [7]). As a result most popular solutions for searching objects of data collections present query results as lists of items, such as Google, or grids of objects, such as Flickr or Facebook with photos. It seems that sophisticated visualization solutions are for experts and straightforward visualisation is for non-expert audiences. This paper tackles the difficult problem of turning an expert visual display to an interesting simple application for novices.

Concept Hierarchies (CH) and Formal Concept Analysis (FCA) (Ganter B. and Wille R. [19]) are examples of such methods which are particularly well grounded on a theoretical point of view, largely experimented in numerous lab applications, and, thanks to Galois lattices and line diagrams called Hasse diagrams, particularly adapted to searching and analysing sets of objects endowed with attributes. However Galois lattices still fall short of managing visual complexity even for medium CHs (Ganter B. and Wille R. [19], Wille [52], Roth et al. [42]): “Representing concept lattices constructed from large contexts often results in heavy, complex diagrams that can be impractical to handle and, eventually, to make sense of” (Kuznetsov [29]).

In this paper we address scalability and expressivity of concept hierarchies for non-expert audiences. A new visualization and interaction paradigm is presented with its key concept: user centred Semantic Probes. Visual results are compared to a traditional Galois lattice and to the proposed Galois Lattice reduction methods. It has been tested in controlled experiments on a benchmark of 127 objects tagged with 245 attributes and on real data, photo albums extracted from Facebook. It has raised the interest of several industrials for which different applications are being developed.

2. RELATED WORKS AND MOTIVATION

Visualizing sets of entities and their properties or relations, such as biological data, multimedia objects or social activity, is an increasingly important issue. The goal is to visually elicit known or hidden organization that mere lists cannot reveal with the intention of extracting specific knowledge or particular objects. A myriad of solutions have been proposed which depend upon the designers’ intention and the type of entities to visualize such as object-object relations (graphs) (Battista G. et al. [3], Herman et al. [24]) or multivariate data (Buja and Swayne [5], Kohonen et al. [28], Platt [39]). In these fields, scalability is often a difficult problem. For example, in graph visualization it is necessary to visualize hundreds and even thousands of usually entangled links between objects. Different strategies have been proposed such as clustering (Fortunato [17], Noack [35], Holten [25], Noack and Lewerentz [36]), interaction techniques (panning, zooming, focus+context, filtering, animation Herman et al. [24], Shneiderman [44], Dwyer et al. [13], Lamping et al. [32]), dynamic

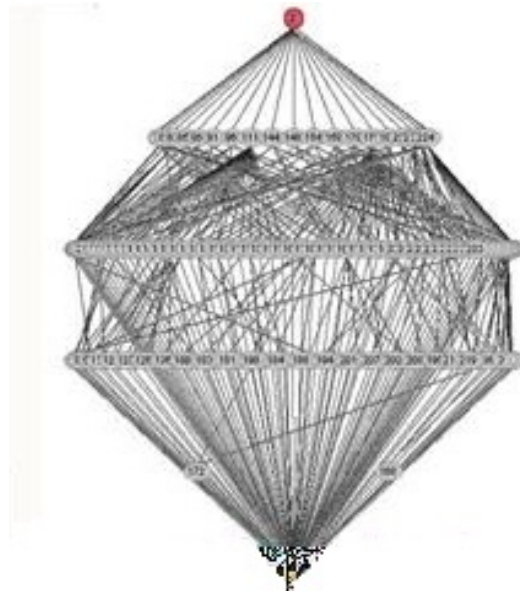


Figure 1: A semantic probe driven Galois sub-hierarchy lattice

local views centred on user’s interest (Alani [1], Van Ham and Perer [50]), or multi-views for linking different complementary views (Streit et al. [47]).

In this paper we consider a different kind of entities, object-attribute databases, which can be found in many areas where objects of some type are tagged with attributes of a different type (e.g. photos with tags, genes linked to their properties, etc.). The problem of displaying and exploring their structure shares with graphs the same difficulty of scalability. But it is even more challenging because relations between objects are linked to attribute ownership which should consequently be visually revealed.

Formally object-attribute sets are equivalent to bipartite graphs. They are graphs whose node set can be partitioned into two disjoint sub-sets, and edges only link nodes from a sub-set to the other. In our case, attribute set and object set are the two sub-sets and links between two nodes represent attribute ownership. Many techniques for visualising bipartite graphs have been proposed mostly focusing upon avoiding as much as possible edge crossings for better interpretation. The most notorious is the two-layer layout with its barycentre method for minimizing edge crossings (Battista G. et al. [3]). However for big bipartite graphs the visual result is still intricate. Recent solutions make use of interaction techniques such as focus+context, Fish Eye and information hiding to handle big data sets (Schulz et al. [43]). The authors argue that the resulting display is usable for experts, but it is far from being simple and straightforward for novices.

Other methods try to catch object-attribute data through node clustering (Fluit et al. [16]) and hypergraphs. In the last case nodes of one of the two

sets become hyper-edges containing the corresponding nodes of the other set. With this respect matrices in (Riche and Fekete [41]) or Euler diagram boxes in (Riche and Dwyer [40]) are used to build hyper-edges. Node duplication is analyzed in both papers to represent hyper-edge intersections. However object duplication does not prove to be visually the most appropriate solution for users in both papers. Testers favour what the authors call Compact Rectangular Euler Diagrams (Riche and Dwyer [40]) where objects have unique visible identities. In this respect underground maps are an interesting hypergraph metaphor. Lines represent hyper-edges and stations stand for nodes which may belong to several lines (Brandes et al. [4]). This original technique still needs experimentation with users to prove its interest.

But in all these above visualization strategies, examples are based on very small data sets and even under that limitation, interpretation is still difficult. Whatever the method, drawing hypergraphs and Euler diagrams is particularly cumbersome on limited data sets and even more on real applications which require the display of large databases.

The most common and formally fruitful approach for object-attribute data visualization is based on Galois lattices (Ganter B. and Wille R. [19], Eklund and Villerd [14]) which are visualized through layered graphs called Hasse diagrams; an example is given in Figure 1. Each node is identified by a subset of objects and a subset of attributes; edges link nodes according to a partial order relation. The detailed state of the art in this domain will be presented after introducing FCA basics in the next section.

The four visualization methods (two-layer layouts, matrices, Euler diagrams and Hasse diagrams) have been deeply studied by researchers with many variations. But usability is still questionable because we see few everyday applications of these methods. Conversely when searching object-attribute databases most common applications display query results as lists (i.e. Google) or grids (i.e. Flickr or Facebook) of objects. Objects may be ordered according to their proximity with the query, but little information is given about the semantics, the ordering, or the structure of the selection. Why are such straightforward visualization methods preferred to more semantically rich approaches? Informal discussions with several non-expert users reveal that the main qualities of visualization applications should be simplicity of interpretation and manipulation. Objects should be easy to identify with their attributes and links should be avoided because they require an effort of concentration. Moreover only contextualised useful information should be displayed. Consequently it is not surprising that all technically sophisticated method, whatever their scientific interest, fall short of being popular in most applications. They could be preferred to traditional lists or grids if 1) their complexity was limited and 2) they could provide new services that balance some still necessary efforts of interpretation on behalf of users.

In this paper we present a new Galois lattice visualization method which tackles this double challenge. For the sake of simplicity it turns Hasse diagrams into object grids without loss of expressivity. The objective is to enrich the popular grid approach with the Hasse diagram power of expression. We still

display Galois lattices as Hasse diagrams, but objects (not concepts) are visible and links are not shown to users. Moreover, this approach provides new interesting services which may enhance its interest for users: it is possible to index objects with objects, and it is possible to spot structures that may be of utmost interest for some applications such as team organisation or document diffusion.

In (Crampes et al. [9]) we already introduced a first version of this visualization method which showed Hasse diagrams as layered grid displays. The goal was to index objects with other already indexed objects which were displayed on the Hasse diagram. But we still made use of links and the display was not contextualized, i.e. the Hasse diagram was incrementally built with all indexed objects each time a new object or group of objects was indexed using other objects. As a result our approach had two limits with regard to users' expectations. Links are still a hurdle for interpretation and since all objects were displayed on the Hasse diagram, scalability was questionable. Our method, like traditional Hasse diagrams, faced the unavoidable problem of complexity and scalability. But it had the quality of providing a good support for fast indexing. The method we present in this paper introduces important improvements which overcome the two problems described above. Contextualisation is obtained through the presence of a virtual probe which represents the user's intention. It is a visual object which by its own presence extracts and organises a subset of objects according to their attributes. It is also possible to load the probe with an object to extract similar objects which are displayed according to a grid based Hasse diagram without links.

The idea of using virtual probes or magnets for visualizing and/or retrieving information is not new. (Miller and Gavosto [34]) introduces an immersive visualization probe for exploring n-dimensional spaces when some scalar function is available depending on n variables. In this solution the probe is not a visible object but a user's 3D view point from which it is possible to project the other dimensions on 3D walls. (de Leeuw and van Wijk [10]) introduce visual probes for the visualization of three-dimensional fluid flow fields. In both papers probes are used to reveal physical phenomena with continuous parameters. In (Spritzer and Freitas [46]), probes are used for extracting sub-graphs from graphs with limited visual capacities. Magnets which play the same role as probes are used in (Yi et al. [54]) to search for multivariate data. Each magnet represents an attribute (possibly valuated) and two or more magnets compete on a 2D screen for attracting dots representing multivariate data. Without being aware of these results we explored such a metaphor a few years ago with a very similar display for building concept maps (Crampes et al. [8]) We then faced a lot of limits among which some are reported by the authors in (Yi et al. [54]) such as the expressivity of the metaphor and the difficulty of interpretation when there are two magnets, i.e. two attributes. It is worse with three or four magnets and the display is meaningless beyond four magnets. We then explored one fixed probe with potentially multiple attributes and Galois lattices to propose expressive hierarchical displays. This strategy turned out fruitful for creating dynamic expressive displays. In this paper we introduce semantic probes in Galois lattices which are complex semantic structures for experts, to

extract Galois sub-hierarchies with rich semantic and interactive capacities for novices.

As far as new services are concerned compared to usual list and grids display, the new method still gives a good support for indexing objects with objects partly inheriting from the method we presented in (Crampes et al. [9]). However the indexing is much improved in this version, particularly as far as scalability is concerned, because it takes advantages of contextualisation and of the probe's presence. As a second new service which differentiates it from trivial list or grid displays, it clearly and simply reveals some interesting structures, particularly in the context of social networks, such as community detection based on Hasse diagrams (which we just introduced in Plantié and Crampes [38]) and social complementarities which are introduced in the present paper.

As a conclusion to this state of the art, it is worth investigating recent developments based on faceted data which present some common features with our approach (Yee et al. [53]) A set of items is tagged with terms. For example scientific papers are tagged with their authors and their subjects. Terms are grouped in orthogonal (i.e. mutually exclusive) subsets called facets in which they can be selected by users. At the starting point, it is possible to see the count of all items in each facet, an item being possibly duplicated in different facets. When selecting a term, the facets are updated with the remaining items that are tagged with this term. In FacetMap and FacetLens facets are graphically and dynamically organized on the screen, each facet occupying an area proportional to its object count (Smith et al. [45], Lee et al. [33]). In faceted data 'terms' are equivalent to 'attributes' (or dually 'objects') in Galois lattices and choosing a subset of terms in different facets is equivalent to selecting a unique 'intent' (or dually an 'extent') in Galois lattices as we shall see below. To compare the technologies we will use the faceted data vocabulary with the words terms and items.

Our approach is different in several respects. First we need not organize terms (in our case attributes or dually objects) in orthogonal facets; they may be of any kind and can be organized in a hierarchy only if it is interesting. Second the screen is mainly occupied in these applications by facets and not by the items that are searched. Our point of view is that users should visually focus on what they are looking for and not the means to get it. Third in FacetMap and FacetLens facets are graphically represented with bubbles which are dynamically reorganized when a term is selected. The equivalent in our application is a traditional hierarchy of terms in alphabetic order because we consider that it is the traditional and most effective way of finding entities. The reported evaluations in both papers mention the attractive effect of the graphical interface, and do not mention usability problems with reorganisation during experiments. But it is also reported that some users do prefer lists in alphabetic order to explore terms. We also observed this users' expectation and this is the reason why we present terms in a hierarchical list. However the main differences with our approach is related to the choice and the presentation of returned items. The facet approach is a way of presenting an 'AND' choice of terms and the selected items are displayed in a list. Thanks to the Galois

lattice theoretical basis our probe approach displays selected items in layers with different levels of match corresponding to all possible combinations of terms (conjunctions and disjunctions) and not only a unique Boolean conjunction. The probe display opens up other functional possibilities such as weighting terms, indexation of items with items, items' complementarities, etc. which are difficult or impossible to obtain with sole conjunctions of terms. Faceted data still remains an interesting approach. Some new functions have been introduced in recent facet driven applications (Lee et al. [33]) such as linear facets and pivoting. They reinforce the interest in this technology. But the set conjunction paradigm remains different from our Galois lattice based paradigm.

2.1. Concept Hierarchies' basics

In Formal Concept Analysis (Ganter B. and Wille R. [19]), a finite set of 'objects' with 'attributes' can be organized in a lattice of 'concepts' that contain these objects according to their attribute commonality. The objects (respectively the attributes) are called formal insofar as they may be real objects or abstract objects (respectively attributes). Many domains are concerned, such as tagged photos, videos or documents, hospitals and patients, social networks, medical data, etc.

The organization process starts with a formal context, i.e. a table with the objects as rows and the attributes as columns. Any entry is marked (e.g. a cross or 1) if the corresponding object possesses the corresponding attribute, and is not marked (e.g. 0) if the object does not possess such an entry. Formally, a formal context is a triple (G, M, I) where G is a set of objects, M a set of attributes and I is a binary relation between the objects and the attributes, i.e. $I \subseteq G \times M$. Table 1 presents a formal context taken from a toy example where the set of objects G is a set of 4 actors, the set of attributes M is a set of 6 films, and the relation (g_i, m_j) is valued 1 if the actor g_i played in the film m_j , and 0 otherwise. We give the same name I to the binary relation and the incidence matrix it defines.

	Film1	Film2	Film3	Film4	Film5	Film6
Brad	1	1	1	0	1	0
Angelina	1	0	1	0	1	0
Cate	1	0	0	1	0	0
Leonardo	0	1	0	1	1	1

Table 1: A small context with films and actors

The next step in building the concept lattice is to define concepts according to Ganter B. and Wille R. [19]. A concept is a pair of subsets: a subset of objects O_i (called the extent) and a subset of attributes A_i (the intent) that the objects share. Two operators both denoted by $'$ connect the power set of objects 2^G and the power set of attributes 2^M :

$$' : 2^G \rightarrow 2^M, O_i' = A_i = \{m \in M | \forall g \in O_i, gIm\}$$

Dually on attributes:

$$' : 2^M \rightarrow 2^G, A_i' = O_i = \{g \in G | \forall m \in A_i, gIm\}$$

Informally applying the operator $'$ to a subset O_i of objects of G extracts the subset A_i of attributes of M that are shared between all objects of O_i and conversely A_i' identifies all objects (the subset O_i of G) who share the same subset of attributes A_i of M . The composition operators $''$ are closure operators (idempotent, extensive, and monotonous), which means that $A_i'' = A_i$ and $O_i'' = O_i$ for any $(O_i, A_i) \subseteq G \times M$. These operators $'$ will be important for the properties of our model below.

For the context presented in table 1, the concepts are nodes in the line diagram shown in figure 2, such as:

$$\text{Concept6} = (\{Angelina, Brad\}, \{Film1, Film3, Film5\})$$

where $\{Film1, Film3, Film5\}$ is the intent and $\{Angelina, Brad\}$ is the extent:

$$\{Film1, Film3, Film5\}' = \{Angelina, Brad\}$$

$$\{Angelina, Brad\}'' = \{Film1, Film3, Film5\}$$

Following the process of concept identification, the next goal is to build a lattice whose elements are the concepts. A partial order on formal concepts is defined as follows:

$$(O_2, A_2) \leq (O_1, A_1) \text{ iff } O_2 \subseteq O_1 \text{ (and consequently } A_1 \subseteq A_2).$$

The ordered concepts form a complete lattice called a "concept lattice". Figure 2 shows the concept lattice of the toy film-actor example as a Hasse diagram. The set of concepts L is completed if necessary by a top concept that contains all objects and a bottom concept that contains all attributes. A Hasse diagram is a graph whose vertices are the concepts, ordered from top to bottom according to their order in the lattice; the edges are drawn between concepts when two concepts are directly ordered without transition through another concept. In our example each concept is a group of films with their common actors. As can be seen in figure 2, an object (an actor) as well as an attribute (a film) may appear in several concepts.

The transpose of the context matrix produces a Galois Lattice which is the dual of the original context. The roles of objects and attributes are reversed. The Hasse diagram's structure is the same; it is just turned upside-down. In our example if we place concept-9 at the top, concept-0 at the bottom and accordingly reorganise the Hasse diagram, films become objects and actors become attributes because tradition applies the object order for the top down hierarchy. This is a slight problem for our presentation. After experimenting with users it appeared that the probe we will introduce below should be placed at the top of the screen. We will see below that when searching for films the probe must be loaded with actors. Consequently, since we want to be sound with Formal Concept Analysis in the present paper, actors must be defined as objects and films as attributes, although the search applies to films. We will choose objects to search attributes. This is purely formal because attributes and objects play dual roles and final users are not concerned by this vocabulary; they only know their domain of application vocabulary, such as films/actors, people/competences, papers/authors, etc.

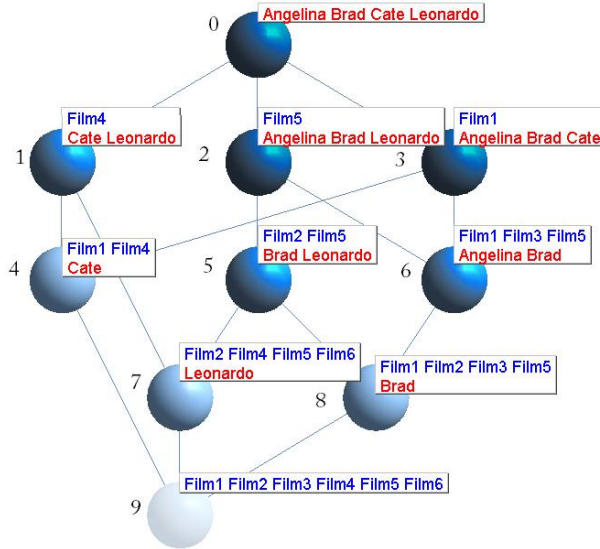


Figure 2: A line diagram with concepts showing intents and extents

2.2. Visualizing Galois lattices

Many methods have been proposed for building line diagrams representing Galois lattices such as incremental building (Godin et al. [22]) or Force Directed Placement (FREESE [18], Hannan and Pogel [23], Kamada and Kawai [27]). Their main goals are algorithmic efficiency and display quality (see (Arévalo et al. [2]) and (Kuznetsov and Obiedkov [30]) for a survey of some algorithms and their performances). Although Galois lattices are mostly targeted to objects with Boolean attributes, they may also be used for organizing multi-valued data (Ganter B. and Wille R. [19]) and even hybrid data (Villerd et al. [51]). Several tools implement these methods among which we used two of the most widely known in the FCA community for illustrating examples in this paper. We took advantage of Lattice Miner (Lahcen and Kwuida L [31]) because it is recent and, beyond a real aesthetic effort, it proposes many visualization options that we use for better illustrations of simple examples. However the number of concepts it can compute is limited. Consequently we also used Galicia (Valtchev et al. [49]), to create the Hasse diagrams in Figure 1.

Even with small examples it is not obvious for non-experts to analyse the Galois lattice conceptual structure and navigation is not easy when looking for particular sets of objects or attributes. However it is possible in Galicia or in Lattice Miner to interact with a concept and display its intent and extent but at the expense of other problems: edges are hidden and information is getting cluttered even in little Galois lattices. Some authors have explored better design and interaction in other FCA environments for helping non-experts browsing

Galois lattice such as in Eklund et al. [15]. Although these authors report positive results, Galois lattices which are tested are small and scalability remains an open question.

Real applications require bigger contexts. To better experiment with scalability we built a benchmark with a medium context containing 127 films (attributes) and 245 actors or directors (objects). The resulting Galois lattice built with Galicia is presented in figure 1. The nice diamond shape with three intermediate layers is exceptional. It reflects the fact that we considered two actors and one director for each film all films with two actors and a director, except for the three films of the Ocean's series for which we considered five actors (two concepts concerning these films are visible on a small fourth layer). Real applications present more complex Galois lattices with no particular symmetry. We built such a simplified benchmark structure for the following reason. Visual analysis is difficult on Galois lattices and their Hasse diagram display using tools like Galicia or Lattice Miner. Conversely the semantic probe is not affected by this problem. As a result in order to build experiments with users and challenge our semantic probe on traditional Hasse diagrams we had to build a simplified benchmark to the detriment of the probe. If experiments give better results on such simplified data with the probe, it would also be the case for more general data.

As far as navigation is concerned, Galois lattices' scalability is even worse. To overcome this problem several approaches have been proposed such as focus & context and fisheye in Lattice Miner (Lahcen and Kwuida L [31]). Only experts can however analyse the resulting display. Other navigation applications which are targeted to novices propose a local concept approach. A user's query is considered as a set of attributes. The corresponding extent is displayed with facilities for removing attributes or adding attributes from the list of descendant concepts.

As a result the user can navigate upward and downward on the Galois lattice without ever seeing it such as in the experiments conducted in Godin et al. [21], the Credo application (Carpineto and Romano [6]) or in the more recent application ImageSleuth (Ducrou et al. [12, 11]). But user's navigation is entirely limited to one concept at a time, and all conceptual structures have disappeared.

Coming back to a global Galois lattice view, several reduction algorithms have been described in the literature for managing scalability. Four of them are frequently applied. They are introduced in the next section.

2.3. Galois lattice reduction and other methods

2.3.1. Nested, iceberg and stability based reductions

Nested line diagrams are constructed when it is possible to extract sub-contexts and partition the attribute set (Ganter B. and Wille R. [19]). Resulting line diagrams are clearer but to the detriment of easiness of navigation and understanding. Iceberg lattices reduce Galois lattices to a subset of concepts whose intent's support count is above a user defined threshold (Stumme et al. [48]). The support count of an attribute set A_i is define as: $support(A_i) = |A_i'|/|G|$.

A concept is frequent if its intent is frequent, i.e. there are many objects with the corresponding set of attributes compared to other concepts. The set of frequent concepts of a context is called the iceberg concept lattice of the context. Reducing a Galois lattice to an iceberg lattice is efficient when looking for association rules to the expense of missing rare information. Another reduction process is based upon stability whose definition is formally less intuitive as support (Kuznetsov [29], Roth et al. [42]). Intuitively, a concept is stable inasmuch its intent is found in many combinations of objects from its extent. This reduction process is particularly interesting for data and knowledge mining (Jay et al. [26]), but its visual efficiency is limited, depending on a user defined threshold, and it loses rare information which may be highly interesting in many applications.

2.3.2. Object or Attribute Galois sub-hierarchies

Extracting Object or Attribute Sub-Hierarchies is another reduction process for pruning Galois lattices. It presents a remarkable advantage: contrary to iceberg or stability driven reduction, there is no loss of information (Godin et al. [22]).

The reduction process is based upon the observation that many objects and attributes belong to several concepts. In our example attribute Film5 for instance belongs to concepts 2, 5, 6, 7, 8 and 9. It is possible to get rid of this redundancy without losing information. The reduced intent (respectively extent) of a concept (O_i, A_i) is the set of objects (respectively attributes) that belong to A_i (respectively O_i) and do not belong to any upper (respectively lower) concept. In the following we will only consider attribute reduction, the same results being dually possible with objects. In the example, the reduced intent of concept6 is {Film3} since Film1 and Film5 belong to lower concepts (respectively concept3 and concept2).

For each attribute (a film in our example), there exists a unique attribute-concept that represents the most specialized concept that contains the attribute. Figure 3 shows the Galois sub-hierarchy which is derived from the Galois lattice in figure 2. Films appear in only one concept although they are implicitly present in other concepts. Since we want attributes to only appear once, only concepts with attributes are kept in the lattice and other concepts can be rebuilt through inheritance. In our example, concept4 which was originally $(\{Cate\}, \{Film1, Film4\})$ is now $(\{Cate\}, \{\})$ with an empty reduced intent, and its original intent $\{Film1, Film4\}$ can be rebuilt through the union of concept3 and concept1's intents. This act of pruning when applied to attributes or objects is the one proposed in Godin et al. [22] under the name PCL/X. The new line diagram is a particular case of what is called a Galois sub-hierarchy (Godin and Mili [20]). It is a lighter visualization of data when only focusing upon the attributes (dually the objects), in our case the films (dually the actors and directors). To our best knowledge, attribute or object driven reduction process has only been applied for building incremental Galois lattices. In Crampes et al. [9] we used it to organize, visualize and index social photos. However the

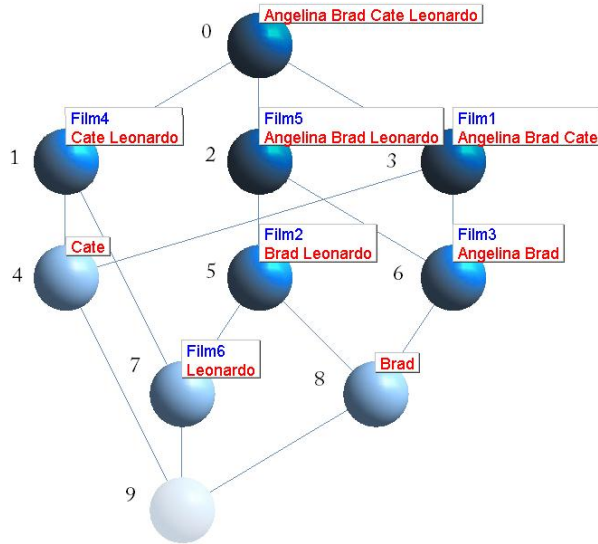


Figure 3: An Attribute Galois Sub-Hierarchy

display was not user centered and scalability was a remaining issue which we address in this paper.

Figure 3 is clearer with no redundancy on films. Thanks to the edges one can see for example that Leonardo played in Film6, Film2, Film5, and Film4. But catching this knowledge is not immediate. Moreover in a realistic context edges would be covered by concept-nodes. In that case a good thing would be to get rid of edges without losing the possibility of identifying concepts. This is what we are going to do with semantic probes that we present in the following section.

3. SEMANTIC PROBES

The semantic probe model and techniques introduce a user centric approach of Galois lattices which is easy to understand for novices because it is not concept oriented and it has no edges. The display clearly shows entities that are searched without repetition and without the necessity of following edges.

Let G be a set of objects and M a set of attributes, each object being characterized by a subset of these attributes. Objects and attributes are represented by words or icons depending on the application domain. We define a semantic probe P as a bag which is loaded with some objects representing a particular focus of interest for a user and with which it is possible to interact. The corresponding objects' attributes react and gather around the probe as if it were a magnet. Remember that the terms 'objects' and 'attributes' are formal and the roles are dual. We chose in this description to load the probe with objects and to attract attributes to comply with the FCA tradition which places all objects

at the top of the hierarchy. If we had placed the probe at the bottom, we would have loaded it with attributes and have attracted objects. This last observation will be of great interest at the end of the paper.

Formally, we define a semantic probe P as follows:

Let (G, M, I) be a formal context with G a set of objects, M a set of attributes, and $I \subseteq G \times M$ an incidence relation,

A probe P is a bag which, when loaded with a set of objects $G = \{g_i\}$, $G \subseteq G$, produces two results: a sub-context and a Galois sub-hierarchy display.

3.1. The sub-context

The probe P 's set of attributes defines a new context (G, M, I) which is a sub-context of the original context (G, M, I) where:

- G is P 's set of objects,
- $M = m_j$ is a subset of M : $M \subseteq M$, $m_j \cap G \neq \emptyset$
- I is an incidence matrix whose rows are the rows of I corresponding to the objects belonging to G and whose columns are the columns of I corresponding to the attributes belonging to M . From this sub-context it is possible to build a Galois lattice G_p and create an original layered display.

3.2. Semantic Probe's object-concept display

In figure 4 the general context is the whole benchmark containing 127 attributes (films) and 245 objects (actors or directors). The probe which is represented by the blue button with a question mark at the top is loaded with the object subset $G = \{AngelinaJolie, BradPitt, CateBlanchett\}$ which defines a sub-context. All attribute-concepts whose extent contains one or more selected objects slide up. Each attribute-concept is a group of attributes (DVD jackets) which share exactly the same objects (actors and directors) in the original context. For the sake of communication with lay users we use the word 'group' rather than 'attribute-concept' or 'concept'. A group is represented by the jacket of one of its DVDs. The figure at the top left of the group's picture indicates the number of DVDs it contains. In this particular benchmark which was created for experimentation all film castings are different but for the three films from the Ocean's trilogy. Consequently all group pictures but one display the number 1 and the Ocean's trilogy displays the number 3.

When clicking a group, a pane opens up at the bottom. It displays the group's DVDs. Since a group represents an attribute-concept from the whole context, clicking actually reveals its intent at the bottom and its extent in the middle right pane. Figure 5 shows the attribute-concept whose intent is the Ocean's trilogy shown in the bottom pane and the extent is a set of 4 people shown in the middle right pane. Two characters are red. They are those that are included in the probe whose extent is shown in the upper right panel. As a result comparing the two upper right panes it is possible to identify the objects that are common to the group and the probe (red), the objects that are in the

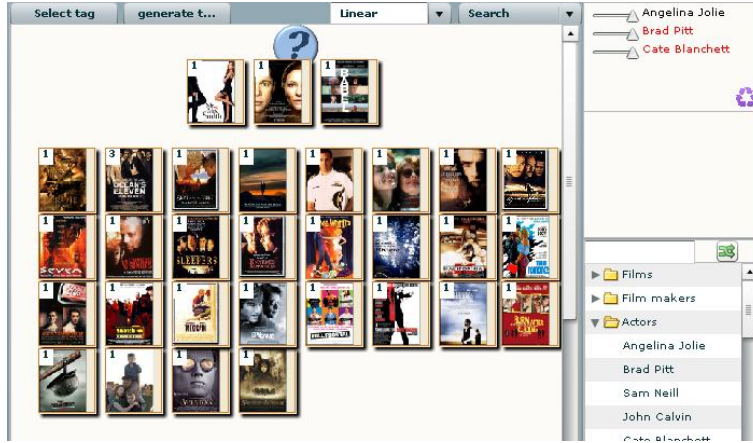


Figure 4: A semantic probe display

probe and are not in the group (black in the upper right pane in figure 6) and the objects that are in the group and not in the probe (black in the middle right pane in figure 6).

The core of the display strategy is to place the groups at a distance from the probe according to their semantics (the extent) and the probe's semantics. Let a_i be a group A 's extent. We define the Semantic Distance between the probe and the group as follows:

Definition 1. $SD(P, A) = \frac{1 - |G \cap \{a_i\}|}{|G|}$
 where G is the probe's extent.

All groups which are at the same semantic distance are put in a common layer. All layers are placed from top to bottom according to their semantic distance, the layer at the top being the one with the smallest distance to the probe. All groups belonging to a layer are then placed in a grid which clearly identifies them. In Figure 4 three groups are visible in the first layer, and 25 groups in layer 2. The probe displays well identified entities, in this example DVDs, when the traditional Galois hierarchies display concepts with no easy means for novices to identify objects or attributes.

The probe is equivalent to the top concept of a Galois lattice as in figure 3: it contains the set of objects from which the hierarchy is built. The ordering from top to bottom is linked to a decreasing number of objects. The result is a balance between the search engines' traditional display and the rich conceptual display of Galois lattices. The core idea is to invite users to interact with Galois lattices as if they were interacting with traditional displays.

3.3. Probe's concept visualization through interactions

Suppose we want to know in which films a particular actor, say Brad Pitt, played. Each group is an original concept with the same subset of attributes

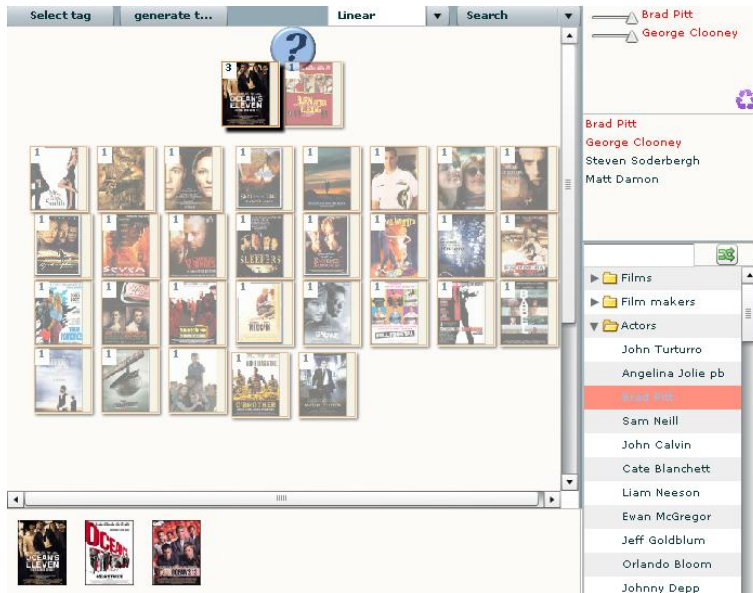


Figure 5: Clicking an object-concept reveals its intent and extent

and the same subset of objects. All groups have different extents and intents. To search for films in which Pitt's acted it is possible to drag the object 'Brad Pitt' to the empty probe to get all groups containing films with at least Brad in the extent. Consequently Brad may be in several groups of films with other actors.

But when the probe is already loaded with several objects like in figure 6 there are groups in the lower layers which contain other objects than the interesting one. Therefore we are only interested in a subset of the visible groups. To reveal this subset we use the fact that the display is a sub Galois lattice with attribute-concepts, the whole objects set being the probe's content: we are looking for the intent of a scattered concept whose extent is Brad.

As it was explained in section 2.3.2, an attribute sub-hierarchy shows intents without attribute redundancy and the concepts' intents of this sub-hierarchy are only visible through inheritance. The concepts' intents we are looking for when searching DVDs with 'Brad Pitt' may be scattered within layers and between layers. To manage this difficulty we apply the following design strategies. First all groups in the same layer belonging to the same probe-filtered concept, i.e. having the same subset of objects common with the probe, are dynamically regrouped side by side (this dynamic regrouping is very spectacular and well appreciated by users). They are optionally separated by blank objects from other groups when probe-filtered extents are different. Second, we showed in section 2.3 that a concept can be rebuilt from an attribute-concept through the union of its parents in the hierarchy. Practically, when the user hovers with the mouse over a group in the probe's induced hierarchy it is possible to reveal the

corresponding probe centered concept to which this group belongs through the following visual effects. The common objects with the probe's objects are turned into red in the two upper right panes to show that they are the probe's driven concept's extent. All other groups which do not belong to the concept are partly turned transparent. The remaining clearly visible groups define the concept's intent whose extent is the intersection between the selected group's extent and the probe's extent. Figure 6 shows the concept extracted from the probe-driven sub-hierarchy when the user hovers with the mouse over the group represented by the film 'Seven'. The only common object between this film and the probe's extent is $\{Brad Pitt\}$. The corresponding probe-driven extent $\{Brad Pitt\}$ is highlighted in red in both side panes. The concept intent reveals three groups in the upper layer, and 26 in the lower layer. All these groups and the probe have $\{Brad Pitt\}$ as a common set of objects. If the user hovers over 'Babel' in the top layer, only the groups represented by 'Benjamin Button' and 'Babel' will appear. 'Brad Pitt' and 'Cate Blanchett' will be highlighted in red. The corresponding concept is $(\{BenjaminButton, Babel\}, \{BradPitt, CateBlanchett\})$.

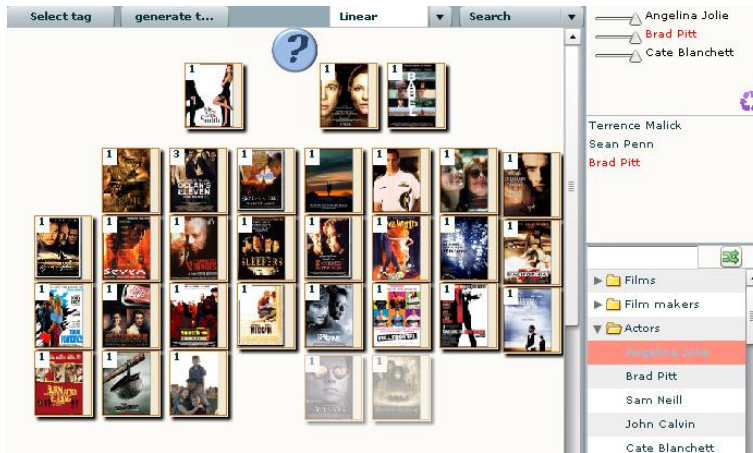


Figure 6: Interaction reveals probe centred concepts

This interaction for revealing a concept is attribute-driven since it is necessary to hover with the mouse over a group. It is also possible to apply an extent oriented way of revealing a concept. The probe's objects in the top right probe pane are endowed with sliders (see Figures). Dragging a slider to 0 turns the corresponding object slide down and all the groups with this object slide down. The remaining groups by the probe define the intent which corresponds to the probe's extent whose objects' weight is equal to 1. The sliders can also be set to a value between 0 and 1. Groups on the same layers are separated into those that do not contain the modified value which stay at their level, and those that slide down but are still on the screen. This advanced interaction was activated by the user searching for personal Facebook photos in figure 15. This is also what is applied to separate Arsenal from Manchester in the industrial proto-

type presented in figure 16. It must be noticed that to our best knowledge this method of weighting objects in concepts' extents (or dually in concepts' intents) to reveal sub-structures in Galois lattices is new even in the Formal Concept Analysis community. Our approach provides a new way of seeing and weighting Hasse diagrams. Moreover this reorganization provokes an impressive animation on the screen which is very appreciated by users.

The interaction and visual effects described above, which reveal concepts' intents avoid edges' visual complexity. Their drawback is that even if they are simple, their interpretation is not so obvious. We do not yet know to which point it is interesting to give these conceptual clues. However several presentations and experiments with users have shown that it is better to use sliders than transparency to reveal concepts. This is an important point for deploying the technology.

3.4. Interactions and navigation

Interacting with traditional Galois lattices is seldom mentioned in the literature although some applications like Lattice Miner offers a few limited possibilities. The probe driven display with explicit intents are not only simple and easy to understand compared to traditional Galois lattices. They are also particularly useful for interacting with all objects and attributes. Users can change a probe's semantic state through different interactions:

1. Adding an object to the probe's semantics by double clicking onto it in the tree of objects at the bottom right, or, after clicking onto a group, selecting an object from the group's object list in the central right panel, then dragging and dropping the new object onto the probe. This second possibility is particularly interesting because a group may suggest new objects for searching other groups.
2. Removing an object from the probe through dragging and dropping it onto the bin in the probe's object pane. Double clicking onto this bin removes all objects from the probe.
3. Adding a group's extent to the probe through dragging and dropping the group's image from the hierarchy. As a result all the group's objects which were not already part of the probe's semantics are added to it (see figure 7). This last interaction is original, particularly useful and well understood by testers and users.
4. Weighting tags in the probe's extent for separating groups in the same layer.

Updating the sub-hierarchy is made after the end of the interaction. If groups must disappear because they have no common objects with the probe, they slide down and hide. If new groups are eligible, they slide up and find their proper place in the hierarchy. Other groups may change smoothly of place in the hierarchy, changing of layer or creating a new layer. All movements are made of fluid aesthetic animation to maintain the user's mental map. These animations are particularly appreciated both during presentations and tests with users.

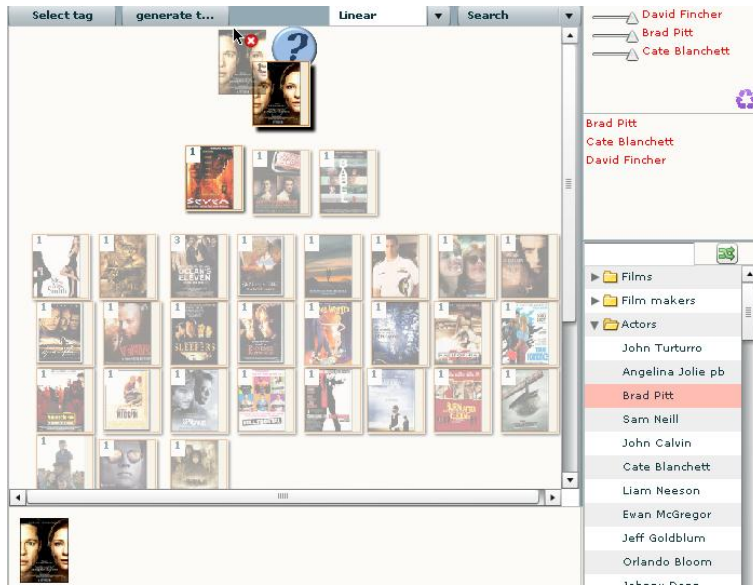


Figure 7: Loading the probe with objects

3.5. Semantic probe's qualities and novelty

Our goal was to define an environment whereby Galois lattices, which are sophisticated experts' tools, are simply used by lay users. The probe's metaphor and display show the usability qualities which are expected by them as explained in section 2:

Contextualization: Only attributes (DVD jackets in the example) that meet totally or partially the probe's semantic profile (actors and directors) are displayed.

Reification: It is possible to easily identify attributes or groups of similar attributes with their objects and without redundancy.

No edges: Contrary to usual Hasse diagrams and the solution we proposed in Crampes et al. [9] there are no visible edges. Edges are difficult to read and understand for lay users. In our application they are replaced by the probe's profile combined with the navigation tools.

These simplification improvements are achieved with little loss of conceptual information which distinguishes our approach from a trivial list or grid:

Conceptual structure: Concepts and concept relations are revealed through the regrouping of attributes and interactions as shown in section 3, or through the use of sliders.

Navigation: The display gives conceptual hints and provides interaction capacities for facilitating navigation when placing objects on to the probe.

Mental map: The soft animation maintains the user's mental map when groups are reorganized after a modification of the probe's profile. This feature is particularly attractive when shown during presentations. Its interest is not

only limited to aesthetic animation. Contrary to a list or a grid presentation of results after a query in traditional search engines, it only shows what is changed in the results and how these changes occur.

All these qualities show first that a probe driven sub Galois lattice display without edges meets most simplicity criteria that lay users are looking for, and second that it provides a better approach for navigation in an object-attribute database. Next section introduces several tests and industrial experiments that have been conducted for verifying the above hypothesis.

4. Tests and applications

Our first goal for conducting tests was to compare the probe paradigm with its two main competitors: Galois lattice based navigation and traditional Boolean querying using index terms, the last one being the most widespread mode of searching databases when indexes are available.

As far as other technologies are concerned such as faceted data, the goal was not to check whether the probe approach is more efficient or more attractive, though some experimental results with these technologies are worth being mentioned. For instance two faceted data applications are compared in Smith et al. [45] using a group of 10 participants well aware of computer interfaces. Memex is a text oriented faceted data browser whereas FacetMap presents adaptive bubbles representing facets on the screen. Results do not reach significant conclusion about the success of a particular technology. But authors are more interested in the formative results given by the testers' observations. Other formative experiments are conducted in Lee et al. [33] with FacetLens, which extends FacetMap. Six people are involved in the test, none of them lay users. Reported usability results are interesting, but no comparison is made with other applications, such as with traditional Boolean search engines.

Focussing on our Galois lattices (GLs) experimental context we mostly find experiments on local navigation around concepts. In Godin et al. [21] local navigation on GLs is compared to two more conventional retrieval methods: hierarchical classification retrieval and Boolean querying with index terms. Their result show that local navigation on GLs outperforms hierarchical classification navigation, but it does not do better compared to Boolean querying. A more recent experiment in Ducrou et al. [12] is conducted with the ImageSleuth application involving 29 testers. GL based local navigation is compared to hierarchical classification navigation. Authors provide similar results: local navigation on a GL gives better results than hierarchical classification navigation. No comparison is given with Boolean querying with index terms when according to Godin et al. [21] this approach is more efficient than hierarchical classification navigation. In our case navigation is performed through extracting a sub hierarchy and organizing it under a probe; it is in between a Hasse diagram search which represents a Galois lattice and local search on individual concepts. Taking into account all these experiments, the conclusion is that Boolean querying is the search method to challenge because it is not clearly outperformed by any of these technologies and it is still the most widespread. However since the probe

rebuilds parts of a Hasse diagrams without lines between concepts, it is also necessary to compare it with traditional Hasse diagrams.

Two phases of tests were conducted. The first phase did not require new developments beyond the prototype and could be organized within the laboratory. The second phase required new developments and the support of an industrial.

The first phase targeted two questions:

- 1) Is it possible for lay users to navigate on a Galois lattice when using our semantic probe compared to traditional Hasse diagrams.
- 2) Does the probe approach equal Boolean search with index terms for traditional tasks and does it outperform it for some tasks.

The second phase had more open goals:

- 3) Application on real data: is the probe interesting for users using their personal data?
- 4) Deployment: for what sort of applications can the probe approach be the most efficient?
- 5) New services: is it possible to imagine new services for which traditional Boolean search engines are not or are poorly adapted?

4.1. *First phase*

Methodology

The first phase required testing our probe environment against navigation on a Hasse diagram, and then against a Boolean search engine. Twelve students studying general engineering ageing from 20 to 23 (including 4 females), were asked to answer questions from the database of 127 films and 245 actors or directors with the support of the three technologies.

The first method consisted in navigating on a whole Hasse diagram of the film database. As it was already mentioned in the paper the database had been built for helping users navigating on such a structure which may be very complex even on limited contexts. The concept hierarchy is very symmetric and there are few layers (see Figure 1). We used Galicia for building this Hasse diagram. Quickly it appeared during the tutoring preparation that explaining what the line diagram meant and how to navigate took a long time. Moreover none of them could properly navigate on the Hasse diagram. We tried using Lattice Miner which proposes advanced filtering and navigation tools. The tool could not build the lattice because there were too many concepts on standard PCs. The test required that the computer had to be of the kind used by lay users and usability study on powerful computers was out of question.

In conclusion, although Eklund et al. [15] suggest that navigation on very simple Hasse diagrams is possible for lay users with the hypothesis that scalability should not be a problem, our tests show a different result. Navigation on medium size Hasse diagram is complicated. We now focus on the second question which assumes that it is possible to search with the probe.

To answer the second question we had to compare navigation using the probe with a Boolean engine. We chose Amazon because it is widely used. There may be more efficient search engines, but our purpose was not performing a general test. We only needed a well known and widely used engine. Moreover the film database which is used in our examples and which is used for testing had been built in the first place using Amazon. We knew that there would not be biases from the data.

Each tester had to answer a set of questions using both environments. The subjects were asked not to tell other testers what was taking place and what questions were asked. The subjects had to draw lots for the order of the environment to assess to avoid any possible biases. The response time for Amazon was also tested prior to commencement to ensure that the two applications were comparable in terms of response. None of the subjects had previously been exposed to the probe application prior to assessment and only a few used Amazon. Consequently, the test started with an explanation read to each subject individually and a short demo on the two applications was provided, even for those who had already had experience with Amazon.

After each test some measures were taken, such as the time for obtaining the answer and the quality of the answer (number of mistakes or failure to give an answer after a time delay). We applied a delay of one minute or two minutes for answering to mimic the fact that lay users are known for abandoning a tool if the service is not quickly given either because the application is too complicated or because they have difficulties to use it. They were also asked to assess the degree of confidence they gave to their answers. At the end of the test, some qualitative questions were asked, comparing the two methods.

We applied “repeated measures t-tests” with unequal variances to the results of the tests for each questions when enough paired values were available (results show that it is not always the case due to the methodology of giving a time delay for the answers). H_0 , the null hypothesis, asserts that the difference between two responses measured on the same statistical unit has a mean value of zero.

Results

Q1: “Cite two films in which Ben Stiller played”

The real objective of this simple first question was to train the subjects on both environments. All subjects managed to give an answer with a mean time of 18.7 seconds for Amazon and 16.9 seconds for the probe. The mean times’ difference is not significant (confidence in H_0 : $p = 0.13$).

After this first question testers were also invited to freely explore the data with other actors and films to get used to the environments. They could do it without any problem on both environments. Consequently as far as the probe is concerned we could conclude that it is possible to navigate on a Galois lattice with the probe when it is difficult and even impossible with the whole Hasse diagram on a medium size database.

The three following questions were of increasing complexity:

Q2: “In how many films have Martin Scorsese and Leonardo Di Caprio acted together?”

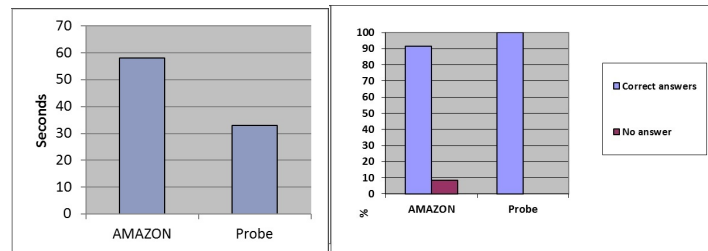


Figure 8: Results for Question2: left) mean time for answering (in seconds), right) number of responses within one minute

The mean times under the two environments for answering is presented in Figure 8 with only 11 measures since one of the subjects failed to provide an answer with Amazon within the minimum delay of one minute. It seems that the probe clearly outperforms Amazon with nearly half mean time (confidence for H_0 : $p = 1.2E-07$)

In fact the difference of mean value is not as instructive as it may look in this experiment; some testers took time for answering, particularly with Amazon, because they knew they had one minute and they did not want to give wrong answers. However this testers' strategy applied for both applications and the difference of mean values is still interesting. The most interesting result is that one tester failed to find the answer with Amazon when all testers succeeded with the probe (it must be noticed that this tester's failure is not taken into account in the mean time for the benefit of Amazon).

Q3: "Here are five actors: Matt Damon, Al Pacino, Julia Roberts, Brad Pitt and Georges Clooney. In how many films have they acted ...

- ... together
- ... four among the five
- ... three among the five
- ... two among the five?"

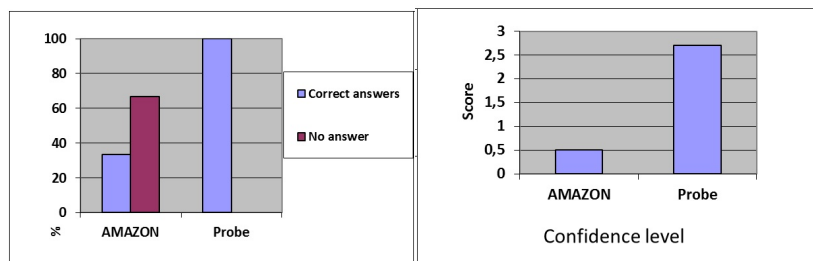


Figure 9: Results for question 3: left) number of answers, right) confidence degree

For this complex question involving more semantics, figure 9 left shows that only one third of the subjects could provide answers using Amazon within a two

minute timeout whilst all answers were provided with the probe. Interestingly, for those who gave an answer, the degree of confidence on a scale of 0 to 3 was low under Amazon and high with the probe (see figure 9 right). t- test cannot be applied because a majority of testers failed to give an answer within the time delay.

The fourth question concerned the capacity of combining two semantic view points.

Q4: “You want to go to the cinema with a friend. You like Brad Pitt and Georges Clooney and your friend likes Julia Roberts and Brad Pitt. What is the best choice for you, for her and the best compromise for both?”

The results are particularly interesting. No one was able to give any answers with Amazon in less than 2 minutes (see figure 10) whilst all answers were given with the probe in a mean time of 64.9 seconds.

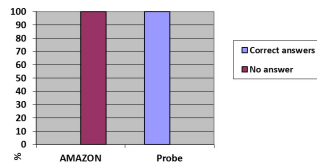


Figure 10: results for question 4

Figure 11 summarises the answers regarding practicality, interest, innovation and enjoyment. Each method is assessed independently by the subjects. The semantic probe method clearly outperforms the more traditional Boolean search method with statistically significant results for all four answers ($p < 0.001$). These results are particularly interesting when going back to those detailed in Godin et al. [22] where a similar test was conducted comparing a query based search with a local Hasse diagram driven navigation search. The authors report equal performances whilst the tests we conducted with semantic probes give much better performances.

The last question was a key assessment.

Q6: “If you had to choose between the two methods, which one would you prefer?”

Figure 12 shows the answer to the question. Nine subjects favoured the probe. Three subjects preferred the traditional Boolean search and its list presentation although they favoured the probe when answering question 5. They were asked the reason for this contradiction. They had the same answer. They were used to buying music or DVDs online and did not expect more from the Internet. They were not concerned with more semantically sophisticated methods as they found they had no use for them.

Analysis of results.

Synthesis of results for the 4 first questions and all 12 testers, i.e. 60 answers is presented in figure 13. This figure demonstrates that all questions could be

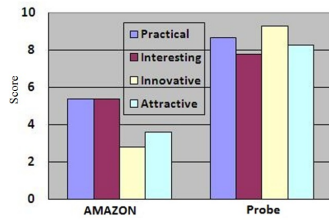


Figure 11: Subjective results for question 5: Satisfaction criteria

answered with the semantic probe within the time limit whilst nearly half could not be answered using Amazon. In particular, nobody was capable to answer complex questions with Amazon in the time period. When there was an answer given, it was 10% faster for simple questions and 50% faster for more complex questions when using the probe. Moreover testers' confidence in answers is low for complex questions with Amazon and high using probes. For instance the mean degree of testers' confidence in question 3 is 2.7 with probes and 0.5 with Amazon on a scale from 0 to 3.

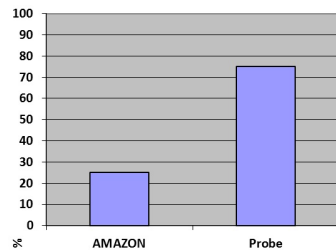


Figure 12: Results for question 6

This set of simple tests on a medium size database shows on the one hand that navigating on a Hasse diagram is difficult for lay users, and on the other hand that the semantic probe outperforms traditional Boolean querying on a standard engine. Satisfaction questions (Q5) confirm these conclusions: the semantic probe is favoured. These results are obtained with a limited number of questions and a limited number of subjects. Our intention was to set up a wider test with more subjects. However there was an important objection to this idea. Answers to the last questions Q5 and Q6 show that there is a contradiction between indoor tests and reality. Conducting other more significant tests would probably confirm the first results and would not be interesting. Conversely since “a common evaluation measure for any technology is adoption by others, and the move into commercial products” (Plaisant [37]) it was decided that the second phase of tests should favour beta testing on real data and move to industrial judgement through presentations and marketing. Moreover considering the hypothesis that users would not adopt a better technology if it does not bring about new useful functional novelty, we explored extended services with

some industrials.

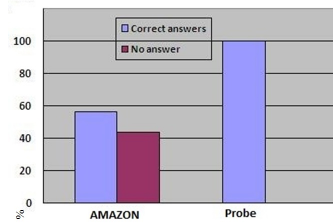


Figure 13: Subjective results

4.2. Second phase

Controlled experiments with real data

Another set of controlled experiments was intended to test the semantic probe approach on data with real users in a real environment. We downloaded three users' photo album from their Facebook profiles, and asked them to navigate in their respective albums looking for photos with some of their friends. Only tagged photos were loaded. We took account with rather big amount of photos, between 1000 and 1200 tagged photos in each album and more than 250 friend tags, nearly 8 times the size of our DVD benchmark. We asked similar questions as those for the DVD benchmark. We added more difficult questions such as "find the photos with the most people". Our intention was to assess scalability on real medium size databases, the interest of users and the ease of use. Results on scalability were excellent (speed and reactions to interactions). Ease of use was as expected: looking for a particular friend took less than 10 seconds with the probe and at least 20 seconds on Facebook.

Some experiments were not possible with Facebook when they were easy with the semantic probe, like finding photos with three particular persons. Tester's interest was high when they rediscovered their photos and confessed they did not know of any applications that could provide these functionalities.

Figure 15 shows a screen capture taken during testing.

Testers' were asked questions similar to Question 4 : "You want photos displaying people that are both known to you and a friend. You like person A and Person B and your friend likes person C and Person A. What is the best choice for you, for her and the best compromise for both?". The results are particularly interesting. No one was able to give any answers with Facebook in less than 2 minutes, whilst all answers were given with the probe in an acceptable time.

Figure 14 shows a screen capture taken during testing.

One of the testers spent about half an hour, using the probe to explore the photo set, and search for photos with his friends.

Moreover, one student having been informed of the probe tool by one of his friend asked to use the tool, to explore his photo albums and sort his collection

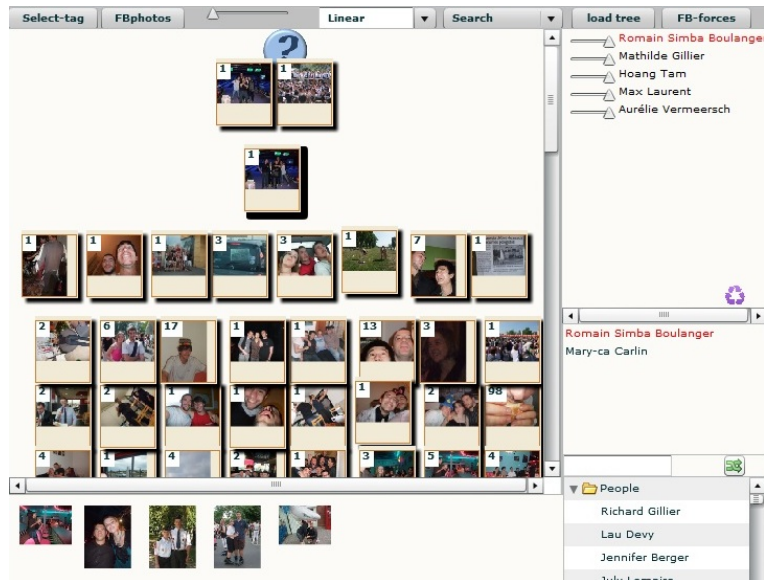


Figure 14: Snapshot of a test with 1288 photos from Facebook

using the semantic probe. He found the tool very useful to browse photos and make complex search tasks.

He was continuously tempted to add photos on the probe and see the computed photo sub-hierarchy, helping him to browse his collection and rediscover his group photos, and the events associated with the photos.

He reported a situation of dropping on to the probe a person and after viewing the photos, he took one of them with persons of interest for him and drop this photo on to the probe. He then discovered very quickly (less than a second) other photos with these persons in a hierarchic tree containing photos with one or more of these persons. He was then able to explore these photos by hovering the mouse on them and seen the persons present on these photos. He then navigate quite a long time putting photos on the probe and exploring his collection.

In Figure 15 this user associated weights to some persons, seen on the upper right zone of the image, in order to favor photos with his best friend while keeping other friends on the photos. The subhierarchy of photos then reorganized according to the given weights.

Several users asked if it was possible to use the probe directly integrated in Facebook, and others asked us when this new tool will be available.

These conclusions confirmed that industrial applications should be considered.

Deployment: Industrial assessment and applications

After patenting, a license agreement was signed with a software distributor.

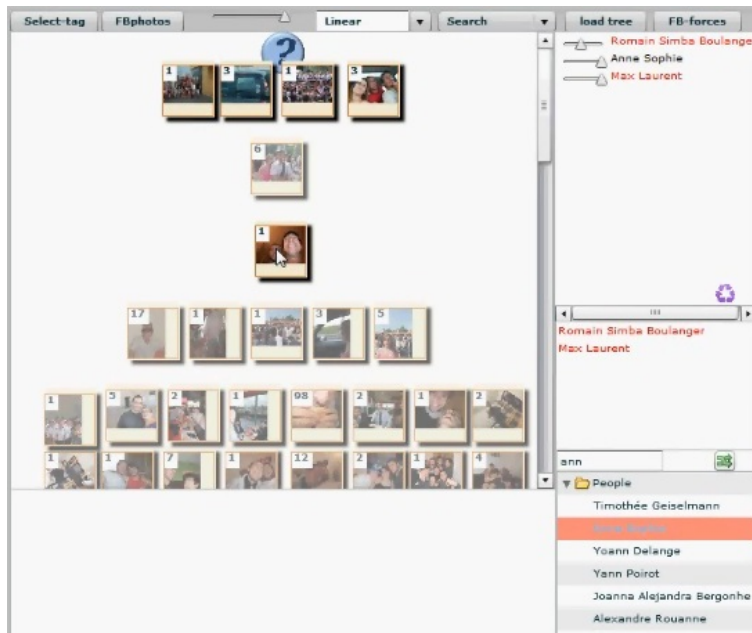


Figure 15: Snapshot of a test with weighted concepts from Facebook photo collection

Many presentations have then been given by our partner to industrials in Europe and in the U.S with excellent feedbacks. Several prototypes are now under development with redesigned interfaces for each industrial target such as a pharmaceutical world leader (drug interactions), TV channels (programs selection), a music major company, a human resource management company, etc. Figure 16 shows an experimental interface for a sports TV channel in the U.K. Performances between soccer clubs (in this case Arsenal and Manchester's victories in competitions) can be distinguished through weighted criteria (on the left). Optimisation allows now the management of thousands of objects, still keeping the display simple, attractive and conceptually rich, far from simple lists or grids. A commercial application is now installed in a show room in Casablanca. This important industrial feed-back shows that we have partly reached our goal of bringing Galois lattices from experts to novices. At the time of writing this paper it is not yet possible to tell how many of these possibilities will turn out into commercial applications due to other industrial considerations than mere functional innovation or aesthetic interest. Industrials are still hesitating to invest in a new technology, whatever its interest, when it competes with well established simple technologies, unless it brings about new interesting services. This is confirmed by the contradiction we observed in the answers of Questions 5 and 6: although the probe was clearly preferred, three testers would continue to use Amazon because they were used to it. It may be a minority of people, but this minority represents industrial reality. To overcome this difficulty a new

technology must offer more. This is what we now explore. The probe approach should offer users a new insight in their data.



Figure 16: Experimental interface for a sports TV channel

New services: indexing and complementarity

We already know from Crampes et al. [9] that a semantic probe is a good means for indexing objects with objects. This is an important new functionality which we have improved a lot but which is not industrially explored yet. Another functionality is shown in this paper because it is particularly interesting in many domains and unveils new insights in databases. The probe can be used for searching for complementary data or objects.

It was drafted by our software partner for the human resource department of a big international electro-mechanics leader company. People from the company are tagged with their competences from a thesaurus; their location and their availability are also defined as properties. Figure 17 shows a snapshot of the mock-up whose database contains a hundred people (eyes and names are barred in this paper and tags are not shown for obvious privacy and confidential reasons). Another interesting feature of this mock-up is that the probe's tags are weighted through the use of sliders presented in Figure 15 and 16. When looking for a particular profile for a project, a human resource manager can load the probe with the expected competences and data. There will hopefully be some people meeting the criteria like the woman just under the probe in the figure. However it is more interesting to see how a team of people can be built from different people's competences. Groups differ according to some competences. Those that come up next to the probe partly meet the requirements. The union of their extents (comptences) may lead to a super-group whose extent matches the probe's profile. This new functionality which was suggested by our industrial visitors leads us to consider complementary concepts, i.e. relations between different concepts which may merge for meeting some overall requirements. This mock-up opens up new research directions for knowledge mining, complementary social networks and information visualization.

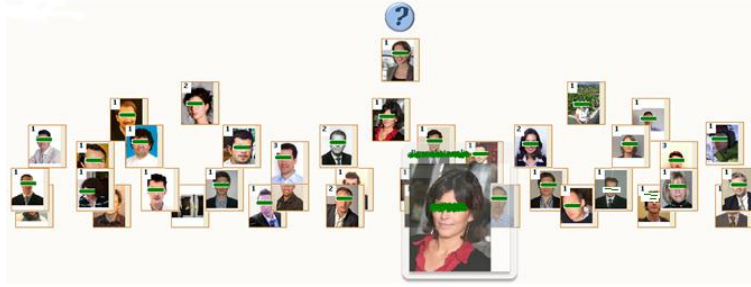


Figure 17: Experimental interface for a company with a hundred people tagged with their competences

4.3. Scalability issues

One of the main arguments in favour of the semantic probe paradigm is its visual scalability for novices compared to Hasse Diagrams. Consequently it is important to analyse how scalability impacts onto semantic probe displays. In a normal Galois lattice, the number of concepts upper bound equals $2N$ for N attributes. In practice it is never the case. First we must consider the number of groups because strictly similar attributes always regroup in the same concepts. Second it is shown in Godin et al. [21] that if K is a fixed upper bound on the number of attributes for each group, the number of nodes $|H|$ in the Hasse diagram is bounded by $2^K n$ where n is the number of attributes. Third these authors also show that in real cases because of the attributes' repartition this upper bound is between $4 \times n$ to $11 \times n$. In the semantic probe case we only display groups and concepts appear as a result of interaction. The upper bound of visible groups is equal to n if the probe were loaded with all objects, which is already 4 times to 10 times less than the number of nodes of a Hasse diagram. In practice, the probe should be loaded with few objects and a little proportion of groups should be visible. In normal usage, scalability is not a problem for semantic probes. Experimentation confirms these results: the Facebook photo tests involved more than 1000 items and most of industrial prototypes involve more than 10000 items.

Beyond this quantitative consideration, the most important visual complexity reduction factor is the absence of edges. Edge crossing is a key problem in graph drawing particularly in the case of object-attribute sets. It is known that a graph containing at least a 5 node clique (K_5) is not planar and necessarily opens up an essential problem of edge crossing visual difficulty. In general Hasse diagrams are far beyond this limit. Probe driven sub-hierarchies ignore this problem. This visual simplification explains the good performances in the controlled experiments and the welcoming by end users and industrialists.

5. CONCLUSION

Irrespective of what visualization strategy is employed, it is difficult to display object-attribute databases with their topological properties. Galois lattices

are good at solving this problem through the use of Hasse diagrams. They provide a powerful tool for knowledge analysis but they fall short from addressing the complexity and scalability bottlenecks for novices.

We proposed an interactive user-centric probe-driven strategy. Our results confirm that this approach, although it does not replace existing ones, improves navigation and is attractive for industrial partners in varying fields. However the issue of providing conceptually enhanced visualization solutions to users at the expense of user acceptance is still on-going. Simple experiments and hesitation among interested industrials show that a new technology must outperform in many ways a simple established technology to become attractive. This is why, although our probe approach shows many qualities, we consider that new services must be provided to reach industrial applications. Some promising experiences are being performed in this direction with assistance to indexing and the original idea of data complementarities.

6. References

References

- [1] Harith Alani. TGVizTab: An Ontology Visualisation Extension for Protég. *Proceedings of Knowledge Capture KCap03 Workshop on Visualization Information in Knowledge Engineering Sanibel Island Florida USA*, 2003.
- [2] Gabriela Arévalo, Anne Berry, Marianne Huchard, Guillaume Perrot, and Alain Sigayret. Sigayret A.: Performances of Galois sub-hierarchybuilding algorithms. In *ICFCA2007, LNCS/LNAI 4390*, pages 166–180, 2007.
- [3] Battista G., Eades, Tamassia, and Tollis. *Graph drawing. Algorithms for the visualisation of graphs*. Prentice Hall, 1999.
- [4] Ulrik Brandes, Sabine Cornelsen, Barbara Pampel, and Arnaud Sallaberry. Path-Based Supports for Hypergraphs. *Order A Journal On The Theory Of Ordered Sets And Its Applications*, pages 1–14, 2010.
- [5] Andreas Buja and Deborah F Swayne. Visualization Methodology for Multidimensional Scaling. *JOURNAL OF CLASSIFICATION*, 19:2002, 2001.
- [6] Claudio Carpineto and Giovanni Romano. Exploiting the Potential of Concept Lattices for Information Retrieval with CREDO. *Journal of Universal Computer Science*, 10(8):985–1013, 2004.
- [7] Chaomei Chen. Top 10 unsolved information visualization problems. *IEEE Computer Graphics and Applications*, 25(4):12–16, 2005.
- [8] Michel Crampes, Sylvie Ranwez, Jean Villerd, Filip Velickovski, Chris Mooney, Andrew Emery, and Nicholas Mille. Concept maps for designing adaptive knowledge maps. *Information Visualization*, 5(3):211–224, 2006.

- [9] Michel Crampes, Jeremy de Oliveira-Kumar, Sylvie Ranwez, and Jean Villerd. Visualizing Social Photos on a Hasse Diagram for Eliciting Relations and Indexing New Photos. *IEEE Transactions on Visualization and Computer Graphics*, 15:985–992, 2009.
- [10] W C de Leeuw and J J van Wijk. A probe for local flow field visualization. In *Visualization, 1993. Visualization '93, Proceedings., IEEE Conference on*, pages 39–45, October 1993.
- [11] Jon Ducrou, Björn Vormbrock, and Peter W Eklund. FCA-Based Browsing and Searching of a Collection of Images. In Henrik Schärfe, Pascal Hitzler, and Peter Ø hrstrøm, editors, *Proceedings of the 14th International Conference on Conceptual Structures (ICCS 2006)*, volume 4068 of *Lecture Notes in Computer Science*, pages 203–214. Springer, 2006.
- [12] Jon Ducrou, Peter Eklund, and Tim Wilson. An intelligent user interface for browsing and searching MPEG-7 images using concept lattices. In *Proceedings of the 4th international conference on Concept lattices and their applications*, CLA'06, pages 1–21, Berlin, Heidelberg, 2008. Springer-Verlag.
- [13] Tim Dwyer, Kim Marriott, Falk Schreiber, Peter Stuckey, Michael Woodward, and Michael Wybrow. Exploration of networks using overview+detail with constraint-based cooperative layout. *IEEE transactions on visualization and computer graphics*, 14(6):1293–300, 2008.
- [14] Peter Eklund and Jean Villerd. A Survey of Hybrid Representation of Concept Lattices in Conceptual Knowledge Processing. In Léonard Kwuida and Baris Sertkaya, editors, *18th International Conference on Formal Concept Analysis - ICFCA 2010*, volume 5986 of *Lecture Notes in Artificial Intelligence*, pages 296–311, Agadir, Maroc, 2010. Springer Berlin / Heidelberg.
- [15] Peter Werner Eklund, Jon Ducrou, and Peter Brawn. Information Visualization using Concept Lattices: Can Novices Read Line Diagrams? In Peter Eklund, editor, *Proc. of the 2nd Int. Conference on Formal Concept Analysis*, volume 2691 of *LNAI*, pages 57–72. Springer-Verlag, 2004.
- [16] Christiaan Fluit, Marta Sabou, and Frank Van Harmelen. Ontology-based Information Visualisation: Towards Semantic Web Applications. In Vladimir Geroimenko, editor, *Visualising the Semantic Web 2nd edition*, pages 36–48. Springer Verlag, 2005.
- [17] Santo Fortunato. Community detection in graphs. *Physics Reports*, 486(3-5):103, June 2009.
- [18] Ralph FREESE. Automated lattice drawing. *Lecture notes in computer science*, pages 112–127, 2004.
- [19] Ganter B. and Wille R. *Formal concept analysis: foundations and applications*. 1999.

- [20] Robert Godin and Hafehd Mili. Building and maintaining analysis-level class hierarchies using Galois Lattices. *SIGPLAN Not.*, 28(10):394–410, 1993.
- [21] Robert Godin, Rokia Missaoui, and Alain April. Experimental comparison of navigation in a Galois lattice with conventional information retrieval methods. *Int. J. Man-Mach. Stud.*, 38(5):747–767, 1993.
- [22] Robert Godin, Guy Mineau, and Rokia Missaoui. Incremental Structuring of Knowledge Bases. In *In Proceedings of the International Knowledge Retrieval, Use, and Storage for Efficiency Symposium KRUSE'95*, 1995.
- [23] Tim Hannan and Alex Pogel. Spring-Based lattice drawing highlighting conceptual similarity. In *Proceedings of the 4th international conference on Formal Concept Analysis, ICFCA'06*, pages 264–279, Berlin, Heidelberg, 2006. Springer-Verlag.
- [24] I Herman, G Melancon, and M S Marshall. Graph visualization and navigation in information visualization: A survey. *IEEE Transactions on Visualization and Computer Graphics*, 6(1):24–43, 2000.
- [25] Danny Holten. Hierarchical edge bundles: visualization of adjacency relations in hierarchical data. *IEEE Transactions on Visualization and Computer Graphics*, 12(5):741–748, 2006.
- [26] Nicolas Jay, François Kohler, and Amedeo Napoli. Analysis of social communities with iceberg and stability-based concept lattices. In *ICFCA'08 Proceedings of the 6th international conference on Formal concept*, pages 258–272, February 2008. ISBN 3-540-78136-6, 978-3-540-78136-3.
- [27] T Kamada and S Kawai. An algorithm for drawing general undirected graphs. *Inf. Process. Lett.*, 31(1):7–15, 1989.
- [28] T Kohonen, M R Schroeder, and T S Huang, editors. *Self-Organizing Maps*. Springer-Verlag New York, Inc., Secaucus, NJ, USA, 3rd edition, 2001.
- [29] Sergei O. Kuznetsov. On stability of a formal concept. *Annals of Mathematics and Artificial Intelligence*, 49(1-4):101–115, June 2007.
- [30] Sergei O Kuznetsov and Sergei A Obiedkov. Comparing performance of algorithms for generating concept lattices. *Journal of Experimental Theoretical Artificial Intelligence*, 14(2-3):189–216, 2002.
- [31] Boumedjout Lahcen and Kwuida L. Lattice Miner: A Tool for Concept Lattice Construction and Exploration. In *Supplementary Proceeding of International Conference on Formal concept analysis (ICFCA'10)*, 2010.
- [32] John Lamping, Ramana Rao, and Peter Pirolli. A focus+context technique based on hyperbolic geometry for visualizing large hierarchies. In *Proceedings of the SIGCHI conference on Human factors in computing systems*,

- CHI '95, pages 401–408, New York, NY, USA, 1995. ACM Press/Addison-Wesley Publishing Co.
- [33] Bongshin Lee, Greg Smith, George G Robertson, Mary Czerwinski, and Desney S Tan. FacetLens: exposing trends and relationships to support sensemaking within faceted datasets. In *Proceedings of the 27th international conference on Human factors in computing systems, CHI '09*, pages 1293–1302, New York, NY, USA, 2009. ACM.
- [34] James R Miller and Estela A Gavosto. The immersive visualization probe for exploring n-dimensional spaces. *IEEE Computer Graphics and Applications*, 24(1):76–85, 2004.
- [35] Andreas Noack. Modularity clustering is force-directed layout. *Physical Review E*, 79(2):9, 2008.
- [36] Andreas Noack and Claus Lewerentz. A space of layout styles for hierarchical graph models of software systems. *Proceedings of the 2005 ACM symposium on Software visualization SoftVis 05*, 1(212):155, 2005.
- [37] Catherine Plaisant. The challenge of information visualization evaluation. *Proceedings of the working conference on Advanced visual interfaces AVI 04*, 11(4):109–116, 2004.
- [38] Michel Plantié and Michel Crampes. From Photo Networks to Social Networks, Creation and Use of a Social Network Derived with Photos. *Proceedings of the ACM international conference on Multimedia , Firenze, Italy, October*, 2010.
- [39] John C Platt. Fastmap, MetricMap, and Landmark MDS are all Nyström algorithms. In *Proceedings of the 10th International Workshop on Artificial Intelligence and Statistics*, volume 25 of *AISTATS '05*, pages 261–268. Society for Artificial Intelligence and Statistics, 2005.
- [40] Nathalie Henry Riche and Tim Dwyer. Untangling euler diagrams. *IEEE Transactions on Visualization and Computer Graphics*, 16(6):1090–1099, 2010.
- [41] Nathalie Henry Riche and Jean-daniel Fekete. Novel Visualizations and Interactions for Social Networks Exploration. *Network*, pages 611–636, 2010.
- [42] Camille Roth, Sergei Obiedkov, and Derrick Kourie. Towards Concise Representation for Taxonomies of Epistemic Communities. In Sadok Ben Yahia, Engelbert Mephu Nguifo, and Radim Bělohlávek, editors, *Concept Lattices and Their Applications*, volume 4923 of *Lecture Notes in Computer Science*, pages 240–255. Springer, Berlin/Heidelberg, 2008.

- [43] Hans-Jörg Schulz, Mathias John, Andrea Unger, and Heidrun Schumann. Visual Analysis of Bipartite Biological Networks. In *Eurographics Workshop on Visual Computing for Biomedicine*, Delft, Pays-Bas, 2008.
- [44] Ben Shneiderman. The eyes have it: a task by data type taxonomy for information visualizations. *Proceedings 1996 IEEE Symposium on Visual Languages*, (UMCP-CSD CS-TR-3665):336–343, 1996.
- [45] G Smith, M Czerwinski, B Robbins Meyers, G Robertson, and D S Tan. FacetMap: A Scalable Search and Browse Visualization, 2006.
- [46] Andre Suslik Spritzer and Carla M D S Freitas. A physics-based approach for interactive manipulation of graph visualizations. *Evaluation*, pages 271–278, 2008.
- [47] Marc Streit, Alexander Lex, Michael Kalkusch, Kurt Zatloukal, and Dieter Schmalstieg. Caleydo: connecting pathways and gene expression. *Bioinformatics*, 25(20):2760–2761, 2009.
- [48] Gerd Stumme, Rafik Taouil, Yves Bastide, Nicolas Pasquier, and Lotfi Lakhal. Computing iceberg concept lattices with TITANIC. *Data and Knowledge Engineering*, 42(2):189–222, 2002.
- [49] Petko Valtchev, David Grosser, Cyril Roume, and Mohamed Rouane Hacene. Galicia: An Open Platform for Lattices. In *In Using Conceptual Structures: Contributions to the 11th Intl. Conference on Conceptual Structures (ICCS'03)*, pages 241–254. Shaker Verlag, 2003.
- [50] Frank Van Ham and Adam Perer. "Search, show context, expand on demand": supporting large graph exploration with degree-of-interest. *IEEE Transactions on Visualization and Computer Graphics*, 15(6):953–960, 2009.
- [51] J Villerd, S Ranwez, M Crampes, and D Carteret. Using concept lattices for visual navigation assistance in large databases. *International Journal of General Systems*, 38(4):405–425, 2009.
- [52] Rudolf Wille. Methods of Conceptual Knowledge Processing. In Rokia Misraoui and Jürg Schmidt, editors, *Formal Concept Analysis*, volume 3874 of *Lecture Notes in Computer Science*, pages 1–29. Springer Berlin / Heidelberg, 2006.
- [53] Ka-Ping Yee, Kirsten Swearingen, Kevin Li, and Marti Hearst. Faceted metadata for image search and browsing. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, CHI '03, pages 401–408, New York, NY, USA, 2003. ACM.
- [54] Ji Soo Yi, Rachel Melton, John Stasko, and Julie A Jacko. Dust & Magnet: multivariate information visualization using a magnet metaphor. *Information Visualization*, 4(4):239–256, 2005.