Otto-von-Guericke
University Magdeburg

Faculty of Computer Science

Data and Knowledge Engineering Group

Bachelor's Thesis

Interactive Visualization of Large Concept Lattices for Exploratory Search

Author:
Johannes Filter

Advisors:
Prof. Dr. Andreas Nürnberger
Otto-von-Guericke University Magdeburg

Prof. Dr. Ana García-Serrano
Universidad Nacional de Educación a Distancia

August 2, 2015

Abstract

Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like “Huardest gefburn”? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

Zusammenfassung

Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like “Huardest gefburn”? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

Acknowledgements

Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like “Huardest gefburn”? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

Contents

Acronyms

DT Dynamic Taxonomy. viii, 13, 23, 25

FCA Formal Concept Analysis. viii, ix, 1, 3, 6, 7, 12–15, 18, 19, 22–26, 33, 34, 36, 43

HCI Human-Computer Interaction. 9, 11

IR Information Retrieval. viii, 3, 6, 7, 13, 15, 17, 24

1. Introduction

The digital revolution is affecting every part of our life. Also the humanities scholars witness a change in their work life when analog collections are digitized. They have to apply computer science methods to organize and analyze huge amount of data. The area “digital humanities” evolved during the last ten years and it can be defined as an “intersection between the humanities and information technology” [46].

The computer science department of the Universidad Nacional de Educación a Distancia (UNED) in Madrid (Spain) cooperates with human scholars to conduct research in the digital humanities. In this project[1], there are historical maps which have been digitized and annotated. To extract knowledge from the collection, the research group advocates for the use of Formal Concept Analysis (FCA) for topic organization [12, 13]. They successfully implemented a FCA algorithm but lack an interactive user interface which will be developed in this thesis. The interface will be evaluated with an user study.

While FCA is a mathematically well-founded principle, the resulting visualization of large concept lattices is a problem. Large concept lattices arose when you apply FCA to large amount of entities (details will be explained in Section 2.1). When applying FCA to a document collection, you are likely to encounter huge amount of entities. That is why alternative visualization techniques are needed.

This thesis focuses on concept lattices that were built from document collection. In this scenario, FCA is especially valuable when the user does not know anything about the collection and wants to browse and explore the data.

[1] `http://linhd.uned.es/p/proyecto-dimh`

This user interface will run in modern web browser to avoid dealing with a heterogeneous operating systems environment. Because of the fast-changing environment of the web, it is important to keep with the latest technologies and techniques to not fall apart. The software utilizes a large amount of frameworks to efficiently create a pleasant user interface.

The remainder of this thesis is structured as follows: The background of Formal Concept Analysis and Interface Design are presented in Section 2. The discussion of related work takes place in Section 3. Inferring from the related work, I will present my work including idea and implementation in Section 4. In Section 5, the evaluation takes place. Built upon the results of the evaluation, a new version is presented in Section 6. Eventually concluding in Section 7.

2. Background

This sections gives background knowledge before we can discuss related work in the following section. First, it gives an introduction into formal concept analysis taken from the work from Ganter and Wille [24]. Second, it introduces some user interface design principles.

2.1 Formal Concept Analysis

Formal Concept Analysis (FCA) is a mathematically well-founded technique to analyze data. FCA creates relationships among objects specified by attributes. It derives from old philosophical ideas and was formalized by Rudolf Wille. First, we describe the formal definitions and explain them with examples. Second, we give examples for the use of FCA in Information Retrieval (IR).

2.1.1 Definition

FCA [24] is constructed on a formal context. A *formal context* is defined as a tripple $K = (G, M, I)$ where G is a set of objects[1], M is a set of attributes[2] and I is a binary relation $I \subseteq G \times M$. I specifies whether an object has an attribute or not[3]. Table 2.1 illustrates an example from David Eppstein [22] where G comprises the integers from 1 to 10 and M comprises the attributes composite, even, odd, prime and square.

Let the operator $'$ for $A \subseteq G$ be defined as following:

$$A' = \{m \subseteq M \mid I(g, m)\ \forall g \in A\}$$

[1] G derives from German "Gegenstände".
[2] M derives from German "Merkmale".
[3] I derives from German "Inzidenzrelation".

Table 2.1: Formal context, integers 1 to 10 as objects with attributes

	composite	even	odd	prime	square
1			×		×
2		×		×	
3			×	×	
4	×	×			×
5			×	×	
6	×	×			
7			×	×	
8	×	×			
9	×		×		×
10	×	×			

A' is the set of those attributes that are present in all objects from given A.

Let the operator $'$ for $B \subseteq M$ be defined as following:

$$B' = \{g \subseteq G \mid I(g, m)\ \forall m \in B\}$$

B' is the set of objects that have at least the attributes given in B.

If for $A \subseteq G$ such that $A = A''$, then A is called *closed*. The same is true for $B \subseteq M$ and $B = B''$.

For example, let a set of objects be defined as $A_1 = \{1, 4\} \subseteq G$. This results into: $A_1' = \{square\}$ and $A_1'' = \{1, 4, 9\}$. A_1 is not closed but $A_2 = \{1, 4, 9\} \subseteq G$ is called closed because $A_2 = A_2''$.

A *formal concept* is a pair of (A, B) where $A \subseteq G$ and $B \subseteq M$ and $A = B' \wedge B = A'$. Informally, all objects in A share exactly the same attributes in B. A is a set of objects called the *extent* of a formal concept. B is a set of attributes called the *intent* of a formal concept. The extent and the intent of all formal concepts are always closed.

From the example in Table 2.1, we can derive several formal concepts. Three randomly chosen concepts are shown in Table 2.2.

Table 2.2: Three formal concepts from the formal context in Table 2.1

Concept	Extent	Intent
C_1	{4,6,8,10}	{composite, even}
C_2	{2,4,6,8,10}	{even}
C_3	{9}	{composite, odd, square}

It is always possible to define an order relation on the formal concepts. Let us introduce the relation $\leq$ as follows:

$$(A_i, B_i) \leq (A_j, B_j) \Longleftrightarrow A_i \subseteq A_j$$

With the help of $\leq$, we can derive relationships from the concepts in Table 2.2. We see that $C_1 \leq C_2$. This means that C_1 is more specific than C_2 and C_2 is more general than C_1. We can also see that C_3 is unrelated to C_1, and that C_3 is unrelated to C_2.

A formal context with $\leq$ is called a *concept lattice* of the context. It can be shown, that for two formal concepts C_i and C_j, there always exists a formal concept C_x such that $C_i \leq C_x \wedge C_j \leq C_x$. That means that there is always a formal concept wich is 'above' in the hierarchy and also related to the two formal concepts C_i and C_j. A formal definition would exceed this section. The interested reader is advices to read "The Basic Theorem on Concept Lattices" as described by Carpineto and Romano on page 13 in their work [9].

In the next section we will take a look at the static visualization of concept lattices.

2.1.2 Static Visualization

It is often said that a picture is worth a thousand words. To convey the abstract insights of a concept lattice, it can be visually represented in a *Hasse diagram* [24]. Figure 2.1 shows the Hasse diagram of the concept lattice derived from the formal context described in Table 2.1.

A Hasse diagram is a graph where the vertices represent formal concepts and edges represent the relation $\leq$ among the formal concepts. An edge between formal concepts C_i and C_j is drawn, when $C_i \leq C_j$ and there does not

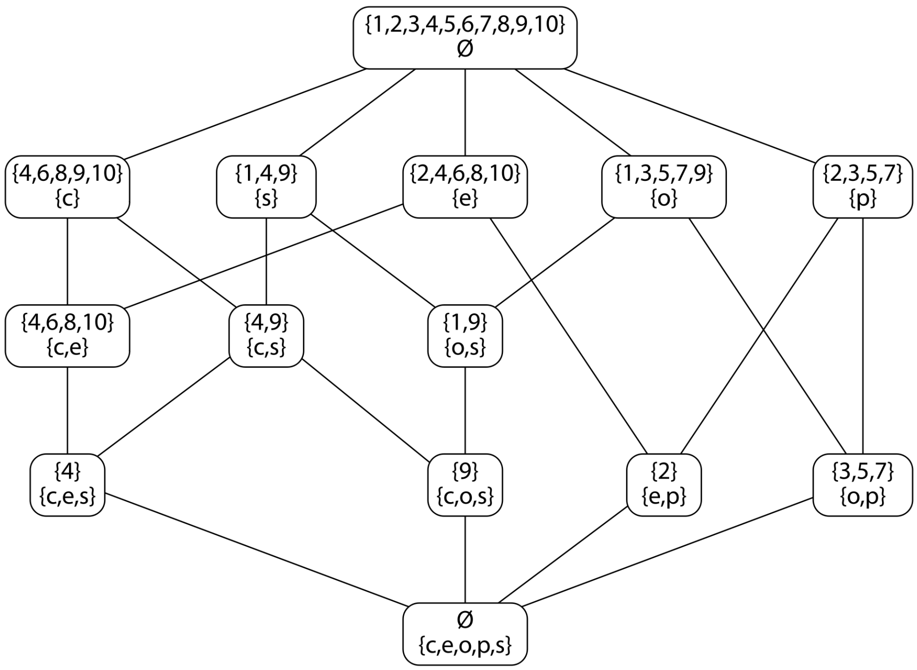

Figure 2.1: Hasse diagram, with the integers 1 to 10 as objects and attributes square (s), prime (p), composite (c), even (e), and odd (o). Drawn by Eppstein [22]

exist a formal concept C_x such as $C_i \leq C_x \leq C_j$. To increase the readability, the nodes are ordered in layers. The concepts in top are more general, the concepts in the bottom are the more specific ones.

There are two special concepts: *supremum* and *infimum*. The supremum is vertex node in the top and the attributes in its intent are those which are present in all objects. In most cases its intent is empty. The infimum is the vertex in the bottom and the objects in its extent are those which have all attributes. In most cases its extent is empty.

After this general introduction, we will describe in the next section how we can apply FCA to IR.

2.1.3 Application of FCA in IR

Up to now, we only showed primitive examples to illustrate the basics of FCA. So where was it applied? According to Poelmans et al. FCA has been "applied in many disciplines such as software engineering, knowledge discovery and IR" [40]. They did two comprehensive surveys on the application of FCA [40, 41].

In the case of IR on a document collection, the objects are the documents. These documents are described by index terms. These index terms are the attributes of the objects. In Table 2.3 are documents described by index terms take from Godin et al. [25]. The concept lattice is visualized in Figure 2.2.

Table 2.3: Documents described by Index Terms, from Godin et al. [25]

Document	Index Terms
1	{animal, bear, canada, child, cow-boy, dream, fantasy, immigration, indian, magic}
2	{animal, cat, child, fantasy, magic, tale}
3	{animal, child, dog, fair, fantasy, love, parade}
4	{child, fantasy, friendship, game, rope}
5	{creativity, child, fantasy, game, music, sound}
6	{animal, child, dream, fair, fantasy, friendship, octupus}

Carpineto and Romano [11] describe that a concept can be seen as a query (intent) with a set of retrieved documents (extent) and that neighboring concepts can be seen as minimal query changes. When the user queries a FCA-based system, the intents of the formal concepts are checked for a match (or a partial match if there is no match in the first place). So for instance with the query "child fantasy friendship" the document 4 and 6 wold be retrieved. You find the concept in the middle of the Hasse diagram. The neighboring concepts are connected with an edge.

Sacco and Tzitzikas [42] describe two information access modes: *focalized search* and*exploratory search*. In focalized search, the user is interested to find relevant information to a given query. In opposite to exploratory, where the user explores relationships among items in a data collection. While FCA provides the user with the ability to do focalized search, it offers sophisti-

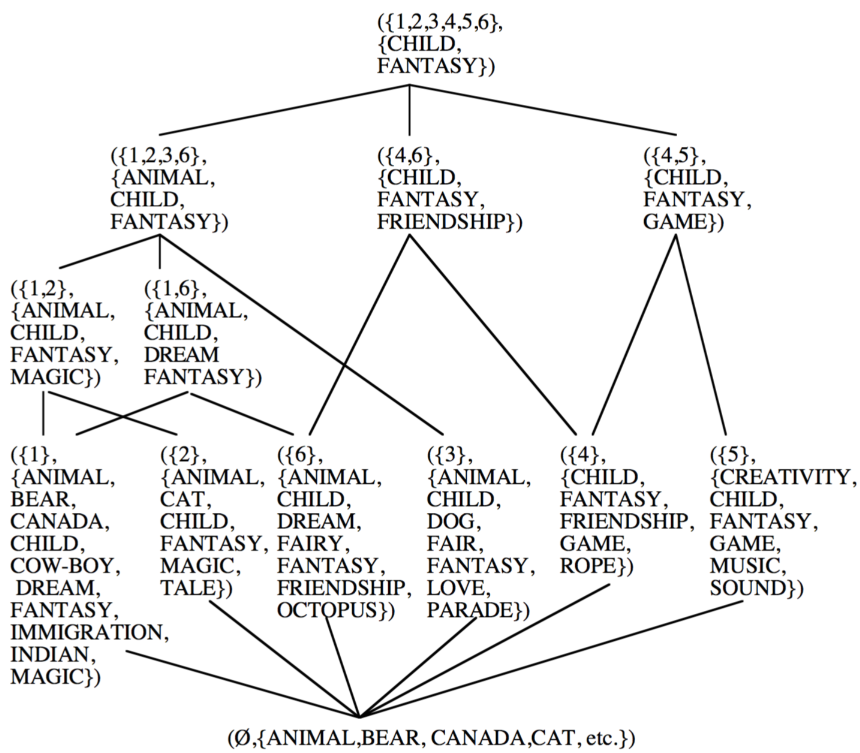

Figure 2.2: Hasse diagram from concept lattice derived from 2.3, from Godin et al. [25]

cated exploratory capabilities. When the user queries the system (as described above), not only the documents of the formal concepts are retrieved, but she can also see the position in the lattice. Is the concept in the upper half and thereby more general? Or is it in the bottom half and thereby more specific? What are the related formal concepts? How do they differ? Would there be a lot more documents if I would remove a term from the query? All this questions can be answered and because the whole lattice is already computed, relatively fast. Without going into detail, a major problem of FCA is the time-consuming creation of a lattice. But this focus of this thesis is the visualization of concept lattices and not the creation of concept lattices. This section should give a slight introduction and the interested reader is advised to study the work from Carpineto and Romano [9] for a detailed investigation.

This work develops an interface for one specific datasets. That is why it is important the lattice structure of this dataset how it was created. The next section described the dataset.

2.1.4 The Data

In the following, I will described how the research group created the data on which this thesis is based.

The research group asked human scholars for a list of 100 interested words. This list is the dictionary. The index terms of a documents specify if this index terms is present in the document. They were 7492 documents with 100 possible attributes, the index terms. They result into a concept lattice with 131379 formal concepts.

Because this thesis focuses on the user interface, let us review user interface design principles before we discussed related work in Section 3.

2.2 Interface Design

The user interface is responsible for the interaction with users. The interaction of humans with computers has its own research area, Human-Computer Interaction (HCI), and one of its pioneers is Ben Shneiderman. In the following, two principles from him are presented: The "Eight Golden Rules of Interface Design" [44] and the "Visual Information Seeking Mantra" [43].

2.2.1 Eight Golden Rules of Interface Design

These rules [44] are genreal advices for user interface designers which should apply to all interfaces. The rules are named and explained with my own remarks.

- Strive for consistency: Use similar actions in similar situations. Use identical terminology, colors, fonts etc. throughout the system.
- Cater to universal usability: Design for the needs of a diverse user group (skill level, age, gender and others)

- Offer informative feedback: Give system feedback for every action.
- Design dialogs to yield closure: Sequences of actions should be grouped. Give feedback on completion of a group.
- Prevent errors: Design the system that the user cannot even do errors in the first place. But if she does some, offer instructions how to recover.
- Permit easy reversal of actions: Actions should be undone. This gives the user confidence to explore the system.
- Support internal locus of control: The user should think that she is in charge of control.
- Reduce short-term memory load: Reduce the number of things the user has to keep in mind while using the system.

There exits alternative principles for instance: Donald Norman's Design Principles [39] or Jakob Nielsen's "10 Usability Heuristics for User Interface Design" [38]. They are very similar and only differ points not worth mentioning.

These principles can be applied to all user interfaces. In the next section, design principles will be presented which are more related to this work.

2.2.2 Visual Information Seeking Mantra

The visual information seeking mantra (the Mantra) was introduced by Ben Shneiderman [43] and is based on his experience with past projects. Albeit the Mantra was intended to be a "descriptive and explanatory" [6], "in effect, the Mantra has become a prescriptive principle for many information visualization designers" write Craft and Cairns [15].

The Mantra describes design principles for interfaces when user view collection of items. This items are described by multiple attributes. The starting principles are: overview first, zoom and filter, and then details on demand. These four principles will be explained below and extended by three other principles.

- Overview: Gain an overview of the entire collection.
- Zoom: Zoom in on items of interest.

- Filter: Filter out uninteresting items.
- Details-on-demand: Select an item or group and get details when needed.
- Relate: View relationships among items.
- History: Keep a history of actions to support undo, replay, and progressive refinement.
- Extract: Allow extraction of sub-collections and of the query parameters.

Some tasks need more explanation.

Zoom and Filter

This task are responsible for reducing the complexity of the data collection. 'Zoom' means that the user focuses on items she wants to see. 'Filter' means that she can hide items which are not interesting for her.

History

It is important to give the user the possibility to easily recover from mistakes or loss of interest. In addition, "it is rare that a single user action produces the desired outcome. Information exploration is inherently a process with many steps, so keeping the history of actions and allowing users to retrace their steps is important." writes Shneiderman [43].

Extract

Once interesting objects are found, the user should have the possibility to extract them from the system. Shneiderman describes printing, emailing or saving the item to the disk as 'extraction'.

2.2.3 Final Remarks

The presented ideas are based mostly on the experience of one person: Ben Shneiderman. The huge number of citations show that his work is influential but Craft and Cairns [15] call for empirical justification of the Mantra. HCI is a young research area and they will come better and more polished guidelines the future. Until then, the work from Ben Shneiderman seems to be valid starting point.

Those principles are up to interpretation and adaption. Every system is different and has its different needs. For this reason, let us review what other people did and how they designed their interface for FCA-based systems.

3. Related Work

After introducing formal concept analysis in Section 2.1, let us review and discuss related work. In the first three sections, we go over different FCA-based approaches. Eventually, we evaluate one FCA-based approach in detail: The Virtual Museum of the Pacific. In Section 3.4, a non-FCA based approach is shown which is related to FCA: DT and Faceted Search.

3.1 Full Hasse Diagrams

The traditional, static visualization of concept lattices are Hasse diagrams as describes in Section 2.1. Eklund et al. [20] conducted user studies and proclaim that non-FCA-experts can read Hasse Diagrams if you fine-tune the Hasse diagram. For instance by choosing appropriate colors, using symbols and carefully positioning the vertices in layer.

But in the domain of IR you get formal contexts with a lot of objects. Those Hasse diagrams scale bad for large concepts lattices. Kuznetsov et al. [33] describe this resulting visualization: "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." Especially the high connectivity of the graph results in enormous edge crossing. The Figure 3.1 shows the first results from data in the digital humanities project. The software [HAS TO BE DETERMINED] was used.

The visualization is useless because it is not even possible to see all the labels. In the next section we describe techniques to improve the situation.

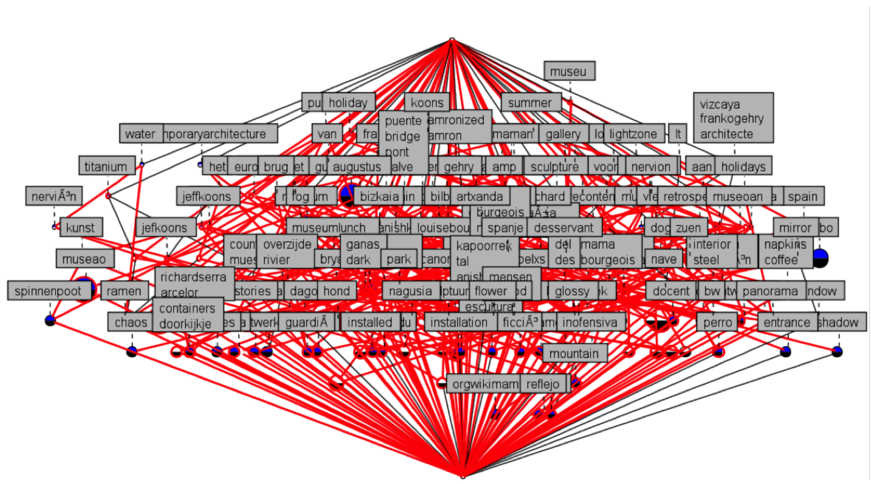

Figure 3.1: First visualization of digital humanist data with traditional FCA software

3.2 Pruned Hasse Diagrams

The Hasse diagrams can be pruned by reducing the number of vertices. The different techniques are discussed in the next section.

3.2.1 Reduce Number of Formal Concepts

One way to reduce the number of vertices is to compute the *iceberg lattices* as described by Stumme et al. [45]. They result after the application of a data mining technique “frequent item-set mining” from Agrawal et al. [1]. Only formal concepts are selected which are considered “frequent”. A formal concept is frequent if its intent, the set of attributes, is frequent. Let B be the intent and $minSupport \in [0, 1]$, then B is frequent if $|B'|/|G| \geq minSupport$. This means the attribute set has to specify a high portion of objects; at least $minSupport$. This approach has some drawbacks as Kuznetsov et al. [33] point out that “exotic” or “emergent” concepts that are not represented by a large number of objects can be interesting too and should not been overlooked. They propose to only select “stable” concepts [33]. The intent of stable concepts does not depend much on each object of the extent. It is also possible to apply traditional cluster techniques like fuzzy K-Means clustering to FCA [2].

While all these techniques undoubtable reduce the number of formal concepts, it is to question if the results are any helpful. In our case of IR, we apply FCA to explore the data and get insights about the lattice structure. When pruning the nodes, you are losing many data relationships, many formal concepts and, consequently, the “power” of FCA as exploratory technique is significantly reduced. When we deal with large concept lattices, the number of nodes has to be very low if we want to represent them with Hasse diagrams. Nowadays, the question is not how do I visualize 16 formal concepts as in Figure 2.1 - it is more how can I analyze 160.000 formal concepts.

Pruning alone is not a proper way to handle large concept lattice. But it can be useful to reduce the clutter. It can be useful in combination with techniques that are presented in Section 3.3.

3.2.2 Nest Formal Concepts

Another approach are *nested line diagrams* - line diagrams are another name for Hasse diagrams. For this, all attributes are partitioned into layers. For example, if you just have two layers, an attribute is either in layer one or two. For the first layer: You built up a Hasse diagram with the attributes *solely* of the first layer. For each vertex in the resulting Hasse diagram, you built up a Hasse diagrams *inside* the vertex. This secondary Hasse diagrams are built from the concept lattice derived from the objects in the vertex (the objects that are the extent of the formal concept). This can be done for an arbitrary number of layers. An example from Carpineto and Romano [9] is shown in Figure 3.2. The general idea should be clear without explaining the context - if not Carpineto and Romano [9] describe it more in detail.

But how to partition the attributes? You have to select the partitions manually. The manual selection might be good idea of small concept larges but in our case it is not feasible.

3.3 Local View

Instead of showing the full Hasse diagram, the user can have a local view on the lattice. We will give an overview about the basic idea and applications before we review one real-world application in detail.

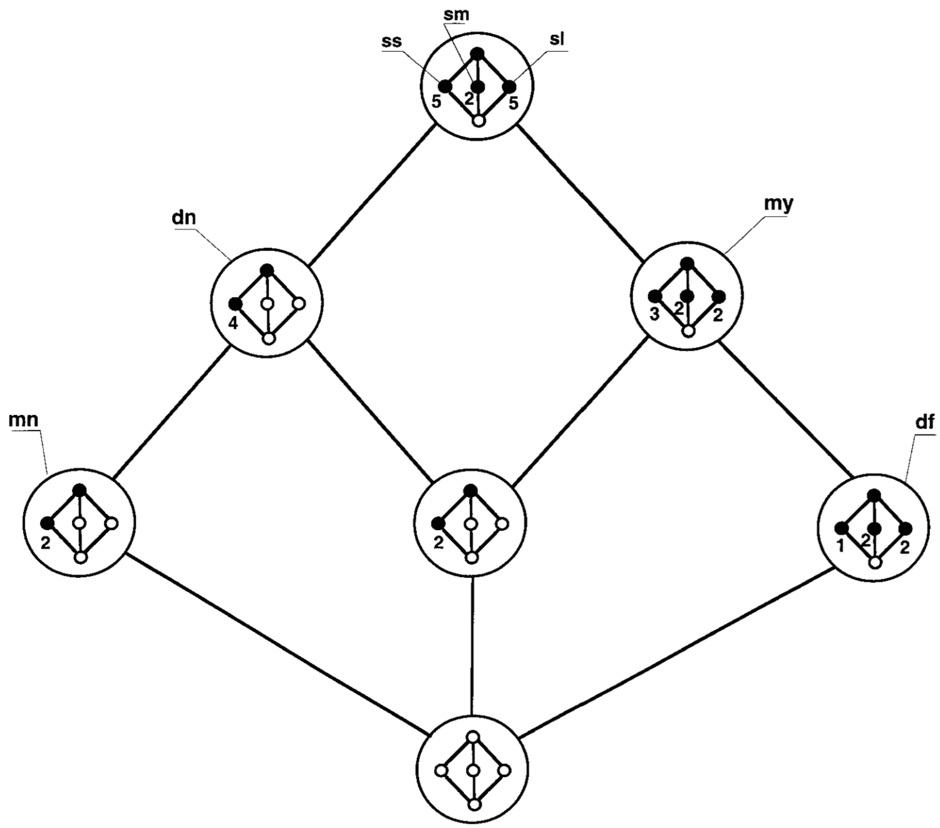

Figure 3.2: Nested line diagram with two layers. [9]

3.3.1 Introduction

One could argue that you just have to visualize everything and then allow to zoom on nodes. This techniques is common among network visualizations [28]. But because of the high connectivity of the graph, this is not helpful to Hasse diagrams. You can see this at the tool 'FCART' presented by Neznanov and Parinov [37]. In Figure 3.3 they visualize a concept lattices comprising more than 20000 concepts.

Even though they chose a grey color for the edges, the screen is almost filled only with the grey. In addition, they labels of the nodes in the top and bottom are overlapping which makes it hard to read them. Let us review some other approaches in the next section which completely break with the Hasse diagram.

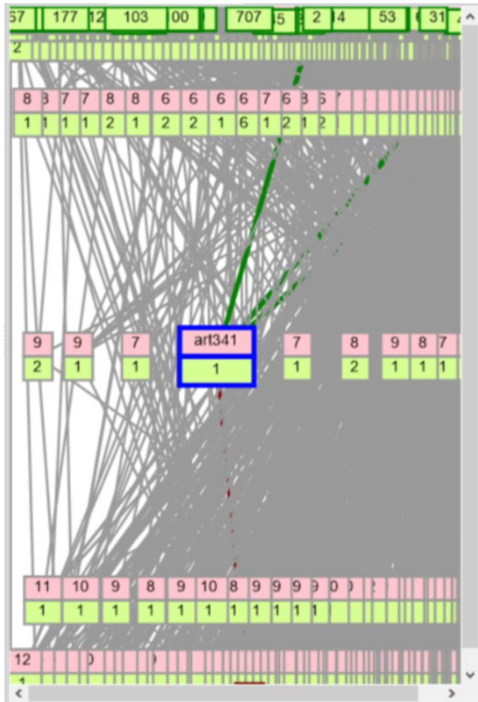

Figure 3.3: FCART, Hasse diagram visualization with over 20000 formal concepts, focus on node with blue borders [37]

The local view on a Hasse diagram is called *conceptual neighborhood* by Eklund et al. [19, 21] or *hybrid navigation* by Carpineto and Romano [8]. The basic idea is always the same: The interface is always focused on exactly *one* formal concept. The user can navigate through the lattice by going up (becoming more general) or going down (becoming more special). Which means removing terms or adding terms. They also offer the possibility to query the system. In most cases, the user would start with a search and focuses on the corresponding formal concept if it exists. From there, the user can fine-tune the search. The idea originated from the IR field and was first proposed by Godin et al. [26].

At least three ideas underly this approach. First, users tend to start with a short query and then refine their needs. Hearst [27] write while referring to [35, 4]:

> "A commonly-observed search strategy is one in which the in-

formation seeker issues a quick, imprecise query in the hopes of getting into approximately the right part of the information space, and then doing a series of local navigation operations to get closer to the information of interest."

Second, it easier for the users to choose from suggestions than to formulate a query. Aula [3] writes that searching is more demanding method for locating information than browsing, as it involves several phases, such as planning and executing queries, evaluating the results, and refining the queries, whereas browsing only requires the user to recognize promising-looking links.

Third, after the initial search, the neighboring concepts are basically suggestions to the users. This prevents them from getting empty results. This is related to the design principle: "Prevent errors" presented in Section 2.2.1. Zero results are not really errors, but they can be seen as a failure in a search process.

Godin et al. [25] evaluated their FCA-based approach in comparison to boolean retrieval and hierarchical retrieval and proclaim that "[their experiment] suggests that retrieval using a Galois lattice structure may be an attractive alternative since it combines a good performance for subject searching along with browsing potential." Galois lattice is a synonym for concept lattice.

3.3.2 Applications

Carpineto and Romano picked up the idea from Godin et al. and developed a FCA search engines ULYSSES [7, 8]. The user can fine-tune what neighboring vertices are displayed by bounding the information seeking space. They are not only showing directly adjacent vertices but also vertices that do not exceed a given distance. It is also possible to restrict the space to vertices which are above, below, left or right of the focus. The system is shown in Figure 3.4.

In their work, CREDO [10], Carpineto and Romano followed the look of ordinary search engines. The presentation of the concept lattice is not oriented at the Hasse Diagram. It looks more like a folder structure. It is shown in 3.5. Work that is similar comes from Koester [32], Dau et al. [16], Nauer and Yannik [36] and Cigarran et al. [14]. In all these cases, FCA

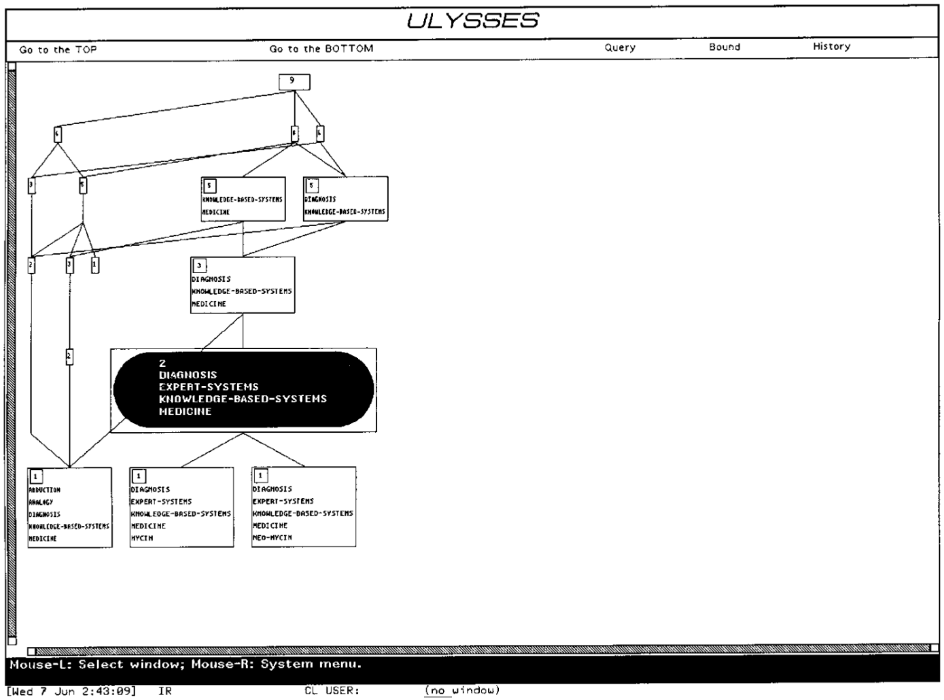

Figure 3.4: ULYSSES with focus on the black node [8]

is applied in slightly different manner. The search is done with ordinary search engines in the background (e.g. Yahoo!) and the concept lattice is built from the results of the search. So for every new search, there is a new concept lattice. This is different from our approach, where there is just one static concept lattice.

Let us now review work of FCA on document collections. Eklund et al. applied FCA to email organization [20], image browsing [18, 17] and a later work is the 'Virtual Museum of the Pacific' [19, 21]. We will focus on the museum because it does exactly what we are trying to do: Visualize a concept lattice built from image metadata. Furthermore it is a rare example of FCA outside of the academic community. It also runs in the browser and it was built in 2009 - so it is fairly recent. In addition, they conducted a usability study with museum experts and non-experts [21].

Figure 3.5: Screenshot of CREDO, after query 'xml' and browsing after 'editor(5)' and 'windows free(2)' [10]

3.3.3 Virtual Museum of the Pacific

The museum was created to give user the possibility to browse images of museum objects. It is available on the web[1] and it is advised to take a look at it before continue reading.

After the login, for a first-time visiter it is not easy to figure out where to start. It would be a great idea to let then start with the 'word cloud' to start at a point. The designers contradicted the first part of the Mantra: Overview first. This can be a problem because potential users do not know where start.

The interface sets the objects, the images, in the focus of the interface. This is, in my opinion, a good decision. The user is probably more interested in the objects than in the concept lattice. The lattice is just an artificial

[1]`http://epoc.cs.uow.edu.au/vmp/` - Credentials are required. Use username: filter and password: 45755

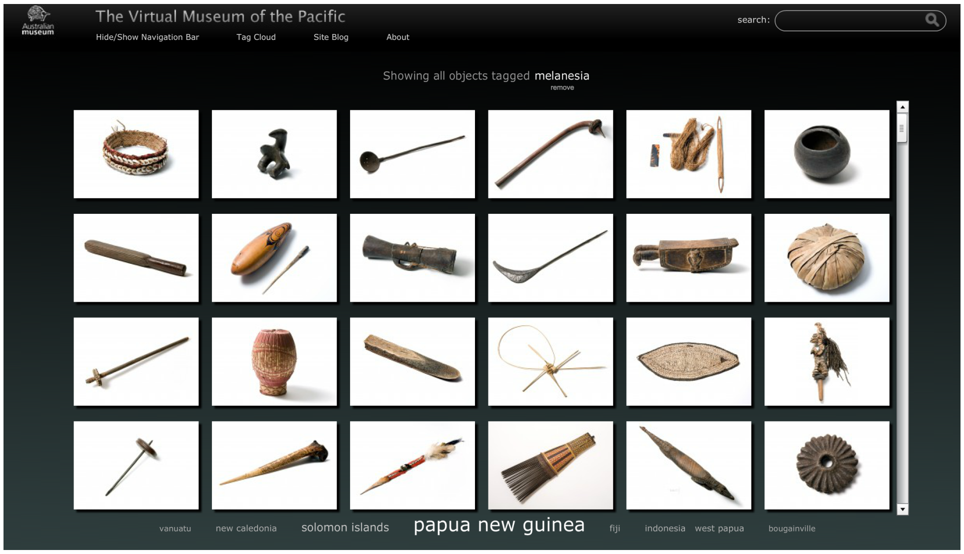

Figure 3.6: The Virtual Museum of the Pacific, focus on concept 'melanesia'

structure that was built around the objects.

Above the objects is the intent of the focused formal concept. It is possible to remove terms from the current selection to generalize (go up in the lattice). Below the objects are terms suggested to specify the information needs (go down in the lattice). The view changes when the users decide to add or remove terms. The sidebar categorizes the attributes into topics. It looks like they have created a taxonomy for their attributes. This allows the users to filter out interesting terms. But the manual selection is time-intensive and it does not scale very well. It is also to question if FCA should be applied to datasets with a taxonomy. More information will follow in section 3.4.

To get details on demand, you have to click on an item. From there you can also find related items. This views offers a good view.

It is possible to search but this search is very rudimentary. They should have stick to well-proven user search interface principles as described by Hearst [27].

The biggest problem is the missing 'home' or 'reset' button and the absence of any form of history. As Shneiderman [43] point out "Information exploration is inherently a process with many steps, so keeping the history of actions and allowing users to retrace their steps is important." FCA is a exploratory technique. Not implementing *any* possible to go back is a huge problem. You can see in the evaluation that users were confused. Eklund et al. write [21]:

> "Users also had difficulty understanding the notion of conceptual navigation as a way of navigating an information space, rather than a fixed hierarchy with a well defined 'home' state. [..] Many users felt 'lost' within [the FCA-based] style of navigation. A recommendation was put forward so that users could at least back track through the navigation sequences (in the form of a 'Back' button) or that users could easily go back to a 'home' or 'reset' state."

They are not addressing this problems - even thought the navigation is *the* problem with FCA. It feels like, they blame the problems on the unfamiliarity of the users [21]: "there were a number of common issues, mostly relating to the unfamiliarity of the interface". This is a weak argument because the users are always unfamiliar with new interfaces.

Combined with meaningless quotes from participants [21]: "There should be more of this kind of thing, but there's room for improvement" or "I think it's really good—it gives specialized information." and a weak summary [21]:

> "Overall, the study concludes that conceptual based browsing can offer significant merit for browsing Pacific collections, and that by resolving the identified issues, the core functionality of searching and browsing via FCA can be considerably easier for users to learn and engage with."

(Easier than what? There was NO comparison among other IR techniques) this paper leaves a bitter aftertaste. It started with a good idea but the implementation - in my opinion - has huge deficits. Sadly, they did draw the right conclusions and did not work out the problems of the implemenation.

3.3.4 Conclusion

The Virtual Museum of the Pacific is one of the view examples of FCA who breach out of traditional static visualizations. They idea to show only the directly related formal concepts is promising because this scales better than Hasse diagrams. My idea is based upon the work from the museum but addresses critical problems: the missing orientation, the missing history and rudimentary search implementation. The idea is described in section 4. Before that, let us review a related non-FCA-based technique, but feel free to skip it.

3.4 Dynamic Taxonomy and Faceted Search

Dynamic Taxonomy (DT) are used for faceted search. This section should only give a small overview about this technique. Completely introducing it is beyond the scope of the technique. Excessive information can be found in the work from Sacco and Tzitzikas [42]. To keep things short, this technique is explained with an example. Figure 3.7 shows a screenshot of Flamenco as described by Yee et al. [48].

This interface is similar to the 'local view' on Hasse diagrams presented in Section 3.3. The main idea: The dataset comprises several dimensions. Incrementally restrict those dimension to narrow down items of interest. In this picture you see the location restricted to 'Asia' and shapes etc. restricted to 'fabrics'. On the left are possible further restrictions suggested. In the center/right are the collections.

Sacco and Tzitzikas [42] state: "Although FCA and DT are apparently two distinct approaches to information modeling and access, and they use a different terminology, they are closely related." Sacco and Tzitzikas [42] compare DT to FCA:

> DT is less informative, but it ensures the results displayed to users is manageable because they are expressed in terms of the original taxonomy. This reduces the cognitive effort of users and is a simpler and more intuitive representation than the line diagrams generally used in FCA.

DT are widely used in eCommerce application. But they can only be applied to dataset with several dimensions. FCA can offer the same possibility to find relationships among arbitrary attributes. This is strength of FCA but it looks like that often people try to create taxonomies around the attributes. In this context I want to introduce a technique called *conceptual*

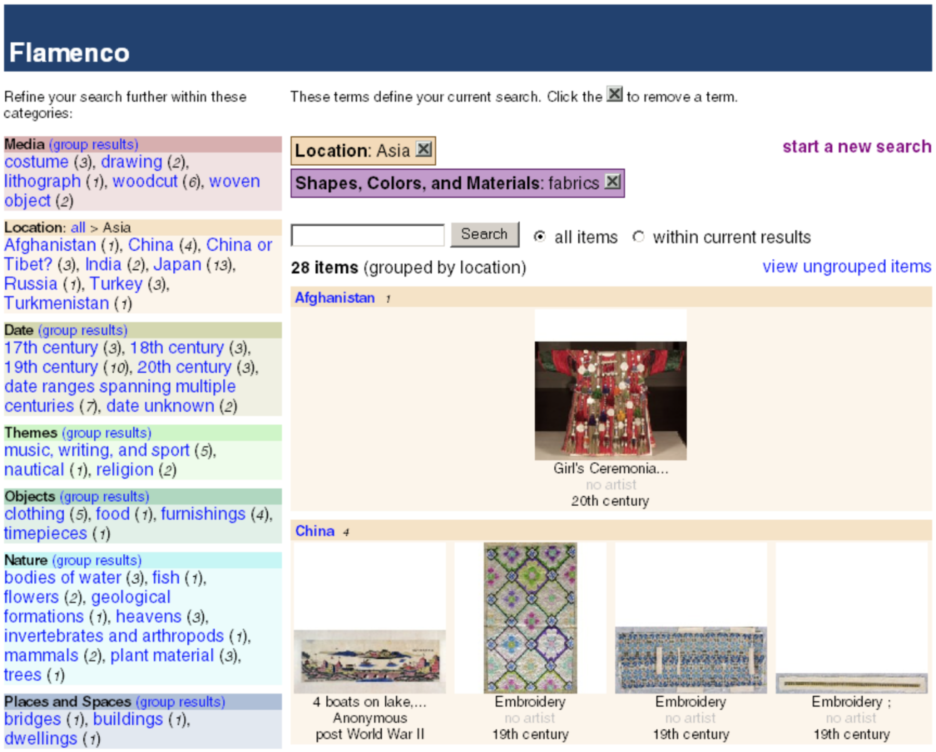

Figure 3.7: Flamenco with ongoing focus

scaling[9]. It is applied to dataset with many dimensions to reduce it into a single dimension. For example take a look at Table 3.1.

The objects are human beings and they are categorized by the attributes name, age and hometown. With conceptual scaling the attributes for the object Johannes are: “Name_Johannes”, “Age_24”, “Hometown_Schwerin” and correspondingly for the other people. Now we have only a single dimension which is needed for FCA.

But this method, albeit it is widespread, is, at least for IR, problematic. From a user-centered perspective, complexity reduction is the key point. How do we break massive, complex data set down to make them understandable for the users? So this transformation sounds like a contradiction - and it is. Zobel describes this kind of research as follows:

> Much research — far too much — is just misguided. People

Table 3.1: People with attributes

Name	Age	Hometown
Johannes	24	Schwerin
Ana	35	Madrid
Andreas	47	Magdeburg

> investigate problems that are already solved and well understood, or solve problems that technology has made irrelevant, or don't realize that the proposed improvement actually makes the method worse.

So for this reason, if you build up taxonomies upon your attributes or you are working on multidimensional data, it is better to drop FCA at all and use DT and faceted search.

4. Fancy FCA 1.0

I start by describing my concept inferred from discussing related work in Section 3. Followed by the description of the implementation which can be split into frontend development, responsible for the visual interface in the browser, and backend development, mostly responsible for handling data on the server.

4.1 Concept

The general idea comes from the Virtual Museum of the Pacific discussed in Section 3.3.3. But it had some problem: Missing orientation, missing history of actions and only a rudimentary implemented search interface. I try to address this problems and a schema of the proposed interface can be seen in Figure 4.1.

First, he interface should remind users of modern search engines: Search bar on top, vertical result list below. Each result has a title with a hyperlinks to a webpage where all the maps are shown in high-quality. Each results has a snippet of the description. The word of the current focus are highlighted. We could not provide thumbnails of the maps because we did not had access to the pictures itself. It was planned to add this feature in later iterations.

Second, there was a lack of history of actions or at least orientation. For this we have two functionalities: Breadcrumbs and Navigation History. Breadcrumbs save the current path of search and allow it to easily backtrack. The navigational history tracks the whole session history. The can create an account and login. If he does this, history will be saved even over several sessions.

Third, the user can save documents which she found interesting.

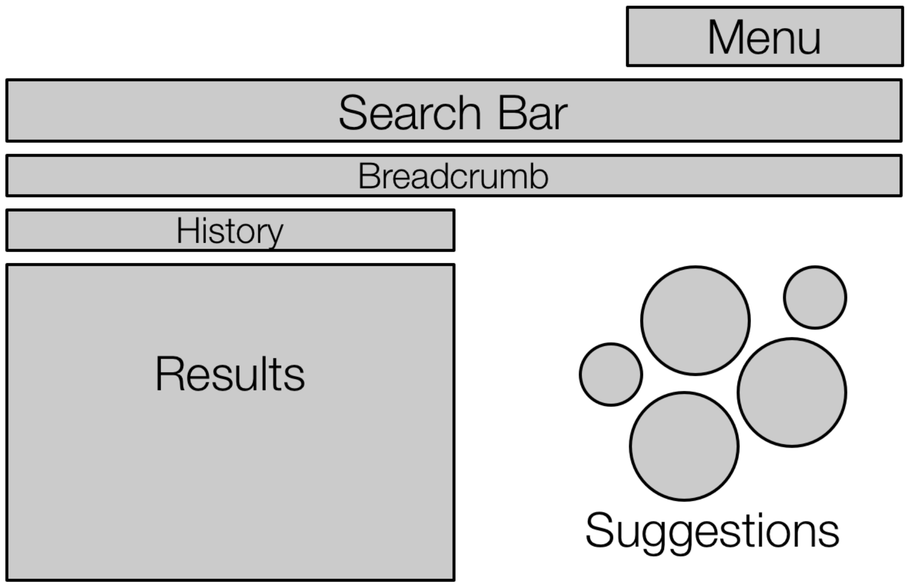

Figure 4.1: Schema of the interface of Fancy FCA 1.0

Fourth, if the user is logged in, all actions, like searches or which linked she clicked on, are logged. They will be saved on the server. This feature was wished by people from the research group to evaluate user behavior.

Fifth, the neighboring concepts (the 'suggestions') are shown right next to the results as a word cloud. The words are presented in a bubble. The font size and bubble radius depend on the number of documents a formal concept comprises.

The UML use case diagram is depicted in Figure 4.2.

4.2 Frontend

The interface was implemented with web techniques. So HTML, CSS, Javascript. Instead of writing pure HTML, we use the Jade templating engine [1]. Instead of writing Javascript, we use CoffeeScript [2] which com-

[1] `http://jade-lang.com`
[2] `http://coffeescript.org`

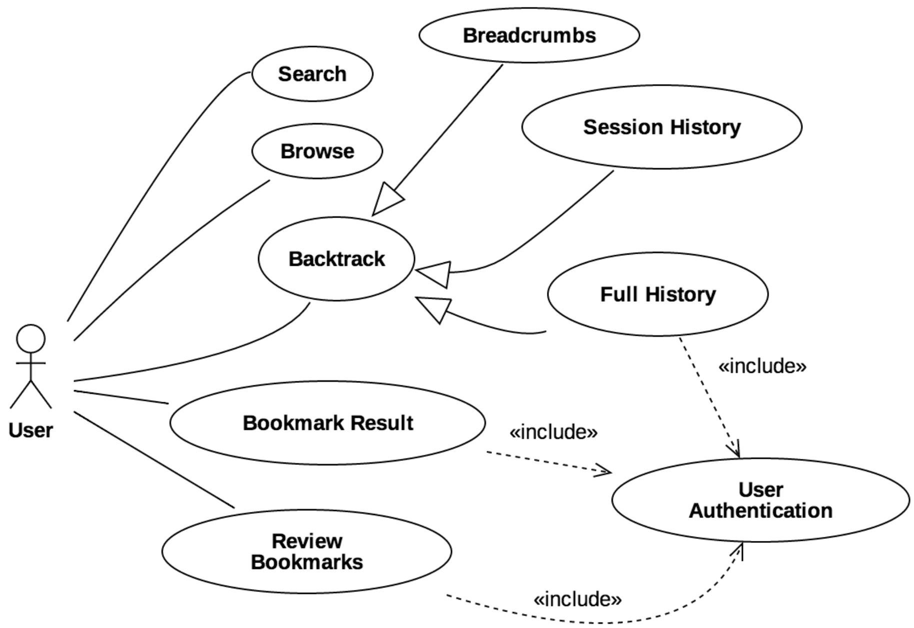

Figure 4.2: UML use case of Fancy FCA 1.0

piles to Javascript but offers semantic sugar. For the the word cloud I use D3[3] [5] as background. D3 utilized SVG to draw elements which are also accessible via the DOM API. The word cloud is a customized graph force-directed layout based on the work from Vallandingham [47]. Bootstrap[4] adds an overall good look to the site.

4.3 Backend

For the backend, Node.JS with the Express Framework [5] was chosen. Reason one, it is good at handling multiple request in parallel. This is important to handle all the log data. Reason two, I can use the same programming language for frontend and backend: Javascript (CoffeeScript). To save data I use MySQL as database because it was already installed on the server. Because of the use of an Object-relational mapping (ORM) framework, Se-

[3]`http://d3js.org`
[4]`http://getbootstrap.com`
[5]`http://expressjs.com`

quelize [6], it is easily possible to change the database.

4.4 Discussion

The whole implementation has one problem: The data is provided as a JSON file. The clients loads the whole file when he starts the session. All the navigation happens in the frontend. While it is positive, that during the navigation the client does not have to communicate with the server, it is a problem that the client has to deal with over 80 megabytes of data. The server compresses the JSON file to eight megabytes. Downloading this file is acceptable especially in an university environment. But it takes up a lot of memory. The design decision was taken before it was clear that the JSON file would be so huge. This should be addressed in further iterations.

That the thumbnails of the maps are not presented is a problem. Albeit all the analysis was done on the metadata, a visual representation would be very useful.

The representation of word cloud is relatively arbitrary. This form was chosen to create an interesting interface. It is to question if it is better than a simple table. The table entries decreasingly ordered after the number of documents the formal concept comprises.

[6] `http://sequelizejs.com`

5. Evaluation

The design of an interface is highly subjective. User studies can help to evaluate an interface but computer scientists are not experts in human studies. For this, a brief review of different techniques for human studies is given first. Then we describe the setup of the study, describe the results and finally conclude.

5.1 Background

We will only scratch the surface here. A comprehensive introduction into "Methods for Evaluating Interactive Information Retrieval Systems with User" gives Kelly [30]. A shorter introduction gives Hearst in Chapter 2 in her book "User Search Interfaces" . [27].

5.1.1 Introduction

So we want to measure usability, but how is it defined? The ISO 9241-11, 1998 [29] defines three aspects of usability:

- Effectiveness: Accuracy and completeness with which users achieve specified goals.
- Efficiency: Resources expended in relation to the accuracy and completeness with which users achieve goals.
- Satisfaction: Freedom from discomfort, and positive attitudes towards the use of the product.

5.1.2 Experiment vs. Evaluation

It is important to distinguish between the terms 'experiment' and an 'evaluation'. Kelly [30] writes:

> "Evaluations are conducted to assess the goodness of a system, interface or interaction technique and can take many forms [..] Experiments have historically been the main method for interactive system evaluation, but experiments can also be conducted to understand behavior [..] Two important characteristics of experiments are that there are at least two things being compared (e.g., system type) and that some manipulation takes place [..] In some types of [interactive information retrieval] studies only a single system is evaluated. This is a weaker form of evaluation since it is not possible to demonstrate how much better users perform or how different their behaviors and inter actions are since there is no point of comparison. Traditional usability tests are examples of this type of evaluation. Traditional usability tests are usually conducted with a single version of a system, with the goal of identifying potential usability problems."

In this thesis, the system is only evaluated to find usability problems. No comparison among other systems are conducted but should be done in further investigations.

5.1.3 Informal Usability Testing

Hearst [27] describes Informal Usability Testing as "Showing designs to participants and recording their responses". It is often used in short iterative cycles to quickly evaluate a design.

We conduct informal usability studies because we do not have proper equipment nor any experience with user studies.

5.1.4 Questionnaires

A questionnaire comprises a set of questions and is cheap and fast way to gather information from people. Kelly et al. [31] describe two types of questions as follows:

> Questionnaires can be comprised of closed questions, open questions or a mixture of both. *Closed questions* are questions that

> provide a fixed set of responses with which subjects must respond. It is common practice for usability questionnaires to include closed questions in the form of statements such as, the system was easy to learn to use. Subjects are typically provided with 5–7-point Likert-type scales for responding, where one scale end-point represents strong agreement and the other represents strong disagreement. [..] *Open questions*, on the other hand, do not provide a response set and subjects are able to provide any type of response they feel is appropriate.

5.1.5 Thinking Aloud

Kelly [30] writes by referring to Ericsson and Simon [23]: "The think-aloud method asks subjects to articulate their thinking and decision-making as they engage in [interactive information retrieval]". The comments from the participants have to be collected. Either by recording the session or by taking notes. It is hoped that the conductors can learn from the thinking process of the participant. There exist variations of this techniques. Because it can be exhausting, challenging and awkward to talk to yourself all the time, participants are encouraged to report either at some fixed times or when the feel the need.

5.2 Setup

The evaluation was done with five people with a background in humanities. First, an introductory presentation was given in Spanish. Then they were split into two groups. Each group interacted with the interface. The people in the group rotated. There were three tasks they should do. The participants were encouraged to talk aloud. Sometimes the instructions asked them what they are doing or what they feel. The one group was recored with an iPhone 6 Plus[1]. The session of the other group was not recored and could not be evaluated at all. After they session they had to fill in a pen-and-paper questionnaire. We use ten closed questions from the USE questionnaire [34] and four commonly asked open questions as asked by Kelly [31]. They can found in appendix C and B. The participants wrote in Spanish which was translated afterwards.

[1] `https://www.apple.com/iphone-6`

5.3 Results

We group thinking aloud, remarks and open questions together because the comments were very similar. After that, we present the results of the closed questions.

5.3.1 Thinking Aloud, Remarks and Open Questions

It turned out, that the evaluation was also a evaluation of the whole research project. All participants mentioned that the interface does not suite their needs. All wanted to refine the search by location and time range and map creators. FCA cannot provide this kind of functionality.

Several participants gave advise how to improve the result list. First, shorten the snippets so that they could read more titles. This could be done by restricting the text length and offering a “show more” button. Second, they mentioned that the thumbnail should be added.

Overall, they give positive comments regarding the interface. They found the navigation with the breadcrumb helpful.

They were not interested in the “Bookmark” feature. Maybe it was because it was an artificial meeting and not a real work environment. They found it a good idea to link to the original site.

There were some problems. They were irritated that they could not search for arbitrary terms. Word for instance like “Valencia” were not in the collection and so also not in the lattice. It was for them not clear that the terms has to be decided before. They suggested that this should get fixed.

They search interface was not used very often. They used the bubble cloud to navigate through the lattice.

One participant mentions that he wanted the see related field to a given field. This would be helpful for his research. FCA can easily provide this method and it is a fail in the interface that it was not included.

5.3.2 Closed Questions

Out of the five participants only three filled out the questionnaire. The results are in the annex B. The results correspond to the comments made by participants. The weakest results got the sentences “It meets my needs” and “It does everything I would expect it to do”. The best result got the sentence “It is easy to use”.

5.4 Accessibility Analysis

In addition to the the evaluation with users, there was a evaluation regarding accessibility features. It was conducted by people from the UNED: Miguel Angel Marqueta and Covadonga Rodrigo San Juan. It can be found in appendix A. They comments were overall positive buy they highlighted some problems with the interface: It is possible to zoom in the bubble cloud, but there are no buttons on the screen for it. In the current state, the zooming can only be done with a mouse or a touchpad.

5.5 Conclusions

Zobel [49] proclaims: “Far too many human studies in computer science are amateurish and invalid” and with this study we proved him true. They overall setup was improvised and all instructors had no experience with user studies. Even though it cannot stand scientific standards, it is useful as informal evaluation.

We can conclude that FCA for this collection in the current implementation is useless. It cannot over the separation between location or time ranges. It looks like that for faceted search as described in Section 3.4 promises better results.

It also suffers from the fact, that only a limited number of terms was selected to built up the concept lattice. If the users searched for terms wich were presented in the documents, they could not explain it why they got no results back. It is impossible to create a concept lattice from all words uses in the documents. It could be promising to combine searching with FCA liked it was done in other systems as described in 3.3. Implement and ordinary search and built up a concept lattice from the results.

For the interface itself, it looks like they liked it. Even though there are some problems like the result list which should be addressed.

After this demolishing results, it was decided not to pursue the idea of FCA for the collection of maps. But the tool itself looks promising that why I transform the tool to visualize arbitrary concept lattices. The changes are described in the following section.

6. Fancy FCA 2.0

Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like “Huardest gefburn”? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

7. Conclusions

Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like "Huardest gefburn"? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

Bibliography

[1] Rakesh Agrawal, Tomasz Imieliński, and Arun Swami. Mining association rules between sets of items in large databases. *ACM SIGMOD Record*, 22(2):207–216, 1993.

[2] Ch Aswani Kumar and S. Srinivas. Concept lattice reduction using fuzzy K-Means clustering. *Expert Systems with Applications*, 37(3):2696–2704, 2010.

[3] Anne Aula. *Studying user strategies and characteristics for developing web search interfaces.* Tampereen yliopisto, 2005.

[4] Marcia J Bates. Where should the person stop and the information search interface start? *Information Processing & Management*, 26(5):575–591, 1990.

[5] Michael Bostock, Vadim Ogievetsky, and Jeffrey Heer. D 3 : Data-Driven Documents. (March), 2011.

[6] Stuart K Card, Jock D Mackinlay, and Ben Shneiderman. Readings in Information Visualization: Using Vision to Think. In *Information Display*, volume 1st, page 686. 1999.

[7] Claudio Carpineto and Giovanni Romano. ULYSSES: a lattice-based multiple interaction strategy retrieval interface. *Human-Computer Interaction*, pages 91–104, 1995.

[8] Claudio Carpineto and Giovanni Romano. Information retrieval through hybrid navigation of lattice representations. *International Journal of Human-Computer Studies*, 45(5):553–578, 1996.

[9] Claudio Carpineto and Giovanni Romano. *Concept data analysis: Theory and applications.* John Wiley & Sons, 2004.

[10] 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.

[11] Claudio Carpineto and Giovanni Romano. Using Concept Lattices for Text Retrieval and Mining. *Formal Concept Analysis*, pages 161–179, 2005.

[12] Ángel Castellanos, Ana García-serrano, Etsi Informática Uned, Juan Cigarrán, and Etsi Informática Uned. Concept-based Organization for semi-automatic Knowledge Inference in Digital Humanities : Modelling and Visualization.

[13] J Cigarrán and A Castellanos. A step forward in Topic Detection Algorithms : An Approach based on Formal Concept Analysis . 16.

[14] Juan M. Cigarrán, Julio Gonzalo, Anselmo Peñas, and Felisa Verdejo. Browsing Search Results via Formal Concept Analysis: Automatic Selection of Attributes. *Concept Lattices*, 2961:201–202, 2004.

[15] Brock Craft and Paul Cairns. Beyond guidelines: What can we learn from the visual information seeking mantra? *Proceedings of the International Conference on Information Visualisation*, 2005:110–118, 2005.

[16] Frithjof Dau, Jon Ducrou, and Peter Eklund. Concept similarity and related categories in SearchSleuth. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 5113 LNAI(October 2007):255–268, 2008.

[17] Jon Ducrou and Peter Eklund. an Intelligent User Interface for Browsing and Searching Mpeg-7 Images Using Concept Lattices. *International Journal of Foundations of Computer Science*, 19(02):359–381, 2008.

[18] Jon Ducrou, Björn Vormbrock, and Peter Eklund. FCA-based browsing and searching of a collection of images. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 4068 LNAI:203–214, 2006.

[19] P. Eklund, P. Goodall, T. Wray, B. Bunt, a. Lawson, L. Christidis, V. Daniel, and M. Van Olffen. Designing the digital ecosystem of the Virtual Museum of the Pacific. *2009 3rd IEEE International Conference on Digital Ecosystems and Technologies, DEST '09*, (June):377–383, 2009.

[20] Peter Eklund, Jon Ducrou, and Peter Brawn. Concept lattices for information visualization: Can novices read line-diagrams? *Concept Lattices. Lecture Notes in Computer Science*, 2961:57–73, 2004.

[21] Peter Eklund, Tim Wray, Peter Goodall, and Amanda Lawson. Design, information organisation and the evaluation of the Virtual Museum of the Pacific digital ecosystem. *Journal of Ambient Intelligence and Humanized Computing*, 3(4):265–280, 2012.

[22] David Eppstein. Hass diagramm, with the integers 1 to 10 as objects and attributes square, prime, composite, even, and odd. `https://commons.wikimedia.org/wiki/File:Concept_lattice.svg`, 2006. Accessed on 16. June 2015.

[23] K A Ericsson and H A Simon. *Protocol analysis: Verbal reports as data*, volume 23. 1993.

[24] Bernhard Ganter and Rudolf Wille. *Formal concept analysis: mathematical foundations*. Springer Science & Business Media, 2012.

[25] Robert Godin, Rokia Missaoui, and Alain April. Experimental comparison of navigation in a Galois lattice with conventional information retrieval methods. *International Journal of Man-Machine Studies*, 38(5):747–767, 1993.

[26] Robert Godin, C Pichet, and J Gecsei. Design of a browsing interface for information retrieval. In *ACM SIGIR Forum*, volume 23, pages 32–39. ACM, 1989.

[27] Marti a Hearst. Search User Interfaces. *Search User Interfaces*, 54(Ch 1):404, 2009.

[28] Ivan Herman, Guy Melançon, and M Scott Marshall. Graph visualization and navigation in information visualization: A survey. *Visualization and Computer Graphics, IEEE Transactions on*, 6(1):24–43, 2000.

[29] ISO. Ergonomic requirements for office work with visual display terminals (vdts) - part 11 : Guidance on usability. ISO 9241-11, International Organization for Standardization, Geneva, Switzerland, 1998.

[30] Diane Kelly. Methods for Evaluating Interactive Information Retrieval Systems with Users. *Foundations and Trends® in Information Retrieval*, 3(1—2):1–224, 2007.

[31] Diane Kelly, David J. Harper, and Brian Landau. Questionnaire mode effects in interactive information retrieval experiments. *Information Processing and Management*, 44(1):122–141, 2008.

[32] Bjoern Koester. Conceptual Knowledge Retrieval with FooCA : Improving Web Search Engine Results with. *Knowledge Creation Diffusion Utilization*, pages 176–190, 2006.

[33] Sergei Kuznetsov, Sergei Obiedkov, and Camille Roth. Reducing the Representation Complexity of Lattice-Based Taxonomies. pages 241–254, 2007.

[34] Arnold M Lund. Measuring usability with the use questionnaire. *Usability interface*, 8(2):3–6, 2001.

[35] Gary Marchionini. Exploratory search: from finding to understanding. *Communications of the ACM*, 49(4):41–46, 2006.

[36] Emmanuel Nauer and Yannick Toussaint. CreChainDo: an iterative and interactive Web information retrieval system based on lattices. *International Journal of General Systems*, 38(4):363–378, 2009.

[37] A A Neznanov and A A Parinov. About Universality and Flexibility of FCA-based Software Tools. *Proceedings of the International Workshop "What can FCA do for Artificial Intelligence?" (FCA4AI 2014)*, page 59, 2014.

[38] Jakob Nielsen. 10 usability heuristics for user interface design. *Fremont: Nielsen Norman Group.[Consult. 20 maio 2014]. Disponível na Internet*, 1995.

[39] Donald A Norman. *The design of everyday things: Revised and expanded edition.* Basic books, 2013.

[40] J Poelmans and So Kuznetsov. Formal concept analysis in knowledge processing: a survey on models and techniques. *Expert systems with ...*, (2003):1–40, 2013.

[41] J Poelmans and So Kuznetsov. Formal concept analysis in knowledge processing: a survey on models and techniques. *Expert systems with ...*, (2003):1–40, 2013.

[42] Giovanni Maria Sacco and Yannis Tzitzikas. *Dynamic taxonomies and faceted search: theory, practice, and experience*, volume 25. 2009.

[43] Ben Shneiderman. The eyes have it: A task by data type taxonomy for information visualizations. *Visual Languages, 1996. Proceedings., IEEE ...*, pages 336–343, 1996.

[44] Ben Shneiderman, Catherine Plaisan, Maxine Cohen, and Steven Jacobs. *Designing the User Interface: Strategies for Effective Human-Computer Interaction*. 5th edition, 2010.

[45] 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.

[46] Patrik Svensson. The Landscape of Digital Humanities. *DHQ: Digital Humanities Quarterly*, 4(1):1–34, 2010.

[47] Jim Vallandingham. Building a bubble cloud, 2012.

[48] Ka-Ping Yee, Kirsten Swearingen, Kevin Li, and Marti Hearst. Faceted metadata for image search and browsing. *Proceedings of the conference on Human factors in computing systems - CHI '03*, (5):401–408, 2003.

[49] Justin Zobel. *Writing for Computer Science*. 2004.

List of Figures

List of Tables

A. Appendix: Accessibility Report

Informe de

Descripción/Objeto	Brief Accessibility Analysis for Visualization Tool
Destinatario	Johannes Filter

Editado	Revisado y Aprobado
Miguel Angel Marqueta	**Covadonga Rodrigo**

Índice

1. Destinatario

2. Documentación/herramientas de referencia

3. Contenido

This a brief accessibility analysis for a visualization tool for NLP group of LSI department of UNED. It is only a list of good and bad things pointed out in a first view. It would be necessary, perhaps, a deeper analysis of this tool but in any case it must be considered as a starting point of a better accessibility document.

Good points:

- Good, nice and clear interface.
- Good color contrast. It must be checked in detail for each component with a specific tool (Contrast Color Analyzer).
- Good size and type font.
- Good layout of interface (Disposition Interface in two parts).
- Good decision of edit box. Simple and clear (with suggestion elements).
- Good navigation with tab key. Except for alternating with two display parts.
- Good choice of “heart icon” for “like button”. But it must be checked with Contrast Color Analyzer.

Bad points (They must be checked):

- It is necessary to add tooltips for 4 main links (Home, login…)
- It is necessary to check in accessibility of "right panel" in order to verify whether screen readers can read JavaScript code for "cloud images". It must be with checked it with tools like "Jaws" or "NVDA" (this last one is free, open source).
- It is necessary to add access key for resizing, (zoom in, and zoom out) now it's only mouse implemented.
- Although keys in "Navigation History" it's a good design with breadcrumbs It is necessary to add keyboard navigation by arrows.
- It is necessary change bold font in "No result" advice. It must be checked in with Contrast Color Analyzer to verify it. It is necessary whether screen reader can read this notice when it appears on screen. Probably it will be possible if focus is set to this notice automatically, but it must be checked.

Conclusion:

In general accessibility of this tools is good, but we propose to check in detail this earlier points and to evaluate the full accessibility of these to ensure a better knowledge of the real accessibility state of this tool. At least we propose some changes and to verify some other to ensure a basic accessibility.

B. Appendix: Closed Questions

Results Closed Questions

	I	II	III	IV	V	Average
Age, Gender		60, F	49, F	44, M		
Background	Historian Expert	Historian Expert	Historian Expert	Historian Expert	Historian Expert	
Usefulness						
It is useful.		7	6	7		6 2/3
It meets my needs.		4	5	4		4 1/3
It does everything I would expect it to do.		4	5	4		4 1/3
Ease of Use						
It is easy to use.		7	7	7		7
It requires the fewest steps possible to accomplish what I want to do with it.		5	6	6		5 2/3
I don't notice any inconsistencies as I use it.		4	5	6		5
Ease of Learning						
I learned to use it quickly.		7	7	6		6 2/3
I easily remember how to use it.		7	7	6		6 2/3
Satisfaction						
I am satisfied with it.		7	5	5		5 2/3
It is fun to use.		4	6	7		5 2/3

C. Appendix: Open Questions

- What were the most positive things about using this system and why?
- What were the most negative things about using this system and why?
- How would you improve this system and why?
- Is there anything else that you would like to tell us about this system and your experiences using it?
- In comparison with other systems (Interface of Catálogo Colectivo de las colecciones de Mapas and previous work of this research group). Do you think it is more useful? If yes, which part of the system is more useful.