

Interactive Visualization of Large Concept Lattices for Exploratory Search

Johannes Filter

June 1, 2015

Contents

1	Introduction	2
2	Background	2
2.1	Formal Concept Analysis	2
2.2	Interface Design Principles	3
2.3	Explorative Search	3
3	Related Work	3
3.1	Pruning Nodes	4
3.2	Local View	4
4	Fancy FCA 1.0	5
4.1	Idea	5
4.2	Implementation	6
5	First Evaluation	6
6	Fancy FCA 2.0	6
7	Second Evaluations	6
8	Conclusions	7

1 Introduction

The digital revolution is affecting every part of our life. Also the humanities scholars stand before a big change in their ways when huge analog collections are digitized. They have to apply computer science methods to organize and analyze huge amount of data. The term "Digital Humanities" evolved during the last 10 years which can be defined as an "intersection between the humanities and information technology" [23].

The Information Retrieval Department of the Universidad Nacional de Educacin a Distancia (UNED) in Madrid (Spain) cooperates with human scholars to conduct research in the Digital Humanities. In this project, 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. [6, 7]

They successfully implemented a FCA algorithm but lack an interactive user interface, which is my task.

The work will be evaluated by a small user study conducted at the UNED with users from different backgrounds.

2 Background

2.1 Formal Concept Analysis

Formal Concept Analysis (FCA) [14] is a mathematically well-funded technique to find relationships among objects and has been applied to areas[20].

Formally it is defined as as a tripple $K = (G, M, I)$ where G is a set of objects, M is a set of attributes and I is a binary relation $G \times M \rightarrow (0, 1)$. I specifies if an object has an attribute or not. Example: we have tree documents(letters) and our attributes categories if a word occurs in this document or not.

	Letter1	Letter2	Letter3	Letter4
Love	0	0	1	1
Hello	1	0	1	0
Apple	0	1	1	1
You	1	0	0	1

HERE WILL BE AN IMAGE

Comments from Angel:

"It is a little bit incomplete and messy.

What you have in the table is a "Formal Context"

When you apply FCA, you obtain a set of "formal concepts", where a formal concept (A, B) is the maximum set of objects (A) sharing all the attributes in B .

These formal concepts can be organized by means of a "partial order relationship"

The set of formal concepts plus the order relationship results in a concept lattice which can be represented as a Hasse Diagram "

Because of the ordered sets, they are a concept lattices. Orders sets can be represented as a Hasse Diagram. [14]

2.2 Interface Design Principles

Well-funded search interface design principles and principles of information visualization will be applied.[15, 21]

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.

2.3 Explorative Search

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.

3 Related Work

While some user studies proclaim that non-experts can read Hasse diagram[12], the study has been conducted on relatively small lattices with a small

amount of objects and attributes. It is easy to see, that the visual representation is hindern when the number of concepts grow. The amount and the edge crossings make understanding impossible. "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." [17]

There exists basically several approaches to reduce complexity of large concept lattices. I describe in the following.

3.1 Pruning Nodes

A prominent approach to prune lattices called "iceberg lattice" [22]. A variation from the frequent item-set mining which specific min-support and min-confidence[1]. It creates a top of the lattice but has some drawbacks because "One should be careful not to overlook small but interesting groups, for example, exotic or emergent groups not yet represented by a large number of objects, or, groups that contain objects who are not members of any other group." [17] The iceberg lattice just focuses on the concepts that contain a lot of documents. That is why an other approaches exist: Stability[18]: "A concept is stable if its intent does not depend much on each particular object of the extent." [17]

It is also possible to apply traditional cluster techniques to FCA.[2]

3.2 Local View

There is been some research done to convey the insights of FCA without visualizing the lattice and just focusing on a local view. Sometimes referes as 'Conceptual Neighborhood'[11, 13], you focus on one concept and show concepts that are directly adjacent to the concept, 'above' and 'below', in the lattice. The workflow is by initialize a concept with the help of a query interface and that to navigate incrementally by removing terms (going up) and adding terms (going down). This approach has been applied to different scenarios. For instance for a 'Digital Museum' [11, 13], image browsing [10, 9] and search engines [8, 16].

There exist small variations: [3] you can bound the search space by limiting the degree of parents, children or siblings to display. So to specify the maximum distance in a lattice to which nodes are displayed. It is also possible to make use of a fisheye view where the nodes in the center of the screen are bigger and the nodes fade out of the screen close to the border

The CREDO system [5] is similar to the Conceptual Neighborhood, you

have to formulate a query first and then a sub lattices is build. But there are also results included that do not directly adjacent. Calculated with a new similar measurement.

Small Variation: Transform to Tree

While transforming the lattice into a tree sounds promising, because you could apply sophisticated tree visualizing techniques to reduce edge crossing, it comes with several drawbacks. One naive approach[4]: Duplicate nodes with more than one parents. While this is a successful transformation, we added a high amount of redundant nodes.

Another approach: just select the 'best' parents and hides all other edges to the other parents[19]. While this technically not breaks the concept, the visual representation does not correspond to the underlying model.

This idea needs some more research to find out how useful it is.

4 Fancy FCA 1.0

4.1 Idea

Comments from Angel: "There is another theoretical problem related to iceberg lattices.

FCA in general, and our work in particular, is intended to be an exploratory technique. That is, I know nothing (or almost nothing) about the data, so I would like to have a data organization (as the one provided by FCA in the concept lattice) that allows me to EXPLORE the data in order to gain some insight about.

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."

Because of huge amount of concepts in our project, a pruned version of a galois lattice can give a good overview but is not suitable for in-depth analysis.

I will develop an interface that comprises two different views: The Local View with an query extension to browse through the lattice and a pruned, global view to give an overview over the lattice.

As I remark in the state of the art review, pruning-based techniques suffer from On the other hand, when you apply local view you can lose the big picture..... To address both problems, we propose an hybrid approach integrating both representations in order to be able to offer a manageable visualization without losing the general view....

4.2 Implementation

The interface should run in modern Web Browser. That is why I utilize mainly the technologies: SVG and Javascript. I use the Javascript Framework d3.js for the visualizing besides several other modern web technology tools and frameworks. d3.js gives a good abstraction to most occurs problems

The data sources comprises two APIs: 1. a precomputed concept lattices, 2. detailed information of a document, specified by a document id.

For an analysis of users, the user behaviour should be logged. That is why an backend has to be implemented offering user registration. The logged can be saved in a relation database. An admin interface is not needed.

5 First Evaluation

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.

6 Fancy FCA 2.0

7 Second Evaluations

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.

8 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.

References

- [1] Rakesh Agrawal, Tomasz Imieliski, 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] Claudio Carpineto and Giovanni Romano. ULYSSES: a lattice-based multiple interaction strategy retrieval interface. *Human-Computer Interaction*, pages 91–104, 1995.
- [4] Claudio Carpineto and Giovanni Romano. *Concept data analysis: Theory and applications*. John Wiley & Sons, 2004.
- [5] 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.
- [6] 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.
- [7] J Cigarrán and A Castellanos. A step forward in Topic Detection Algorithms : An Approach based on Formal Concept Analysis . 16.
- [8] 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.
- [9] 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.
- [10] 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.

- [11] 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.
- [12] 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.
- [13] 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.
- [14] Bernhard Ganter and Rudolf Wille. *Formal concept analysis: mathematical foundations*. Springer Science & Business Media, 2012.
- [15] Marti a Hearst. Search User Interfaces. *Search User Interfaces*, 54(Ch 1):404, 2009.
- [16] Mihye Kim and Paul Compton. The perceived utility of standard ontologies in document management for specialized domains. *International Journal of Human Computer Studies*, 64(1):15–26, 2006.
- [17] Sergei Kuznetsov, Sergei Obiedkov, and Camille Roth. Reducing the Representation Complexity of Lattice-Based Taxonomies. pages 241–254, 2007.
- [18] Sergei Kuznetsov, Sergei Obiedkov, and Camille Roth. Reducing the Representation Complexity of Lattice-Based Taxonomies. pages 241–254, 2007.
- [19] Cassio Melo, Bénédicte Le-Grand, Marie Aude Aufaure, and Anastasia Bezerianos. Extracting and visualising tree-like structures from concept lattices. *Proceedings of the International Conference on Information Visualisation*, 6:261–266, 2011.
- [20] J Poelmans and So Kuznetsov. Formal concept analysis in knowledge processing: a survey on models and techniques. *Expert systems with ...*, (2003):1–40, 2013.

- [21] Ben Shneiderman. The eyes have it: A task by data type taxonomy for information visualizations. *Visual Languages, 1996. Proceedings., IEEE ...*, pages 336–343, 1996.
- [22] 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.
- [23] Patrik Svensson. The Landscape of Digital Humanities. *DHQ: Digital Humanities Quarterly*, 4(1):1–34, 2010.