

Steps Towards Interactive Formal Concept Analysis with LatViz

Mehwish Alam, Thi Nhu Nguyen Le, Amedeo Napoli

▶ **To cite this version:**

Mehwish Alam, Thi Nhu Nguyen Le, Amedeo Napoli. Steps Towards Interactive Formal Concept Analysis with LatViz. Proceedings of the 5th International Workshop "What can FCA do for Artificial Intelligence"? co-located with the European Conference on Artificial Intelligence ECAI 2016, Aug 2016, The Hague, Netherlands. hal-01420753

HAL Id: hal-01420753
https://inria.hal.science/hal-01420753

Submitted on 21 Dec 2016

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Steps Towards Interactive Formal Concept Analysis with LatViz

Mehwish Alam[1], Thi Nhu Nguyen Le[2], and Amedeo Napoli[2]

1. Laboratoire d'Informatique de Paris-Nord, Université Paris 13, Paris, France
2. LORIA (CNRS – Inria Nancy Grand Est – Université de Lorraine)
BP 239, Vandoeuvre-lès-Nancy, F-54506, France
{alam@lipn.univ-paris13.fr}{thi-nhu-nguyen.le,amedeo.napoli@loria.fr}

Abstract. With the increase in Web of Data (WOD) many new challenges regarding exploration, interaction, analysis and discovery have surfaced. One of the basic building blocks of data analysis is classification. Many studies have been conducted concerning Formal Concept Analysis (FCA) and its variants over WOD. But one fundamental question is, after these concept lattices are obtained on top of WOD, how the user can interactively explore and analyze this data through concept lattices. To achieve this goal, we introduce a new tool called `LatViz`, which implements several algorithms for constructing concept lattices and allows further navigation over lattice structure. `LatViz` proposes some remarkable improvements over existing tools and introduces various new functionalities such as interaction with expert, visualization of Pattern Structures, AOC posets, concept annotations, filtering concept lattice based on several criteria and finally, an intuitive visualization of implications. This way the user can effectively perform an interactive exploration over a concept lattice which is a basis for a strong user interaction with WOD for data analysis.

Keywords: Lattice Visualization, Interactive Exploration, Web of Data, Formal Concept Analysis.

1 Introduction

In the last decade, there has been a huge shift from the Web of Documents to the Web of Data (WOD). Web of Documents represents data in the form of HTML pages which are linked with other HTML pages through hyperlinks. This web of documents has evolved into WOD where all the information contained is represented in the form of entities and relations allowing the semantics to be embedded in the representation of these data. These data are in the form of a (node-arc) labeled graph belonging to several domains such as newspapers, publications, biomedical data etc. The growth in the publication of data sources in WOD has made it an important source of data, which has led towards many challenges pertaining to effective utilization of this data. WOD mainly represents data in the form of Resource Description Framework (RDF)[1]. There are several

[1] `http://www.w3.org/RDF/`

ways such as data dumps and SPARQL queries to access these data, which can be utilized for several applications, one of which is visualization and interactive exploration for data analysis purposes. Several visualization tools have been developed for this purpose, one of which is LODLive[2] [6], where the user can choose data-sets such as DBpedia and Freebase and specify an entity as a starting point for browsing the node-arc labeled graph. Another tool based on graphical display is RelFinder [15], where given several entities the tool automatically finds the paths connecting these entities. The major drawback of LODLive is that after two hops the number of nodes increase and it is hard to visualize the data. Moreover, these tools are good for getting an insight into what RDF graph contains but they are not built for the purpose of knowledge discovery.

In order to provide the user with the ability to perform data analysis and knowledge discovery over such kind of data, there is a need to perform classification, where the obtained classes are further made available to the user for exploration and subjective interpretation. In the current study we use Formal Concept Analysis as the basis for classification. Several studies have already been conducted using FCA and its variants over RDF graphs or its generalization to knowledge graphs. Out of these studies so far RV-Xplorer [3] is the only tool that actually allows interactive exploration of clustered RDF data [2]. The purpose of this paper is to enhance the functionalities discussed in the previous studies. In this study we introduce a new tool `LatViz` which increases the interpretability of a concept lattice by remarkably improving the user interaction with the concept lattice as compared to existing tools. Various new functionalities have been introduced such as the visualization of Pattern Structures and AOC-posets, concept annotation, filtering concept lattice and pattern concept lattice based on several criteria and finally, an intuitive visualization of implications. This way the user can effectively perform an interactive exploration over a concept lattice which in turn gives a basis for a strong user interaction with WOD for knowledge discovery purposes. In this paper, we detail the important interaction operations implemented in `LatViz`. In the rest of this paper we refer to "user" as an "expert" as (s)he having basic knowledge about the lattice structure.

The paper is structured as follows: Section 2 introduces a motivating example. Section 3 introduces some of the important functionalities of `LatViz`, while in Section 4, we discuss some of the related tools already developed and finally Section 5 details the future perspectives of the current work.

2 Motivating Example

Let us consider that an expert is searching for papers published by a particular team in conferences or journals related to his/her field of research. In order to locate the papers of his/her interest (s)he has to search for specific keywords or authors in the local portal. For getting the view of which kind of papers are contained (s)he has to run a broad query and then narrow down his/her

[2] `http://en.lodlive.it/`

query to obtain papers on specific keywords or authors or group of keywords or authors. The expert will end up running several queries to get what (s)he wants. Moreover, if the expert wants to know the collaborations of the team with other members of the research community outside the team, as well as the diversity and the specialization of the team members, this cannot be directly obtained by simple querying. To obtain such kind of knowledge there is a need to introduce a support for data analysis. Based on this scenario, we show how the expert can be guided thanks to an adapted visualization tool to obtain such information of interest with the help of concept lattices.

3 `LatViz` for Interactive Exploration of Concept Lattices

3.1 User Interface

The display of `LatViz` resembles Conexp[3], which provides basic functionalities for building a concept lattice. `LatViz` implements two algorithms for building a concept lattice from a binary context, one of which is introduced in [14]. Another, efficient algorithm for building a concept lattice is AddIntent [20]. A demo of LatViz is available on-line through `http://latviz.loria.fr/latviz/`.

The concept lattice for the scenario in section 2 was created by mapping the RDF data to a formal context $\mathcal{K} = (G, M, I)$. The subjects of the triples were considered as the set of objects G, the objects in the RDF triples i.e., keywords and authors were considered as the set of attributes M. In this example, the RDF triples were created from the publications of the Knowledge Discovery (KDD) team in LORIA. The number of objects in the context are 343 and attributes are 1516. Very often huge concept lattices are obtained based on the context size. `LatViz` provides several interactive operations allowing for reduction of exploration space of the expert. To-date this is the most interactive tool having many unique functionalities such as handling numeric data with the help of interval pattern structures, AOC-posets, filtering concept lattice and implications which provides support for data analysis. Other functionalities such as annotating the lattice, level-wise display of a concept lattice etc. are discussed in many contexts but are not yet directly implemented in the commonly used tools. In the following we detail each of these functionalities for data analysis.

3.2 AOC-Posets

AOC-poset is a partially ordered set of the attribute and object concepts, first introduced in [18, 19]. The object and attribute concept are referred to as *introducers* in [5]. Once an attribute is introduced in a concept it is inherited from top to bottom while, dually, an introduced object is "inherited" from bottom to top. During this study, we implement the Hermes Algorithm introduced in [5] for building AOC-Poset from binary context. Figure 1 show the highlighted AOC-Posets of the concept lattice built for the running scenario.

[3] `http://conexp.sourceforge.net/`

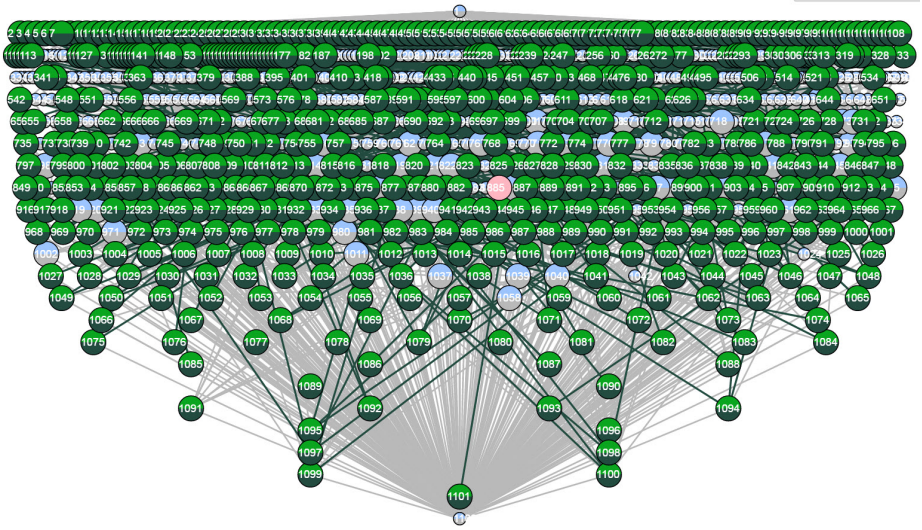


Fig. 1: AOC-Posets.

3.3 Displaying Concept Lattice Level-wise

AOC-Posets actually reduce exploration space but still a huge number of concepts remain to be observed. `LatViz` allows the creation of concept lattice level-wise by interaction. When an expert clicks on the top concept, `LatViz` computes and displays the first level. After that the expert can select the concept for continuing the exploration, then `LatViz` computes the next level for that concept. In Figure 2, the top image shows the first level of the concept lattice built by selecting the top concept. Then the expert can view the contents of each concept on mouseover. In the running example, expert locates the concept with all the papers of Amedeo Napoli (i.e., K#2), which shows that the total number of documents written by Amedeo Napoli are 152. On selecting this concept, the next level of the lattice originating from the selected concept is computed (shown in the bottom image in Figure 2).

3.4 Display Sub/Super-Concepts of a Concept

In case of huge concept lattices sometimes it is hard to keep track of the ordering relations between the concepts. `LatViz` allows the expert to only highlight sub/super-concepts of a concept. For example, if the expert wants to display all the publications along with the collaborations of the author Amedeo Napoli, (s)he can highlight the associated sub-lattice of the attribute concept of "Amedeo Napoli". Figure 3 shows the highlighted sub-lattice in brown. An expert can highlight the super-concepts connected to a concept. If the expert is looking for all the papers having some keywords common with the paper *Knowledge Organization and Information Retrieval Using Galois Lattices* having one or more of the

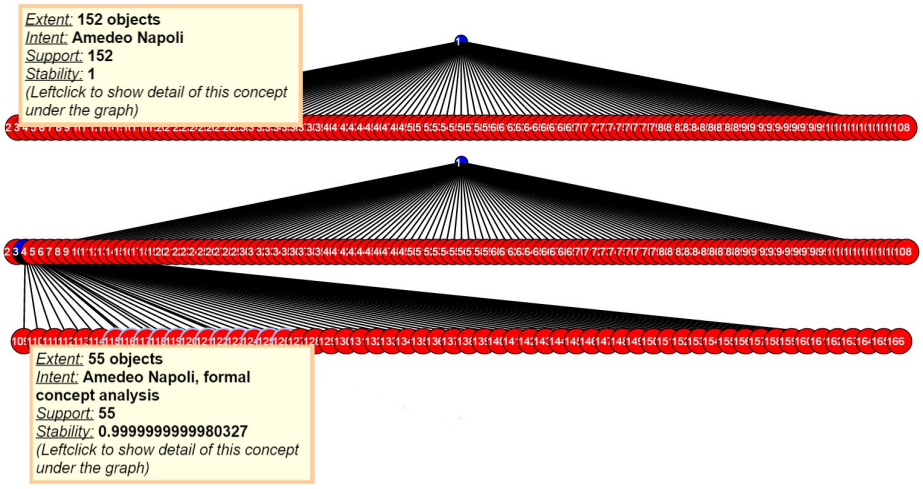


Fig. 2: Top image shows the first level of the concept lattice, the bottom image shows the second level built by interaction.

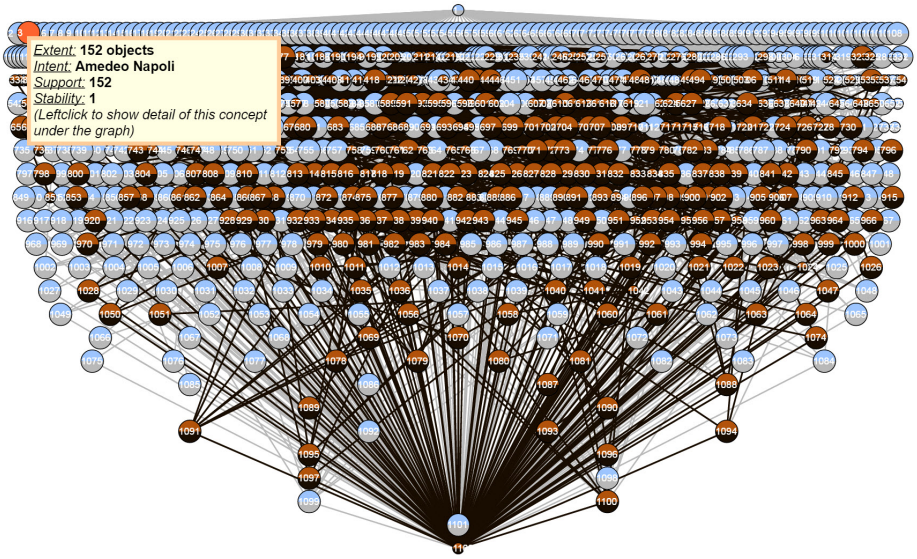


Fig. 3: The sub-lattice highlighted for the author “Amedeo Napoli”.

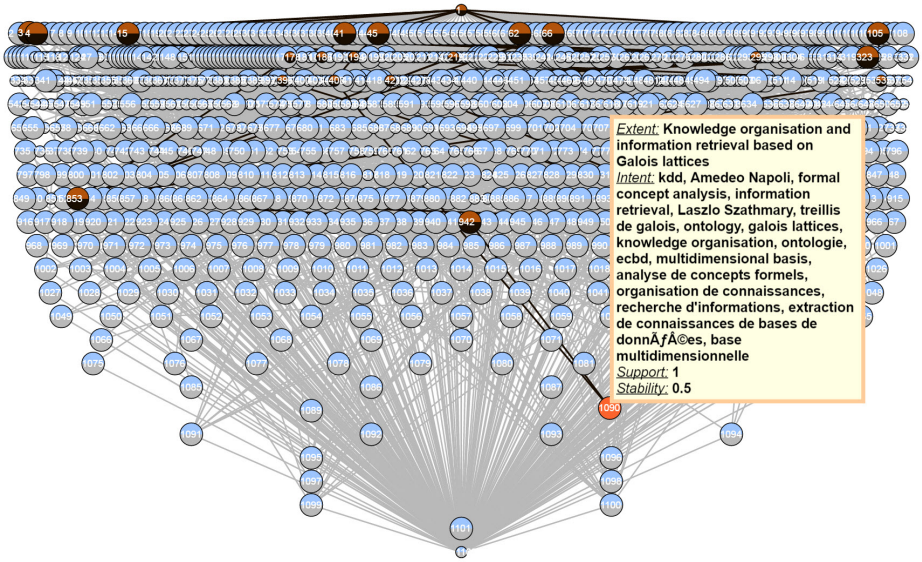


Fig. 4: Highlighting super-concepts of a concept.

keywords in the intent of the concept then (s)he can highlight the sub-lattice of super-concepts associated to it (see Figure 4).

3.5 Display/Hide the Sub-lattice

This functionality was partially implemented in RV-Xplorer [3] to reduce the interaction space of the expert. Here the expert can only show the part of the concept lattice in which (s)he is interested. The expert can locate the interesting concept by navigation, containing some intent or extent. If an intent is interesting and the expert marks the concept as interesting then only the sub-concepts of this concept are shown to the expert as the intents are inherited from top to bottom. Dually, if an extent is interesting for the expert then all the super-concepts are shown to the expert as the extent is inherited bottom-top. Previously, the expert highlighted the sub-lattice of the concept containing all the papers of Amedeo Napoli, now if the expert is interested in only the papers of Amedeo Napoli on Knowledge Representation then (s)he can navigate downwards and only see this part of concept lattice by marking it interesting (see Figure 5). Similarly, we previously highlighted all the super-concepts of the concept having the paper entitled *Knowledge Organization and Information Retrieval Using Galois Lattices* in the extent, Figure 6 only shows the associated sub-lattice to have a clearer view (see Figure 6).

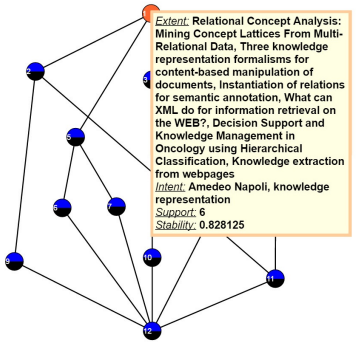


Fig. 5: Showing only sub-lattice of the interesting concept.

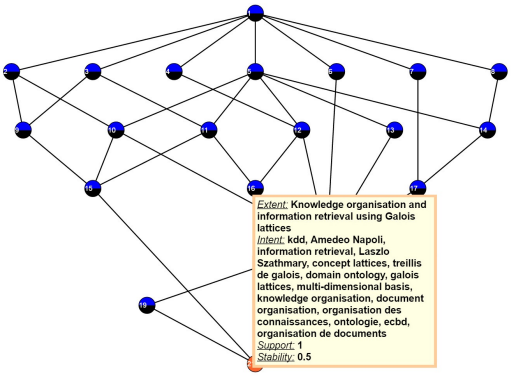


Fig. 6: Showing only super-concepts of the interesting concept.

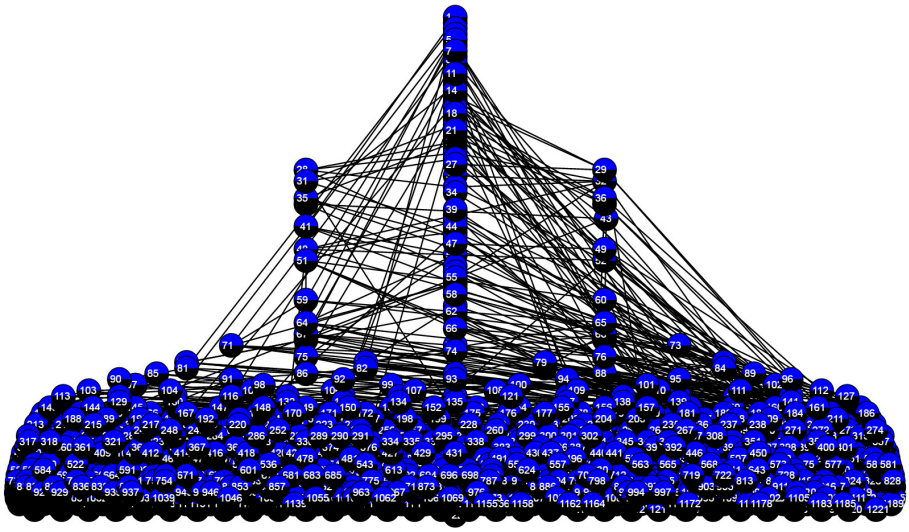


Fig. 7: Interval pattern concept lattice for publications.

3.6 Interval Pattern Structures

Interval Pattern Structures were first introduced in [16] for dealing with numerical data. Consider two descriptions $\delta(g_1) = \langle [l_i^1, r_i^1] \rangle$ and $\delta(g_2) = \langle [l_i^2, r_i^2] \rangle$, with $i \in [1..n]$ where n is the number of intervals used for the description of entities. The similarity operation $\sqcap$ and the associated subsumption relation $\sqsubseteq$ between descriptions are defined as the convex hull of two descriptions as follows: $\delta(g_1) \sqcap \delta(g_2) = \langle [min(l_i^1, l_i^2), max(r_i^1, r_i^2)] \rangle$. Based on this similarity measure interval pattern concept lattice can be built. In the running scenario, three numerical attributes for the papers were used i.e., year of publications, rank of the conference and the number of pages. The ranks of the conferences were con-

sidered based on COmputing Research and Education (CORE) rankings[4]. The ranks were A*, A, B, C and other which were coded as 1, 2, 3, 4 and 5 respectively. The final concept lattice generated for the last five years of publications of Knowledge Discovery Team is shown in Figure 7.

3.7 Lattice Filtering Criteria

There are two categories of filtering provided by `LatViz`; one is for the concept lattice created with the binary data and the other one is provided for the pattern concept lattice built with the help of interval pattern structures.

Filtering Concept Lattice. After a concept lattice is built by applying FCA, expert is allowed to set several filtering criteria such as stability, lift, extent size, intent size and finally specific object or attribute names. Let us consider that in the running example, the expert is looking for the papers published by Amedeo Napoli on the topic of pattern structures and FCA. A filter on the number of attributes in the intent is set to 3. The filtered concept lattice obtained over the complete lattice is shown in Figure 8. It further shows the authors with who Amedeo Napoli has worked i.e., Sergei O. Kuznetsov and Mehdi Kaytoue. This part of concept lattice shows the community of authors working with Amedeo Napoli on the topic of pattern structures.

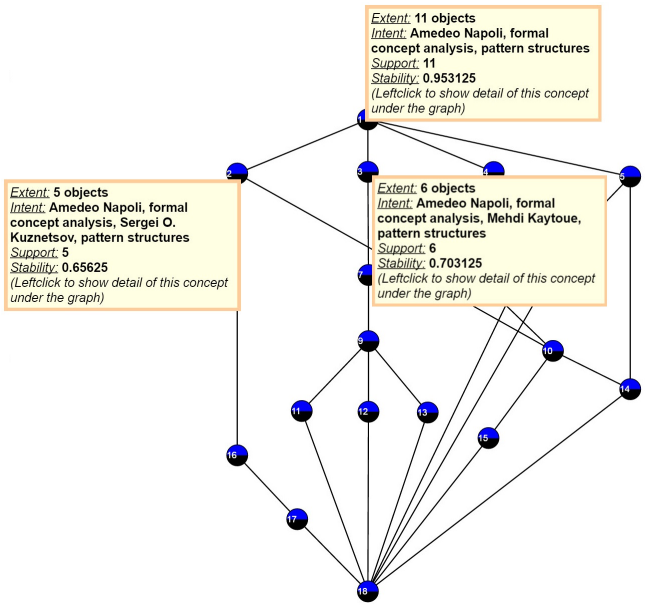


Fig. 8: Filtered concept lattice obtained from binary context.

[4] http://portal.core.edu.au/conf-ranks/

Filtering Pattern Concept Lattice. Interval Pattern Concept Lattices can also be filtered by specifying the number of attributes to be considered, the upper and the lower limits for the intervals in the intent of each attribute along with stability, lift and extent size. Let us consider the pattern concept lattice in Figure 7 which is hard to interpret. To make it more readable based on what an expert wants, (s)he is allowed to specify filters. For example, if the expert is looking for a paper published in a conference of a rank 1-4 in the year 2012 - 2015 and has the number of pages not less than 2 and no more than 42 then the respective filters can be set for the values of all three attributes. The filtered pattern concept lattice will then only contain the part of lattice needed by the user. Figure 7 shows the concept containing group of papers published from 2014-2015 in conferences with rank 2 having number of pages 2-42.

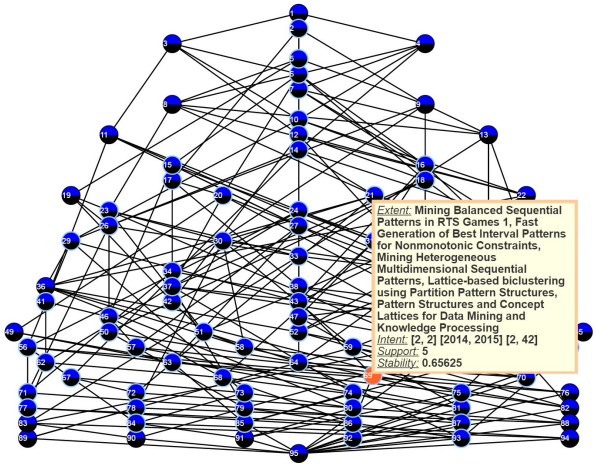


Fig. 9: Filtered Pattern Concept Lattice.

3.8 Attribute Implications

One of the many proposed visualization techniques for implications includes *table-based views*. The columns in the table represent rule ID, LHS and RHS of the rule, support and confidence measures. These views were used because of the simplicity of storage. However, as the number of rules can be too many it is not very evident for the expert to focus on interesting rules at a simple glance. Another way of visualizing association rules are *Matrix Views*, where rows represent the LHS and columns represent the RHS of the rules. Support and confidence are displayed by different colors in the intersection of the LHS and RHS. In case of a formal context, the number of objects/attributes can be very big leading to problems in displaying the matrix. By carefully taking into account the above drawbacks, we finally settle on visualizing implications with the help of scatter plots, where the x-axis shows the increasing support

and the y-axis shows the increasing lift (as we are considering implications the confidence of the rule is always 100%). Such kind of display helps the expert to single-out the rules (s)he wants to visualize based on the values of support and lift. Figure 10 shows implications of the running example, x-axis keeps the support in percentage and y-axis keeps lift. The number on top of the circle shows the number of rules existing in the same point in the plot. On mouse over, expert can view the implications.

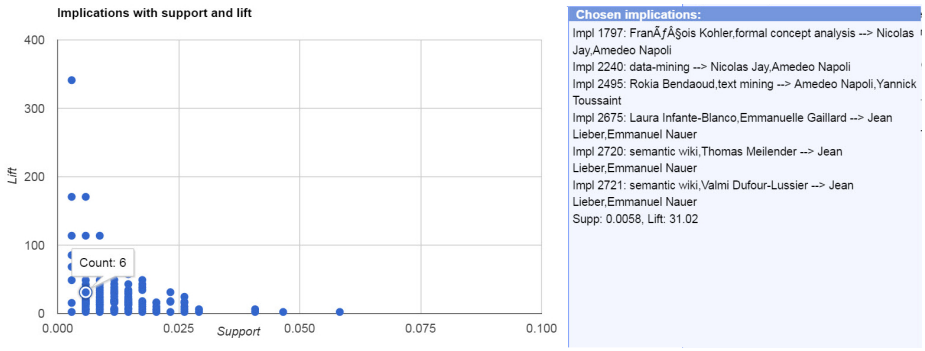


Fig. 10: Attribute implications for the running example.

4 Related Tools

In [2], the authors focus mainly on interactive data exploration over RDF data for interactive knowledge discovery. It clusters RDF triples based on RDF Schema and then allows interactive exploration with the help of RV-Xplorer (Rdf View eXplorer) [3]. It is a tool for visualizing views over RDF graphs mainly for identifying interesting parts of data and allow data analysis. It has also been extended for clustering SPARQL query answers. To-date there have been many other tools developed for reducing the effort of expert in observing and interpreting a concept lattice. Many of the tools have been developed for more specific purposes. CREDO [8] and FooCA [17] are the Web Clustering Engines [7] which take the answers from queries posed against search engines and create a concept lattice which is then displayed to the expert for interaction. CREDO allows only limited interaction, however, FooCA allows the expert to edit the context and iteratively build the concept lattice. CEM [10] is an email manager which allows quick search through the e-mails and usually deals with smaller concept lattices. Camelis [11] is a system based on FCA for the organization of documents allowing several navigation operations. Another set of tools such as Sewelis [13] and Sparklis [12] allows navigation/interaction over knowledge graphs. Many other tools such as Galicia[5], ConExp and ToscanaJ[6] are developed for academic purposes. `LatViz` takes the basic functionalities of ConExp and takes it to the another level by

[5] `https://sourceforge.net/projects/galicia/`
[6] `http://toscanaj.sourceforge.net/`

providing visualization for many algorithms introduced over time to increase the readability. Moreover, it re-uses the source-code for building concept lattice with the help of the algorithm in [14] from ToscanaJ [4]. It can not only be applied to WOD but it has been extended for interpreting any kind of data.

5 Discussion and Future Improvements

`LatViz` is a tool built for allowing expert interaction for data analysis purposes. It provides many new functionalities for reducing the exploration space of the expert and enable him to interpret the results. As a future perspective, we also want to implement other variations of pattern structures such as Pattern Structures introduced for structured set of attributes discussed in [1] and Heterogeneous Pattern Structures [9]. We also want to extend the implementation of implications to association rules. Finally, we also want to take into account matrix factorization.

References

1. Mehwish Alam, Aleksey Buzmakov, Amedeo Napoli, and Alibek Sailanbayev. Revisiting pattern structures for structured attribute sets. In *Proceedings of the Twelfth International Conference on Concept Lattices and Their Applications.*, pages 241–252, 2015.
2. Mehwish Alam and Amedeo Napoli. Interactive exploration over RDF data using formal concept analysis. In *2015 IEEE International Conference on Data Science and Advanced Analytics, DSAA.*, pages 1–10, 2015.
3. Mehwish Alam, Matthieu Osmuk, and Amedeo Napoli. RV-Xplorer: A way to navigate lattice-based views over RDF graphs. In *Proceedings of the Twelfth International Conference on Concept Lattices and Their Applications.*, pages 23–34, 2015.
4. Peter Becker and Joachim Hereth Correia. The ToscanaJ suite for implementing conceptual information systems. In *Formal Concept Analysis, Foundations and Applications*, pages 324–348, 2005.
5. Anne Berry, Alain Gutierrez, Marianne Huchard, Amedeo Napoli, and Alain Sigayret. Hermes: a simple and efficient algorithm for building the aoc-poset of a binary relation. *Ann. Math. Artif. Intell.*, 72(1-2):45–71, 2014.
6. Diego Valerio Camarda, Silvia Mazzini, and Alessandro Antonuccio. Lodlive, exploring the web of data. In *I-SEMANTICS 2012 - 8th International Conference on Semantic Systems, I-SEMANTICS '12, Graz, Austria, September 5-7, 2012*, 2012.
7. Claudio Carpineto, Stanislaw Osiński, Giovanni Romano, and Dawid Weiss. A survey of web clustering engines. *ACM Comput. Surv.*, 41(3):17:1–17:38, 2009.
8. Claudio Carpineto and Giovanni Romano. Exploiting the potential of concept lattices for information retrieval with CREDO. *J. UCS*, 10(8):985–1013, 2004.
9. Víctor Codocedo and Amedeo Napoli. A proposition for combining pattern structures and relational concept analysis. In *Formal Concept Analysis - 12th International Conference, ICFCA 2014, Proceedings*, pages 96–111, 2014.
10. Richard Cole and Gerd Stumme. CEM - A conceptual email manager. In *8th International Conference on Conceptual Structures, ICCS, Proceedings*, 2000.

11. Sébastien Ferré. Camelis: a logical information system to organise and browse a collection of documents. *Int. J. General Systems*, 38(4):379–403, 2009.
12. Sébastien Ferré. SPARKLIS: a SPARQL endpoint explorer for expressive question answering. In *Proceedings of the Posters & Demonstrations Track a track within the 13th International Semantic Web Conference, ISWC.*, 2014.
13. Sébastien Ferré and Alice Hermann. Reconciling faceted search and query languages for the semantic web. *IJMSO*, 7(1):37–54, 2012.
14. Bernhard Ganter and Rudolf Wille. *Formal Concept Analysis: Mathematical Foundations*. Springer, Berlin/Heidelberg, 1999.
15. Philipp Heim, Steffen Lohmann, and Timo Stegemann. Interactive relationship discovery via the semantic web. In *The Semantic Web: Research and Applications, 7th Extended Semantic Web Conference, ESWC 2010, Heraklion, Crete, Greece, May 30 - June 3, 2010, Proceedings, Part I*, pages 303–317, 2010.
16. Mehdi Kaytoue, Sergei O. Kuznetsov, and Amedeo Napoli. Revisiting numerical pattern mining with formal concept analysis. In *Proceedings of the 22nd International Joint Conference on Artificial Intelligence.*, pages 1342–1347, 2011.
17. Bjoern Koester. Conceptual knowledge retrieval with FooCA: Improving web search engine results with contexts and concept hierarchies. In *Advances in Data Mining, Applications in Medicine, Web Mining, Marketing, Image and Signal Mining, 6th Industrial Conference on Data Mining, ICDM Proceedings*, 2006.
18. R. Osswald and W. Petersen. A logical approach to data-driven classification. In *KI*, volume 2821 of *Lecture Notes in Computer Science*. Springer, 2003.
19. Wiebke Petersen. A set-theoretical approach for the induction of inheritance hierarchies. *Electr. Notes Theor. Comput. Sci.*, 53:296–308, 2001.
20. Dean van der Merwe, Sergei A. Obiedkov, and Derrick G. Kourie. Addintent: A new incremental algorithm for constructing concept lattices. In *ICFCA*, pages 372–385, 2004.