

An introduction to Conceptual Exploration

A bouquet of Concept Lattice Applications and Tools

Bernhard Ganter, Tom Hanika, Johannes Hirth, Sergei Obiedkov

HHA124

10. 6. 2024

Let me show you two very different
applications

Why ?

- We present a short list of domains where *Formal Concept Analysis* was successfully applied.
- Every domain where FCA can be applied, *Conceptual Exploration* is applicable.
- Hence, we want to give you an idea how universal this approach is.

The Drama Corpus

Mäster Olof

Skådespel i fem akter

Q7724
Q1765367
1881

Swe DraCor

August Strindberg

b. 1849, Storkyrkoförsamlingen
d. 1912, Adolf Fredriks parish

This tab shows a co-occurrence network. If characters appear in the same scene or act, they are linked.

Network Speech distribution Full text Downloads

Network properties

Segments: 9

All-in at segment 8 (at 89%)

Network size: 35

Density: 0.58

Diameter: 2

Average path length: 1.42

Average clustering coefficient: 0.87

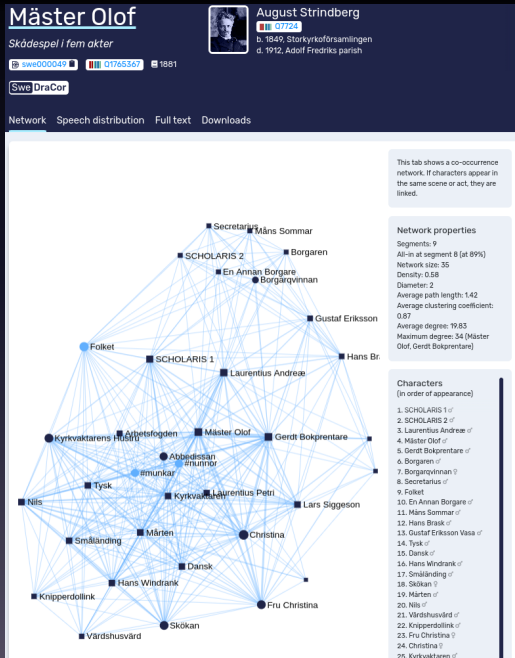
Average degree: 19.83

Maximum degree: 34 (Mäster Olof, Gerdt Bokrentare)

Characters

(In order of appearance)

1. SCHOLARIS 1 of
2. SCHOLARIS 2 of
3. Laurentius Andreæ of
4. Mäster Olof of
5. Gerdt Bokrentare of
6. Borgaren of
7. Borgartvinnan of
8. Secretarius of
9. Folket
10. En Annan Borgare of
11. Mäns Sommar of
12. Hans Brask of
13. Gustaf Eriksson Vasa of
14. Tyask of
15. Dansk of
16. Hans Windrank of
17. Småländing of
18. Skökan of
19. Märten of
20. Nils of
21. Världshusvärd of
22. Knipperdöllink of
23. Fru Christina of
24. Christina of
25. Kyrkvaktare of



“The strategy of a play”

(Brainerd and Neufeldt 1974)

Marcus defines characters p and q as

CONCOMITANT if $A(p) = A(q)$,

INDEPENDENT if $A(p) - A(q) \neq \emptyset \neq A(q) - A(p)$,

ALTERNATIVE if $A(p) \cap A(q) = \emptyset$,

COMPLEMENTARY if p and q are alternative and $A(p) \cup A(q)$ contains all the scene numbers.

Finally, if $A(p) \subset A(q)$, then p IS DOMINATED BY q .

“The strategy of a play”

(Brainerd and Neufeldt 1974)

Marcus defines characters p and q as

CONCOMITANT if $A(p) = A(q)$,

INDEPENDENT if $A(p) - A(q) \neq \emptyset \neq A(q) - A(p)$,

ALTERNATIVE if $A(p) \cap A(q) = \emptyset$,

COMPLEMENTARY if p and q are alternative and $A(p) \cup A(q)$ contains all the scene numbers.

Finally, if $A(p) \subset A(q)$, then p IS DOMINATED BY q .

in FCA notation

concomitant $p' = q'$

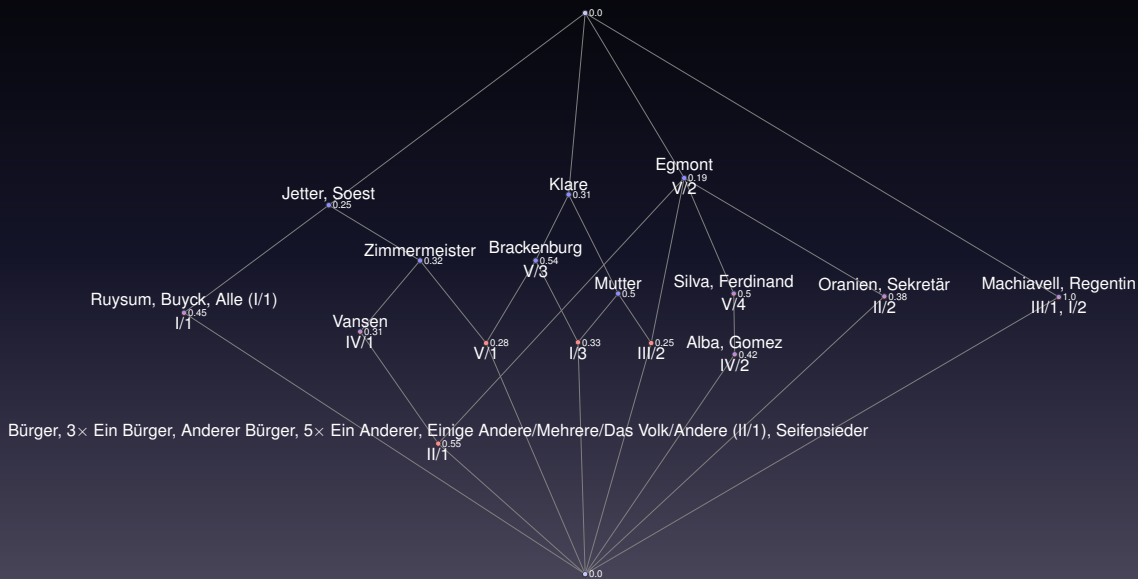
independent $p' \setminus q' \neq \emptyset \neq q' \setminus p'$

alternative $p' \cap q' = \emptyset$

complementary $p' \cap q' = \emptyset$ and $p' \cup q' = G$

p is dominated by q $p' \subset q'$

Characters and Scenes in Goethe's play "Egmont"

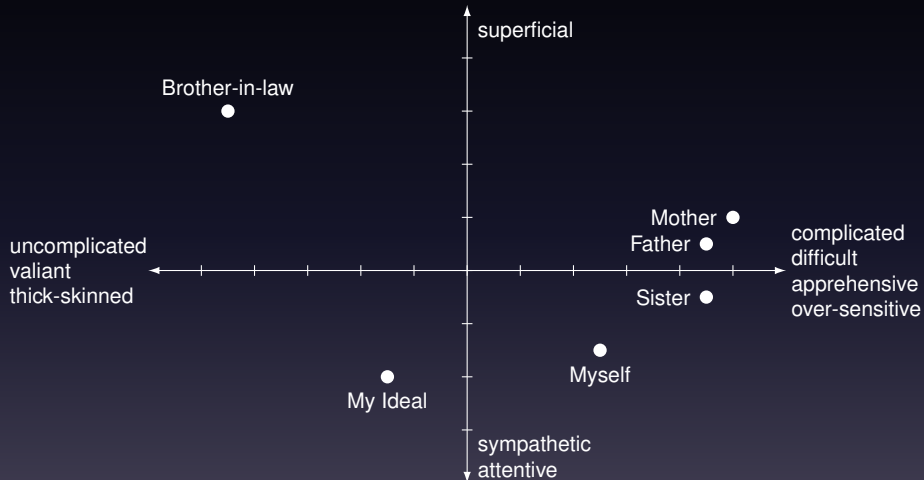


From a treatment of *Anorexia nervosa*

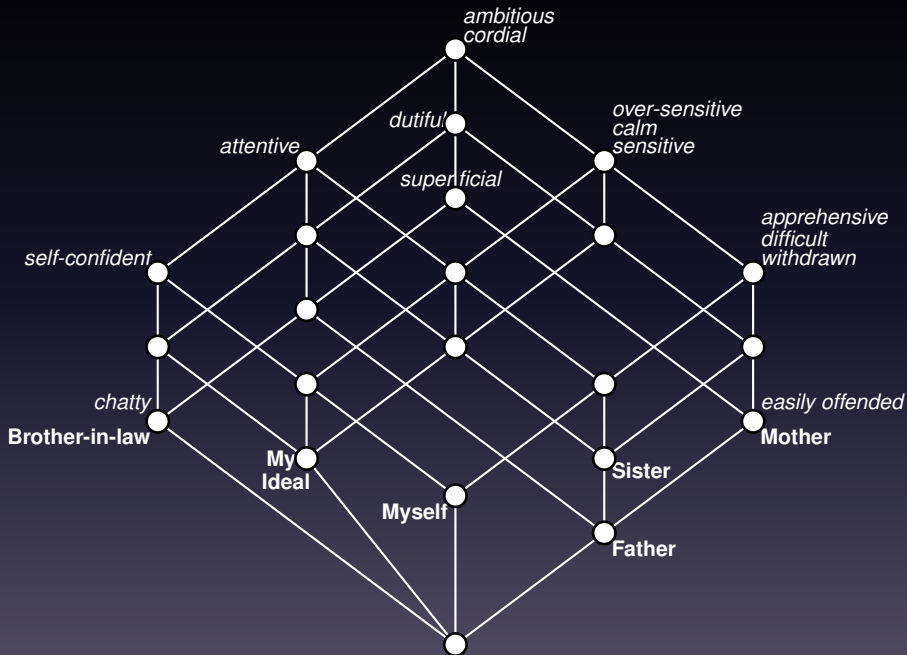
Wolff and Gabler 1998

	over-sensitive	withdrawn	self-confident	dutiful	cordial	difficult	attentive	easily offended	calm	apprehensive	chatty	superficial	sensitive	ambitious
Myself	×	×	×		×	×	×		×	×			×	×
My Ideal	×		×	×	×		×		×				×	×
Father	×	×		×	×	×	×	×	×	×		×	×	×
Mother	×	×		×	×	×		×	×	×		×	×	×
Sister	×	×		×	×	×	×		×	×			×	×
Brother-in-law			×	×	×		×				×	×		×

A biplot of the interview data



A concept lattice of Anorexia data



...and some tools tools ...

fcaR — Formal Concept Analysis with R

Cordero, Enciso, López-Rodríguez and Mora 2022

fcaR: Formal Concept Analysis

Provides tools to perform fuzzy formal concept analysis, presented in Wille (1982) < It provides functions to load and save a formal context, extract its concept lattice and build recommendation systems.

Version: 1.2.2
Depends: R (≥ 4.1)
Imports: [dplyr](#), [forcats](#), [fractional](#), [ggplot2](#), [glue](#), [grDevices](#), [Matrix](#), [meth](#)
LinkingTo: [Rcpp](#)
Suggests: [arules](#), [covr](#), [hasseDiagram](#), [knitr](#), [markdown](#), [rmarkdown](#), [testthat](#)
Published: 2023-11-30
Author: Domingo Lopez Rodriguez  [aut, cre], Angel Mora [aut], Jes
Maintainer: Domingo Lopez Rodriguez <dominlopez@uma.es>
BugReports: <https://github.com/Malaga-FCA-group/fcaR/issues>
License: [GPL-3](#)
URL: <https://github.com/Malaga-FCA-group/fcaR>
NeedsCompilation: yes
Materials: [README NEWS](#)
CRAN checks: [fcaR results](#)

Documentation:

Reference manual: [fcaR.pdf](#)
Vignettes: [Integration with the arules package](#)
[Using FormalContexts and ConceptLattices](#)
[Conceptual Scaling](#)
[Extending fcaR: Equivalence Rules for Implications](#)
[Working with ImplicationSets](#)

CONTRIBUTED RESEARCH ARTICLE

fcaR, Formal Concept Analysis with R

by Pablo Cordero, Manuel Enciso, Domingo López-Rodríguez, and Ángel Mora

Abstract Formal concept analysis (FCA) is a solid mathematical framework to manage information based on logic and lattice theory. It defines two explicit representations of the knowledge present in a dataset as concepts and implications. This paper describes an R package called *fcaR* that implements FCA's core notions and techniques. Additionally, it implements the extension of FCA to fuzzy datasets and a simplification logic to develop automated reasoning tools. This package is the first to implement FCA techniques in R. Therefore, emphasis has been put on defining classes and methods that could be reusable and extensible by the community. Furthermore, the package incorporates an interface with the *arules* package, probably the most used package regarding association rules, closely related to FCA. Finally, we show an application of the use of the package to design a recommender system based on logic for diagnosis in neurological pathologies.

1 Introduction

The main goal of knowledge retrieval and knowledge discovery systems is to extract hidden patterns, trends, behaviours, or rules to solve large-impact real-world problems. Usually, these problems present heterogeneous data sources and are at the core of more general decision processes.

A fundamental principle is that the extracted knowledge should provide some understanding of the analyzed data. As computational systems get in critical and sensitive areas such as medicine, the justice system or financial markets, the knowledge becomes much more relevant to make predictions or recommendations or to detect common interest groups or leaders. However, in many cases, the inability of humans to understand the extracted patterns seems problematic.

To represent and retrieve knowledge from datasets, it has become more important to use formal methods based on logic tools. The formal representation of knowledge and the use of logic tools are more suitable for providing understandable answers. Therefore, it can help avoid the lack of interpretability and explainability of the results.

In particular, formal concept analysis (FCA) (Wille, 1982; Ganter and Wille, 1999) is a well-founded mathematical tool, based on lattice theory and logic, which can retrieve and store the knowledge in the form of concepts (analogous to closed itemsets in transactional databases) and implications (association rules with confidence 1). From this perspective, FCA constitutes a framework that complements and extends the study of exact and approximate association rules.

The origins of FCA were devoted to the study of binary datasets (formal contexts) where variables are called attributes. A relevant extension of FCA uses fuzzy sets (Beloňáček and Vychodil, 2016, 2017) to model real-world problems since datasets may contain imprecise, graded or vague information that is not adequately represented as binary values. The fuzzy extension can also model problems with numerical and categorical attributes since these can be scaled to a truth value describing the degree of fulfilment of the attribute.

Some authors have considered the use of FCA in machine learning. Kuznetsov (2004) relates FCA to some mathematical models of machine learning. Igratov (2017) summarizes the main topics in machine learning and data mining where FCA has been applied: frequent itemset mining and

<https://cran.rstudio.com/web/packages/fcaR/index.html>

Concepts — A Python Library for FCA

<https://github.com/xflr6/concepts>

[README](#) [MIT license](#)

Concepts

python 3.6-3.7

license MIT

python 2.7 | 3.5 | 3.6 | 3.7 | 3.8

wheel yes

Build passing

codecov 100%

docs passing

docs passing

Concepts is a simple Python implementation of Formal Concept Analysis (FCA).



concepts
Formal Concept Analysis

FCA provides a mathematical model for describing a set of **objects** (e.g. King Arthur, Sir Robin, and the holy grail) with a set of **properties** (e.g. human, knight, king, and mysterious) which each of the objects either has or not. A table called *formal context* defines which objects have a given property and vice versa which properties a given object has.

Links

- GitHub: <https://github.com/xflr6/concepts>
- PyPI: <https://pypi.org/project/concepts/>
- Documentation: <https://concepts.readthedocs.io>
- Changelog: <https://concepts.readthedocs.io/en/latest/changelog.html>
- Issue Tracker: <https://github.com/xflr6/concepts/issues>
- Download: <https://pypi.org/project/concepts/#files>

Installation

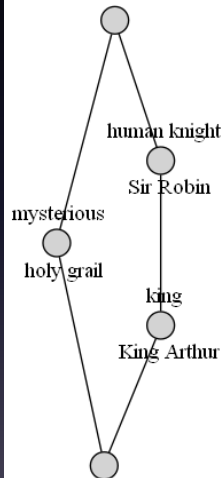
```
>>> context['Sir Robin', 'holy grail']  
(('King Arthur', 'Sir Robin', 'holy grail'), ())  
  
>>> context['king',]  
(('King Arthur',), ('human', 'knight', 'king'))
```

Iterate over the **concept lattice** of all objects-properties pairs:

```
>>> for extent, intent in context.lattice:  
...     print(extent, intent)  
() ('human', 'knight', 'king', 'mysterious')  
('King Arthur',) ('human', 'knight', 'king')  
('holy grail',) ('mysterious',)  
('King Arthur', 'Sir Robin') ('human', 'knight')  
('King Arthur', 'Sir Robin', 'holy grail') ()
```

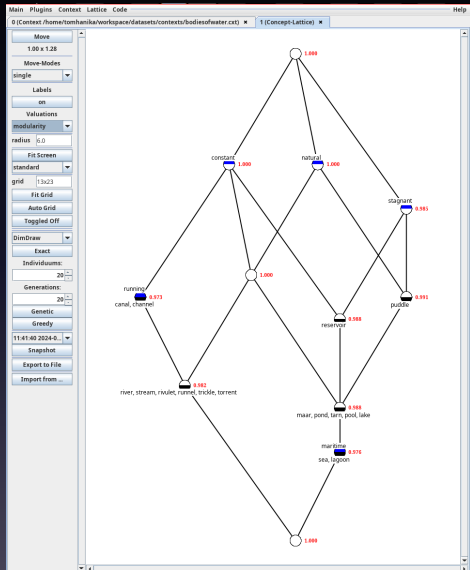
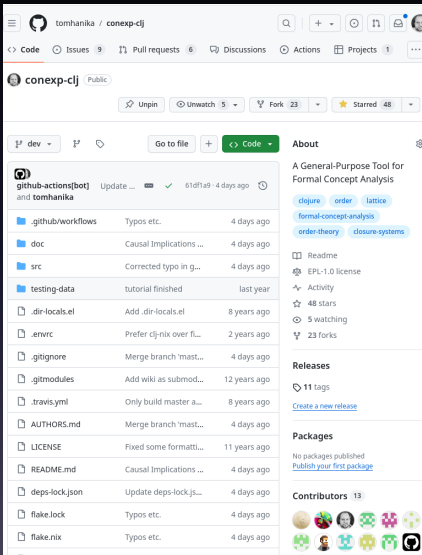
Make a Graphviz visualization of the lattice (use `.graphviz(view=True)` to directly render it and display the resulting PDF):

```
>>> context.lattice.graphviz() # doctest: +ELLIPSIS  
<graphviz.graphs.Digraph object at 0x...>
```



Hanika and Hirth 2019

Hanika and Hirth 2019



<https://github.com/tomhanika/conexp-clj>

A good overview: Uta's Webiste

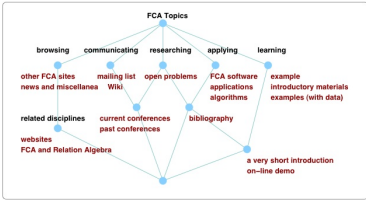
[Formal Concept Analysis Homepage](#) -- [Table of Contents/Sitemap](#) -- [Uta Priss](#)



Formal Concept Analysis Homepage

Formal Concept Analysis (FCA) is a mathematical theory of data analysis using formal contexts and concept lattices. The contents of this site can be accessed by clicking on one of the red **objects** in the concept lattice below or via the [table of contents](#).

On-line demos: [FCA Demo](#), Concept Neighbourhoods in [Roget's Thesaurus](#) and [WordNet](#).



Current Conferences on Formal Concept Analysis

- ICFCA, ICCS and CLA will be held as a joint conference called **CONCEPTS 2024** in Cadiz, Spain.
- Caution: there are some FAKE conferences that pretend to be ICFCA or ICCS conferences. It should be easy to recognise that they are fake because they do not have any FCA researchers on their committees. Furthermore, an internet search for the website domain name and "scam" or "fake" will provide evidence of the fake nature.
- There will be no CLA conference in 2023. ([permanent CLA website](#)).
- [FCA4AI](#) Macau, co-located with IJCAI, August 2023. (Caution: I am not sure whether the link is safe)

Conferences with special sessions/workshops on FCA:

Also possibly of interest: [Diagrams conference series CARLA workshop series](#)

Mailing List

Information about the [Formal Concept Analysis mailing list](#)

Links

- [Bibliography](#) (links to on-line publications, other bibliographies and the tables of contents of conference proceedings) and [Past conferences](#)
- Links to [Other FCA websites](#), [News and sites in related disciplines](#)
- [Introductory materials and examples](#)
- The participants of ICFCA'06 compiled a [List of open problems](#). The community invites anybody who has a solution to any of these to submit it to any of the next FCA conferences.

FCA Software

<https://upriss.github.io/fca/fca.html>

Request for Data!

Should you decide to explore your data using *Conceptual Exploration* or techniques from *Formal concept analysis*:

Consider contributing your data / your examples to the *FCA-Repository*
(<https://fcarepository.org>)

The screenshot shows the GitHub repository page for 'fcertools / contexts'. The repository is public and has 4 stars, 4 watchers, and 2 forks. The main branch is 'main'. The repository contains a collection of formal contexts (datasets). The file list shows:

File	Description	Last Commit
contexts	cleanup whitespace	3 months ago
CNAME	Create CNAME	3 months ago
LICENSE	Initial commit	2 years ago
README.md	using absolute paths such that lin...	3 months ago
contexts.yaml	fixed yaml	2 months ago

The repository is described as 'A Collection of Formal Contexts (Datasets)' and is licensed under CC0-1.0. It includes a README, a CNAME file, a LICENSE file, and a contexts.yaml file.

References

- [BN74] Barron Brainerd **and** Victoria Neufeldt. “On Marcus’ method for the analysis of the strategy of a play”. **in** *Poetics*: 10 (1974), **pages** 31–4.
- [Cor+22] Pablo Cordero **and** others. “fcaR, Formal Concept Analysis with R.”. **in** *R Journal*: 14.1 (2022).
- [HH19] Tom Hanika **and** Johannes Hirth. “Conexp-Clj-A Research Tool for FCA.”. **in** *ICFCA (Supplements)*: 2378.70-75 (2019), **page** 346.
- [WG98] Karl Erich Wolff **and** Siegfried Gabler. “Comparison of visualizations in formal concept analysis and correspondence analysis”. **in** *Visualization of categorical data*: Elsevier, 1998, **pages** 85–97.