

Formal Concept Analysis in R

Motivation, success stories and future work with the **fcaR** library

Domingo López-Rodríguez

Department of Applied Mathematics
Universidad de Málaga

Personal homepage: <https://dominlopez.netlify.app>
fcaR homepage: <https://malaga-fca-group.github.io/fcaR>

Motivation

Why to develop an R package for FCA?

- R, together with Python, are the two most widely used programming languages in Machine Learning and Data Science.
- In R there are already libraries for association rule mining that have become standard: **arules**.
- There is no library in R that implements the basic ideas and functions of FCA and allows them to be used in other contexts.

Our purpose

- To help disseminate FCA as a knowledge discovery tool.
- To be able to perform rapid testing of new ideas, algorithms, etc., both from a theoretical and practical point of view.
- Rapid prototyping of new solutions that can be integrated into more complex computational systems.
- To enable the application of FCA to real problems: automatic reasoning and recommender systems.

Design principles

Usability

- Direct execution of most classical algorithms (even in the fuzzy setting).
- Provide methods to operate on contexts, concept lattice and implications.
- **Logic**: include the SL_{FD} logic to compute closure wrt implication sets.
- Interoperability:
 - Read/write datasets in various formats (CSV, CTX, ...).
 - Import and export to **arules**.
- Allow reproducible research.
- Provide lots of documentation with examples.

Implementation

- Modern programming paradigms (object-oriented).
- Classes representing entities: contexts, lattices, implications. . .
- Allow for extensions: new algorithms, new ideas. . .
- Use base R for the interface, but bottlenecks implemented in C.

The **fcaR** library

Library availability



Available Packages

Currently, the CRAN package repository features 18994 available packages.

Contributed Packages

The package is in a stable phase in a repository on Github and on CRAN.

- Unit tests
- Vignettes with demos
- Status:
 - lifecycle: stable
 - CRAN version: 1.1.1
 - downloads: ~19K

Classes and methods

Classes

Class name	Use
"Set"	A basic class to store a fuzzy set using sparse matrices
"Concept"	A pair of sets (extent, intent) forming a concept for a given formal context
"ConceptLattice"	A set of concepts with their hierarchical relationship. It provides methods to compute notable elements, sublattices and plot the lattice graph
"ImplicationSet"	A set of implications, with functions to apply logic and compute closure of attribute sets
"FormalContext"	It stores a formal context, given by a table, and provides functions to use derivation operators, simplify the context, compute the concept lattice and the Duquenne-Guigues basis of implications

Table 1: Main classes found in **fcaR**.

Main methods

Formal Contexts

intent
extent
closure
clarify
reduce
standardize
find_concepts
find_implications

Concept Lattice

supremum
infimum
sublattice
meet_irreducibles
join_irreducibles
subconcepts
superconcepts
lower_neighbours
upper_neighbours

Implication Set

closure
recommend
apply_rules
to_basis

A remark on the Simplification Logic

SL_{FD}	Equivalence rules
[Ref] $\frac{A \supseteq B}{A \Rightarrow B}$	
[Frag] $\frac{A \Rightarrow B \cup C}{A \Rightarrow B}$	$\{A \Rightarrow B\} \equiv \{A \Rightarrow B \setminus A\}$
[Comp] $\frac{A \Rightarrow B, C \Rightarrow D}{A \cup C \Rightarrow B \cup D}$	$\{A \Rightarrow B, A \Rightarrow C\} \equiv \{A \Rightarrow BC\}$
[Simp] $\frac{A \Rightarrow B, C \Rightarrow D}{A(C \setminus B) \Rightarrow D \setminus B}$	$A \subseteq C \Rightarrow \{A \Rightarrow B, C \Rightarrow D\} \equiv \{A \Rightarrow B, A(C \setminus B) \Rightarrow D \setminus B\}$ $A \subseteq D \rightarrow \{A \Rightarrow B, C \Rightarrow BD\} \equiv \{A \Rightarrow B, C \Rightarrow D\}$

The SL_{FD} closure algorithm makes use of the above equivalence rules to compute the closure X^+ of a set X using a set of implications Σ , and return a simplified Σ' where the attributes in X^+ do not appear, and such that:

$$\{\emptyset \Rightarrow X\} \cup \Sigma \equiv \{\emptyset \Rightarrow X^+\} \cup \Sigma'$$

Practical example of the functionalities

Context and derivation operators

Concept lattice

Implications and logic

Conceptual scaling

Reproducible research with `fcaR` and interoperability

All classes have a `to_latex()` method to export in a suitable form to a \LaTeX document:

- Tables (for formal contexts):

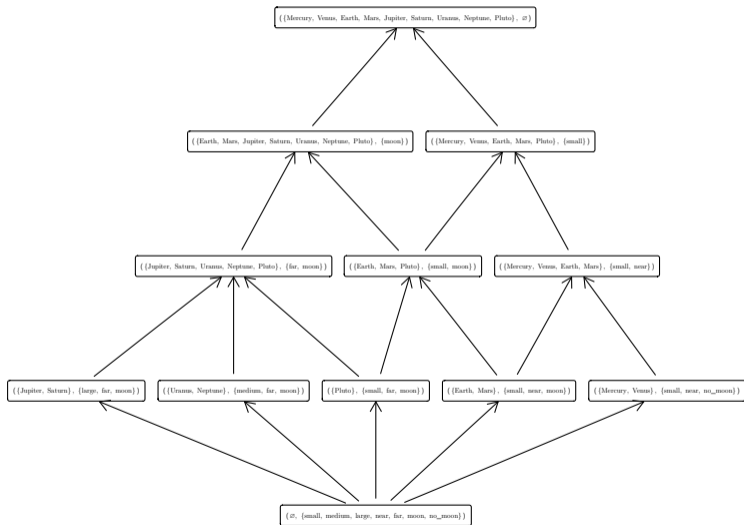
	small	medium	large	near	far	moon	no_moon
Mercury	×			×			×
Venus	×			×			×
Earth	×			×		×	
Mars	×			×		×	
Jupiter			×		×	×	
Saturn			×		×	×	
Uranus		×			×	×	
Neptune		×			×	×	
Pluto	×				×	×	

Table 2

- Listings (for concepts, implications...):

1:	{no_moon}	⇒	{small, near}
2:	{far}	⇒	{moon}
3:	{near}	⇒	{small}
4:	{large}	⇒	{far, moon}
5:	{medium}	⇒	{far, moon}
6:	{medium, large, far, moon}	⇒	{small, near, no_moon}
7:	{small, near, moon, no_moon}	⇒	{medium, large, far}
8:	{small, near, far, moon}	⇒	{medium, large, no_moon}
9:	{small, large, far, moon}	⇒	{medium, near, no_moon}
10:	{small, medium, far, moon}	⇒	{large, near, no_moon}

- Plots (for formal contexts, lattice):



fcaR code can be embedded in RMD files (plain text + code + results) and produce a presentation (such as this one!) or a complete paper:

fcaR, Formal Concept Analysis with R

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

Implications and logic

The knowledge stated in a formal context can also be represented as a set of implications, which are expressions of the form $A \Rightarrow B$ where A and B are sets of attributes or items, indicating that, for every object in which the set of attributes A is present, also B is present. This interpretation is similar to the one defined in data mining (machine learning over the so-called association rules. The confidence (a well-known estimate of the rule's quality) has value 1 in all implications.

For instance $\{P_1, P_2\} \Rightarrow \{P_4\}$ is a valid implication in the previous example, having the following interpretation: when the attribute P_1 has degree at least 0.5 then we have P_4 with degree 1.

The Duquenois-Gallager basis of implications (Duquenois and Gallager, 1986) is a set of valid implications from which all other valid implications can be deduced. The Duquenois-Gallager basis in our example is given by:

1. $\emptyset \Rightarrow \{P_1, P_2, P_3, P_4\}$
2. $\{P_1, P_2\} \Rightarrow \{P_4\}$
3. $\{P_2, P_3\} \Rightarrow \{P_4\}$
4. $\{P_1, P_3\} \Rightarrow \{P_4\}$
5. $\{P_1, P_2, P_3\} \Rightarrow \{P_4\}$
6. $\{P_1, P_2, P_3, P_4\} \Rightarrow \{P_4\}$

In Cordero et al. (2012), the simplification logic, denoted as SL_{P_1} , was introduced as a method to manipulate implications (functional dependencies or if-then rules), removing redundancies or computing closures of attributes. This logic is equivalent to Armstrong's Axioms (Armstrong, 1978), which are well known from the 80s in databases, artificial intelligence, formal concept analysis, and others. The axiomatic system of this logic considers reflexivity as the axiom scheme

$$[A] \Rightarrow \frac{A}{A}$$

together with the following inference rules called *fragmentation*, *composition* and *simplification*, respectively, which are equivalent to the classical Armstrong's axioms of augmentation and, more importantly, transitivity:

$$[A \Rightarrow B] \text{ and } [B \Rightarrow C] \Rightarrow [A \Rightarrow C] \quad [A \Rightarrow B \text{ and } C] \Rightarrow [A \Rightarrow B \text{ and } C] \quad [A \Rightarrow B \text{ and } C] \Rightarrow [A \Rightarrow B]$$

The main advantage of SL_{P_1} with respect to Armstrong's Axioms is that the inference rules may be considered as equivalence rules (see the work by [Llorens et al. \(2017\)](#) for further details and proofs), that is, given a set of implications Σ , the application of the equivalence transforms it into an equivalent set. In the package presented in this paper, we develop the following equivalences:

1. **Fragmentation Equivalency [Frag]** $[A \Rightarrow B] \Rightarrow [A \Rightarrow B \text{ and } A]$.
2. **Composition Equivalency [Comp]** $[A \Rightarrow B \text{ and } A \Rightarrow C] \Rightarrow [A \Rightarrow B \text{ and } C]$.
3. **Simplification Equivalency [Simp]** $[A \Rightarrow C]$ and $[A \Rightarrow B \text{ and } C] \Rightarrow [A \Rightarrow B]$.
4. **Right-Simplification Equivalency [RSimp]** $[A \Rightarrow C]$ and $[A \Rightarrow B \text{ and } C] \Rightarrow [A \Rightarrow B]$.

Usually, many axioms, the implications have always atomic attributes on the right-hand side. We emphasize that this logic can manage aggregated implications, i.e. implications' consequents do not have to be singletons. This represents an increase of the logic efficiency.

The logic removes attribute redundancies in some of the implications in the Duquenois-Gallager basis presented before. Particularly, the implications with numbers 2, 3, 4, 5 and 6 are simplified to:

2. $\emptyset \Rightarrow \{P_4\}$
3. $\{P_1\} \Rightarrow \{P_4\}$
4. $\{P_2\} \Rightarrow \{P_4\}$
5. $\{P_3\} \Rightarrow \{P_4\}$
6. $\{P_1, P_2, P_3\} \Rightarrow \{P_4\}$

One of the primary uses of a set of implications is computing the closure of a set of attributes, the maximal fuzzy set that we can arrive at from these attributes using the given implications.

Derivation operators

The methods that implement the derivation operators are named after their: `intext()`, `extint()` and `closure()`. They can be applied on objects of type "fct", representing fuzzy sets of objects or attributes:

```
> s <- fctrow(fchobjctx, obj = 1, obj = 1)
> s
[0], [0]
> fctintext(s)
[0] [0.5], [0.5]
> fctextint(s)
[0] [0.5], [0.5]
> fctclosure(s)
[0] [0.5], [0.5]
> fctclosure(T)
[0] [0.5], [0.5], [0.5]
```

In addition, we can perform derivations on the formal context, by using `fctderiv()`, giving:

```
fctderiv(context with 3 objects and 4 attributes)
[0] [0.5] [0.5] [0.5]
[0] [0.5] [0.5]
[0] [0.5] [0.5]
[0] [0.5] [0.5] [0.5]
```

The duplicated rows and columns in the formal context have been collapsed, and the corresponding attributes and objects' names are grouped together between brackets, e.g. $\{P_2, P_4\}$.

Concept Lattices

The command to compute the concept lattice for a "formalcontext" `f` is `fctint_concepts()`. The lattice is stored in `fctconcepts`, which is of the "ConceptLattice" class.

fctconcepts

```
A set of 8 concepts:
1: ([0], [0], [0], [0], [0]) [0.5], [0.5]
2: ([0], [0], [0], [0], [0]) [0.5], [0.5]
3: ([0], [0], [0], [0], [0]) [0.5], [0.5], [0.5]
4: ([0], [0], [0], [0], [0], [0]) [0.5], [0.5], [0.5], [0.5]
5: ([0], [0], [0], [0], [0], [0]) [0.5], [0.5], [0.5]
6: ([0], [0], [0], [0], [0], [0]) [0.5], [0.5], [0.5]
7: ([0], [0], [0], [0], [0], [0]) [0.5], [0.5], [0.5]
8: ([0], [0], [0], [0], [0], [0]) [0.5], [0.5], [0.5]
```

In order to know the cardinality of the set of concepts (that is, the number of concepts), we can use `fctconceptslen()`, which gives 8 in this case. The complete list of concepts can be printed with `fctconceptsintext()`, or simply `fctconcepts`. Also, they can be translated to RDS using the `to_rds()` method, as mentioned before.

The typical subsetting operation in R with brackets is implemented to allow specific concepts from the lattice, giving their names as a boolean vector indicating which concepts to keep. The same rules for subsetting as in R base apply:

```
> fctconcepts[c(1, 3, 5, 8)]
A set of 4 concepts:
1: ([0], [0], [0], [0], [0]) [0.5], [0.5]
2: ([0], [0], [0], [0], [0]) [0.5], [0.5], [0.5]
3: ([0], [0], [0], [0], [0]) [0.5], [0.5], [0.5], [0.5]
5: ([0], [0], [0], [0], [0], [0]) [0.5], [0.5], [0.5], [0.5]
8: ([0], [0], [0], [0], [0], [0]) [0.5], [0.5], [0.5]
```

In addition, the user can compute concepts' support (the proportion of objects whose set of attributes contains the intent of a given concept) by means of `fctconceptsupper()`:

```
> fctconceptsupper()
[1] 1.00 0.50 0.25 0.50 0.00 0.50 0.00 0.00
```



Figure 4: Hasse diagram for a sublattice of the fcaR22 formal context.

an individual and returns the degree of the diagnosis attributes using the implications extracted from the formal context as an inference engine.

Next, we use the `NEXTLUTEM` algorithm to extract implications and compute the set of concepts, using `fctint_concepts()`.

The concept lattice is big (9476 concepts), therefore, it cannot be plotted here for space and readability reasons. For this reason, we only plot a sublattice of small size in Figure 4.

There is an aggregate of 965 implications extracted. Let us compute the average cardinality of the LHS and the RHS of the extracted rules:

```
> calcMeans(fctimplicationsize())
LHS RHS
2.47301 1.61648
```

Note that our paradigm can deal with non-collid implications, that is, where there is more than one attribute in the RHS of the implication. This feature is an extension of what is usual in other paradigms, for example, in transactional databases.

We can use the simplification logic to remove redundancies and reduce the LHS and RHS size of the implications. The reason to do this is to decrease the computational cost of computing closures:

```
> fctimplicationsimplify(rules = c("simplification", "reinfication"))
calcMeans(fctimplicationsize())
LHS RHS
1.98988 1.62741
```

We can see that the average cardinality of the LHS has been reduced from 2.473 to 1.988 and that the size of the RHS, from 1.616 to 1.627.

With the simplified implication logic, we can build a recommender system by simply wrapping the `recommend()` method inside a function:

```
> diagnose <- function(x) {
+   fctint_concepts_recomm(x) = 5,
+   attribute_fillter =
+     c("diagnose", "diagnose"))
+ }
}
```

This function can be applied to "fct"s that have the same attributes as those of the formal context. The `attribute_fillter` argument specifies which attributes are of interest. In our case, the diagnosis attributes.

Let us generate some sets of attributes and get the recommendation (`diagnose()`) for each one:

```
> s1 <- fctrow(attributes = fctattributes,
+             CGAS_1 = 1/2, CGAS_2 = 1, CGAS_3 = 1/2,
+             CGAS_4 = 1/4, CGAS_5 = 1/2, CGAS_6 = 1)
> diagnose(s1)
diagnose
diagnose_diagnose
1 0
```

Integration with **arules**

arules is a standard in R to compute and manage (unit) association rules.

The implications computed *or simplified* using **fcaR** can be exported to the **arules** format.

- Since, in many cases, other people know about association rules mining and its terminology, it is easy to introduce FCA and its methods with this integration.

Where to find help

<https://malaga-fca-group.github.io/fcaR/>



fcaR: Tools for Formal Concept Analysis

The aim of this package is to provide tools to perform fuzzy formal concept analysis (FCA) from within R. It provides functions to load and save a Formal Context, extract its concept lattice and implications. In addition, one can use the implications to compute semantic closures of fuzzy sets and, thus, build recommendation systems.

Stories of success

Where have we used **fcaR**?

The ways in which we have used **fcaR** so far are:

- From a theoretically point of view:
 - Rapid development and checking of new ideas: **fcaR** allows for a fast iteration of the cycle **theory** - **practice** - **theory**.
- With practical purposes:
 - Use the simplification logic for automated reasoning and creation of recommender systems.
 - Explore the concept lattice in real-world problems to model and extract knowledge.

Recommender systems



Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa

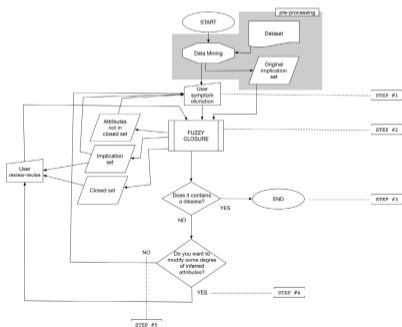


A conversational recommender system for diagnosis using fuzzy rules

P. Cordero^a, M. Enciso^b, D. López^{a,*}, A. Mora^a

^aDept. of Applied Mathematics, Universidad de Málaga, Andalucía Tech, Málaga, Spain

^bDept. of Computer Science, Universidad de Málaga, Andalucía Tech, Málaga, Spain



Comparison of the current proposal to other recommender systems and machine learning methods.

	Accuracy	Sensitivity	Specificity	Precision
ALS	0.360	0.333	0.380	0.290
IBCF (Cosine)	0.555	0.475	0.615	0.483
IBCF (Pearson)	0.770	0.466	1.000	1.000
LIBMF	0.491	0.901	0.181	0.455
SVD	0.376	0.515	0.271	0.349
SVDF	0.431	1.000	0.000	0.431
UBCF (Cosine)	0.608	0.967	0.335	0.524
UBCF (Pearson)	0.525	0.783	0.330	0.470
C5.0	0.674	0.636	1.000	1.000
PART	0.883	0.847	0.950	0.970
JRip	0.752	0.814	0.688	0.731
Random Forest	0.953	0.924	1.000	1.000
xgboost	0.818	0.963	0.713	0.706
k-nn	0.589	0.603	0.544	0.815
Proposal	0.982	0.996	0.948	0.955

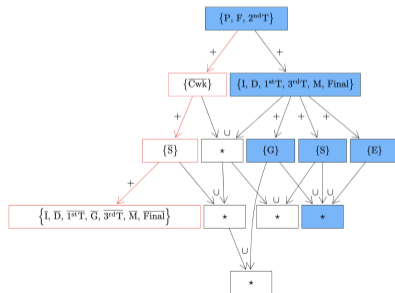
Mixed attributes

Used for knowledge spaces and learning paths (IJCIS, submitted)

Minimal generators from positive and negative attributes: analysing the knowledge space of a maths course






Closed set	Min. gen.	Closed set	Min. gen.
$\{c, \bar{a}, \bar{d}\}$	$\{\bar{a}, \bar{d}\}$	$\{b, d, \bar{a}, \bar{c}\}$	$\{\bar{a}, \bar{c}\}$
$\{c, \bar{a}, \bar{b}, \bar{d}\}$	$\{\bar{a}, \bar{b}\}$	$\{c, \bar{b}, \bar{d}\}$	$\{\bar{b}, \bar{d}\}, \{c, \bar{b}\}$
$\{b, c, d, \bar{a}\}$	$\{c, d\}$	$\{b, d, \bar{a}\}$	$\{\bar{d}, \bar{a}\}, \{b, d\}$
$\{b, c, \bar{a}\}$	$\{b, c\}$	$\{a, d, \bar{b}, \bar{c}\}$	$\{\bar{b}, \bar{c}\}, \{d, \bar{b}\}, \{a, d\}$
$\{a, c, \bar{b}, \bar{d}\}$	$\{a, c\}$	$\{a, b, c, \bar{d}\}$	$\{\bar{c}, \bar{d}\}, \{a, b\}$

Closed set	Minimal generators
$M \cup \bar{M}$	$\{\bar{b}, \bar{c}, \bar{d}\}, \{\bar{a}, \bar{c}, \bar{d}\}, \{\bar{a}, \bar{b}, \bar{c}\}, \{d, \bar{a}, \bar{b}\}, \{d, \bar{d}\}, \{c, d, \bar{b}\}, \{c, \bar{c}\}, \{a, c, d\}, \{b, \bar{b}\}, \{a, b, d\}, \{a, b, c\}, \{a, \bar{a}\}$



Article

Simplifying Implications with Positive and Negative Attributes: A Logic-Based Approach

Francisco Pérez-Gómez , Domingo López-Rodríguez , Pablo Cordero , Ángel Mora 
and Manuel Ojeda-Aciego 

Departamento Matemática Aplicada, Universidad de Málaga, 29071 Málaga, Spain;
franciscoperezgomez@uma.es (F.P.-G.); dominlopez@uma.es (D.-L.R.); pcordero@uma.es (P.C.);
amora@uma.es (Á.M.)

* Correspondence: aciego@uma.es

Theorem 3. Consider $A, B, C, D \subseteq M\bar{M}$:

[KeyEq'] If there exist $x \in A \cap D$, $y \in B \cap \bar{C}$ with $A \setminus x = C \setminus \bar{y}$, then

$$\{A \rightarrow B, C \rightarrow D\} \equiv \{A \rightarrow B \setminus y, C \setminus \bar{y} \rightarrow y\} \equiv \{A \rightarrow B \setminus y, C \rightarrow M\bar{M}\}.$$

[KeyEq''] If $A \subseteq C \neq \emptyset$ and $B \cap \bar{D} \neq \emptyset$, for any $x \in C$ we have that then

$$\{A \rightarrow B, C \rightarrow D\} \equiv \{A \rightarrow B, C \setminus x \rightarrow \bar{x}\}.$$

[RedEq'] If $D \subseteq B$ and there exists $x \in A \cap \bar{C}$ such that $A \setminus x = C \setminus \bar{x}$, then

$$\{A \rightarrow B, C \rightarrow D\} \equiv \{A \rightarrow B \setminus D, C \setminus \bar{x} \rightarrow D\}.$$

[RftEq] If there exist $x \in A$, $y \in B \cap \bar{C}$ and $A \setminus x = C \setminus \bar{y}$, then

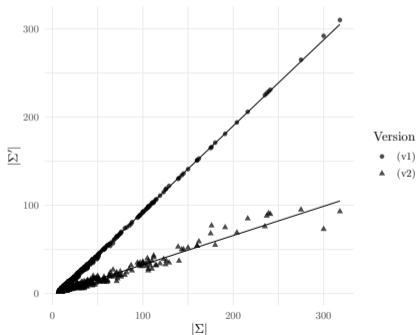
$$\{A \rightarrow B, C \rightarrow D\} \equiv \{A \rightarrow B \setminus y, C \rightarrow D\bar{x}\}.$$

[RftEq'] If there exist $x \in A \cap \bar{D}$, $y \in B \cap \bar{C}$ and $A \setminus x \subseteq C \setminus \bar{y}$, then

$$\{A \rightarrow B, C \rightarrow D\} \equiv \{A \rightarrow B, C \rightarrow D \setminus \bar{x}\}.$$

[MixUnEq] If there exist $x \in A$, $y \in C$ such that $A \setminus x = C \setminus y$ and $b \in D$, then

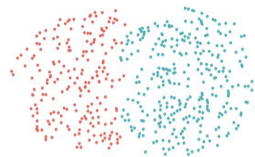
$$\{A \rightarrow b, C \rightarrow D\} \equiv \{(A \setminus x)\bar{b} \rightarrow \bar{x}\bar{y}, C \rightarrow D \setminus b\}.$$



Clustering of implications applied to Social Network Analysis

Clustering and Identification of Core Implications

Domingo López-Rodríguez^(✉)^{ID}, Pablo Cordero^{ID}, Manuel Enciso^{ID},
and Ángel Mora^{ID}



Sentiment Analysis



Positive



Negative



Neutral

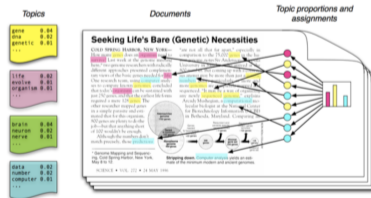


Figure source: Blei, D. M. (2012). Probabilistic topic models. *Communications of the ACM*, 55(4), 77-84.

Collaborations

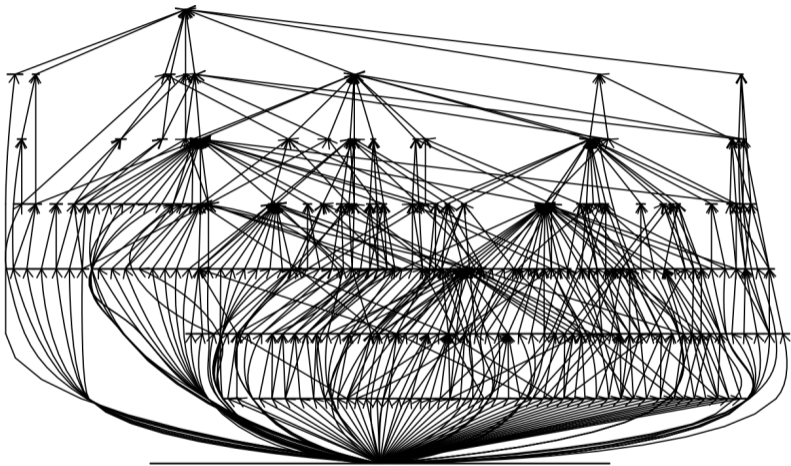
- Construction companies: Integration of FCA-based recommendation systems within the *Building Information Modelling* (BIM) methodology.

Panel de Control

Datos			Resultados		
Título Obra:	<input type="text"/>		Baja:	Importe/Oferita:	Fecha:
Organismo:	Tipo Contratación:	Importe Licitación:	<input type="text"/>	<input type="text"/>	<input type="text"/>
<input type="text"/>	<input type="text"/>	<input type="text"/>	Tipo de Obra:	Organismo:	Tipo Contratación:
Propuesta económica:	Propuesta técnica:	Puntuación total:	<input type="text"/>	<input type="text"/>	<input type="text"/>
<input type="text"/>	<input type="text"/>	<input type="text"/>	Valor TM:	Es Baja Temeraria:	Ganador:
Baja:	Fecha:		<input type="text"/>	<input type="text"/>	<input type="text"/>
<input type="text"/>	<input type="text" value="dd/mm/aaaa"/>		Prop. económica:	Prop. técnica:	Punt. total:
			<input type="text"/>	<input type="text"/>	<input type="text"/>



- Cybersecurity company: Creation of an ontology of malware threats.



{Sophos = bckdr-rxm} \Rightarrow {Avast = metasploit-g, Kaspersky = heur:downloader.os.

- Other research groups:
 - a. Application of fuzzy FCA to neuroimage processing and understanding.
 - b. Use of recommender systems and logic tools to analyse and reduce the Urban Heat Island (UHI) effect for urban planning.
 - c. FCA for concept drift detection in online unsupervised machine learning.

Future works

Future works

We are exploring several lines:

- Improve the usability by non-experts.
- Better integration for external tools (both in R and other languages or via web).
- Develop new algorithms and extensions.

Future developments

Web application

- Web app for **fcaR** (demo)

Some extensions

- Integrate association rules in the library (Luxenburger's basis).
- Logic for mixed attributes: new algorithms to compute bases of mixed implications, iterative closure algorithm. . .
- Other extensions: $\{\circ, +, -, \imath\}$.

Other algorithms

- Concept lattice (InClose, FastCbO, NextNeighbour)
- Canonical basis of implications
- Direct bases and minimal generators.
- Parallelization of the above.

THANK YOU VERY MUCH