# Characterization of Order-like Dependencies with Formal Concept Analysis

Victor Codocedo[1] Jaume Baixeries[2]
Mehdi Kaytoue[1] Amedeo Napoli[3]

[1]INSA Lyon, LIRIS. UMR5205, F-69621, France
[2]Universitat Politcnica de Catalunya. 08032, Barcelona. Catalonia
[3]LORIA (INRIA, CNRS), Villers-ls-Nancy, F-54600, France

August 30 2016 – FCA4AI at ECAI 2016
The Hague, Netherlands

INSTITUT NATIONAL DES SCIENCES APPLIQUÉES LYON

UNIVERSITAT POLITÈCNICA DE CATALUNYA BARCELONATECH

Loria Laboratoire lorrain de recherche en informatique et ses applications

# Functional Dependencies & FCA

This is the last of a trilogy of contributions regarding the relations of Formal Concept Analysis and Functional Dependencies which include:

- Computing **Functional Dependencies** with Pattern Structures (CLA 2012) [Baixeries et al., 2012]
- Computing **Similarity Dependencies** with Pattern Structures (CLA 2013,AMAI 2014) [Baixeries et al., 2013, Baixeries et al., 2014]
- Characterization of **Order-like Dependencies** with FCA (CLA 2016)

Consequently, in what follows we will define:

- Functional Dependencies
- Functional Dependencies & FCA
- Functional Dependency Generalizations
  - Similarity Dependencies
  - Order-like Dependencies
  - Restricted Order Dependencies
  - Lexicographical Ordered Dependencies

# Functional Dependencies: Notations

Let $\mathcal{U} = \{a, b, c, d\}$ be a set of attributes and $Dom = \{1, 3, 4, 7, 8\}$ be a set of numerical values.

| id | a | b | c | d |
|---|---|---|---|---|
| $t_1$ | 1 | 3 | 4 | 1 |
| $t_2$ | 4 | 3 | 4 | 3 |
| $t_3$ | 1 | 8 | 4 | 1 |
| $t_4$ | 4 | 3 | 7 | 3 |

**Notation**

- A tuple $t$ is a function $t : \mathcal{U} \mapsto Dom$ and a data table $T$ is a set of tuples.
- $T$ is represented as a matrix where is $T = \{t_1, t_2, t_3, t_4\}$.
- Given a tuple $t \in T$ and $X = \{x_1, x_2, \ldots, x_n\} \subseteq \mathcal{U}$, we have:

$$t[X] = \langle t[x_1], t[x_2], \ldots, t[x_n] \rangle$$

e.g. $t_2[\{a, c\}] = \langle t_2[a], t_2[c] \rangle = \langle 4, 4 \rangle$

# Functional Dependencies: Definition

Let $T$ be a data table (i.e. a set of tuples), $\mathcal{U}$ a set of attributes and $X, Y \subseteq \mathcal{U}$

| id | a | b | c | d |
|---|---|---|---|---|
| $t_1$ | 1 | 3 | 4 | 1 |
| $t_2$ | 4 | 3 | 4 | 3 |
| $t_3$ | 1 | 8 | 4 | 1 |
| $t_4$ | 4 | 3 | 7 | 3 |

**Definition**

A **Functional Dependency (FD)** $X \rightarrow Y$ holds in $T$ if:

$$\forall t_i, t_j \in T : t_i[X] = t_j[X] \Rightarrow t_i[Y] = t_j[Y]$$

**Example**

Functional dependencies $a \rightarrow d$ and $d \rightarrow a$ hold whereas $a \rightarrow c$ does not hold since $t_2[a] = t_4[a]$ but $t_2[c] \neq t_4[c]$.

# Functional Dependencies: an FCA characterization

Let us consider the tuple set $T$ and attributes in $\mathcal{U}$, we can define formal context $\mathbb{K} = (\mathcal{B}_2(T), \mathcal{U}, I)$, where

- $\mathcal{B}_2(T) = \{(t_i, t_j) \mid i < j \text{ and } t_i, t_j \in T\}$ is an ordered set of pairs of tuples from $G$,
- $((t_i, t_j), x) \in I \Leftrightarrow t_i[x] = t_j[x], \text{ for } x \in \mathcal{U}$

## Example

| id | a | b | c | d |
|---|---|---|---|---|
| $t_1$ | 1 | 3 | 4 | 1 |
| $t_2$ | 4 | 3 | 4 | 3 |
| $t_3$ | 1 | 8 | 4 | 1 |
| $t_4$ | 4 | 3 | 7 | 3 |

| $\mathbb{K}$ | a | b | c | d |
|---|---|---|---|---|
| $(t_1, t_2)$ | | × | × | |
| $(t_1, t_3)$ | × | | × | × |
| $(t_1, t_4)$ | | × | | |
| $(t_2, t_3)$ | | | × | |
| $(t_2, t_4)$ | × | × | | × |
| $(t_3, t_4)$ | | | | |

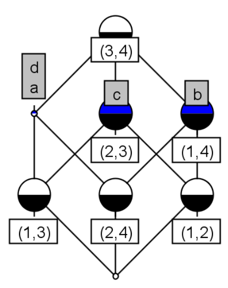
# Functional Dependencies: an FCA characterization

## Functional Dependencies as Implications

A Functional Dependency $X \rightarrow Y$ **holds** in a table $T$ iff $\{X\}'' = \{X, Y\}''$ in the formal context $\mathbb{K} = (\mathcal{B}_2(T), \mathcal{U}, I)$.

## Example

| $\mathbb{K}$ | a | b | c | d |
|---|---|---|---|---|
| $(t_1, t_2)$ | | × | × | |
| $(t_1, t_3)$ | × | | × | × |
| $(t_1, t_4)$ | | × | | |
| $(t_2, t_3)$ | | | × | |
| $(t_2, t_4)$ | × | × | | × |
| $(t_3, t_4)$ | | | | |




- $ac \rightarrow d$ holds since $\{a, c\}'' = \{a, c, d\}$ and $\{a, c, d\}'' = \{a, c, d\}$,
- $a \rightarrow b$ does not hold since $\{a\}'' = \{a, d\}$ and $\{a, b\}'' = \{a, b, d\}$.

# Functional Dependency Generalizations

**Definition**

A **Functional Dependency (FD)** $X \rightarrow Y$ holds in $T$ if:

$$\forall t_i, t_j \in T : t_i[X] = t_j[X] \Rightarrow t_i[Y] = t_j[Y]$$

Two types of Generalizations [Caruccio et al., 2016]:

- Extent Relaxing: **not really for all** in $\forall t_i, t_j \in T$
- Attribute Relaxing: **not really equal** in $t_i[X] = t_j[X] \Rightarrow t_i[Y] = t_j[Y]$

**In this article we cover Attribute Relaxing Dependencies**

- Similarity Dependencies (already covered - just a recall)
- Order-like Dependencies
- Restricted Order Dependencies
- Lexicographical Ordered Dependencies

# Similarity Dependencies [Baixeries et al., 2013]

## Functional Dependency Generalizations

When tuple values for a given attribute (or a set of them) are subject to *noise*, **Similarity Dependencies** are proposed to relax the former definition for FDs.

### Definition

Similarity Dependency $X \Rightarrow Y \mid \theta_X, \theta_Y$ holds with $a \simeq_\theta b \iff |a - b| \leq \theta$, $\forall a, b \in \mathbb{R}$ iff:

$$\forall t_i, t_j \in T : t_i[X] \simeq_{\theta_X} t_j[X] \Rightarrow t_i[Y] \simeq_{\theta_Y} t_j[Y]$$

### Example

| id | Month | Year | Av. Temp. | City |
|---|---|---|---|---|
| $t_1$ | 1 | 1995 | 36.4 | Milan |
| $t_2$ | 1 | 1996 | 33.8 | Milan |
| $t_3$ | 5 | 1996 | 63.1 | Rome |
| $t_4$ | 5 | 1997 | 59.6 | Rome |
| $t_5$ | 1 | 1998 | 41.4 | Dallas |
| $t_6$ | 1 | 1999 | 46.8 | Dallas |
| $t_7$ | 5 | 1996 | 84.5 | Houston |
| $t_8$ | 5 | 1998 | 80.2 | Houston |

Similarity Dependency $City, Month \Rightarrow Av.Temp \mid \theta_{Month} = 0, \theta_{Av.Temp} = 6$ holds

# Order-like Dependencies [Ginsburg and Hull, 1983]

## Functional Dependency Generalizations

Analogously, we can consider the scenario when two (sets) of attributes are related through their *structure*.
In such a case, **Order-like Dependencies have been proposed.**

### Definition

Order-like Dependency $X \rightarrow Y$ holds with $\sqsubseteq_X, \sqsubseteq_Y \in \{<, >, \leq, \geq\}$, iff:

$$\forall t_i, t_j \in T : t_i[X] \sqsubseteq_X t_j[X] \Rightarrow t_i[Y] \sqsubseteq_Y t_j[Y]$$

### Example

| id | Month | Year | Av. Temp. | City |
|---|---|---|---|---|
| $t_1$ | 1 | 1995 | 36.4 | Milan |
| $t_2$ | 1 | 1996 | 33.8 | Milan |
| $t_3$ | 5 | 1996 | 63.1 | Rome |
| $t_4$ | 5 | 1997 | 59.6 | Rome |
| $t_5$ | 1 | 1998 | 41.4 | Dallas |
| $t_6$ | 1 | 1999 | 46.8 | Dallas |
| $t_7$ | 5 | 1996 | 84.5 | Houston |
| $t_8$ | 5 | 1998 | 80.2 | Houston |

Order-like Dependency $Month \rightarrow Av.Temp$ holds, with $\sqsubseteq_{Month} = \sqsubseteq_{Av.Temp.} = <$

# Order-like Dependencies [Ginsburg and Hull, 1983]

## FCA Characterization

For each attribute $x \in \mathcal{U}$ we define the **general ordinal scale** $\mathbb{K}_x = (T, T, \sqsubset_x)$

$$\sqsubset_x = \{(t_i, t_j) \mid t_i, t_j \in T,\ t_i \sqsubseteq_x t_j\}$$

For a set of attributes $X \subseteq \mathcal{U}$ we have: $\mathbb{K}_X = (T, T, \bigcap_{x \in X} \sqsubset_x)$

**Proposition 1**

The Order-like Dependency $X \to Y$ holds iff $\mathbb{K}_X = \mathbb{K}_{XY}$

**Example:** $ab \to c$ with $\sqsubseteq_{ab} = \sqsubseteq_c = <$

| id | a | b | c |
|---|---|---|---|
| $t_1$ | 1 | 3 | 1 |
| $t_2$ | 2 | 7 | 2 |
| $t_3$ | 3 | 4 | 4 |
| $t_4$ | 5 | 3 | 9 |
| $t_5$ | 4 | 2 | 5 |
| $t_6$ | 3 | 8 | 4 |

| $\sqsubset_{ab}$ | $t_1$ | $t_2$ | $t_3$ | $t_4$ | $t_5$ | $t_6$ |
|---|---|---|---|---|---|---|
| $t_1$ | | × | × | | | × |
| $t_2$ | | | | | | × |
| $t_3$ | | | | | | |
| $t_4$ | | | | | | |
| $t_5$ | | | | × | | |
| $t_6$ | | | | | | |

| $\sqsubset_c$ | $t_1$ | $t_2$ | $t_3$ | $t_4$ | $t_5$ | $t_6$ |
|---|---|---|---|---|---|---|
| $t_1$ | | × | × | × | × | × |
| $t_2$ | | | × | × | × | × |
| $t_3$ | | | | × | × | × |
| $t_4$ | | | | | | |
| $t_5$ | | | | × | | |
| $t_6$ | | | | × | × | |

| $\sqsubset_{abc}$ | $t_1$ | $t_2$ | $t_3$ | $t_4$ | $t_5$ | $t_6$ |
|---|---|---|---|---|---|---|
| $t_1$ | | × | × | | | × |
| $t_2$ | | | | | | × |
| $t_3$ | | | | | | |
| $t_4$ | | | | | | |
| $t_5$ | | | | × | | |
| $t_6$ | | | | | | |

# Restricted Order Dependencies [Caruccio et al., 2016]



When considering **cyclical phenomena**, attributes may contain *local structures* relating them with others.
This is the case with **Restricted Order Dependencies.**

## Definition

Restricted Order Dependency $X \rightarrow Y$ holds with
$t_i \sqsubseteq^*_X t_j \iff \forall x \in X : 0 \leq t_j[x] - t_i[x] \leq \theta_x$, iff:

$$\forall t_i, t_j \in T : t_i[X] \sqsubseteq^*_X t_j[X] \Rightarrow t_i[Y] \sqsubseteq^*_Y t_j[Y]$$

## Example

| | *Time* | *People_waiting* |
|---|---|---|
| $t_1$ | 10:00 | 101 |
| $t_2$ | 10:20 | 103 |
| $t_3$ | 10:40 | 105 |
| $t_4$ | 11:00 | 77 |
| $t_5$ | 11:20 | 80 |
| $t_6$ | 11:40 | 85 |

Order-like Dependency *People_waiting* $\rightarrow$ *Time* holds, with $\theta_{People\_waiting} = 10$ and $\theta_{Time} = 02:00$.

# Restricted Order Dependencies [Caruccio et al., 2016]

## FCA Characterization

Similarly, for each attribute $x \in \mathcal{U}$ we define $\mathbb{K}_x^* = (T, T, \sqsubset_x^*)$

$$\sqsubset_x^* = \{(t_i, t_j) \mid t_i, t_j \in T,\ t_i \sqsubseteq_x^* t_j\}$$

For a set of attributes $X \subseteq \mathcal{U}$ we have: $\mathbb{K}_X^* = (T, T, \bigcap_{x \in X} \sqsubset_x^*)$

**Proposition 2**

The Restricted Order Dependency $X \to Y$ holds iff $\mathbb{K}_X^* = \mathbb{K}_{XY}^*$

**Example:** *People_waiting* $\to$ *Time* with $\theta_{People\_waiting} = 10$ and $\theta_{Time} = 02:00$.

| | *Time* | *People_waiting* |
|---|---|---|
| $t_1$ | 10:00 | 101 |
| $t_2$ | 10:20 | 103 |
| $t_3$ | 10:40 | 105 |
| $t_4$ | 11:00 | 77 |
| $t_5$ | 11:20 | 80 |
| $t_6$ | 11:40 | 85 |

| $\sqsubseteq_{Tm}^*$ | $t_1$ | $t_2$ | $t_3$ | $t_4$ | $t_5$ | $t_6$ |
|---|---|---|---|---|---|---|
| $t_1$ | × | × | × | × | × | × |
| $t_2$ | | × | × | × | × | × |
| $t_3$ | | | × | × | × | × |
| $t_4$ | | | | × | × | × |
| $t_5$ | | | | | × | × |
| $t_6$ | | | | | | × |

| $\sqsubseteq_{Pp}^*$ | $t_1$ | $t_2$ | $t_3$ | $t_4$ | $t_5$ | $t_6$ |
|---|---|---|---|---|---|---|
| $t_1$ | × | × | × | | | |
| $t_2$ | | × | × | | | |
| $t_3$ | | | × | | | |
| $t_4$ | | | | × | × | × |
| $t_5$ | | | | | × | × |
| $t_6$ | | | | | | × |

| $\sqsubseteq_{Tm,Pp}^*$ | $t_1$ | $t_2$ | $t_3$ | $t_4$ | $t_5$ | $t_6$ |
|---|---|---|---|---|---|---|
| $t_1$ | × | × | × | | | |
| $t_2$ | | × | × | | | |
| $t_3$ | | | × | | | |
| $t_4$ | | | | × | × | × |
| $t_5$ | | | | | × | × |
| $t_6$ | | | | | | × |

# Lexicographical Ordered Dependencies [Ng, 1999]

## Functional Dependency Generalizations

When there is no only a structure in tuple values w.r.t $X \subseteq \mathcal{U}$, but there is also an **attribute structure** in $X$ to be found.
This is the case with **Lexicographical Ordered Dependencies (LODs).**

### Example

| | *Hour* | *Minute* | People_waiting |
|---|---|---|---|
| $t_1$ | 1 | 30 | 10 |
| $t_2$ | 1 | 45 | 13 |
| $t_3$ | 2 | 00 | 20 |
| $t_4$ | 2 | 15 | 35 |

- In this case we have a dependency which indicates that as *Time* **passes**, there is **more** *People waiting*.
- The notion of *Time* is a structure built over attributes *Hour* and *Minute*:
  - $t_i[Hour] < t_j[Hour] \Rightarrow Time(t_i) < Time(t_j)$
  - $t_i[Hour] = t_j[Hour], t_i[Minute] < t_j[Minute] \Rightarrow Time(t_i) < Time(t_j)$
- **The order of Hour and Minute in the Time structure are fixed!**

# Lexicographical Ordered Dependencies [Ng, 1999]

## Functional Dependency Generalizations

When there is no only a structure in tuple values w.r.t $X \subseteq \mathcal{U}$, but there is also an **attribute structure** in $X$ to be found.
This is the case with **Lexicographical Ordered Dependencies (LODs).**

### Example

| | *Hour* | *Minute* | People_waiting |
|---|---|---|---|
| $t_1$ | 1 | 30 | 10 |
| $t_2$ | 1 | 45 | 13 |
| $t_3$ | 2 | 00 | 20 |
| $t_4$ | 2 | 15 | 35 |

- The notion of *Time* is a structure built over attributes *Hour* and *Minute*:
  - $t_i[Hour] < t_j[Hour] \Rightarrow Time(t_i) < Time(t_j)$
  - $t_i[Hour] = t_j[Hour], t_i[Minute] < t_j[Minute] \Rightarrow Time(t_i) < Time(t_j)$
- We say that there is a **lexicographical ordering** of tuples w.r.t. the ordered set $\langle\, Hour, Minute\,\rangle$ which we will denote:

$$t_i <^{l}_{\langle\, Hour, Minute\,\rangle} t_j,\; i,j \in [1,4],\; i < j$$

## Lexicographical Order

Let $X \subseteq \mathcal{U}$ be an ordered set s.t. $X = \langle x_1, x_2, \ldots, x_n \rangle$. We say that $t_i$ is lexicographically lower or equal than $t_j$ in $X$ ($t_i \leq^l_X t_j$) iff:

- $\exists k \in [1, n]$ s.t. $t_i[x_k] < t_j[x_k]$ and $t_i[x_m] = t_j[x_m]$, $m \in [1, k[$, or
- $t_i[x_k] = t_j[x_k]$, $k \in [1, n]$

## Lexicographical Ordered Dependency

Given the ordered sets $X, Y \subseteq \mathcal{U}$, the LOD $X \leadsto Y$ holds iff:

$$\forall t_i, t_j \in T : t_i \leq^l_X t_j \Rightarrow t_i \leq^l_Y t_j$$

An Order-like Dependency corresponds to **(several)** Lexicographical Ordered Dependencies!

# Lexicographical Ordered Dependencies [Ng, 1999]

## Functional Dependency Generalizations

### Example

| | *Hour* | *Minute* | People_waiting |
|---|---|---|---|
| $t_1$ | 1 | 30 | 10 |
| $t_2$ | 1 | 45 | 13 |
| $t_3$ | 2 | 00 | 20 |
| $t_4$ | 2 | 15 | 35 |

In the previous example, the LOD $Hour, Minute \leadsto People\_waiting$ holds.

# Lexicographical Ordered Dependencies [Ng, 1999]

## FCA Characterization

**There is none!**
Since LODs are established between ordered attribute sets, intersections of general ordinal scales cannot be used.

e.g. $\mathbb{K}_{\langle\, Hour, Minute\,\rangle}$ and $\mathbb{K}_{\langle Minute, Hour\rangle}$ should be different, but they are not.

Instead, LODs can be found algorithmically by defining the following scales for an attribute $x \in \mathcal{U}$:

- $\mathbb{K}_x^{=} = (T, T, \{(t_i, t_j),\ t_i[x] = t_j[x]\})$
- $\mathbb{K}_x^{<} = (T, T, \{(t_i, t_j),\ t_i[x] < t_j[x]\})$
- $\mathbb{K}_x^{\leq} = (T, T, \{(t_i, t_j),\ t_i[x] \leq t_j[x]\})$

**Algorithm available in the article.**

# Summary

## Functional Dependency Generalizations

| | |
|---|---|
| **Functional Dependency**<br>$t_i[X] = t_j[X] \Rightarrow t_i[Y] = t_j[Y]$ | Equivalence Relation |
| **Similarity Dependency**<br>$t_i[X] \simeq_\theta t_j[X] \Rightarrow t_i[Y] \simeq_\theta t_j[Y]$ | Tolerance Relation<br>Non-Transitive |
| **Order-like Dependency**<br>$t_i[X] \sqsubseteq_X t_j[X] \Rightarrow t_i[Y] \sqsubseteq_Y t_j[Y]$ | Transitive, Anti-symmetric Relation<br>Not Necessarily Reflexive |
| **Restricted Order Dependency**<br>$t_i[X] \sqsubseteq^*_X t_j[X] \Rightarrow t_i[Y] \sqsubseteq^*_Y t_j[Y]$ | Non-Transitive, Anti-symmetric Relation<br>Not Necessarily Reflexive |
| **Lexicographical Ordered Dependency**<br>$t_i[X] \leq^l_X t_j[X] \Rightarrow t_i[Y] \leq^l_Y t_j[Y]$ | Reflexive,Transitive, Anti-symmetric Relation |

# Conclusions & Perspectives

- New generalizations of FDs require covering two different aspects: **axiomatization** and **computation**.
- FCA allows characterizing FDs providing a framework for both aspects.
- Some Attribute Relaxing FDs (Similarity, Order-like and Restricted Order) adapt easily tho this framework.
- Other Attribute Relaxing FDs (Lexicographical) require a heuristic approach.
- In both cases, the FCA framework is flexible and robust enough to provide a mining environment for Attribute Relaxing FDs.

# References I


Baixeries, J., Kaytoue, M., and Napoli, A. (2012).
Computing functional dependencies with pattern structures.
In Szathmary, L. and Priss, U., editors, *CLA*, volume 972 of *CEUR Workshop Proceedings*, pages 175–186. CEUR-WS.org.

Baixeries, J., Kaytoue, M., and Napoli, A. (2013).
Computing similarity dependencies with pattern structures.
In Ojeda-Aciego, M. and Outrata, J., editors, *CLA*, volume 1062 of *CEUR Workshop Proceedings*, pages 33–44. CEUR-WS.org.

Baixeries, J., Kaytoue, M., and Napoli, A. (2014).
Characterizing functional dependencies in formal concept analysis with pattern structures.
*Annals of Mathematics and Artificial Intelligence*, 72(1-2):129–149.

# References II


- Caruccio, L., Deufemia, V., and Polese, G. (2016).
  Relaxed functional dependencies - A survey of approaches.
  *IEEE Trans. Knowl. Data Eng.*, 28(1):147–165.

- Ginsburg, S. and Hull, R. (1983).
  Order dependency in the relational model.
  *Theoretical Computer Science*, 26(1):149 – 195.

- Ng, W. (1999).
  Ordered functional dependencies in relational databases.
  *Information Systems*, 24(7):535 – 554.