

Computing Formal Concepts on Hadoop Demo

Johannes Gosda and Stefan Wehrmeyer

Advanced M/R Seminar, HPI

15.12.2009

Challenge

Finding groups of similar articles.

A T T R I B U T E S

ARTICLES

	Drehkreuz	Studenten	Abschluss	Gründung	Lehrstühle
Lufthansa	X			X	
IFI		X	X		X
GermanWings	X			X	
HPI		X	X	X	X

Operators \downarrow and \uparrow

$B\downarrow$ = Set of all **objects** sharing all attributes from B.

$A\uparrow$ = Set of all **attributes** shared by all objects from A.

Examples:

$\{\}\downarrow = \{L, I, G, H\}$

$\{1\}\downarrow = \{I, H\}$

$\{I, H\}\uparrow = \{1, 2, 4\}$

$(\{I, H\}, \{1, 2, 4\})$

	0	1	2	3	4
L	X			X	
I		X	X		X
G	X			X	
H		X	X	X	X

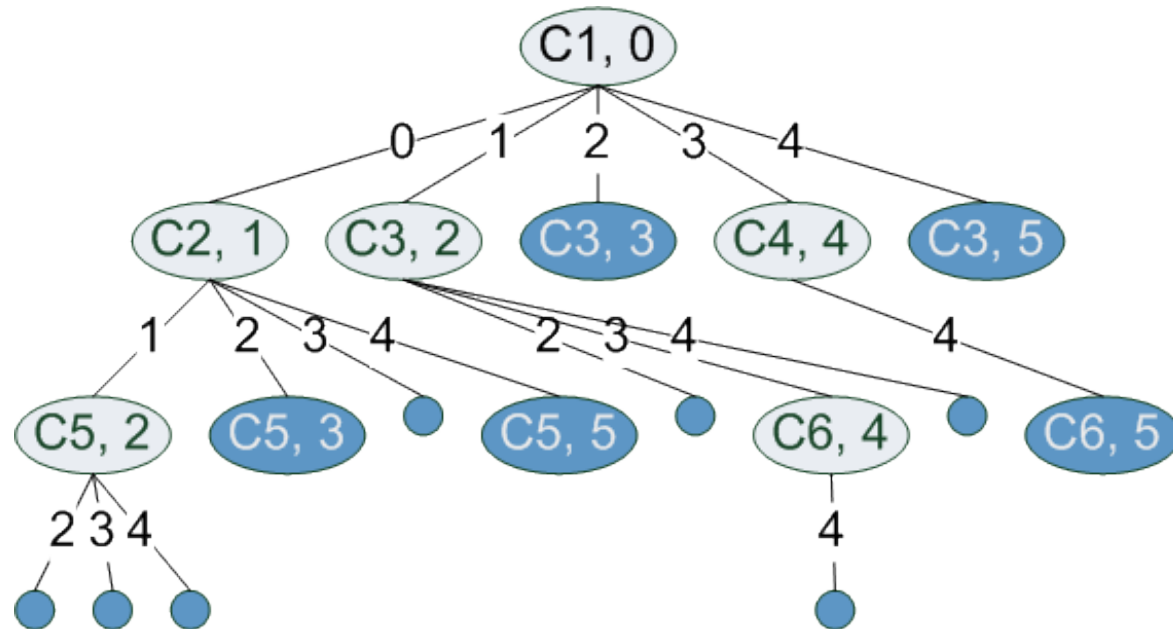
	0	1	2	3	4
0	X			X	
1		X	X		X
2	X			X	
3		X	X	X	X

Adjacency List:

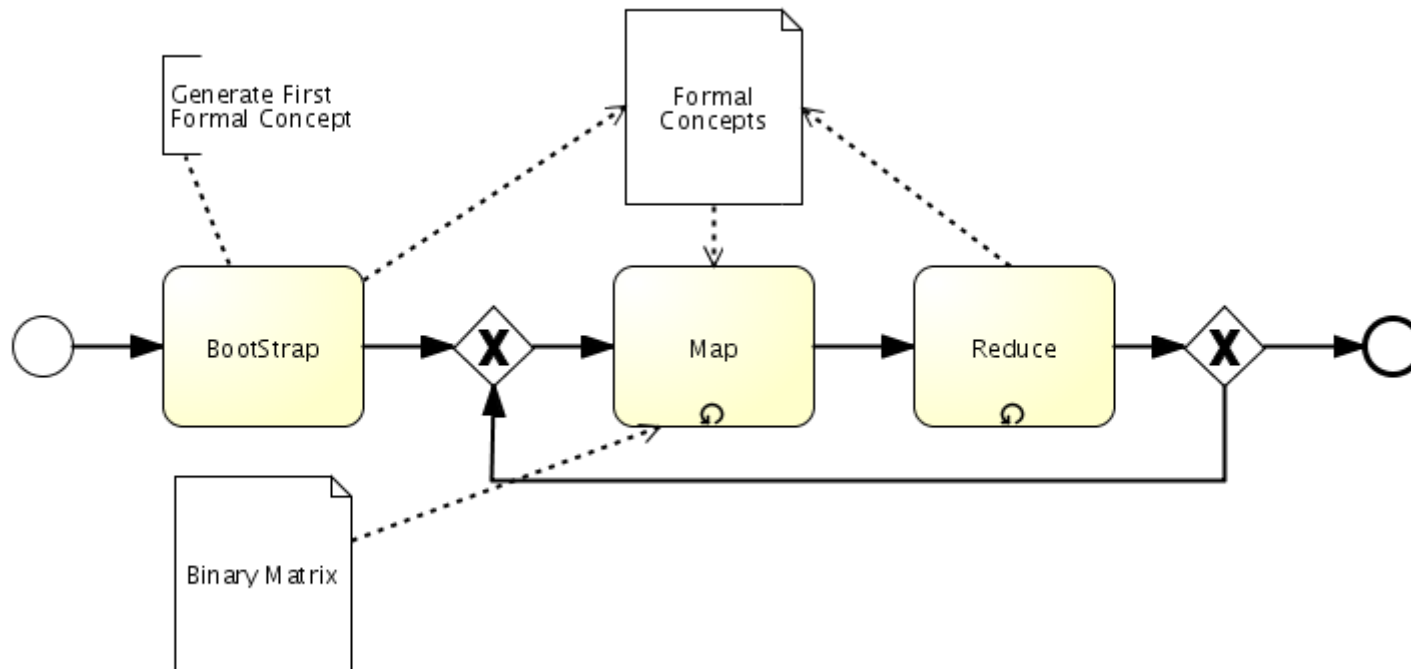
0	3			
1	2	4		
0	3			
1	2	3	4	

All Formal Concepts:

- C1 = ({0,1, 2, 3}, { })
- C2 = ({0,2}, {0,3})
- C3 = ({1,3}, {1,2,4})
- C4 = ({0,1,3}, {3})
- C5 = ({ }, {0,1,2,3,4})
- C6 = ({3}, {1,2,3,4})



Program Flow



Demo