Exploring Attributes with Domain Knowledge in Formal Concept Analysis

Jonnalagadda Annapurna[1] and Aswani Kumar Cherukuri[2]

[1] School of Computing Science and Engineering, VIT University, Vellore, India
[2] School of Information Technology, VIT University, Vellore, India

Recent literature reports the growing interests in data analysis using Formal Concept Analysis (FCA), in which data is represented in the form of object and attribute relations. FCA analyzes and then subsequently visualizes the data based on duality called Galois connection. Attribute exploration is a knowledge acquisition process in FCA, which interactively determines the implications holding between the attributes. The objective of this paper is to demonstrate the attribute exploration to understand the dependencies among the attributes in the data. While performing this process, we add domain experts' knowledge as background knowledge. We demonstrate the method through experiments on two real world healthcare datasets. The results show that the knowledge acquired through exploration process coupled with domain expert knowledge has better classification accuracy.

Keywords: association rules, attribute exploration, background knowledge, concept lattice, formal concept analysis

1. Introduction

Formal Concept Analysis (FCA) is an applied mathematical method of data analysis. Emerging from the order and lattice theory, FCA analyzes the data which describes the relationship between a set of objects and a set of attributes of a particular domain (Davey and Priestley, 2002). The objects and attributes are structured into formal abstractions called formal concepts, which together form a hierarchically ordered conceptual structure called concept lattice and collection of attribute implications. The process of concept formation in FCA is generally considered as a knowledge discovery from data and constructing the concept set constitutes the mining phase of data (Poelmans et al., 2010; Valtchev et al., 2004). With its ability to unfold different views of data for interpretations and finding patterns in the data, FCA is well suited in different areas where data is to be analyzed at several levels of detail and from different view points.

However, there are a few issues in using FCA for data analysis which includes representing the domain knowledge, handling incomplete and redundant attribute or object details, knowledge reduction while maintaining structure consistency etc (Sergio Mariano and Newton Jose, 2010; Wei and Jian-Jun, 2010; Wu et al, 2009). Literature reports mathematical or heuristic techniques to handle these issues (Aswani Kumar and Srinivas, 2010; Aswani Kumar and Srinivas, 2010a; Aswani Kumar, 2011; Aswani Kumar, 2011b; Mi et al, 2010; Prem Kumar Singh and Aswani Kumar, 2012; Snasel et al., 2007; Snasel et al., 2008).

However, it is desirable to understand dependencies among the attributes before applying any such technique. One way of addressing this situation is to perform attribute exploration process. Generally, this exploration process is regarded as a tool for knowledge acquisition and discovery. Attribute exploration process in FCA determines a minimal set of implicational dependencies between attributes that hold for all objects of the domain of study. Initially, the exploration starts by selecting the objects and attributes that describe these objects. From these attributes, the exploration process computes hypothetical implications. These impli-

cations are validated by a human, who is an expert of the domain of the objects. The output of the exploration process is the set of implications which are true for the chosen set of attributes and a representative set of examples of the domain of study (Cynthia, 2012; Jascke and Rudolph, 2013). The objective of this paper is to demonstrate the attribute exploration as a process of understanding the dependencies between the attributes. Also, the knowledge acquired in this process is integrated with the domain experts' knowledge. The remainder of this paper is organized as follows. Section 2 provides a detailed background and related work. Section 3 provides the problem description and motivation for attribute exploration. Section 4 demonstrates the attribute exploration on two healthcare datasets and analyzes the results.

2. Background

This section focuses on the notions and terminology of FCA. Introduced by Wille in 1982, FCA is an order-theoretic mathematical framework which represents lattice as a conceptual hierarchy of the data and each element of lattice as a formal concept (Ganter, 1999; Ganter and Wille, 1999; Stumme, 2009; Wille, 2008).

Definition 1. A formal context $C = (G, M, I)$ consists of two sets G, M and binary relation I between G and M. The elements of G, M are called the objects and attributes of the context, respectively.

Definition 2. A formal concept of the context (G, M, I) can be defined as an ordered pair (A, B) with $A \subseteq G, B \subseteq M, A' = B, B' = A$. We call A as extent and B as intent of the concept (A, B).

The set of formal concepts is ordered by partial order '$\leq$' such that for any two formal concepts (A_1, B_1) and (A_2, B_2), $(A_1, B_1) \leq (A_2, B_2)$ if and only if $A_1 \subseteq A_2$ and $B_2 \subseteq B_1$. The set of concepts ordered by this partial order '$\leq$' constitutes a complete lattice termed concept lattice (Davey and Priestley, 2002). With this notion of partial order, lattices provide a clear structure for knowledge representation. Each node of the lattice structure represents a concept. Each concept is linked by a descending path to all the concepts that are labeled by objects belonging to the extent of the concept, and by an ascending path to all concepts that are labeled by attributes

belonging to the intent of the concept. The most general concepts are at the top of the hierarchy.

An object g is attached to a node representing the smallest concept with g in its extent and an attribute m is associated with the node representing the largest concept with m in its intent. Hence if a node has an object g, then all the nodes above it also contain the object g. The smallest concept for an object g is called the object concept of g. Similarly, every attribute will have attribute concepts in the lattice. In the lattice structure, instead of labeling the elements with all their objects and attributes, we label the object and attribute concepts with their generating objects and attributes.

Definition 3. An attribute implication is an expression $P \rightarrow Q$, where $P, Q \subseteq M$, is true in C if each object which has all attributes from P has also all attributes from Q.

The basic assumption is that $P \cap Q = \emptyset$ i.e. the attributes in the premise P are discarded from the conclusion Q.

Definition 4. A set T of attribute implications is called sound and complete with respect to a formal context $C = (G, M, I)$, if T is true in C and each implication true in C follows from T.

Definition 5. A set T of non-redundant attribute implications which is sound and complete with respect to a formal context C is a base with respect to context C.

Attribute implications are closely related to functional dependencies in the database field and hence have made their way into Association Rules Mining (ARM) problem in data mining (Aswani Kumar, 2012; Dias et al., 2013; Li et al., 2013; Pasquier et al., 1999). The basis for the rules with 100% confidence is called as *Duquenne-Guigues (DG) basis* and the basis for the rules with confidence less than 100% is called as *Luxenburger basis*. From the DG basis all the implications of the context can be derived in a canonical way so that all the implications of the context semantically follow from the basis (Bazhanov and Obiedkov, 2013). As lattices are algebraic structures, it is natural to consider canonical, those maps between lattices which preserve the operations join and meet. For more detailed introductory information on FCA, interested readers can refer to the literature including (Kuznetsov and Poelmans, 2013; Poelmans et al, 2013a; Stumme, 2009; Valtchev et al, 2004; Wille, 2008);

FCA has attracted several applications in a wide variety of disciplines (Aswani Kumar et al., 2012; Aswani Kumar, 2013). Priss (2006) has provided an exhaustive overview of FCA applications. Very recently Poelmans et al. (2013b) have provided a survey of applications of FCA based methods in different domains, including software mining, web analytics, medicine, biology and chemistry etc.

3. Problem Description and Exploring Attributes

Formal contexts, concept lattices and implication bases represent the same structural information of a dataset. From the formal context representation, concept lattice and an implication base can be derived. Alternatively, with the help of implication base, all possible intents can be defined from which a suitable formal context and a lattice structure can be obtained (Wille, 2008). However, there are cases where the set of objects of a domain is either incomplete or too large to be listed completely. Also, the attribute implications from a context generally hold for the objects from that context and do not hold for all objects of a domain. To address these issues, FCA provides a method called attribute exploration for incremental construction of formal contexts. Given a set of objects belonging to a subject domain and their descriptions in the form of presence or absence of certain attributes, attribute exploration aims to build a set of implications that hold for all the objects in the entire domain and a representative set of its objects. Similar to the attribute exploration, FCA also supports object, rule and concept explorations (Stumme, 1995).

The main purpose of attribute exploration over a context C is to generate the DG basis of implications and an associated context. The process of attribute exploration is interactive, which suggests the attribute implications to the domain expert. The role of the expert in this process is to validate the implication. At each step in the process, an implication base for the context representing the domain data at that step is generated and shown to the expert. Given an implication $P \rightarrow Q$, the expert can either accept the implication or refute it with a counterexample. The counterexample is an object O, where $O \in G$, that has all attributes from P, but there exists at least one attribute from Q that the object O does not have. With these counterexamples, object set of the context can be obtained. These objects or counterexamples are sufficient to determine the structure of the concept lattice. Since the decision on validity of an implication cannot be reversed, the counterexamples cannot contradict with the already confirmed implications. This interactive and iterative method completes both the logical specification and the basis for the constructed context. The concept lattice of the domain is isomorphic to the concept lattice generated from the relatively small set of objects. The algorithm proposed by Ganter and Wille (1999) computes non-redundant and complete set of implications. Practical implementation of this algorithm can be found in some of the FCA tools such as ConExp, Conexp-clj. However, this implementation does not consider any available background knowledge.

This exploration process is regarded as the first and well known FCA based procedure for knowledge discovery (Poelmans et al., 2013a). Obiedkov et al. (2009) have shown that attribute exploration can be used to create lattice-based access control models by considering one by one dependencies between security labels. Revenko and Kuznetsov (2010) have proposed an approach based on attribute exploration for studying the relations between properties of functions on ordered sets. Very recently Jaschke and Rudolph (2013) have proposed an approach for supporting attribute exploration process by Web information retrieval. Their approach has potential to speed up the attribute exploration process. In another interesting work, Aswani Kumar (2013) has performed attribute exploration for designing role based access control.

4. Experimental Results and Discussion

We demonstrate here the attribute exploration on two healthcare datasets which are part of consumer healthcare informatics project of Medical Research Council of South Africa (Horner, 2007). The diseases which were studied in the project are Tuberculosis (TB), Chronic Bronchitis (CBr) and Hypertension (HP). However, in our analysis we consider only TB and CBr diseases. We have conducted the experiments in two ways. First, from the formal context

corresponding concept lattice and its implication base are derived. Alternatively, with the help of attribute exploration, all possible intents are identified and by using them, suitable formal context and its concept lattice structure are obtained. For the purpose of exploration, we have considered treating doctors' diagnosis and experts' rules as domain knowledge. We have tested quality of the rules by comparing them with the rules originally given by the domain experts and with a test dataset. Domain experts team includes a chest specialist, a gastroenterologist who are both the faculty members at Medunsa medical campus of the University of Limpopo in South Africa and a nursing staff member with 30 years of experience in primary healthcare.

4.1. Chronic Bronchitis

Chronic Bronchitis (CBr) dataset contains the data about 7 patients for various symptoms of CBr and experts' rules for determining the disease. Table 1 shows the list of various CBr symptoms. Table 2 lists experts' opinions in the form of rules applied in determining the disease using the symptoms listed in Table 1. Table 3 shows the formal context, also known as object-attribute binary incidence matrix of CBr data with details of 7 patients. The last column of the matrix indicates treating doctor's conclusions on presence or absence of CBr. From the context it is clear that there is only one positive example in the dataset, namely Obj 4.

No	Symptom	Abbreviation
1	Persistent Cough	PC
2	Sputum Production	SP
3	Sputum produced is Muco-Purulent	MC
4	Chest Tightness	CT
5	Shortness of Breath	SB
6	Wheezing Chest	WC
7	Smoking	SM

Table 1. Chronic Bronchitis (CBr) symptoms.

Figure 1 shows the concept lattice obtained by applying FCA on the CBr incidence matrix given in Table 3. Concept lattice shown in Figure 1 is of height 6 and contains 10 concepts with

12 edges. As discussed in Section 2, nodes in the concept lattice structure indicated the object and attribute concepts only. From the formal context shown in Table 3, along with the concept lattice, FCA has produced 9 implications in the DG basis i.e. the implications with 100% confidence. Implications which make positive conclusions about the CBr are of interest in this study. Table 4 lists all such implications derived from FCA. If the antecedent of an implication which has the target attribute in its consequent is a subset of the antecedents of an expert's rule, then we can consider that the expert's rule is subsumed by the implication. From Table 4, we can understand that antecedent of the implication 2 is subset of the antecedents of the expert's rules. Implication 1 is not part of the experts' rules. However none of the implications in the DG basis, shown in Table 4, is overlapped exactly with any of the experts' rules. Hence these implications are considered as new knowledge about the domain.

Sl. No	Expert Rules for Tuberculosis
1	PC SM SP CT SB → CBr
2	PC SM SP WC → CBr
3	PC SM SP MC → CBr

Table 2. Expert's rules for CBr.
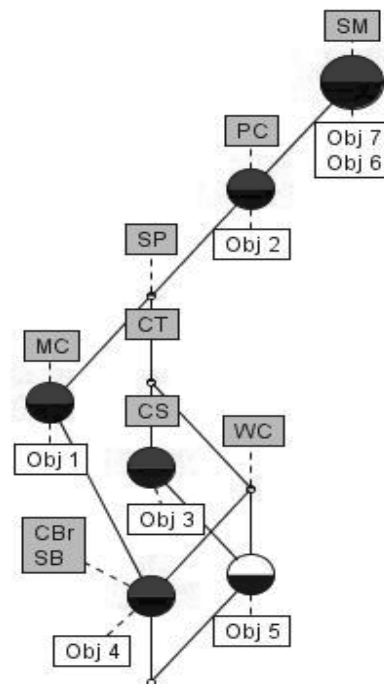
Figure 1. Concept lattice of Bronchitis training context.

	PC	SP	MC	CS	CT	SB	WC	SM	CBr
Obj 1	X	X	X					X	
Obj 2	X							X	
Obj 3	X	X		X	X			X	
Obj 4	X	X	X		X	X	X	X	X
Obj 5	X	X		X	X		X	X	
Obj 6								X	
Obj 7								X	

Table 3. Incidence matrix from original Bronchitis dataset.

Sl. No	Implications	No. of objects implication holds
Implications **NOT** part of expert rules		
1	PC SP MC CT SM → CBr	1
Implications part of expert rules		
2	SB SM → CBr	1

Table 4. Implications obtained from CBr context.

Next we perform the exploration of CBr attributes by considering all the symptoms of CBr disease. We have also considered the experts' rules and treating doctors' diagnosis given in Tables 2 and 3 as domain knowledge of the disease. We start with the object empty context $(\emptyset, \{PC, SP, MC, CS, CT, SB, WC, SM, Cbr\})$. Exploration starts with the question whether all

patients (objects) have all the symptoms (attributes). Domain expert provides a counterexample of patient 4 having symptoms {PC, SP, MC, CT, SB, WC, SM, CBr} and adds this example as object 1 to the new context. Then the second question will explore whether all patients have at least all the symptoms of object 1. Since this is not the case, the expert provides a counterexample of patient 5 having the symptoms {PC, SP, CS, CT, WC, SM} and adds the example as object 2 to the new context. Proceeding as above, we accept 9 more implications and provide 4 counterexamples before the attribute exploration stops. Table 5 lists all the explored questions and corresponding counterexamples by the expert.

From the counterexamples and the experts' rules listed in Table 2 as domain knowledge, we get the resultant formal context as shown in Table 6. From Table 6 we can observe that object 1 to object 6 are obtained from the counterexamples provided by the domain expert. Objects 7, 8 and 9 are the domain expert rules of CBr disease as mentioned in Table 2. Figure 2 shows the concept lattice obtained from the resultant context. The new lattice structure is of height 6 and contains 15 concepts with 23 edges. Along with the lattice the resultant context has produced 13 implications in the DG basis. Implications that conclude CBr are listed in Table 7. We can observe that the support count of the implication

#	Question	Answer	Counter-examples
1	Whether all objects (patients) have all attributes (symptoms)	No	Obj 4
2	Whether all objects have attributes PC, SP, MC, CT, SB, WC, SM, CBr	No	Obj 5
3	Whether all objects have attributes PC, SP, CT, WC, SM	No	Obj 3
4	Whether all objects have attributes PC, SP, CT, SM.	No	Obj 1
5	Whether all objects have attributes PC, SP, SM.	No	Obj 2
6	Whether all objects have attributes PC, SM.	No	Obj 6
7	Whether all objects have attributes SM.	Yes	–
8	Is it SM, CBr → PC, SP, MC, CT, SB, WC?	Yes	–
9	Is it WC, SM → PC, SP, CT?	Yes	–
10	Is it SB, SM → PC, SP, MC, CT, WC, CBr?	Yes	–
11	Is it CT, SM → PC, SP?	Yes	–
12	Is it CS, SM → PC, SP, CT?	Yes	–
13	Is it MC, SM → PC, SP?	Yes	–
14	Is it SP, SM → PC?	Yes	–
15	Is it PC, SP, MC, CT, SM → SB, WC, CBr?	Yes	–

Table 5. Exploring the attributes of CBr disease.

{SB, SM → CBr} is increased due to the fact that the experts' rules are being considered as domain knowledge. From Table 7 it is clear that the context obtained after exploration has derived more knowledge in the form of implications about the disease than the original context. Another interesting point is that attribute exploration has clarified objects 6 and 7 that have equal sets of attributes in the original context as single object in resultant context.

	PC	SP	MC	CS	CT	SB	WC	SM	CBr
Obj 1	X	X	X		X	X	X	X	X
Obj 2	X	X		X	X		X	X	
Obj 3	X	X		X	X			X	
Obj 4	X	X	X					X	
Obj 5	X							X	
Obj 6								X	
Obj 7	X	X			X	X		X	X
Obj 8	X	X					X	X	X
Obj 9	X	X	X					X	X

Table 6. Resultant context after exploration and combining experts' knowledge.
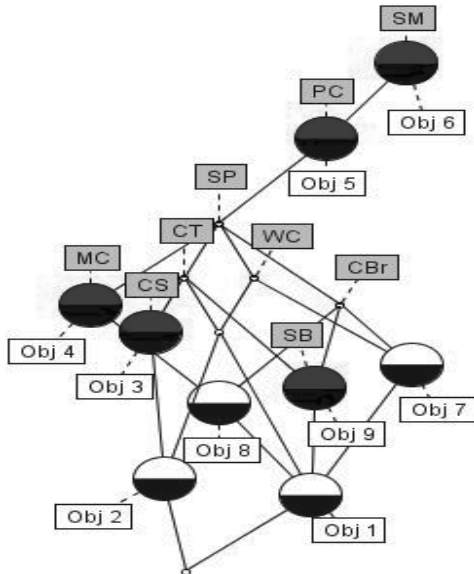
Figure 2. Concept lattice obtained after exploration.

Sl. No	Implications	No. of objects implication holds
Implications **NOT** part of expert rules		
1	PC SP MC CT SM → CBr	1
Implications part of expert rules		
2	SB SM → CBr	1

Table 7. Implications obtained from CBr context after attribute exploration.

No	Symptom	Abbreviation
1	Persistent Cough	PC
2	Sputum Production	SP
3	Sputum produced is Muco-Purulent	MC
4	Sputum Bloody	BS
5	Clear Sputum	CS
6	Weight Loss	WL
7	Extreme Night Sweats	NS
8	No Appetite	NA
9	Chest Pain	CP
10	Shortness of Breath	SB
11	Tuberculosis Contact	TC
12	Tiredness	TN

Table 8. TB symptoms.

Sl. No	Expert Rules for Tuberculosis
1	PC SP BS WL → TB
2	PC SP BS NS → TB
3	PC WL NS → TB
4	PC SP BS TC → TB
5	PC SP BS CP NA → TB
6	PC SP BS SB → TB
7	PC WL CP SB → TB
8	PC SP BS CP TN → TB

Table 9. Expert rules for TB.

4.2. Tuberculosis

We have conducted experiments on Tuberculosis (TB) dataset which contains the details of 21 patients for various symptoms of TB as a training data, symptoms of 10 patients as testing data and experts' rules for determining the disease. Table 8 shows the list of various TB symptoms. Table 9 lists experts' opinions, in determining the disease using the symptoms listed in Table 8. Table 10 shows the formal context, also

known as object-attribute binary incidence matrix of TB training data. The last column of the matrix indicates presence or absence of TB as diagnosed by the treating doctors. Figure 3 shows the concept lattice obtained by applying FCA on the TB incidence matrix given in Table 10. Each node of the lattice structure shown in Figure 3 represents a concept. Concept lattice shown in Figure 3 is of height 11 and contains 101 concepts with 253 edges. From the formal context shown in Table 10, along with the concept lattice, FCA has produced 33 implications in DG basis. Implications which make positive conclusions about the TB are of interest in this study. Table 11 lists all such implications derived from FCA.

From Table 11 we can understand that antecedents of the implications 4 to 9 are subsets of the antecedents of the experts' rules. Also, we can observe that these implications subsume all the experts' rules. Implications from 1 to 3 are not part of the experts' rules. However none of the implications in the DG basis shown in Table 11 are overlapped exactly with any of the experts' rules. Hence we treat these implications as new knowledge about the domain.

Sl. No	Implications	Support Count
Implications **NOT** part of expert rules		
1	NS CP→ TB	8
2	WL TN → TB	11
3	PC SP MC CS CP TN → TB	0
Implications part of expert rules		
4	NA CP → TB	9
5	BS → TB	1
6	PC SP NS → TB	6
7	WL NS → TB	10
8	WL CP → TB	10
9	TC → TB	3

Table 11. Implications obtained from TB training data using FCA.

	PC	SP	MC	BS	CS	WL	NS	NA	CP	SB	TC	TN	TB
Obj 1	X	X	X			X			X	X			X
Obj 2	X	X	X			X	X	X	X				X
Obj 3	X	X	X			X	X	X	X		X	X	X
Obj 4	X					X		X		X		X	X
Obj 5						X	X	X	X	X		X	X
Obj 6	X					X	X	X	X	X	X	X	X
Obj 7	X	X			X	X		X	X	X		X	X
Obj 8						X	X					X	X
Obj 9							X						X
Obj 10	X	X	X			X	X	X	X			X	X
Obj 11	X	X	X			X		X		X			
Obj 12	X	X	X			X		X					
Obj 13	X	X			X				X			X	
Obj 14	X	X		X		X	X	X	X			X	X
Obj 15									X				
Obj 16	X	X	X			X	X	X	X		X	X	X
Obj 17	X					X	X	X	X	X		X	X
Obj 18	X	X			X	X	X					X	X
Obj 19	X						X	X				X	
Obj 20	X	X	X						X			X	
Obj 21									X			X	

Table 10. Incidence matrix of TB training dataset.

To perform the attribute exploration on the TB dataset, we have considered all the symptoms of TB. Also, we have considered the rules elicited by the TB experts and treating doctors' diagnosis as domain knowledge of the disease. We start with the object empty context ($\emptyset$, {PC, SP, MC, BS, CS, WL, NS, NA, CP, SB, TC, TN, TB}). Attribute exploration starts with the question whether all the patients have all the symptoms. The obvious answer to this question is, "no". The human expert provides a counterexample by mentioning the first patient from the formal context who is having the symptoms {PC, SP, MC, WL, CP, SB, TB} and adds the example to the new context. Then the second question will explore whether all patients have at least all the same symptoms as patient 1. Since this is not the case, the expert provides a counter example of patient 2 having the symptoms {PC, SP, MC, WL, NS, NA, CP, TB} and adds patient 2 to the new context as an object. The common symptoms between patient 1 and patient 2 are PC, SP, MC, WL, CP, and TB. Hence the next question will explore whether all objects have these symptoms. The human expert provides patient 7 as a counterexample having the symptoms {PC, SP, CS, WL, NA, CP, SB, TN, TB} and adds the same to the new context.
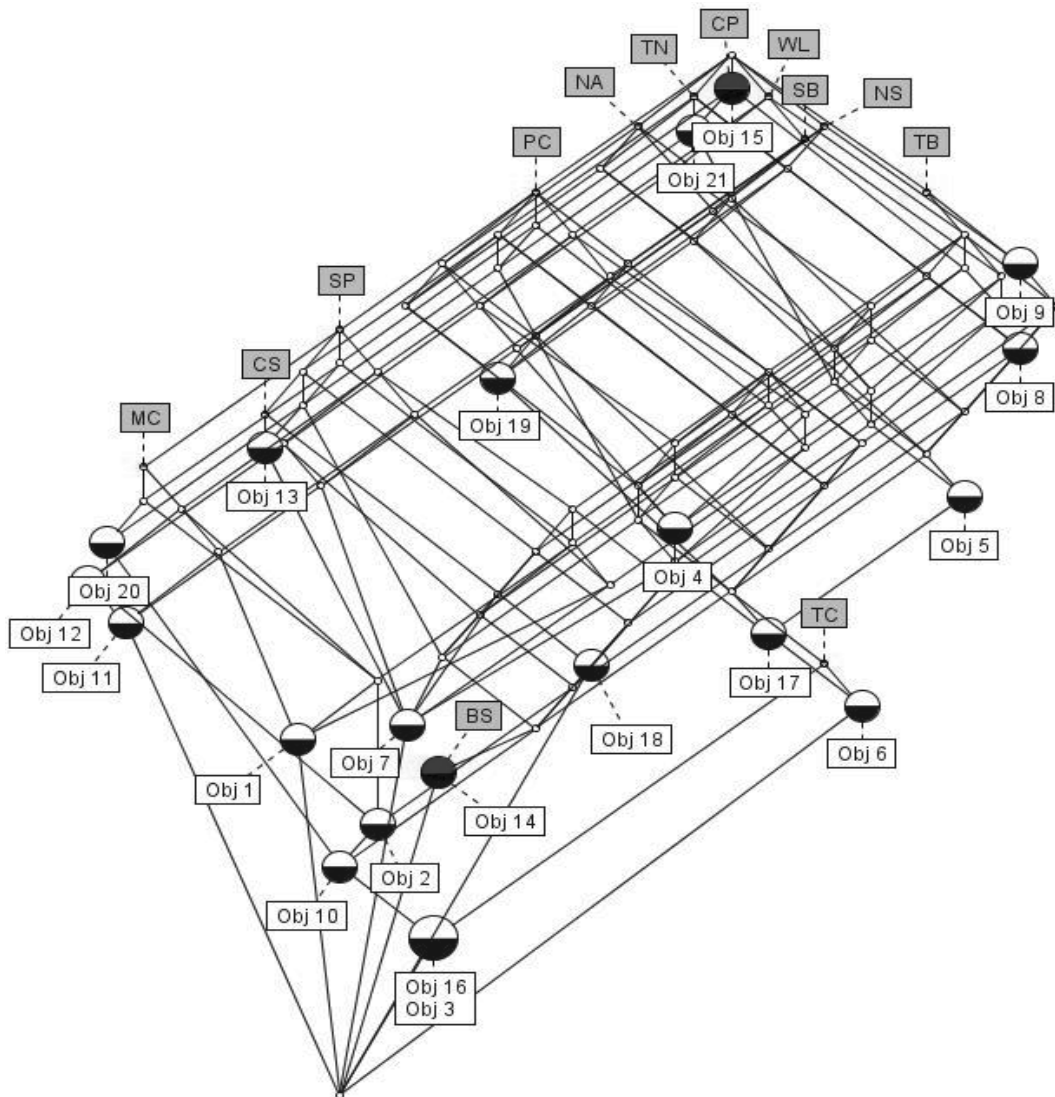

Figure 3. Concept lattice of the TB original context.

#	Question	Reply	Counter Examples
1	Whether all objects (patients) have all the attributes (symptoms)?	No	Obj 1
2	Whether all objects have attributes PC, SP, MC, WL, CP, SB, TB	No	Obj 2
3	Whether all objects have attributes PC, SP, MC, WL, CP, TB	No	Obj 7
4	Whether all objects have attributes PC, SP, WL, CP, TB	No	Obj 18
5	Whether all objects have attributes PC, SP, WL, TB.	No	Obj 4
6	Whether all objects have attributes PC, WL, TB.	No	Obj 8
7	Whether all objects have attributes WL, TB.	No	Obj 9
8	Whether all objects have attributes TB.	No	Obj 11
9	Is it TN → WL, TB?	No	Obj 20
10	Is it TN, TB → WL?	Yes	
11	Is it TC → PC, SP, MC, BS, CS, WL, NS, NA, CP, SB, TN, TB?	No	Obj 3
12	Is it TC → PC, SP, MC, WL, NS, NA, CP, TN, TB?	No	Obj 6
13	Is it TC → PC, WL, NS, NA, CP, TN, TB?	Yes	
14	Is it SB → PC, WL?	No	Obj 5
15	Is it SB → WL?	Yes	
16	Is it TN, TB → WL?	Yes	
17	Is it TC → PC WL NS NA CP TN TB?	Yes	
18	Is it SB → WL?	Yes	
19	Is it CP, TB → WL?	Yes	
20	Is it NA → WL?	No	Obj 19
21	Is it NA, TB → WL?	Yes	
22	Is it NA, CP → WL, TB?	Yes	
23	Is it NS, CP → WL. NA TB?	Yes	
24	Is it WL, TN → TB?	Yes	
25	Is it WL, SB, TN, TB → NA?	Yes	
26	Is it WL, CP → TB?	Yes	
27	Is it WL, CP, TN, TB → NA?	Yes	
28	Is it WL, NA, SB, TB → TN?	Yes	
29	Is it WL, NS → TB?	Yes	
30	Is it WL, NS, SB, TB → NA, CP, TN?	Yes	
31	Is it WL, NS, NA, TB → CP?	Yes	
32	Is it CS → PC, SP, WL, TN, TB?	No	Obj 13
33	Is it CS → PC, SP, TN?	Yes	
34	Is it BS → PC, SP, MC, CS, WL, NS, NA, CP, SB, TC, TN, TB?	No	Obj 14
35	Is it BS → PC, SP, WL, NS, NA, CP, TN, TB?	Yes	
36	Is it MC → PC, SP?	Yes	
37	Is it SP → PC?	Yes	
38	Is it PC, TB → WL?	Yes	
39	Is it PC, WL, NS, NA, CP, SB,TN, TB → TC?	Yes	
40	Is it PC, SP, NA → WL?	Yes	
41	Is it PC, SP, NS → WL, TB?	Yes	
42	Is it PC, SP, WL, SB, TB → CP?	Yes	
43	Is it PC, SP, WL, NA, TB → CP?	Yes	
44	Is it PC, SP, WL, NA, CP, SB, TN, TB → CS?	Yes	
45	Is it PC, SP, WL, NS, NA, CP, TC, TN, TB → MC?	Yes	
46	Is it PC, SP, CS, WL, NA, CP, TN, TB → SB?	Yes	
47	Is it PC, SP, MC, TN, → CP?	Yes	
48	Is it PC, SP, MC, WL, TB → CP?	Yes	
49	Is it PC, SP, MC, WL, NA, CP, TB, → NS?	Yes	
50	Is it PC, SP, MC, WL, NS, NA, CP, TN, TB, → TC?	No	Obj 10
51	Is it PC, SP, MC, CS, CP, TN → BS, WL, NS, NA, SB, TC, TB?	Yes	
52	Is it PC, SP, BS, WL, NS, NA, CP, TN, TB → CS, SB, TC?	Yes	

Table 12. Exploring the TB attributes.

	PC	SP	MC	BS	CS	WL	NS	NA	CP	SB	TC	TN	TB
Obj 1	X	X	X			X			X	X			X
Obj 2	X	X	X			X	X	X	X				X
Obj 3	X	X			X	X		X	X	X		X	X
Obj 4	X	X			X	X	X					X	X
Obj 5	X					X		X		X		X	X
Obj 6						X	X					X	X
Obj 7							X						X
Obj 8	X	X	X			X		X		X			
Obj 9	X	X	X					X				X	
Obj 10	X	X	X			X	X	X	X		X	X	X
Obj 11	X					X	X	X	X	X	X	X	X
Obj 12						X	X	X	X	X		X	X
Obj 13	X						X	X				X	
Obj 14	X	X			X				X			X	
Obj 15	X	X		X		X	X	X	X			X	X
Obj 16	X	X	X			X	X	X	X			X	X
Obj 17	X	X		X		X							X
Obj 18	X	X		X			X						X
Obj 19	X					X	X						X
Obj 20	X	X		X							X		X
Obj 21	X	X		X				X	X				X
Obj 22	X	X		X						X			X
Obj 23	X					X		X	X				X
Obj 24	X	X		X					X			X	X

Table 13. Formal context obtained after attribute exploration.

Proceeding as above, we accept 36 implications and provide 13 more counterexamples before attribute exploration stops. Table 12 lists all the explored questions and corresponding expert acceptance or counterexamples. From these counterexamples and the experts' rules listed in Table 9, a new resultant formal context is obtained as shown in Table 13. From the resultant context we can understand that the new formal context represents the original knowledge (shown in Table 10 with 21 objects) with 16 objects. Also, from Table 13 we can understand that the objects 1 to 16 are obtained through exploration process and objects 17 to 24 are obtained from the experts' knowledge listed in Table 9. Figure 4 shows the concept lattice obtained from the resultant context, having a concept count of 135 with 356 edges and height of 11. Along with the lattice, FCA has produced 44 implications in the DG basis. From this list, 13 implications that are inferring TB are shown in Table 14. We can observe that the implica-

Sl. No	Rule	Support
	Rules **NOT** part of expert rule	
1	NS CP→ TB	6
2	WL TN→ TB	9
3	PC SP MC CS CP TN → TB	0
4	NS SB→ TB	2
5	CP SB→ TB	5
6	PC SP NA TN → TB	4
7	SB TN→ TB	4
	Rules part of expert rules	
8	NA CP → TB	8
9	BS → TB	7
10	PC SP NS → TB	6
11	WL NS → TB	9
12	WL CP → TB	9
13	TC → TB	3

Table 14. Implications obtained from the resultant context.

tions 4 to 7 are newly obtained implications. A point to recall here is that FCA on the original context has produced only 9 implications concluding TB. Since these 13 implications are not exactly overlapped with any of the experts' rules, we treat all these implications as new knowledge acquired about the domain.

Table 15 summarizes the number of concepts, edges, implications in DG basis, new rules, height of the concept lattice and number of expert rules subsumed using FCA, resultant context after attribute implications. From this summary, we can understand that the context obtained after combining exploration process with domain experts' knowledge has produced more concepts than the original context.

The next step of our analysis is to verify the accuracy of the implications obtained in the analysis. Classification accuracy of the implications is measured by identifying the number of times implications from FCA and implications after attribute exploration have same conclusion as of the treating doctor on a test dataset. Table 16 shows the test dataset which contains symptoms of 10 patients and the treating doctor's conclusion on the presence of TB for each of the patient. Experts' rules, implications produced from FCA and FCA of the resultant context are compared on the test dataset and their classification accuracy details are summarized in Table 17. From this analysis, we can understand

	Concepts	Edges	# implications in DG basis	Height of lattice	# new rules	# rules subsumed
FCA	101	253	33	11	9	8
Context after exploration	135	356	44	11	13	8

Table 15. Summary of the results on TB data.
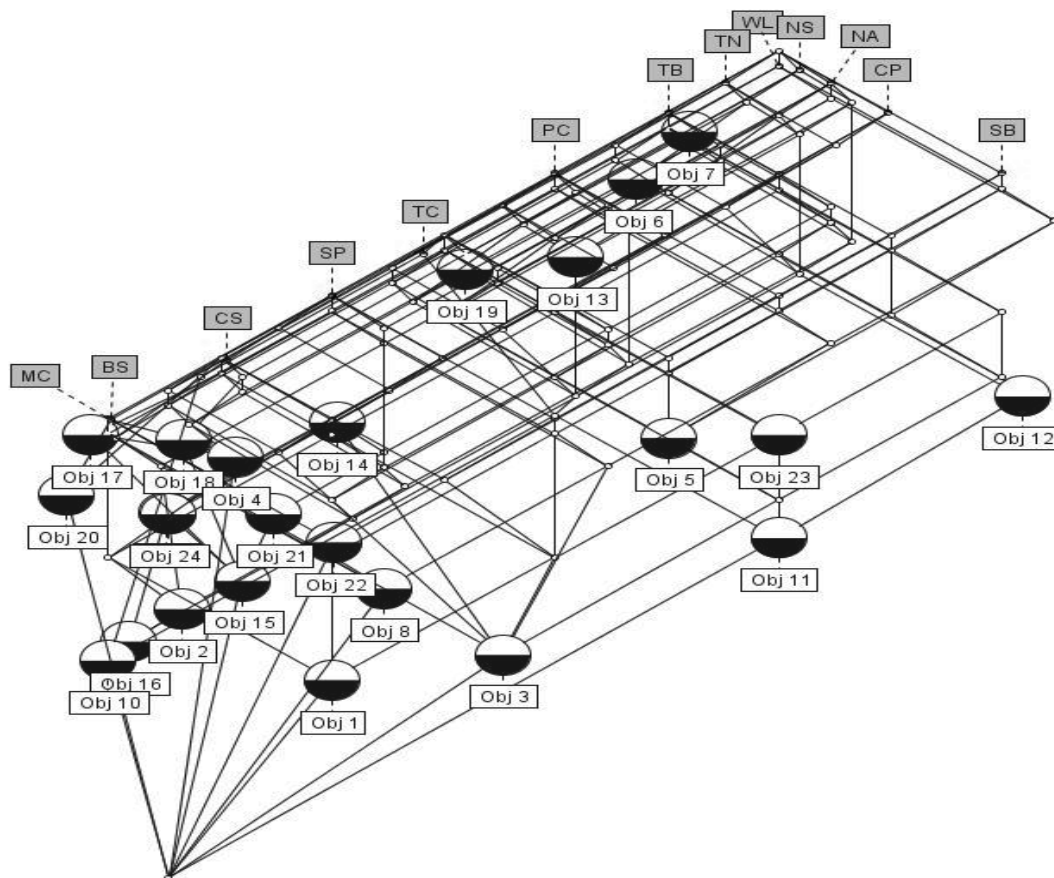
Figure 4. Concept lattice obtained with attribute exploration on TB attributes.

	PC	SP	MC	BS	CS	WL	NS	NA	CP	SB	TC	TN	TB
Obj 22	X								X			X	
Obj 23	X	X	X			X	X		X	X		X	X
Obj 24	X	X	X			X	X	X	X		X	X	X
Obj 25	X	X	X			X		X	X	X	X	X	X
Obj 26	X	X	X			X	X	X			X	X	X
Obj 27	X	X	X			X	X		X	X	X	X	X
Obj 28	X	X			X	X	X	X				X	X
Obj 29	X					X	X		X			X	X
Obj 30						X							X
Obj 31						X	X	X					X

Table 16. Incidence matrix of TB test dataset.

that the attribute implications obtained from the resultant context are able to diagnose TB better than experts' rules and are similar to FCA. Patient 30 is having only one symptom – Weight Loss (WL). However the doctor's diagnosis has confirmed the TB for this patient. Implications from FCA and attribute exploration have failed to confirm the disease for this patient due to the reason that there is no knowledge available in the domain to represent this dependency.

Patient	Doctor's Assessment	Original expert's rules	FCA on TB training context	FCA with attribute exploration
Obj 22	–	–	–	–
Obj 23	TB	TB	TB	TB
Obj 24	TB	TB	TB	TB
Obj 25	TB	TB	TB	TB
Obj 26	TB	TB	TB	TB
Obj 27	TB	TB	TB	TB
Obj 28	TB	TB	TB	TB
Obj 29	TB	TB	TB	TB
Obj 30	TB	–	–	–
Obj 31	TB	–	TB	TB
Classification Accuracy		80%	90%	90%

Table 17. Classification accuracy on TB test data.

Further, we have performed specificity and sensitivity analysis of the results shown in Table 17. Specificity measures the proportion of true negatives being correctly classified while sensitivity measures the proportion of true positives being correctly classified. Table 18 summarizes these results. Specificity analysis on

expert rules, implications from FCA and implications obtained by combining attribute exploration with domain experts' knowledge are successful in diagnosing all healthy people as healthy. Sensitivity analysis indicates that implications from FCA and from exploration process have identified more true positives than experts' rules.

	expert rules	Implications from FCA	Attribute exploration
Specificity	100%	100%	100%
Sensitivity	78%	89%	89%

Table 18. Specificity and Sensitivity analysis on TB data.

We have used two real world healthcare datasets for experimental analysis. All of these experiments are conducted using ConExp (http://conexp.sourceforge.net/) in which a general purpose implementation of the attribute exploration without background knowledge is available. Hence, to this process we have combined domain experts' knowledge as the background knowledge (Belohlavek and Vychodil, 2009; Ganter and Wille, 1999). The analysis has concentrated on the implications with 100% confidence due to the fact that the formal context is of medical domain. An implication in the DG basis can have a low support and can still be valid if it does not contradict to any example of the context. Identifying the dependencies existing in a domain allows efficient data analysis. Attribute exploration is a way to identify these dependencies. From the analysis we can observe that the attribute exploration process merges the objects

having equal set of attributes. Such merging of the objects can be performed as a data pre-processing stage. Generally, the attribute exploration would be performed in the conditions where objects are infinite or unknown. Hence our analysis also started with an object empty context in the exploration process. So this merging can be regarded as a natural outcome of attribute exploration process. The number of necessary exploration steps depends on the counterexamples provided by the expert. However, the result of attribute exploration is not a minimal set of objects needed for determining the structure of the concept lattice. From a formal concept lattice, we can find a minimal concept lattice so that it can avoid redundancy while maintaining the structure consistency. Future work can also focus upon extending the relation between conditional functional dependencies and FCA (Medina and Nourine, 2010). Exploration of fuzzy attributes in FCA is also an interesting research (Cynthia, 2012).

5. Conclusions

Attribute exploration process in FCA provides a means to acquire knowledge and transform it into a formal model. Through this process we understand the dependencies between the attributes of the model. However, the available implementation of this process does not consider the background knowledge of the domain. In this paper we have combined the knowledge obtained from the attribute exploration process with the knowledge available with domain experts, so as to better understand the dependencies between the attributes. Our analyses on two real world healthcare datasets conclude that this integration resulted in better classification accuracy than the experts' knowledge.

6. Acknowledgment

References

[1] CH. ASWANI KUMAR, S. SRINIVAS, Concept lattice reduction using fuzzy k means clustering. *Expert Systems with Applications*, **37** (2010), 2696–2704.

[2] CH. ASWANI KUMAR, S. SRINIVAS Mining associations helathcare data using formal concept analysis and singular value decomposition. *Biological Systems*, **18** (2010a), 787-807.

[3] CH. ASWANI KUMAR, Knowledge discovery in data using formal concept analysis and random projections. *International Journal of Applied Mathematics and Computer Science*, **21**(4) (2011), 745–756.

[4] CH. ASWANI KUMAR Mining association rules using non-negative matrix factorization and formal concept analysis. *Proceedings of 5th International conference on information processing Springer CCIS*, 157(2011b), 31–39.

[5] CH. ASWANI KUMAR, M. RADVANSKY, J. ANNAPURNA Analysis of vector space model, latent semantic indexing and formal concept analysis for information retrieval. *Cybernetics and Information Technology*, **12**(1) (2012), 34–48.

[6] CH. ASWANI KUMAR, Fuzzy clustering based formal concept analysis for association rule mining. *Applied Artificial Intelligence*, **26**(3) (2012), 274–301.

[7] CH. ASWANI KUMAR, Designing role based access control using formal concept analysis *Security and communication networks*, **6** (2013), 373–383.

[8] K. BAZHANOV, S. OBIEDKOV, Optimizations in computing the Dequenne-Guigues basis of implications, *Annals of mathematics and artificial intelligence*, (2013).

[9] R. BELOHLAVEK, V. VYCHODIL, Formal concept analysis with background knowledge: Attribute priorities. *IEEE Transactions on Systems, Man and Cybernetics*, **39** (2009), 399–409.

[10] V. G. CYNTHIA, Attribute exploration in fuzzy settings. Presented in the *Proceedings of 10th International Conference on Formal Concept Analysis, ICFCA*, (2012), 114–130.

[11] B. A. DAVEY, H. A. PRIESTLEY, *Introduction to Lattices and Order*. 2nd Edition, Cambridge University Press, 2002.

[12] S. M. DIAS, L. E. ZARATE, N. J. VIEIRA, Using concept lattices and implication rules to extract knowledge from ANN. *Intelligent automation and soft computing*, (2013).

[13] B. GANTER, Attribute exploration with background knowledge. *Theoretical Computer Science*, **217** (1999), 215–233.

[14] B. GANTER, R. WILLE, *Formal Concept Analysis: Mathematical Foundations*. Springer-Verlag, 1999.

[15] V. HORNER, Developing a consumer health informatics decision support system using formal concept analysis. Masters' Thesis, University of Pretoria 2007.

[16] R. JASCHKE, S. RUDOLPH, Attribute exploration on web, Contributions to the *11ᵗʰ International conference on formal concept analysis*, (2013), 19–34.

[17] S. O. KUZNETSOV, J. POELMANS, Knowledge representation and processing with formal concept analysis. *WIREs Data mining and knowledge discovery*, **3** (2013), 200–215.

[18] J. LI, C. MEI, CH. ASWANI KUMAR, X. ZHANG, On rule acquisition in decision formal contexts, *International journal of machine learning and cybernetics*, (2013).

[19] J. S. MI, Y. LEUNG, W. Z. WU, Approaches to attribute reduction in concept lattices induced by axialities. *Knowledge Based Systems*, **23** (2010), 504–511.

[20] R. MEDINA, L. NOURINE, Conditional functional dependencies. Presented in the *Proceedings of 8ᵗʰ International Conference on Formal Concept Analysis*, (2010), Springer-Verlag, pp. 161–176.

[21] S. OBIEDKOV, D. G. KOURIE, J. H. P. ELOFF, Building access control models with attribute exploration. *Computers and Security*, **28** (2009), 2–7.

[22] N. PASQUIER, Y. BASTIDE, R. TAOUIL, L. LAKHAL, Efficient mining of association rules using closed itemset lattices. *Information Systems*, **24** (1999), 25–46.

[23] J. POELMANS, P. ELZINGA, S. VIAENE, G. DEDENE, Formal concept analysis in knowledge discovery: A survey. In *Conceptual Structures: From Information to Intelligence*, Springer-Verlag (M. Croitoru et al., Ed.), (2010), pp. 139–153.

[24] J. POELMANS, S. O. KUZNETSOV, D. I. IGNATOV, G. DEDENE, Formal concept analysis in knowledge processing: A survey on models and techniques. *Expert systems with applications*, **14** (2013a), 6601–6623.

[25] J. POELMANS, S. O. KUZNETSOV, D. I. IGNATOV, G. DEDENE, Formal concept analysis in knowledge processing: A survey on applications, *Expert systems with applications*, **14** (2013b), 6538–6560.

[26] P. KUMAR SINGH, CH. ASWANI KUMAR, A method for reduction of fuzzy relation in fuzzy formal context. Presented in the *Proceedings of International Conference on Mathematical Modelling and Scientific Computation*, CCIS, Springer-Verlag, 283(2012), pp. 343-350

[27] U. PRISS, Formal concept analysis in information science. *Annual Review of Information Science and Technology*, **40** (2006), 521–543.

[28] A. REVENKO, S. O. KUZNETSOV, Attribute exploration of functions of ordered sets. Presented in the *Proceedings of 7ᵗʰ International Conference on Concept Lattices and Their Applications*, (2010), pp. 313–324.

[29] D. SERGIO MARIANO, V. NEWTON JOSE, Reducing the size of concept lattices: The JBOS approach. Presented in the *Proceedings of 7ᵗʰ International Conference on Concept Lattices and their Applications*, (2010), pp. 80–91.

[30] V. SNASEL, H. M. D. ABDULLA, M. POLOVINCAK, Behavior of the concept lattice reduction to visualizing data after using matrix decompositions. Presented in the *Proceedings of 4ᵗʰ International Conference on Innovations in Information Technology*, (2007), pp. 392–396.

[31] V. SNASEL, M. POLOVINCAK, H. M. DAHWA, Z. HORAK, On concept lattices and implication bases from reduced contexts. Presented in the *Proceedings of ICCS Supplement*, (2008), pp. 83–90.

[32] G. STUMME, Exploration tools in formal concept analysis. In *Ordinal and Symbolic Data Analysis. Studies in Classification* (OPITZ et al., Ed) Data Analysis and Knowledge Organization, (1995), pp. 31–44.

[33] G. STUMME Formal concept analysis. In *Handbook on Ontologies* (S. Staab, R. Studer, Ed.) (2009), pp. 177–200.

[34] P. VALTCHEV, R. MISSAOUI, R. GODIN, Formal concept analysis for knowledge discovery and data mining: The new Challenges. Presented in the *Proceedings of 2ⁿᵈ International Conference on Formal Concept Analysis*, (2004), pp. 352–371.

[35] L. WEI, L. QI, JIAN-JUN, Relation between concept lattice reduction and rough set reduction. *Knowledge Based Systems*, **23** (2010), 934–938.

[36] W. Z. WU, Y. LEUNG, J. S. MI, Granular computing and knowledge reduction in formal contexts. *IEEE Transactions on Knowledge and Data Engineering*, **21** (2009), 1461–1474.

[37] R. WILLE, Formal concept analysis as applied lattice theory. In *Proceedings of Concept Lattices and their Applications* (S. Ben Yahia et al. Ed.) Springer-Verlag (2008), pp. 42–67.

Contact addresses:

Jonnalagadda Annapurna
School of Computing Science and Engineering
VIT University
Vellore
India
e-mail: jannapurna@gmail.com

Aswani Kumar Cherukuri
School of Information Technology
VIT University
Vellore
India
e-mail: cherukuri@acm.org

JONNALAGADDA ANNAPURNA has received her Master's degree in Comp. Science & Engineering from VIT University, Vellore, India. Presently she is working as Asst. Professor (Senior) in School of Computing Sciences & Engineering, VIT University. She has research interests in theoretical computer science, social networks.

ASWANI KUMAR CHERUKURI is Professor of Network and Information Security Division, School of Information Technology and Engineering, VIT University, Vellore, India. Ch. Aswani Kumar holds a PhD degree in Computer Science from VIT University, India. His current research interests are data mining, formal concept analysis, information security, and machine intelligence. Aswani Kumar has published 50 refereed research papers in various national and international journals and conferences. Aswani Kumar was principal investigator to a major research project sponsored by the Department of Science and Technology, Govt. of India, during 2006 - 2008. Presently he is the principal investigator to a major research project funded by National Board of Higher Mathematics, Dept of Atomic Energy, Govt. of India. Aswani Kumar is a senior member of ACM and is associated with other professional bodies including ISC, CSI, ISTE.