# Chapter 11
# Knowledge Representation Using Formal Concept Analysis: A study on Concept Generation

**Ch. Aswani Kumar**
*VIT University – Vellore-632014, India*

**Prem Kumar Singh**
*VIT University – Vellore-632014, India*

## ABSTRACT

*Introduced by Rudolf Wille in the mid-80s, Formal Concept Analysis (FCA) is a mathematical framework that offers conceptual data analysis and knowledge discovery. FCA analyzes the data, which is represented in the form of a formal context, that describe the relationship between a particular set of objects and a particular set of attributes. From the formal context, FCA produces hierarchically ordered clusters called formal concepts and the basis of attribute dependencies, called attribute implications. All the concepts of a formal context form a hierarchical complete lattice structure called concept lattice that reflects the relationship of generalization and specialization among concepts. Several algorithms are proposed in the literature to extract the formal concepts from a given context. The objective of this chapter is to analyze, demonstrate, and compare a few standard algorithms that extract the formal concepts. For each algorithm, the analysis considers the functionality, output, complexity, delay time, exploration type, and data structures involved.*

## INTRODUCTION

Formal Concept Analysis (FCA) is a mathematical framework based on mathematical order and lattice theory which supports knowledge discovery in databases. The unique aspect of FCA is the integration of several components of conceptual data and knowledge processing. These components include discovery and reasoning with concepts and dependencies in data, visualization of concepts and dependencies. With such integration, FCA has been successfully applied in different domains

DOI: 10.4018/978-1-4666-4936-1.ch011

including gene expression data analysis, ontology design, software code analysis, psychology etc (Kaytoue et al. 2011, Krohn et al. 1999, Mephu & Ngiwoua, 1998, Priss 2006, Priss & Old 2004, Priss et al., 2007). Knowledge representation in FCA is shown in the form of formal concepts (Boulicaut & Besson 2008, Ganter & Wille 1999, Jurekevicius & Vasilecas 2009, Stumme 2009, Stumme et al. 2002), concept lattices(Stumme et al. 1998, Valtchev et al. 2002, Wille 1982) and association rules(Agrawaal & Srikant, 1994, Aswani Kumar, 2011, Aswani Kumar & Srinivas, 2010, Stumme et al. 2001, Zhang and Wu 2011). FCA starts the analysis on a data matrix, known as formal context, specifying a set of objects, a set of attributes and the relation between them (Wille, 1982). FCA has been successfully extended into fuzzy settings(Ghosh et al, 2010, Prem Kumar and Aswani Kumar, 2012a, 2012b, Maio et al., 2012), however this study is focused on FCA in crisp setting. From the formal context, FCA finds the natural clusters of objects that share a common subset of attributes and natural clusters of attributes that are shared by natural object clusters (Shi et al., 2007). A formal concept is a pair containing object cluster and corresponding attribute cluster. Concept lattice structure visualizes all the concepts (Belohlavek and Vychodil, 2009). The notion "formal" emphasizes that concepts are mathematical objects. The sets of objects and attributes in a formal concept mutually relate each other through a Galois connection which induces closure operator. The set of all the concepts of a given context is partially ordered and form a complete lattice. Main features of FCA include mathematical background, algorithmic methods that can perform conceptual clustering through concept lattice, rule mining through attribute implications, data apposition and concatenation makes FCA a suitable paradigm for KDD (Maddouri, 2005, Stumme, 1995, Zhang & Wu, 2011). Due to the closure properties and mathematical order theory that FCA follows, only the patterns of maximal size are extracted which reduces the exploration and increases the efficiency while mining the data. Through attribute implications, FCA provides a compact representation of knowledge. These attribute implications are closely associated with functional dependencies in the database field (Poelmans et al. 2010, Stumme 2002a, Wu et al. 2009).

Unlike other data mining techniques where highly iterative approaches are used, FCA organizes knowledge as a conceptual hierarchy. The basic notions of FCA are formal context, formal concept and concept lattice. Among the tasks of FCA, computing the formal concepts from the large binary matrices is a complex one. Several algorithms have been designed that compute all the concepts from a given context (Carpineto & Romano 2004, Hermann & Sertkaya, 2008, Kuznetsov & Obiedkov 2000, Kuznetsov & Obiedkov 2002). In this study we concentrate on Bordat, Next Neighbor, Object intersection, Next Closure algorithm. Generally the formal contexts are of four types: Average density, Small and Sparse, Large and dense and linearly incremental. It has been established in literature that these algorithms works better on these contexts(Carpineto & Romano 2004, Kuznetsov & Obiedkov 2000). Hence our analysis is focused on concept generation from these algorithms. The study is concentrated on the step by step demonstration of each algorithm, procedure for generating the concepts, and building the line diagrams. In the next section we present the brief background about FCA and its issues.

## BACKGROUND

Analysis of concepts, concept formation and conceptual learning are central to cognitive informatics. Key notions of FCA are formal context, formal concept and concept lattice. A formal context represents data in the form of triplet such as $\mathbf{K} = (\mathbf{G}, \mathbf{M}, I)$, where $\mathbf{G}$ is a finite set of objects and $\mathbf{M}$ is a finite set of attributes, and $I$ ($I \subseteq \mathbf{G} \times \mathbf{M}$) is a relation between the objects and its attributes.

## Related Content

Modeling and Forecasting the Daily Number of Emergency Department Visits Using Hybrid Models
Görkem Saryerand Ceren Öcal Taar (2020). *Computational Intelligence and Soft Computing Applications in Healthcare Management Science (pp. 19-41).*
www.irma-international.org/chapter/modeling-and-forecasting-the-daily-number-of-emergency-department-visits-using-hybrid-models/251966

An Evolving System in the Text Classification Problem
Elias Oliveira, Patrick Marques Ciarelliand Evandro Ottoni Teatini Salles (2013). *Handbook of Research on Computational Intelligence for Engineering, Science, and Business (pp. 467-486).*
www.irma-international.org/chapter/evolving-system-text-classification-problem/72504

Application of Statistical Modelling and Evolutionary Optimization Tools in Resin-Bonded Molding Sand System
Ganesh R. Chate, Manjunath Patel G. C., Mahesh B. Parappagoudarand Anand S. Deshpande (2018). *Handbook of Research on Investigations in Artificial Life Research and Development (pp. 123-152).*
www.irma-international.org/chapter/application-of-statistical-modelling-and-evolutionary-optimization-tools-in-resin-bonded-molding-sand-system/207202

Hierarchies of Architectures of Collaborative Computational Intelligence
Witold Pedrycz (2012). *Software and Intelligent Sciences: New Transdisciplinary Findings (pp. 32-46).*
www.irma-international.org/chapter/hierarchies-architectures-collaborative-computational-intelligence/65121

A Method for Identifying Fatigue State of Driver's Face Based on Improved AAM Algorithm
Zuojin Li, Jun Peng, Liukui Chen, Ying Wuand Jinliang Shi (2017). *International Journal of Software Science and Computational Intelligence (pp. 31-49).*
www.irma-international.org/article/a-method-for-identifying-fatigue-state-of-drivers-face-based-on-improved-aam-algorithm/181047