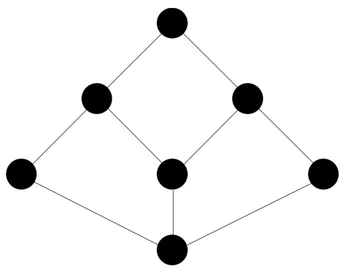
A FIRST COURSE

IN

FORMAL CONCEPT ANALYSIS

HOW TO UNDERSTAND LINE DIAGRAMS



Karl Erich Wolff

Fachhochschule Darmstadt

Forschungsgruppe Begriffsanalyse der Technischen Hochschule Darmstadt

Ernst Schröder Zentrum für Begriffliche Wissensverarbeitung

In: Faulbaum, F. (ed.) SoftStat'93

Advances in Statistical Software **4**, 429-438.

A FIRST COURSE
IN
FORMAL CONCEPT ANALYSIS

HOW TO UNDERSTAND LINE DIAGRAMS

Karl Erich Wolff

SUMMARY

This paper introduces the main ideas of Formal Concept Analysis in a very elementary way without using formal mathematical definitions. It leads the reader to an understanding of an important technique of graphical knowledge representation namely the line diagrams of concept lattices. As a short application some aspects of conceptual learning using attribute exploration are described. In the second part we introduce the method of conceptual scaling by an example and the use of nested line diagrams, the main tool for the graphical representation of multivariate data.

1 INTRODUCTION

Remarks

Formal Concept Analysis has been introduced by WILLE (82) and applied in many quite different realms like psychology, sociology, anthropology, medicine, biology, linguistics, computer sciences, mathematics and industrial engineering. During the last years I had the opportunity to use Formal Concept Analysis in cooperation with scientists from all the realms mentioned. They often asked me for a paper which should explain the main ideas just like in an informal talk without using the strange mathematical formal language. This has the advantage that the main ideas are understood quite quickly, but the disadvantage that some important connections cannot be explained since the language chosen is too inexact. In the following I try to explain the main ideas from examples using the transfer ability of the reader. Those readers which are interested in the mathematical definitions of contexts, formal concepts and concept lattices should read the papers of WILLE(82) and WOLFF(88).

Concepts and formal concepts

In the following we should distinguish between our informal understanding of concepts like *preying (animal)* and our understanding of the formal concept *preying* as we use it in our first example. But what is our understanding of the concept *preying*? It is obvious that different people use the same word *"preying"* in different contexts. Do they mean the same concept? What does this question mean and what is a concept?
From a philosophical point of view a *concept* is a unit of thoughts consisting of two parts, the *extension* and the *intension*. The extension covers all objects belonging to this concept and the intension comprises all attributes valid for all those objects (WAGNER 73). Hence *objects* and *attributes* play a prominent role together with several relations like e.g. the hierarchical *"subconcept-superconcept" relation* between concepts, the *implication* between attributes, and the *incidence relation* "an object has an attribute".
It was the idea of WILLE (82) to combine in a first step objects, attributes and the incidence relation in a mathematical definition of a *formal context* as a tool for the description of those very elementary linguistic situations when some of the most simple verbal utterances of the form "an object has an attribute" are given.The second and crucial step was then the definition of a *formal concept* for a given formal context. We describe this using the following example.

2 THE CONCEPTS OF A CONTEXT

A context of animals

The following Table 1 describes for some animals which of the mentioned attributes they have.This is indicated by crosses. An empty cell in Table 1 indicates that the corresponding animal doesn't have the corresponding attribute. In other examples an empty cell might also mean that it is not known whether this object has the attribute or not.

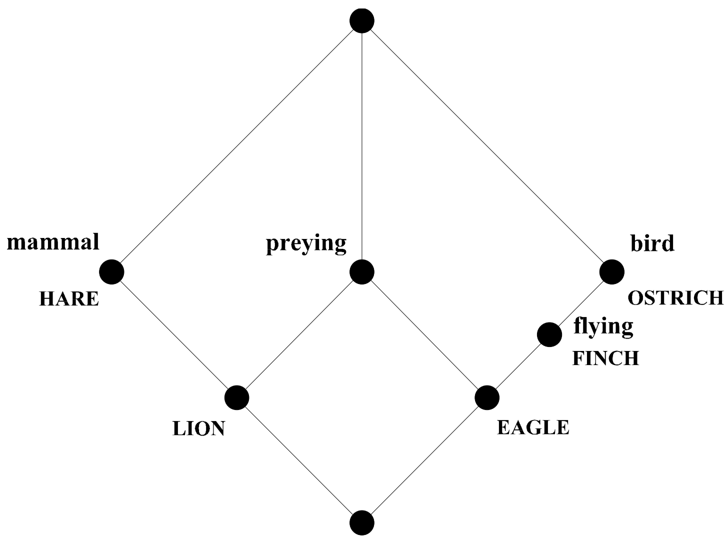
TABLE 1:

ANIMALS	preying	flying	bird	mammal
LION	X			X
FINCH		X	X	
EAGLE	X	X	X	
HARE				X
OSTRICH			X	

A table of crosses represents a very simple and often used data type. The mathematical structure which is used to describe formally these tables of crosses is called a *formal context* (or briefly a *context*). To explain the notion of a *formal concept* of a context we look at the

attributes of the FINCH and ask for all those animals (of this context) which have all the attributes of the FINCH. Hence we get the set A consisting of FINCH and EAGLE. This set A of objects is closely connected to the set B consisting of the attributes flying and bird: A is the set of all objects having all the attributes of B, and B is the set of all attributes which are valid for all the objects of A. Each such pair (A,B) is called a *formal concept* (or briefly a *concept*) of the given context. The set A is called the *extent*, the set B the *intent* of the concept (A,B). One should mention that the extent of a concept determines the intent and the intent determines the extent. Hence the notion of a formal concept contains some redundant information. But this is very useful since one can choose which of both parts of the concept should be used in a given situation. Between the concepts of a given context there is a natural hierarchical order, the *"subconcept-superconcept" relation*. E.g. the preying flying birds describe a subconcept of the concept of the flying birds. The extent of this subconcept consists only of the EAGLE, the intent consists of the three attributes preying, flying and bird. The extent of the given superconcept consists of FINCH and EAGLE, the intent only of flying and bird. In general a concept c is a *subconcept* of a concept d (and d is called a *superconcept* of c) if the extent of c is a subset of the extent of d (or equivalently: if the intent of c is a superset of the intent of d).
Just like in the example of the FINCH we can construct for each object g its *object concept* (A,B), where B is the set of all attributes of g and A is the set of all objects having all the attributes of B. In the same way each attribute m determines its *attribute concept* (C,D), where C is the set of all objects of m and D is the set of all attributes valid for all objects of C.
In the following *line diagram* we represent the conceptual hierarchy of all concepts of the context ANIMALS:

DIAGRAM 1:



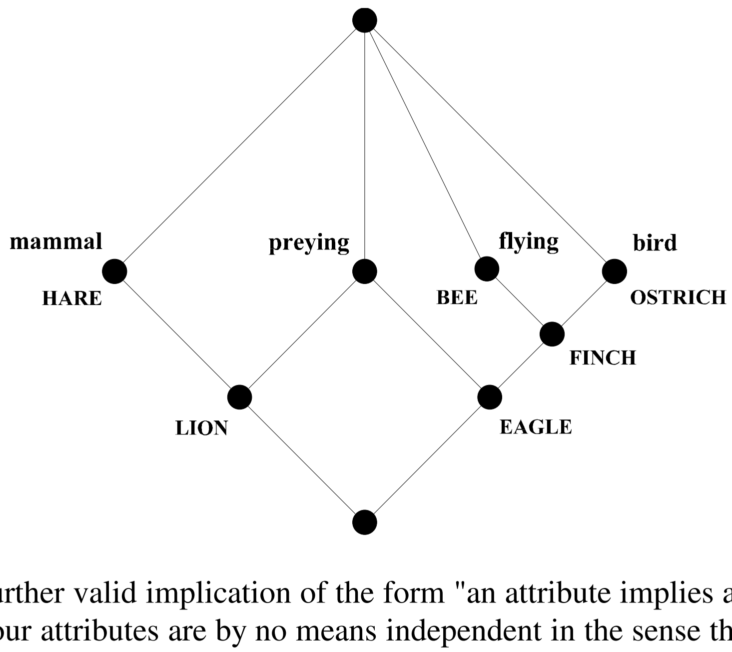
3 HOW TO READ A LINE DIAGRAM

A line diagram consists of circles, lines and the names of all objects and all attributes of the given context. The circles represent the concepts and the information of the context can be read from the line diagram by the following simple *reading rule*:
An object g has an attribute m if and only if there is an upwards leading path from the circle named by "g" to the circle named by "m".
Hence we recognize from the line diagram above that the FINCH has exactly the attributes flying and bird. As a consequence of the reading rule we can easily read from the line diagram

the extent and the intent of each concept by collecting all objects below respectively all attributes above the circle of the given concept. Hence the object concept "FINCH" has the extent FINCH and EAGLE and the intent flying and bird. The extent of the top concept is (always) the set of all objects while the intent of it does not contain any attribute (in this context). But in other contexts it may occur that the intent of the top concept is not empty, e.g. if we add to the given context the attribute "animal" with crosses in each row then the top concept would be the attribute concept of "animal" and the intent of the top concept would contain just the attribute "animal".

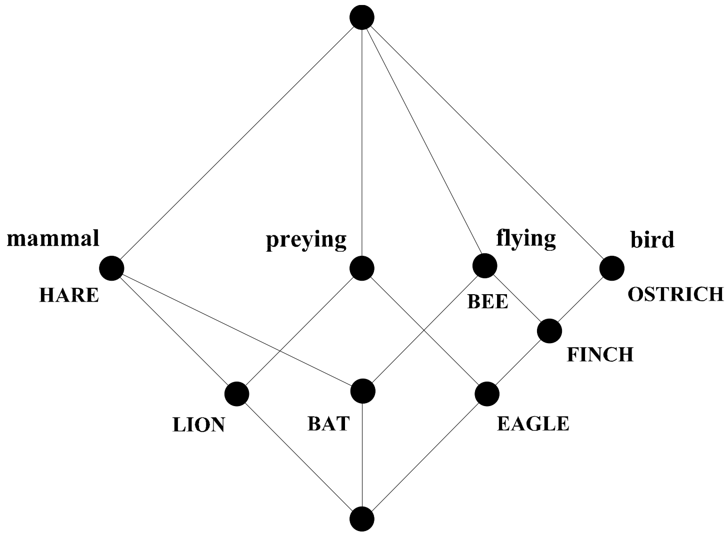
4 CONCEPTUAL LEARNING

To model the conceptual structure of a simple learning situation, e.g. that a son learns something from his father, we represent the knowledge of the father by a context U (called the universe) and the actual knowledge of the son by a partial context K of U, i.e. the objects and attributes of K are objects resp. attributes of U and if an object has an attribute in K then it has this attribute also in U, hence we assume that the son learns without mistakes (with respect to the knowledge of the father).
Let's assume that the son is only interested in some attributes between animals, say the given four attributes. Then we can assume that the context U has also just these attributes.Let's assume that the son has just learned from his father the context ANIMALS. He then realizes that in this small context each flying animal is a bird. Hence he asks his father whether this implication is valid also in U. The father answers no and gives a counterexample, e.g. a bee, which has among the four given attributes of K only the attribute flying. The new knowledge of the son is represented in the following line diagram.
DIAGRAM 2:



Now there is no further valid implication of the form "an attribute implies another one" in this context. But the four attributes are by no means independent in the sense that there is no valid implication between the four attributes, since the implication "preying and bird implies flying" is valid in the new context. Is this implication also valid in the universe? The father says yes (and I think the reader agrees with respect to his own knowledge). But the clever son also recognizes in his context the valid implication "mammal and flying implies preying and bird". Is this one also valid in the universe? The father says no, giving the bat as a counterexample which is a flying mammal but not preying (in his sense of preying animal) and not a bird. The new knowledge of the son is represented in the following line diagram.

DIAGRAM 3:



Now a very remarkable situation occurs: The father sees that the concept lattice of his son has the same structure as his own concept lattice (in mathematical terms both lattices are isomorphic). The concept lattice of U has much more objects, but they are all arrangable in the pattern of DIAGRAM 3. E.g. the object concept of butterfly is the same as that of the bee, the object concept of alligator is the attribute concept of preying and the goldfish has none of the given attributes, hence its object concept is the top concept. Now the son has reached the knowledge state of understanding the conceptual structure of his father's knowledge, i.e. of the given universe.

This knowledge acquisiton procedure has been implemented in the program CONIMP4 by P. Burmeister (cf. BURMEISTER 87). This is an interactive program which learns (like the son) the knowledge of an expert (corresponding to the father). Usually the expert does not know the conceptual structure of his knowledge, but he is interested to understand it. Working with the program he has to answer implications between the attributes he is interested in. If an implication is not valid in his universe, then he has to put in a counterexample. In the actual context of the counterexamples the program calculates valid implications and asks the expert whether they are valid also in the universe. If all valid implications of the actual context are also valid in the universe then the program stops, since in this case the concept lattices of the actual context and the universe are isomorphic. Hence it is not necessary that the expert knows the concept lattice of his universe. In practice the expert likes to understand his universe by seeing the connections between his concepts in a line diagram. Indeed the program does not ask all valid implications of the actual context, but only a certain minimal basis of implications from which all other valid implications can be deduced (cf. DUQUENNE, GUIGUES 86).

This so-called attribute exploration can be dualized by interchanging objects and attributes to explore the conceptual structure between given objects with respect to a certain universe. This program has been applied in psychology, linguistics and mathematics.

For a more detailled desciption the reader is referred to WILLE (89a, 92), WOLFF(88) and FISCHER, SPANGENBERG, WOLFF (93).

5 SCALING: THE TRANSFORMATION OF DATA INTO CONTEXTS

Data are often given in a table of the following form:

TABLE 2:

K_0	sex	age
ADAM	m	21
BETTY	f	50
CHRIS	/	66
DORA	f	88
EVA	f	17
FRED	m	/
GEORGE	m	90
HARRY	m	50

This table might be a part of a questionaire. Usually there are some missing values in the data which are indicated in Table 2 by the slash "/". We transform this so-called many-valued context into a formal context:

TABLE 3:

	sex		age				
K	m	f	<18	<40	≤65	>65	≥80
ADAM	×			×	×		
BETTY		×			×		
CHRIS						×	
DORA		×				×	×
EVA		×	×	×	×		
FRED	×						
GEORGE	×					×	×
HARRY	×				×		

This transformation fixes the view we wish to look at our data. In the beginning of a data exploration one should first choose a rough view to get an overview over the data. Later on one can unfold step by step any detail given in the original data.

In our example we like to look at the values of the age in a rough manner, using some new attributes (the so-called scale attributes) given in the second line of Table 3 under the header "age". The rule how to transform the two many-valued columns into the (seven) columns of a formal context are given by the following two formal contexts, called *scales*:

TABLE 4:

S_1	m	f
m	×	
f		×
/		

TABLE 5:

S_2	<18	<40	≤65	>65	≥80
17	×	×	×		
21		×	×		
50			×		
66				×	
88				×	×
90				×	×
/					

These scales represent the language describing the chosen view. The meaning of the scale attributes can be visualized by the following line diagrams of the scales:

DIAGRAM 4:

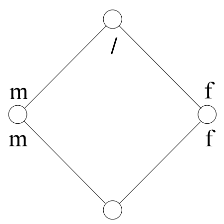
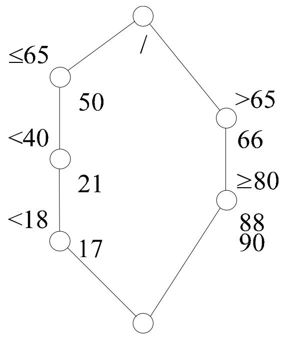


DIAGRAM 5:



In both scales the slash "/" as an object of the scale doesn't have any scale attribute, hence its object concept is the top concept. The line diagram of the age scale shows that the attributes of this scale divide the age values into two classes, say the young and the old ones, and each class is ordered by a chain. Therefore this scale is called a biordinal scale.
Now we are ready to unfold the structure of the scaled context in Table 3. First of all we look at the two subcontexts spanned by the scale attributes of sex and age. The corresponding concept lattices are given by the following line diagrams.

DIAGRAM 6:

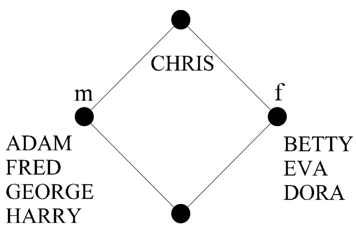
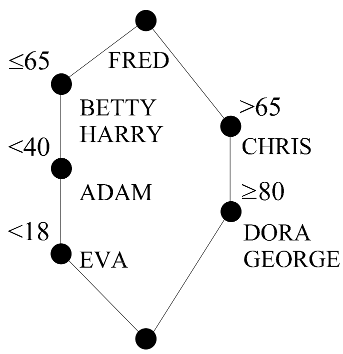


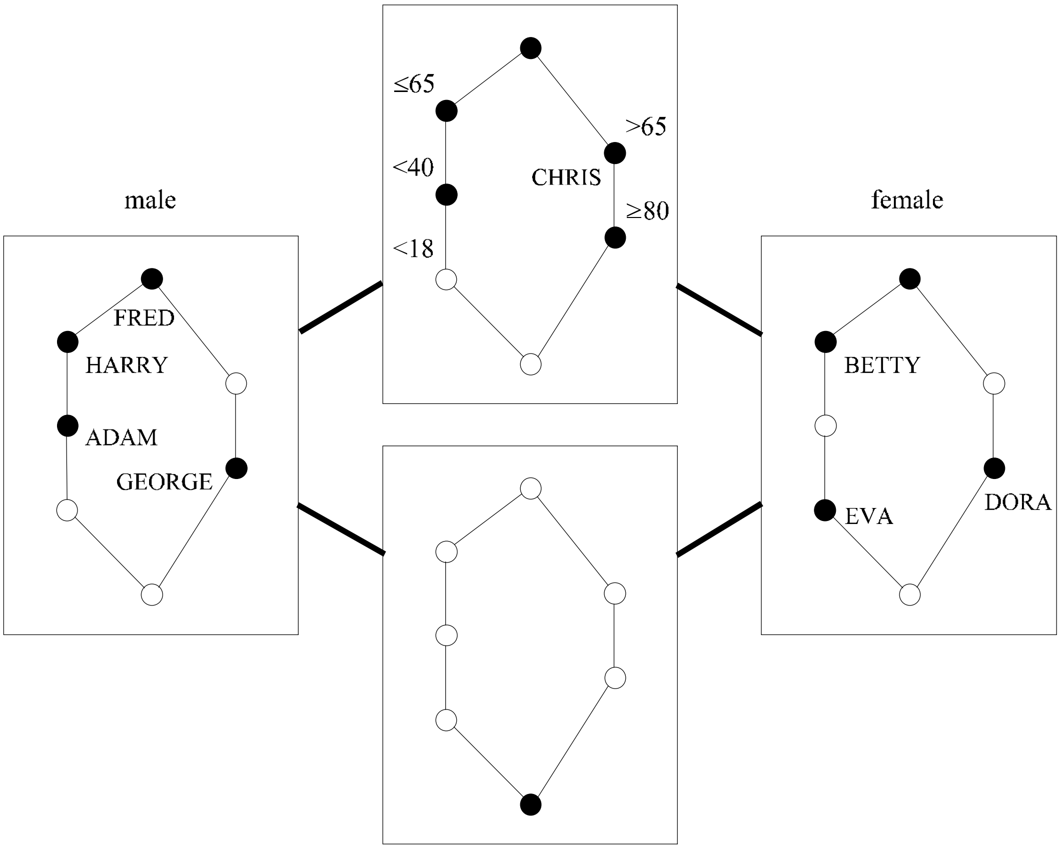
DIAGRAM 7:



These diagrams can be obtained from the scale diagrams above by replacing each value (e.g. the age value 50) by the set of all people with this value (e.g. BETTY, HARRY).
This very simple sorting procedure gives us for each many-valued attribute the distribution of the objects in the line diagram of the chosen scale. The well-known histograms for one variable arise as special cases from line diagrams.
Usually we are interested in the interaction between two or more many-valued attributes. These interactions can be visualized using the so-called *nested line diagrams.*

5 A NESTED LINE DIAGRAM

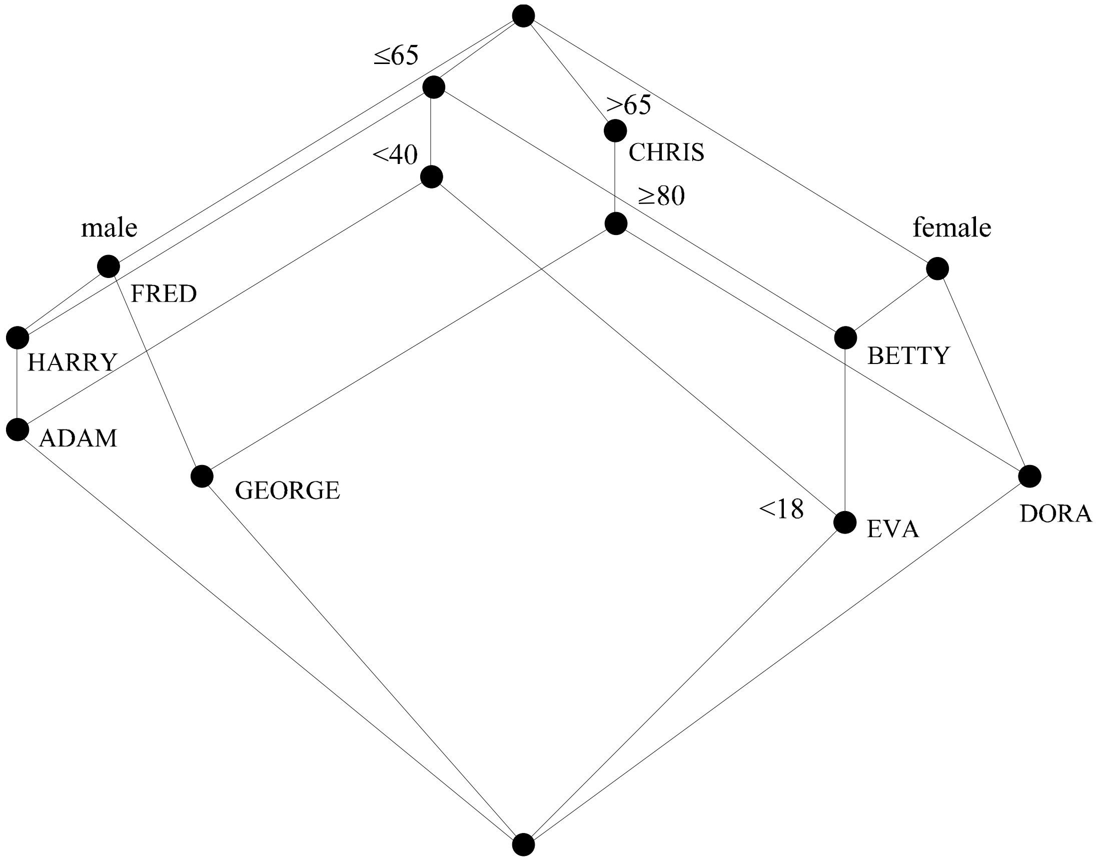
DIAGRAM 8:



This nested line diagram is constructed from Diagram 6 and Diagram 7 in a simple way:
Since we are interested to see for each of the concepts of Diagram 6 how its people are distributed in the age scale, we blow up each circle of Diagram 6 and insert the line diagram of the age scale.
Hence Diagram 8 represents all pairs (c,d) of concepts c from the first and concepts d from the second lattice. This structure is called the *direct product* of the two given lattices.
How to read the conceptual hierarchy in a nested line diagram? Answer: Read at first in the first factor (the rough line diagram) and then in the second. Example: The concept of DORA is a subconcept of the concept of CHRIS, since firstly the rough concept of DORA, namely female, is a subconcept of the rough concept of CHRIS, and secondly the fine concept of DORA, namely ≥ 80, is a subconcept of the fine concept of CHRIS.

From the nested line diagram we can easily draw a nice usual line diagram of the same context.

DIAGRAM 9:



The last two diagrams demonstrate the general situation that the concept lattice of a scaled context can be embedded into the direct product of the concept lattices of the scales. Usually the direct product has some elements not occupied by the embedded lattice. These elements are indicated by white circles in Diagram 8 and have the same meaning as empty cells in contingency tables. Indeed all contingency tables can be represented as special concept lattices.

To demonstrate just by an example another meaning of the white circles we formulate the implication which is given by the white circle above the circle of the object concept of EVA: Each woman younger than fourty is younger than eighteen (in this context!). In larger concept lattices these *context implications* enable us to describe regions of empty cells quite easily.

Finally it should be mentioned that the scaling procedure, i.e. the transformation of the given data into a table of crosses might look at the first glance as a rough description of the original data, but it is indeed a very powerful method to represent the given data with respect to many different views just in the language of the user. It's especially possible to represent the data without any loss of information. The direct product of the concept lattices of the chosen scales gives us a meaningful frame for the representation of the data under this view. Therefore this direct product might be denoted as the *cognitive space* of the given view.

6 APPLICATIONS AND SOFTWARE

There are two main branches of applications of Formal Concept Analysis: conceptual data analysis and conceptual knowledge systems, including knowledge representation, acquisition and inference. In conceptual data analysis one should mention the following applications: SPANGENBERG and WOLFF use Formal Concept Analysis for the evaluation of psychoanalytic data, often given in the form of repertory grids. They have demonstrated that line diagrams show up remarkable advantages in comparison with biplots (based on Principal Component Analysis). One of several applications of conceptual data analysis in industrial engineering is described by WOLFF and STELLWAGEN (93). For an introduction into the application of Formal Concept Analysis in the social sciences the reader is referred to an evaluation of the "SPIEGEL" data on german universities (cf. WOLFF 93) and a conceptual representation of differences between people in the eastern and western part of Germany - based on the ALLBUS-Baseline 1991 (cf. WOLFF, GABLER, BORG).

For an introduction into the theory and application of conceptual knowledge systems the reader is referred to WILLE (92).

The software concerning Formal Concept Analysis contains the following programs: CONIMP4 (cf. BURMEISTER 87) calculates contexts and implications, ANACONDA (written by M. SKORSKY) draws line diagrams and TOSCANA (written by V. ABEL, P. REISS, M. SKORSKY, F. VOGT, A. WOLF) scales many-valued contexts and draws nested line diagrams. The central algorithm in these programs is the "NEXT CLOSED SET" algorithm by B. GANTER (cf. GANTER 87).

REFERENCES

Burmeister, P. (1987) Programm und Programmbeschreibung zur formalen Begriffsanalyse einwertiger Kontexte (unter Mithilfe von A. Rust, P. Scheich und N. Newrly). Fachbereich Mathematik, Technische Hochschule Darmstadt, neueste Fassung : CONIMP4 (1993).

Duquenne, V. & Guigues, J.-L. (1986) Familles minimales d'implication informatives resultant d'un tableau de donnees binaires. *Math. Sci. Humaines 95*, 5-18.

Fischer, R. & Spangenberg, N. & Wolff, K.E. (1993) Implicit knowledge structures of therapists concerning the indication of psychoanalysis versus psychotherapy: an empirical strategy of conceptual research. Preprint of the Sigmund-Freud-Institut Frankfurt.

Ganter, B. (1987) Algorithmen zur formalen Begriffsanalyse. In: Ganter, B. & Wille, R. & Wolff, K.E. (eds.) (1987) *Beiträge zur Begriffsanalyse*.241-254.Mannheim/Wien/Zürich, BI-Wissenschaftsverlag.

Ganter, B. & Wille, R. (1989). Conceptual scaling. In: F.Roberts (ed.) *Applications of combinatorics and graph theory to the biological and social sciences*. 139-167. New York, Springer-Verlag.

Spangenberg, N. & Wolff, K.E. (1991a) Interpretation von Mustern in Kontexten und Begriffsverbänden. *Actes 26e Séminaire Lotharingien de Combinatoire*, 93-113. Strasbourg, Publication de l' Institut de Recherche Mathématique Avancée.

Spangenberg, N. & Wolff, K.E. (1991b) Comparison between Biplot Analysis and Formal Concept Analysis of Repertory Grids. In: H.H.Bock, P.Ihm (eds.) *Classification, Data Analysis, and Knowledge Organization*, 104-112, Berlin-Heidelberg, Springer-Verlag.

Spangenberg, N. & Wolff, K.E. (1993) Datenreduktion durch die Formale Begriffsanalyse von Repertory Grids. J.W.Scheer, A.Catina (eds.) *Einführung in die Repertory Grid-Technik, Vol. 2, Klinische Forschung und Praxis*, 38-54. Bern, Verlag Hans Huber.

Wagner H. (1973) Begriff. In: *Handbuch philosophischer Grundbegriffe*. 191-209. München, Kösel.
Wille, R. (1982) Restructuring lattice theory: an approach based on hierarchies of concepts. In: Rival, I. (ed.) *Ordered Sets*.445-470. Dordrecht-Boston, Reidel.
Wille, R. (1984) Liniendiagramme hierarchischer Begriffssysteme. In: H.H.Bock (ed.): *Anwendungen der Klassifikation: Datenanalyse und numerische Klassifikation*.32-51. Frankfurt, INDEKS-Verlag ; engl. translation: Line diagrams of hierarchical concept systems. *International Classification 11*, 77-86.
Wille, R. (1987) Bedeutungen von Begriffsverbänden. In: Ganter, B. & Wille, R. & Wolff, K.E. (eds.) (1987) *Beiträge zur Begriffsanalyse*.161-211. Mannheim/Wien/Zürich, BI-Wissenschaftsverlag.
Wille, R. (1989a) Knowledge Acquisition by Methods of Formal Concept Analysis. In: E. Diday (ed.) *Data Analysis, Learning Symbolic and Numeric Knowledge*. 365-380. Nova Science Publishers, New York - Budapest.
Wille, R. (1989b) Lattices in data analysis: how to draw them with a computer. In: Rival, I. (ed.): *Algorithm and order*. 33-58, Dordrecht-Boston, Kluver.
Wille, R. (1992) Concept lattices and conceptual knowledge systems. In: *Computers and Mathematics with Applications*, 23 , 493-515.
Wolff, K.E. (1988) Einführung in die Formale Begriffsanalyse. *Actes 19e Séminaire Lotharingien de Combinatoire*. 85-96. Strasbourg, Publication de l' Institut de Recherche Mathématique Avancée.
Wolff, K.E. & Stellwagen, M. (1993) Conceptual optimization in the production of chips. In: Janssen, J. & Skiadas, C.H. (eds.) *Applied Stochastic Models and Data Analysis*, Vol. 2, 1054-1064. World Scientific Publishing Co. Pte. Ltd.
Wolff, K.E. (1993) Begriffsanalytische Visualisierung mehrdimensionaler Daten aus einer SPIEGEL-Umfrage an westdeutschen Universitäten. Preprint Fachhochschule Darmstadt.
Wolff, K.E. & Gabler, S. & Borg, I. (1994) Formale Begriffsanalyse von Arbeitswerten in Ost- und Westdeutschland. ZUMA-Nachrichten 34, Jg.18, S. 69-82.