TOSCANA-Systems Based on Thesauri

Bernd Groh¹, Selma Strahringer², and Rudolf Wille²

¹ School of Information Technology
Griffith University
PMB 50 Gold Coast Mail Centre
QLD 9726
Australia
e-mail: b.groh@gu.edu.au

² Fachbereich Mathematik
Technische Universität Darmstadt, Schloßgartenstr. 7
64289 Darmstadt, Germany
e-mail: {strahringer, wille}@mathematik.tu-darmstadt.de

Abstract. TOSCANA is a computer program which allows an online interaction with data bases to analyse and explore data conceptually. Such interaction uses conceptual data systems which are based on formal contexts consisting of relationships between objects and attributes. Those formal contexts often have attributes taken from a thesaurus, which may be understood as ordered set and be completed to a join-semilattice (if necessary). The join of thesaurus terms indicates the degree of resemblance of the terms and should therefore be included in the formal contexts containing those terms. Consequently, the formal contexts of a conceptual data system based on a thesaurus should have join-closed attribute sets. A problem arises for the TOSCANA-system implementing such conceptual data system because the attributes in a nested line diagram produced by TOSCANA might not be join-closed, although its components have join-closed attribute sets. In this paper we offer a solution to this problem by developing a method for extending line diagrams to those whose attributes are join-closed. This method allows to implement TOSCANA-systems based on thesauri which respect the join-structure of the thesauri.

Keywords: Conceptual Knowledge Processing, Formal Concept Analysis, Drawing Line Diagrams, Thesauri

1 TOSCANA

TOSCANA is a computer program which allows an online interaction with databases to analyse and explore data conceptually. TOSCANA realizes conceptual data systems [17],[20] which are mathematically specified systems consisting of a data context and a collection of formal contexts, called conceptual scales, together with line diagrams of their concept lattices. There is a connection between formal objects of the conceptual scales and the objects in the data context that
can be activated to conceptually represent the data objects within the line diagrams of the conceptual scales. This allows thematic views into the database (underlying the data context) via graphically presented concept lattices showing networks of conceptual relationships. The views may even be combined, interchanged, and refined so that a flexible and informative navigation through a conceptual landscape derived from the database can be performed (cf. [21]).

For an elementary understanding of conceptual data systems, it is best to assume that the data are given by a larger formal context \( K := (G, M, I) \). A conceptual scale derived from data context \( K \) can then be specified as a sub-context \( (G, M_j, I \cap (G \times M_j)) \) with \( M_j \subseteq M \). A basic proposition of Formal Concept Analysis states that the concept lattice of \( K \) can be represented as a \( \bigvee \)-subsemilattice within the direct product of the concept lattices of subcontexts \( (G, M_j, I \cap (G \times M_j)) \) \( (j \in J) \) if \( M = \bigcup_{j \in J} M_j \) [3; p. 77]. This explains how every concept lattice can be represented by a nested line diagram of smaller concept lattices. Figure 1 shows a nested line diagram produced with the assistance of TOSCANA from a database about environmental literature. One concept lattice is represented by the big circles with their connecting line segments, while the line diagram of the second is inserted into each big circle. In such a way the direct product of any two lattices can be diagramed, where a non-nested line diagram of the direct product may be obtained from a nested diagram by replacing each line segment between two big circles by line segments between corresponding elements of the two line diagrams inside the two big circles. In Fig. 1\(^3\), the black little circles represent the combined concept lattice which has the two smaller concept lattices as \( \bigvee \)-homomorphic images. Since larger data contexts usually give rise to an extensive object labelling, TOSCANA attaches first to a node the number of objects which generate the concept represented by that node; after clicking on that number, TOSCANA presents all names of the counted objects.

TOSCANA-systems have been successfully elaborated for many purposes in different research areas, but also on the commercial level. For example, TOSCANA-systems have been established: for analyzing data of children with diabetes [17], for investigating international cooperations [11], for exploring laws and regulations concerning building constructions [13], for retrieving books in a library [12], [15], for assisting engineers in designing pipings [19], for inquiring flight movements at Frankfurt Airport [10], for inspecting the burner of an incinerating plant [9], for developing qualitative theories in music esthetics [14], for studying semantics of speech-act verbs [8], for examining the medical nomenclature system SNOMED [16] etc. In applying the TOSCANA program, the desire often arises to extend the program by additional functionalities so that TOSCANA is still in a process of further development (cf. [18]).

\(^3\) Translation of the labels in Fig. 1: Fluss/river, Oberflaechengewaesser/surface waters, Talsperre/impounded dam, Stausee/impounded lake, Staugewaesser/back water, Stauanlage/reservoir facilities, Staudamm/storage dam, Stausufe/barrage weir with locks, Seen/lakes, Teich/pond