Reuse in the Development Process of TOSCANA Systems

Martin Skorsky,[1] Gerd Stumme,[2] Rudolf Wille,[2] Uta Wille[3]

[1] NaviCon Gesellschaft für Begriffliche Wissensverarbeitung mbH, Heinrichstrasse 9, 60327 Frankfurt/Main, Germany; skorsky@navicon.de
[2] Technische Universität Darmstadt, Fachbereich Mathematik, D–64289 Darmstadt, Germany; {stumme, wille}@mathematik.tu-darmstadt.de
[3] Jelmoli AG, Information Systems, Postfach 3020, CH–8021 Zürich, Switzerland; wille_u@jelmoli.ch

Abstract. TOSCANA is a graphical tool that supports the human-centered interactive processes of conceptual knowledge processing. The generality of the approach makes TOSCANA a universal tool applicable to a variety of domains. Only the so-called conceptual scales have to be designed for new applications. The presentation shows how the use of *abstract scales* allows the reuse of formerly defined conceptual scales. Furthermore it describes how thesauri and conceptual taxonomies can be integrated in the generation of conceptual scales.

1 Introduction

The software system TOSCANA is a graphical tool that supports the human-centered interactive processes of conceptual knowledge processing (see [14], [15], [13]). Each TOSCANA System implements a *conceptual data system* that consists of a relational or dimensional database, a collection of *conceptual scales* describing different views to the underlying data, and line diagrams representing the *concept lattices* of the chosen conceptual scales. The generality of the approach makes Conceptual Knowledge Processing applicable to a variety of domains and TOSCANA a universal tool for knowledge processing.

TOSCANA Systems have been developed for various purposes in research and practice, for example, for assisting engineers in designing pipings [11], for developing qualitative theories in music esthetics [5], for studying semantics of speech-act verbs [1], for analyzing data of children with diabetes [7], for investigating international cooperations [2], for exploring laws and regulations in civil engineering [4], for retrieving books in a library [3], [6]. The role of TOSCANA Systems for On-Line Analytical Processing (OLAP) and within the Knowledge Discovery in Databases process (KDD) is discussed in [8] and [9].

TOSCANA Systems are developed in an interactive process with a domain expert. In the presentation, we describe this process, analyze and discuss the process with respect to reuse, and illustrate individual steps by means of examples.

2 Reusing conceptual scales in developing TOSCANA Systems

Focusing on the principal aspects of the development of TOSCANA Systems rather then on technical or implementational details, the development of TOSCANA Systems basically can be broken down to the following five steps:

1. Formulation of the central business or research issues
2. Assessment of available data and definition of basic contexts
3. Definition and implementation of the underlying data model
4. Definition and preparation of conceptual scales
5. Test and review of the TOSCANA System

The definition of conceptual scales is a key step of the development process and most relevant with regard to reuse. In this step, provisional query themes which are formulated in the first step are refined to conceptual scales. Depending on the amount of theory that the domain expert has on the attributes, we have to consider two cases: theory- and data-driven development of the conceptual scales.

2.1 Preparation of theory-driven scales with Anaconda

The first step in theory-driven preparation of conceptual scales is to choose, for each database attribute, derived attributes which are meaningful to the user. The derived attributes are combined to a conceptual scale. The knowledge engineer brings in his knowledge about typical conceptual structures where standard scales (e. g., nominal, ordinal, interordinal, Boolean scales) are often used. Together with the domain expert he chooses one of the standard scales or – where necessary – creates a new, more specific scale.

For each conceptual scale, the structure of its concept lattice, the layout of its line diagram, the SQL-queries providing the reference to the database, and its attributes are described in the description language Conscript ([12]). Only the last two of them are domain-specific, while the structure of the concept lattice and the layout of the diagram can be reused. They are stored as an *abstract scale*. For a specific TOSCANA System, the conceptual scales are generated as *concrete scales* which link the abstract scales with the domain-specific attributes and SQL-queries.

The knowledge engineer and the domain expert interactively define suitable scales for the intended system. If there is a scale of the same structure that was created for an earlier application, then the old abstract scale is taken and will be combined with the new attributes and SQL queries to form a new concrete scale. It is also possible to copy the abstract scale and modify it to fit better the new application.

The preparation of abstract and concrete scales is supported by the preparator software Anaconda. At the moment, each scale needs at least a minimal effort by the knowledge engineer. It is planned to provide a set of standard scales for the standard data types of the database system. Then a first prototype of a TOSCANA System can be generated automatically ([10]), and the knowledge engineer will only be needed for fine-tuning the system and for creating individual scales.

2.2 Establishment of data-driven scales with DOKUANA

If a part of the database consists only of Boolean attributes, then usually 5–10 of them are grouped together in order to form one conceptual scale. Often one does not know which combinations of these 5–10 attributes can be realized by objects in the database. Hence the only scale which can be applied is a Boolean scale providing all combinations. The big disadvantage of Boolean scales is their exponential growth which restricts their applicability to at most five attributes. But if there are large dependencies between the attributes in the data, then more attributes may fit into a scale. DOKUANA is a tool which automatically extracts the structure of such a scale from the database. It is the task of the knowledge engineer and the domain expert to find a suitable grouping of the attributes.

Updates in the database may provide new combinations of the attributes and hence require to extend the scale by new concepts. Small changes (up to ten new concepts) can be handled automatically by TOSCANA; larger changes cause a warning message, and the scale has to be regenerated with DOKUANA.

In the data-driven design one does not refer to formerly established abstract scales. However, further research may use them for supporting the layout of data-driven scales. As long as satisfying automatic layout algorithms for line diagrams for concept lattices do not exist, one could search the existing abstract scales for a similar scale with a “readable” layout.

DOKUANA also allows the use of knowledge about hyponomy (i. e., subconcept–superconcept relation) between the database attributes. This knowledge might be derived from a thesaurus or a conceptual taxonomy. Then the resulting data-driven scales respect these relationships.

3 Conclusion

We considered three different levels of reuse in the development process of TOSCANA Systems:

1. The management system TOSCANA is a domain independent universal tool, and can be used without being adapted or reconfigured.
2. Abstract scales support the reuse of conceptual hierarchies in different domains.
3. Knowledge provided by thesauri and conceptual taxonomies can be considered for the data-driven generation of conceptual scales.

The use of the generic model of concept lattices has the advantage of being applicable in almost every domain. This universality is obtained by paying the price of not using more refined knowledge representation techniques and more specific visualization tools which might be better fitted to specialised applications. Therefore, a future research topic is to cautiously extend the knowledge representation and visualization techniques to often used data types.

References

1. Grosskopf, A., Harras, G.,*Eine TOSCANA-Anwendung für Sprechaktverben des Deutschen.* In: G. Stumme and R. Wille (eds.): *Begriffliche Wissensverarbeitung: Methoden und Anwendungen.* Springer, Berlin-Heidelberg (in preparation)
2. Kohler-Koch, B., Vogt F., *Normen und regelgeleitete internationale Kooperationen.* FB4-Preprint 1632, TU Darmstadt, 1994.
3. Kollewe, W., Sander, C., Schmiede, R., Wille, R., TOSCANA als Instrument der der bibliothekarischen Sacherschließung. In: H. Havekost and H.J. Wätjen (eds.), *Aufbau und Erschließung begrifflicher Datenbanken.* (BIS)-Verlag, Oldenburg, 1995, 95-114.
4. Kollewe, W., Skorsky, M., Vogt, F., Wille, R., TOSCANA — ein Werkzeug zur begrifflichen Analyse und Erkundung von Daten. In: R. Wille and M. Zickwolff (eds.), *Begriffliche Wissensverarbeitung – Grundfragen und Aufgaben.* B.I.-Wissenschaftsverlag, Mannheim, 1994, 267-288.
5. Mackensen, K., Wille, U., *Qualitative Text Analysis Supported by Conceptual Data Systems.* Preprint, ZUMA, Mannheim, 1997.
6. Rock, T., Wille, R., Ein TOSCANA-System zur Literatursuche. In: G. Stumme and R. Wille (eds.): *Begriffliche Wissensverarbeitung: Methoden und Anwendungen.* Springer, Berlin-Heidelberg (in preparation)
7. Scheich, P., Skorsky, M., Vogt, F., Wachter, C., Wille, R., Conceptual Data Systems. In: O. Opitz, B. Lausen, R. Klar (eds.): *Information and Classification.* Springer, Berlin-Heidelberg, 1993, 72–84.
8. Stumme, G.: On-Line Analytical Processing with Conceptual Information Systems. *Proc. of the 5th Intl. Conf. of Foundations of Data Organization.* Kobe, November 12-14, Kluwer 1999 (to appear)
9. Stumme, G., Wille, R., Wille, U.: Conceptual Knowledge Discovery in Databases Using Formal Concept Analysis Methods. In: J. M. Żytkow, M. Quafofou (eds.): *Principles of Data Mining and Knowledge Discovery.* Proc. 2nd European Symposium, PKDD '98, LNAI 1510, Springer, Heidelberg 1998, 450–458
10. Stumme, G., Wolff, K.E., Computing in Conceptual Data Systems with Relational Structures. In In: *Proc. of KRUSE '97.* Vancouver, August 11–13, 1997, 206–219
11. Vogel, N., *Ein begriffliches Erkundungssystem für Rohrleitungen.* Diplomarbeit, TU Darmstadt, 1995.
12. Vogt, V.: *Datenstrukturen und Algorithmen zur Formalen Begriffsanalyse: Eine C++-Klassenbibliothek.* Springer, Heidelberg 1996
13. Vogt, F., Wille, R., TOSCANA — A Graphical Tool for Analyzing and Exploring Data. In R. Tamassia, I. G. Tollis (eds.): *Graph Drawing '94.* LNCS **894**. Springer, Berlin-Heidelberg, 1995, 226–233.

14. Wille, R., Begriffliche Datensysteme als Werkzeug der Wissenskommunikation. In H. H. Zimmermann, H.-D. Luckhardt, A. Schulz (eds.): *Mensch und Maschine – Informationelle Schnittstellen der Kommunikation.* Univ.-Verl. Konstanz, 1992, 63–73.
15. Wille, R., Conceptual Landscapes of Knowledge: A Pragmatic Paradigm for Knowledge Processing. In: *Proc. of KRUSE '97.* Vancouver, August 11–13, 1997, 2–13.