

# Using Subgroup Mining for the Refinement of Knowledge Systems

Martin Atzmueller<sup>1</sup>, Joachim Baumeister<sup>1</sup>, Achim Hemsing<sup>2</sup>, Ernst-Jürgen Richter<sup>2</sup>, and Frank Puppe<sup>1</sup>

<sup>1</sup> Institut für Informatik, Univ. Würzburg, D-97074 Würzburg

<sup>2</sup> Poliklinik für Zahnärztliche Prothetik, Univ. Würzburg, D-97070 Würzburg

**Abstract.** The development of knowledge systems is still a complex and difficult task. When systems are used in a real-world environment the profiling and maintenance of the knowledge is a crucial success factor for the effectiveness of such systems, e.g., as discussed in Atzmueller et al. (2003), and Atzmueller et al. (2005).

In this paper we introduce a novel approach based on data mining techniques for refining knowledge systems. The presented method provides a semi-automatic process for improving the formalized knowledge using existing experiences with the system. Subgroup mining, c.f., Kloesgen (2002), is a method to discover 'interesting' subgroups of individuals concerning a certain target property of interest, e.g., subgroups with a high share of 'incorrectly' solved cases.

A knowledge-intensive subgroup mining approach, c.f., Atzmueller et al. (2004), is applied to identify causes of incorrect behavior of the knowledge base. Subgroup mining is especially applicable to this task, because it can be used to discover *local* patterns, i.e., sets of attribute-value pairs, which describe subgroups of erroneous sets of cases. Then, the local patterns can be improved first instead of trying to derive a global model for the refinement task, which may be quite difficult. Furthermore, in a subsequent refinement step proposals for corrections, i.e., refinements of the knowledge, can be applied. In a semi-automatic approach the expert can apply the suggested corrections step by step, using the subgroup mining results. If the refinements were helpful, then the mining and refinement process continues with a new iteration focussing on the new situation of the system. We provide a case study of the presented work with a fielded system in the medical domain.

## References

- ATZMUELLER, M., BAUMEISTER, J. and PUPPE, F. (2003): Evaluation of two Strategies for Case-Based Diagnosis handling Multiple Faults In: *Proc. 2nd Conf. Professional Knowledge Management (WM2003)*, Luzern, Switzerland
- ATZMUELLER, M., BAUMEISTER, J., HEMSING, A., RICHTER, E.-J. and PUPPE, F. (2005): Profiling Examiners using Intelligent Subgroup Mining In: *Proc. 10th Intl. Workshop on Intelligent Data Analysis in Medicine and Pharmacology*, 46–51, Aberdeen, Scotland.
- ATZMUELLER, M., PUPPE, F. and BUSCHER, H.-P. (2005): Subgroup Mining for Interactive Knowledge Refinement In: *Proc. 10th Conference on Artificial Intelligence in Medicine*, LNAI 3581, 453-462 Springer, Berlin, Germany.
- ATZMUELLER, M., PUPPE, F. and BUSCHER, H.-P. (2004): Towards Knowledge-Intensive Subgroup Discovery In: *Proc. LWA 2004, FGML*, Berlin
- KLOESGEN, W. (2002): Subgroup Discovery. In: *Handbook of Data Mining and Knowledge Discovery*. Oxford University Press, New York. Chapter 16.3