Semantic Knowledge Engineering – Main Concepts

Grzegorz J. Nalepa
AGH University of Science and Technology, Department of Automatics

Abstract: The Semantic Knowledge Engineering approach aims at providing new design and analysis methods for rule-based intelligent systems. It uses the XTT2 knowledge representation for building modularized rule bases that form decision networks. The representation is formalized, thus allowing for the anayslys of the designed system with respect to its qualitative properties. The visual design is supported by practical tools.

Keywords: knowledge-based systems, rules, reasoning, semantic knowledge engineering

Building intelligent systems (IS) is the main goal of Artificial Intelligence [1]. *Rule-based systems* [2] (RBS) are an important class of IS that is oriented towards decision support and intelligent control. There are number of RBS tools available, including CLIPS, Jess, or recently Drools. They found number of appliactions in decision support and business systems.

However, the practice of building such systems has not been fully mastered. In fact, effective design of RBS [3] still faces important problems. They become apparent in case of large rule bases. In such cases, maintaing appropriate quality of the knowledge base as well as providing effective knowledge processing (reasoning) algorithms is hard.

The *main issues* in the design of RBS are as follows [4]: 1) systematic design process, 2) knowledge acquisition facilities, 3) modularization of the knowledge base, 4) quality assurence procedures, 5) scalable inference control mechanisms. Existing methods tools approach only some of these problems. The approach discussed here is called Semantic Knowledge Engineering (SKE). It aims at addressing these general problems in a coherent manner.

1. Concepts of SKE

In a recent book [5] a sythesis of extensive studies with respect to state-of-the-art of rule-based representations has been presented. Based on them, the following important *problem areas* constitute a field open for research in RBS:

1) Informal and sparse knowledge representation,
2) Blind inference algorithms,
3) Unstructured design approach, limited quality analysis,
4) Lack of integration on the semantic level.

The first problem is related to the fact that in existing systems rules are mostly informally described, with no formal semantics. Moreover, single rules have low knowledge processing capabilities. Therefore, for practical applications a higher level of abstraction is needed.

The second problem stems from the observation, that commonly used rule inference algorithms assume a flat rule base, with no internal structure. In such a case a brute-force (blind) search for solution is used. However, numerous practical applications are goal-oriented.

The third problem is the lack of a widely available practical methodology for consecutive development of RBS which would be acceptable to engineers. Moreover, most of existing design solutions consider system quality analysis only as a small step in the system life-cycle, not as an integral part of the design procedure. This often makes the design of high quality RBS a challenging task.

The last problem is related to the observation that the rule-based approach needs to be aligned with other design and implementation methods and tools on the programming level. In todays software engineering practice, these are not conceptually compatible with rule-specific semantics. Because of this incompatibility, the rule technologies can not be integrated directly with software systems.

In the context of these problems, the main motivation for the *Semantic Knowledge Engineering* approach presented in [5] was to provide a coherent formal notation for a rule-based knowledge representation language that can be used is various domains. The formalization of the language includes its syntax and semantics. Moreover, practical translations of this representation on the semantic level were provided. The SKE approach is based on sveral important concepts briefly presented next.

The approach addresses the first issue by providing a *formalization of the rule language* [6]. The formalization uses the ALSV(FD) (Attributive Logic with Set Values over Finite Domains) logic [7]. It is more expressive than the classic (mostly propositional) rule languages, e.g. it allows for formal specification of non-atomic values in rule conditions. Moreover, a *custom rule representation* formalism called XTT2 (eXtended Tabular Trees Version 2) [7, 8] is introduced. It is based on the the ALSV(FD). It ensures high density and transparency of visual knowledge representation. This representation combines decision trees and decision tables. XTT2 forms a transparent and hierarchical visual structure composed of the decision tables linked into an *inference network*. This solution addresses the issue of low processing capabilities of single rules.

Efficient inference in the XTT2 network is provided by *custom inference methods*. Only the rules necessary for achieving the goal are fired by selecting the desired output tables and identifying the tables necessary to be fired first. Multiple inference modes are provided, including forward-chaining and goal-driven ones [9].

The SKE approach is based on a *systematic design procedure*. It covers all of the main phases of the system life-cycle, from the conceptual design, through the logical formulation, to the physical implementation. It is a top-down design methodology based on successive refinement of RBS. The procedure involves constant verification of the system model with respect to important formal properties.

The approach also includes *quality analysis solutions* during the design. What is important for RBS is their logical description. The formal rule description in ALSV(FD) allows for the use of formal quality properties tests of the rule base. Thanks to the logical formulation, such an analysis should be an integral part of the design [10].

SKE is oriented towards a *practical integration of rule-based approach* with important existing methods to software design. So far, not much has been done in the area of integrating the semantics of rule-based and object-oriented methods. More recently the Semantic Web initiative [11] set ambitious goals of developing rule reasoning for an intelligent Web. However, the semantics of logical methods that are used (ontologies) is incompatible with rules. The SKE approach proposes integration solutions for both object-oriented in [12] as well as Semantic Web design methods on the semantic level [13–15].

2. Tool Support

The original tool for designing rule bases with XTT2 was presented in [16]. Currently, the HQEd (HeKatE Qt Editor) [17] provides support for the logical design with XTT2. It allows for editing the XTT2 structure with support for syntax checking at the table level.

HQEd can automatically generate the textual executable representation of the XTT2 knowledge base in the textual HMR format. HMR files can be directly executed by a custom inference engine called HeaRT (HeKatE RunTime) [18]. The role of the engine is to run the XTT2 rule logic designed with the use of the editor, and provide on-line formal analysis of the rule base using the HalVA (HeKatE Verification and Analysis) framework [10]. The engine supports multiple inference modes [9], including forward chaining and goal-oriented.

HaThoR (HeKatE Translation Framework for Rules) [19] provides rule import and export modules for other languages and formats, including Semantic Web languages as well as Drools [20]. It is mainly implemented in XSLT with certain extra plugins integrated with HeaRT implemented in Prolog. An experimental module allows for the translation of visual XTT2 representation to a dedicated UML representation using an XMI-based (XML Metadata Interchange) serialization.

3. Future Work

Several directions for the future development of the Semantic Knowledge Engineering approach are considered. A number of extensions of the XTT2 rule language can be proposed. This include temporal formalisms, as well as support for fuzzy rules. Moreover, semi-automatic optimizations of the structure of decision tables are also considered. The inference algorithms can be further modified and extended.

Concerning the verification of the XTT2 rule bases, some other procedures are also evaluated, such as automatic minimization of a rule base. Research on global verification of the model is in progress.

Finally, a semantic translation of the XTT2 model to the BPMN (Business Process Model and Notation) [21] representation is being prepared [22]. It should allow for an integration method for Business Rules and Processes using the Semantic Knowledge Engineering approach.

Acknowledgements

Research supported by the PARNAS (NCN N516 481240) project.

Bibliography

1. Tadeusiewicz R. (2011): *Introduction to intelligent systems*, [in:] Wilamowski B.M., Irwin J. D. (eds.): *Intelligent systems, The Electrical Engineering Handbook Series*. The Industrial Electronics Handbook, 1–1–1–12, Boca Raton, London, New York, CRC Press Taylor & Francis Group, second edition.
2. Giurca A., Gasevic D., Taveter K.(eds.) (2009): *Handbook of Research on Emerging Rule-Based Languages and Technologies: Open Solutions and Approaches*, Information Science Reference, Hershey, New York.
3. Nalepa G.J. (2009): *Languages and Tools for Rule Modeling*, [in:] Giurca A., Gasevic D., Taveter K. (eds.): *Handbook of Research on Emerging Rule-Based Languages and Technologies: Open Solutions and Approaches*, IGI Global, Hershey, New York, 596–624.
4. Ligęza A., Nalepa G.J. (2011): *A study of methodological issues in design and development of rule-based systems: proposal of a new approach*, "Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery" 1(2), 117–137.
5. Nalepa G.J. (2011): *Semantic Knowledge Engineering. A Rule-Based Approach*, Wydawnictwa AGH, Kraków.
6. Nalepa G., Ligęza A., Kaczor K. (2011): *Overview of Knowledge Formalization with XTT2 Rules*, [in:] Bassiliades N., Governatori G., Paschke A. (eds.): *Rule-Based Reasoning, Programming and Applications*, "Lecture Notes in Computer Science" volume 6826, Springer Berlin/Heidelberg, 329–336.
7. Nalepa G.J., Ligęza A. (2010): *HeKatE Methodology, Hybrid Engineering of Intelligent Systems*, "International Journal of Applied Mathematics and Computer Science" 20(1), 35–53.
8. Nalepa G.J., Ligęza A., Kaczor K. (2011a): *Formalization and Modeling of Rules Using the XTT2 Method*, "International Journal on Artificial Intelligence Tools" [in press].
9. Nalepa G., Bobek S., Ligęza A., Kaczor K. (2011b): *Algorithms for Rule Inference in Modularized Rule Bases*, [in:] Bassiliades N., Governatori G., Paschke A. (eds.): *Rule-Based Reasoning, Programming and Applications*, "Lecture Notes in Computer Science" volume 6826, Springer Berlin/Heidelberg, 305–312.

10. Nalepa G., Bobek S., Ligęza A., Kaczor K. (2011c): *HalVA - Rule Analysis Framework for XTT2 Rules*, [in:] Bassiliades N., Governatori G., Paschke A. (eds.): *Rule-Based Reasoning, Programming and Applications*, "Lecture Notes in Computer Science" volume 6826, Springer Berlin/Heidelberg, 337–344.
11. Hitzler P., Krötzsch M., Rudolph S. (2009): *Foundations of Semantic Web Technologies*, Chapman & Hall/CRC.
12. Nalepa G.J. (2009): *XTT Rules Design and Implementation with Object-Oriented Methods*, [in:] Lane H.C., Guesgen H.W. (eds.): FLAIRS-22: Proceedings of the twenty-second international Florida Artificial Intelligence Research Society conference, 19–21 May 2009, Sanibel Island, Florida, USA, FLAIRS, AAAI Press, Menlo Park, California, 390–395.
13. Nalepa G.J., Furmańska W.T. (2010): *Integration Proposal for Description Logic and Attributive Logic – Towards Semantic Web Rules*, [in:] Nguyen N.T., Kowalczyk R. (eds.): Transactions on Computational Collective Intelligence II, "Lecture Notes in Computer Science" volume 6450, Springer Berlin/Heidelberg, 1–23.
14. Nalepa G.J. (2010): *Collective Knowledge Engineering with Semantic Wikis*, "Journal of Universal Computer Science" 16(7), 1006–1023.
15. Nalepa G.J. (2011): *Loki – Semantic Wiki with Logical Knowledge Representation*, [in:] Nguyen N.T. (ed.): Transactions on Computational Collective Intelligence III, "Lecture Notes in Computer Science" volume 6560, Springer, 96–114.
16. Nalepa G.J., Ligęza A. (2005): *A visual edition tool for design and verification of knowledge in rule-based systems*, "Systems Science" 31(3), 103–109.
17. Kaczor K., Nalepa G.J. (2009): *Extensible design and verification enviroment for XTT rule bases*, [in:] Tadeusiewicz R., Ligęza A., Mitkowski W., Szymkat M. (eds.): CMS'09: Computer Methods and Systems: 7th conference, 26–27 November 2009, Kraków, Poland, AGH University of Science and Technology, Cracow, Oprogramowanie Naukowo-Techniczne.
18. Nalepa G.J. (2010): *Architecture of the HeaRT Hybrid Rule Engine*, [in:] Rutkowski L. et al. (eds.): Artificial Intelligence and Soft Computing: 10th International Conference, ICAISC 2010, Zakopane, Poland, June 13–17, 2010, Pt. II, "Lecture Notes in Artificial Intelligence" volume 6114, Springer, 598–605.
19. Nalepa G.J., Ligęza A., Kaczor K., Furmańska W.T. (2009): *HeKatE Rule Runtime and Design Framework*, [in:] Giurca A., Nalepa G.J., Wagner G. (eds.): Proceedings of the 3rd East European Workshop on Rule-Based Applications (RuleApps 2009) Cottbus, Germany, September 21, 2009, 21–30.
20. Kluza K., Nalepa G.J., Łysik Ł. (2010): *Visual Inference Specification Methods for Modularized Rulebases. Overview and Integration Proposal*, [in:] Nalepa G.J., Baumeister J. (eds.): 6th Workshop on Knowledge Engineering and Software Engineering (KESE2009) at the 32nd German conference on Artificial Intelligence, September 21, 2010, Karlsruhe, Germany, 6–17.
21. OMG (2011): *Business Process Model and Notation (BPMN): Version 2.0 Specification*, Technical Report formal/2011-01-03, Object Management Group.
22. Kluza K., Maślanka T., Nalepa G.J., Ligęza A. (2011): *Proposal of Representing BPMN Diagrams with XTT2-based Business Rules*, [in:] Brazier F.M., Nieuwenhuis K., Pavlin G., Warnier M., Badica C. (eds.): Intelligent Distributed Computing V. Proceedings of the 5th International Symposium on Intelligent Distributed Computing – IDC 2011, Delft, Netherlands – October 2011, "Studies in Computational Intelligence" volume 382, Springer.

Podstawowe założenia podejścia semantycznej inżynierii wiedzy

Streszczenie: Celem podejścia semantycznej inżynierii wiedzy jest dostarczenie nowych metod projektowania i analizy systemów inteligentnych wykorzystujących reprezentację regułową. Podejście to bazuje na metodzie XTT2 służącej do budowania na poziomie logczinym zmodularyzowanej bazy reguł stanowiącej sieć wnioskującą. Metoda ta jest sformalizowana, co pozwala na przeprowadzenie analizy systemu pod kątem jego jakości.

Słowa kluczowe: systemy z bazą wiedzy, systemy regułowe

Grzegorz J. Nalepa, PhD

He holds a position of assistant professor in AGH UST in Krakow, Poland, Department of Automatics. Since 1995 he has been actively involved in number of research projects, including Regulus, Mirella, Adder, HeKatE, and recently INDECT and BIMLOQ. In 2007-9 he coordinated the HeKatE project (hekate.ia.agh.edu.pl). Since 2010 he has been cooridinating the BIMLOQ project (bimloq.ia.agh.edu.pl). He is the author of over one hundred papers from the domains of knowledge and software engineering, and intelligent systems. His fields of interest also include computer security and operating systems. He formulated a new design and implementation approach for rule-based systems called XTT (eXtended Tabular Trees). He is involved many conferences and workshops. Since 2008 he's been co-organizing the Knowledge and Software Engineering Workshop (KESE) at KI, the German AI conference. He is the President of the Polish Artificial Intelligence Society (PSSI). In 2011 he published a monograph "Semantic Knowledge Engineering. A Rule-Based Approach".

e-mail: gjn@agh.edu.pl