Advanced Tools for Knowledge Model Construction: Lessons Learned from an Intelligent System for Emergency Management

Martin Molina, Gemma Blasco
Department of Artificial Intelligence
Universidad Politecnica de Madrid, Spain
mmolina@fi.upm.es,gblasco@isys.dia.fi.upm.es

Abstract. This paper shows general conclusions about advanced general tools for knowledge modeling, derived from our experience in the construction and maintenance of a complex real-world intelligent system for emergency management called SAIDA. The paper first describes the characteristics of the knowledge model of SAIDA and summarizes the problems that we found during the development of such type of model. The paper describes the CAM-Hidro tool, a software application that we have designed to help in this task to be used by users who are not experts in computer programming. Finally, the paper shows a generalization of the features derived from this experience to be provided by tools for model construction and shows a comparison of existing approaches, both from knowledge acquisition tools and from the recent approaches for knowledge sharing.

1 Introduction

One of the key factors in the development of intelligent systems is the *knowledge acquisition* process. This is an area in which historically AI researchers have pay attention and which has evolved together with the advances in knowledge representation. Since the proposal of *knowledge level* [Newell, 82] the acquisition process is considered as a modeling task, which is a more appropriate view that also takes the advantage of reusing abstract models to guide the development of new applications.

However, despite that the recent advances for knowledge acquisition have facilitated this process significantly, the current experience shows that there is still an important gap between the way end-users describe their expertise and the existing tools. In particular, this issue has received recently attention from AI researchers in the context of web-based applications. Thus, the need of knowledge development tools usable by non-experts in knowledge engineering has been recently exposed within the *semantic web* context as one of the challenges for the twenty-first century AI research [Hendler, Feigenbaum, 01].

According to this need, this paper presents a contribution in this direction based on our recent experience in the development and maintenance of a complex real-world model for an intelligent system (with more than 140 knowledge bases). This paper analyzes the services that a software tool should provide to help in the development of such a model by exploring the characteristics of the hydrologic model for emergency management. The paper also presents a particular software tool called CAM-Hidro that we designed to facilitate the development of hydrologic models. Finally, the paper summarizes the set of features that should be included in this type of software tools and describes a comparison of the existing general approaches.

2 General view of the knowledge model for decision support

This section summarizes the characteristics of the knowledge model of the SAIDA system, which was developed to provide assistance in making decisions about hydraulic actions during floods. This model is a good example of the complexity that a real-world model may present and will be analyzed later from the point of view of the development and maintenance.

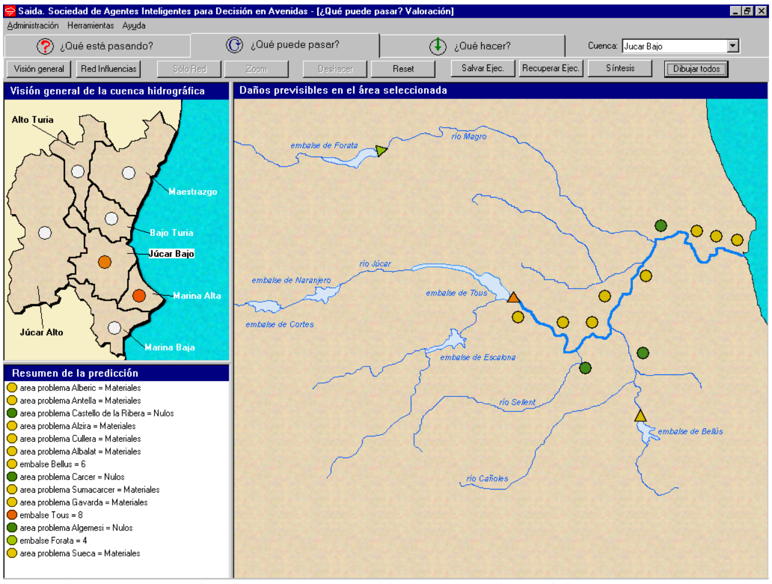

Figure 1: Example of screen presented by user interface of the SAIDA system

The SAIDA (Spanish acronym for Intelligent Agents Society for Decision-making during Floods) system is a computerized system based on artificial intelligence techniques that provides assistance in flash flood situations for basin control centers [Cuena, Molina, 99] [Molina, Garrote, 02]. The system was developed in a project developed during more than three years promoted by the Spanish Ministry of the Environment with the purpose of having it installed and used in connection with the information hydrologic systems in several Spanish basins. SAIDA receives as input the available data provided by sensors about discharge, water level and rainfall at different locations in the river basin. The answers are produced with time constraints and the conclusions are justified at a reasonable level of abstraction given that the operator must take the final responsibility of decisions. Figure 1 shows an example of screen presented by the SAIDA user interface. This interface provides answers to the following types of questions: (1) what is happening? (2) what may happen in the future?, and (3) what can be done?. With SAIDA he operator can quickly understand the current situation, identify the main problems that have to be solved and is briefed on the actions that could be taken to reduce the problems and minimize the risks.

The design of the SAIDA knowledge base followed *model-based* approach according to the modern knowledge engineering methodologies. This approach provided patterns of reasoning methods that were useful as building blocks for the development of the application. To implement such a model, the KSM environment was used [Cuena, Molina, 00] following and a methodology that use some concepts similar to CommonKADS [Schreiber et al. 00]. SAIDA also followed a multiagent approach to facilitate the development of the complex knowledge base. Figure 2 shows the basic structure of the knowledge model of SAIDA according to the three main tasks provided by the system: (1) evaluation, (2) prediction and (3) recommendation. The goal of the evaluation task is to determine the level of severity of the current situation by identifying problematic scenarios. This goal corresponds to a classification task that from a set of observations (partial information and data with noise from sensors) selects a category (a type of problem) within a prefixed set of categories. Here, we applied a version of the *heuristic classification* method [Clancey, 85] with extensions to cope with spatial distribution and temporal dimension.

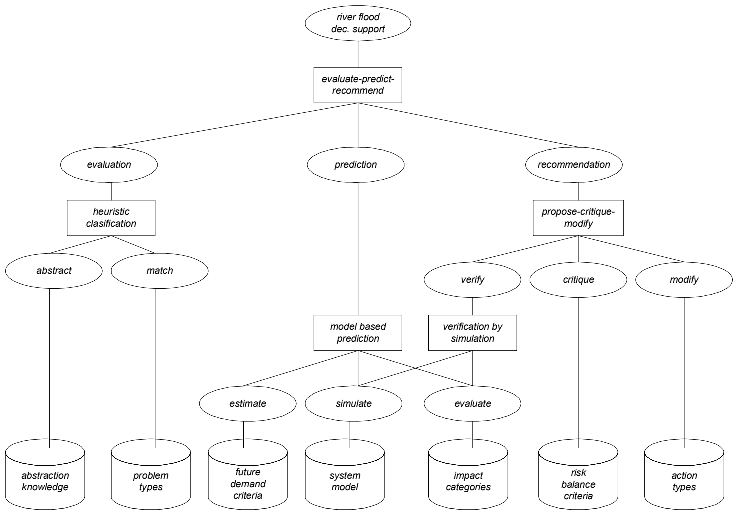

Figure 2: The basic structure of the model for decision support in the field of emergency management during floods (circle = task, square = method, cylinder = type of knowledge base).

The goal of the prediction task is to estimate the expected damages that are consequence of the current hydrologic situation. This task receives as input the result of the evaluation task, together with the recent behavior recorded by sensors. The prediction task is performed with the following steps: (1) *estimate future rain,* which generates hypotheses of future rain based on heuristic knowledge (2) s*imulate the river behavior,* which uses a model of the river basin using a set of bayesian networks [Pearl, 88], (3) *estimate potential damages* by using empirical knowledge that relates water levels and flows with qualitative ranges of severity for each particular location.

The recommendation task distinguishes between two possibilities: (1) hydraulic actions, in order to find reasonable discharge policies at the dams to avoid undesirable impacts and (2) defensive actions, to protect the problematic areas when it is not possible to avoid the flood such as population alert, evacuation procedures, etc. The task is performed by a method that can be viewed as a particular of *propose-critique-modify* strategy for design tasks [Brown, Chandrasekaran, 89]. The method proposes a set of hydraulic actions that potentially could solve the problem, then, these actions are tested by simulation, and modified with additional knowledge in a loop until a satisfactory set of control actions together with defensive actions are found

The distributed nature of the decisions and the spatial location of certain components makes very appropriate using the multiagent approach as a complementary design approach to organize the knowledge model. Within each type of agent, the knowledge bases were adequately organized and implemented by using additional modular approaches (details about this solution can be found at [Molina, Cuena, 02]). According to different types of decisions, we identified four types of agents: (1) *hydraulic agents* that are responsible to give answers about the behavior of the physical process, (2) *problem detection agents*, responsible of evaluating the flood risk in a particular geographical area, (3) *reservoir management agents*, which contain criteria for exploitation strategy for each reservoir, and (4) *civil protection agents,* responsible to provide with resources of different types according to the needs of the problem detection agents. For each type of agent, there are several instances according to the geography of the river basin. The main three tasks (evaluation, prediction and recommendation) were distributed among the different agents, in such a way that they communicate partial results to complete their individual goals.

Agents	**N. of agents**	**Knowledge Bases**	**Knowledge Representation**	**N. of KBs**
Problem detection agents	15	Abstraction	Functional + temporal represent.	15
		Problem types	Frames with uncertainty degrees	15
		Future Demand	Rules	15
		Impact categories	Bayesian network	15
		Risk balance criteria	Rules	1
		Action types: agent relations	Horn Logic Clauses	15
Reservoir management agents	4	Abstraction	Functional + temporal represent.	4
		Problem types	Frames with uncertainty degrees	4
		Future Demand	Rules	4
		Impact categories	Bayesian nets	4
		Risk balance criteria	Rules	1
		Action types: discharge strategies	Rules	1
Hydraulic agents	2	Abstraction	Functional + temporal represent.	2
		System model: influence diagram	Temporal causal network	2
		System model: infiltration	Bayesian network	12
		System model: discharge	Bayesian network	12
		System model: reservoir discharge	Bayesian network	4
		System model: junction	Bayesian network	11
Protection agents	2	Action types: transport network	Rules	2
		Action types: population	Rules	2
		Action types: constructions	Rules	2
TOTAL	23 Ag.		TOTAL	143 KBs

Figure 3: Summary of knowledge bases corresponding to a particular model for a river basin in the case of the Júcar River Basin (East of Spain).

A model for a particular river basin is constructed formulating a set of knowledge bases. Figure 3 shows a summary of a complete model for the case of a river basin in Spain (the Júcar river basin). This includes a total of 23 agents, one for each specific decision point at certain location in the river basin depending on its nature (problem area, reservoir, river channel or protection). For each agent, there is a set of types of knowledge bases, each one with its particular language representation, with a total of 143 knowledge bases.

3 CAM-Hidro: A software tool for knowledge model construction

According to the previous description, the characteristics of a knowledge model for hydrologic decision presents a significant complexity, which is inherent to the physical phenomena in which the decision is based. The following list summarizes the problems reported by users that were responsible of model construction:

- *Dimension and complexity*. The first problem to manage the knowledge model is that it presents a high level of complexity with different interrelated types of knowledge for different purposes (e.g., 143 KBs for the Júcar river model).
- *Heterogeneity of symbolic representation*. Each type of knowledge base has its own symbolic representation (frames, rules, uncertainty, temporal and spatial dimensions, etc.). Despite they are based on natural and declarative representation, this factor increases the difficulty of understanding the complete model.
- *Low level of certain representations.* For certain types of knowledge, civil engineers use certain common sense usual in their professional area, but the corresponding knowledge base may use low level representations to represent such a knowledge with excessive detail about implicit terms that makes the model more artificial and difficult to understand.
- *Abstract computer-oriented terminology*. The SAIDA model follows a general methodology based on the KSM tool. This introduces an additional terminology closer to process information, different from hydrology, that sometimes is too abstract for end-users and increases the difficulty to understand the complete model.
- *Consistency between modules*. The model presents a distributed organization of knowledge, following a multiagent architecture and, for each agent there is a set of knowledge bases with different inference procedures. This modular organization makes easier to understand and validate parts of the whole model but, still, sometimes it is difficult anticipate the dependencies between such modules in order to keep the global consistency.
- *False idea of procedural representation*. The edition of the content of knowledge bases uses text processors with the corresponding language for each case. We found that this may give the false idea that the user writes a kind of procedure (following a conventional programming style) instead of a set of expertise criteria with a declarative approach with more freedom to add or remove sentences.
- *Different procedures for knowledge acquisition*. Another problem is that the user must combine different non-integrated software tools to cope with different sources of knowledge. Thus, for example, certain knowledge can be manually represented using symbolic formalisms, but another type of knowledge can be learned with machine learning procedures.
- *Low level of guidance*. The user has the possibility of editing and modifying any part of the model with certain freedom. However, this freedom should be complemented with certain guidance in the model development, to suggest to the developer about what are the next steps to be done.

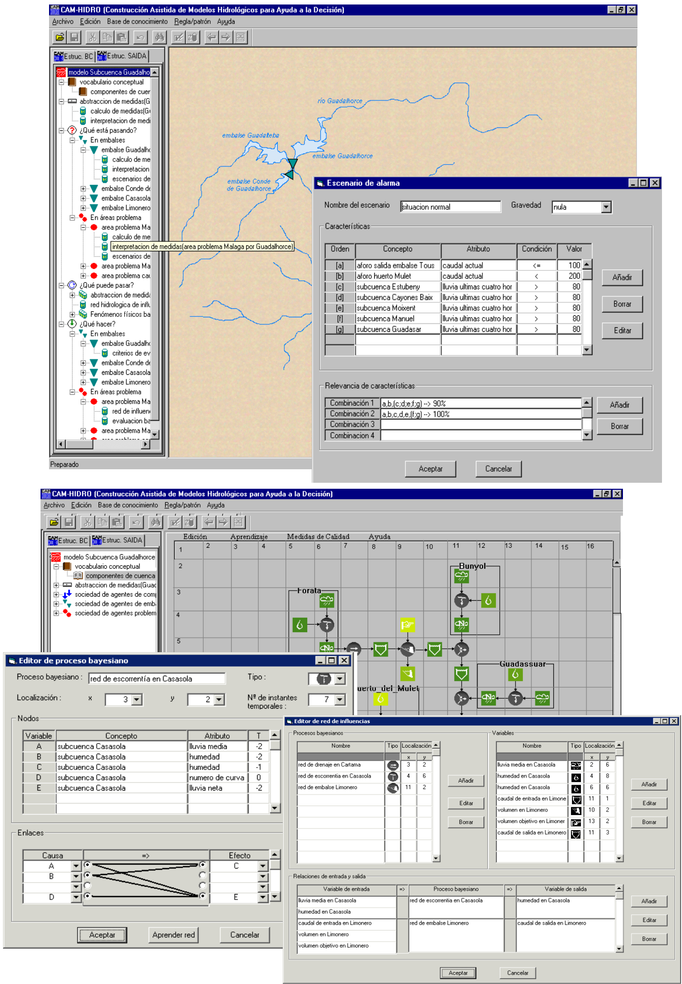

Figure 4: Example of windows presented by the CAM-Hidro user interface.

In order to give an answer to the previous needs, we have designed a software tool called CAM-Hidro that assists developers in the construction of distributed hydrologic models for the SAIDA system. CAM-Hidro integrates three main components: (i) a user interface, (ii) procedures for edition support and (iii) procedures for machine learning. The first component, the user interface, is oriented to present a user-friendly image of the model together with visual facilities for edition. The features of this interface include: (1) user interaction based on edition standards (the user of this application is an expert in hydrology who is used to apply conventional software tools generalized within the professional area of civil engineers such as databases, graphical editors, formula-based editors, spreadsheets, geographical information systems, etc.), (2) communication language using hydrologic terminology with the minimum set of abstract computer-based concepts together with the management of common representations in civil engineering such as maps, tables, histograms, time series, math formula, etc., (3) multiple complementary views of the model at different levels of abstraction showing also the role that the knowledge play in the inference process, (4) consistency checking to verify local and global coherence, and (5) edition guide to help the user to decide what are the next steps during the model construction process. Figure 4 shows an example of screens presented by the user interface of CAM-Hidro.

The second component is a set of automatic procedures to support the edition process. In particular, this includes processes to interpret and translate the domain-oriented languages to general representations such as rules, frames, etc. There are also processes for consistency checking between different knowledge bases and procedures for edition assistance in order to guide the user through the complete process of development.

Finally, the third component includes the set of procedures for machine learning corresponding to the abstract river basin model. This model uses a set of bayesian networks extended for different time steps. The development of this model is based on an automatic process that starts from a basic general structure given by the user that is refined in an iterative process that uses cases generated by simulation. A statistical process abstracts the set of cases and a heuristic module evaluates the quality of the model and decides to produce new cases by simulation until the model reaches an acceptable level of quality.

4 Discussion

This section summarizes a set of general requirements of general tools for knowledge model construction and includes a comparison of existing tools based on these requirements. Basically this type of software tools provide the following three services: (1) *user communication*, with a user interface with visual media and edition facilities for model construction, (2) *construction assistance*, which means that the software tool includes knowledge about the development process of the model, and (3) o*perational translation*, to translate the model to a formal computational version tractable by inference engines. We have generalized a set of features to be included in this type of software tools based on our experience in the development of the SAIDA hydrologic model presented in this paper (see figure 5). We present here a comparison of the existing approaches based on these features. It shows that none of them totally satisfy these requirements, thus, in order to have complete solutions, it suggest that further research should be done in this direction. The comparison includes three types of the most advanced software tools for knowledge modeling: (1) method-based knowledge acquisition tools, (2) general knowledge modeling tools, and (3) the recent ontology management tools. We also include separately the case of the CAM-Hidro tool presented in this paper. Figure 6 shows a summary of the comparison of these approaches.

The category of method-based knowledge acquisition tools includes a type of software tool that assists in the development of a knowledge-based system for a prefixed kind of task and problem-solving method. Examples of such tools are MORE [Eshelman et al., 87] for diagnosis systems with the cover-and-differentiate method, or

SALT [Marcus, McDermott, 89] for design systems with the propose-and-revise method. This category also includes other more specific tools in certain domains such as SIRAH [Alonso et al., 90] for prediction tasks in hydrology. The advantages of method-based knowledge acquisition tools are derived from the fact that the organization of the knowledge is prefixed, so they have a good level of support for model construction and efficiency in the generation of the operational version. However the range of applicability is significant lower than other approaches.

Category	***Requirement***	***Comments***
User communication	*User language*	It is important to follow the language of the professional field of the user using certain communication primitives with which the user is familiar (maps, graphs, etc.).
	Abstraction	It is useful to present global views of the model based on categories at different levels of abstraction to facilitate a complete understanding.
	Explicit role	Information about the role that the knowledge plays in the global inference process facilitates the comprehension of the model.
	Edition standards	To make easy to use the application, it is important to follow the communication standards used in conventional software.
Construction assistance	*Local consistency*	It is important to include procedures to check the local consistency of each knowledge base using certain syntax and semantic properties of symbolic representation .
	Global consistency	It is useful to provide methods to check the consistency between categories of knowledge (e.g., consistency between different knowledge bases, agents, etc.).
	Guidance	Automatic processes should suggest what are the next steps to be done, providing acceptable levels of freedom.
	Professional common sense	The tool should share standard ontologies and procedures about usual concepts of the user terminology in different professional domains (hydrology, road traffic, etc.).
	Generality	The tool should be useful for the development of a wide range of specific applications with different domain knowledge.
	Integration	The tool may integrate manual edition with automatic machine learning procedures, keeping a homogeneous operation.
Operational translation	*Efficiency*	The model is translated to certain formal computational constructs that offer an acceptable level of efficiency to be tractable by inference engines.
	Reusability	The model is translated to general representations (rules, frames, constraints, etc.) usable by general inference procedures.

Figure 5: Requirements for model construction tools, generalized from the case of SAIDA.

The category of general knowledge modeling tools includes a type software tool that assists to the developer in the application of a modeling methodology. For example, MIKE [Angele et al., 93] follows the KADS methodology [Schreiber et al., 00] and allows a partial validation of the knowledge model using a computational language, KARL [Fensel et al, 91]. Other approaches such as KREST [Macintyre, 93] or KSM [Cuena, Molina, 00] also follow a modeling methodology somehow similar to KADS but, in addition to that, they produce the final operational version using preprogrammed constructs. Another interesting approach derived from the EXPECT system, takes advantage of the explicit representation of problem-solving methods to guide the knowledge acquisition process [Blythe, 01]. The knowledge modeling tools are more general compared to the previous approach because the developer can formulate any kind of problem solving method. However they introduce certain abstract terminology that can be difficult to be understood by users non-programmers.

The category of ontology management tools has been developed the recent years within the field of knowledge sharing and reuse, especially in the context of Internet. Examples of these tools are Protégé-2000 [Grosso et al., 99], WebOnto [Domingue, 98], OntoSaurus [ISX, 91], Ontolingua/Chimaera [Farquhar et al., 97], [McGuinness et al., 00]. In general, these tools are easier to be operated by users who are not expert in programming, compared to the knowledge modeling tools. They also provide an interesting solution to the need of having certain professional common sense, by reusing standard ontologies that have been previously formulated in different domains. In addition to that, they provide certain advanced services that facilitate knowledge sharing such as cooperative construction, merging assistance or internationalization. However, they follow general knowledge representations (frames, relations, production rules, etc.) that can be limited in certain complex domains such as the case of hydrology presented in this paper, and these tools are not able to show the role that the knowledge plays in inference processes.

		Method based KA tools	Knowledge modeling tools	Ontology Manag. Tools	CAM-Hidro
User communication	User language	+	-	0	+
	Abstraction	+	+	0	+
	Explicit role	+	+	-	+
	Edition standards	0	0	0	+
Construction assistance	Local consistency	+	0	+	+
	Global consistency	+	0	0	+
	Guidance	+	0	0	+
	Prof. common sense	0	-	+	0
	Generality	-	+	+	-
	Integration	0	-	-	+
Operational translation	Efficiency	+	0	0	+
	Reusability	0	+	+	0

Figure 6: Comparison of approaches of software tools for knowledge model construction.

Finally, the case of the CAM-Hidro software environment described in this paper presents some similarities to the first approach given that it follows a prefixed knowledge organization (decision support in hydrologic emergency management). However CAM-Hidro presents a higher level of complexity because it includes a number of complex tasks (diagnosis, prediction, planning) with a multiagent organization. It also presents a user interface that follows the visual standards oriented to the type of user of the application and good level of integration of manual and automatic procedures (machine learning) for model construction.

4. Conclusions

In summary, the paper shows general conclusions derived from our recent experience in the development and maintenance of a complex real-world knowledge model in the field of decision support for hydrologic emergency management. The paper presents the characteristics of the hydrologic model and identifies a set of problems for model construction. As a first answer to the need of assistance in the development and maintenance of such a model by users who are not programmers, the paper describes a software tool called CAM-Hidro that we designed for this purpose. The paper also presents a generalization of desired services to be provided by such type of tools and presents a comparison among the existing approaches in this field. The comparison shows that none of the existing approaches totally covers the needs, which suggests that further research should be done by integrating and extending the current solutions.

Acknowledgements. The development of the SAIDA system was supported by the Ministry of Environment of Spain (*Dir. General de Obras Hidráulicas y Calidad de las Aguas*) with the participation of local public organizations from river basins (*Conf. Hidrográfica del Júcar* and *Conf. Hidrográfica del Sur de España.)*. The development of the CAM-Hidro tool was supported by the Ministry of Science and Technology of Spain within the RIADA project.

References

Alonso M., Cuena J., Molina M."SIRAH: An Architecture for a Professional Intelligence", ECAI Conference 1990. (L.Carlucci Ed.) Pitman, 1990.

Angele J., Fensel D., Landes D., Neubert S., Studer R.: "Model-Based and Incremental Knowledge Engineering: The MIKE Approach" en *Knowledge Oriented Software Design*, J.Cuena (ed.). Elservier, 1993.

Blythe, J.; Kim J.; Ramachandran, S.; Gil Y.: "An Integrated Environment for Knowledge Acquisition". Proc. International Conference on Intelligent User Interfaces, 2001.

Brown D., Chandrasekaran B.: "Design Problem-solving: Knowledge Structures and Control Strategies", Morgan Kaufman, 1989.

Clancey W.J.: "Heuristic Classification". *Artificial Intelligence*, vol 27, pp. 289-350, 1985.

Cuena J., Molina M., "A Multiagent System for Emergency Management in Floods". En "Multiple Approaches to Intelligent Systems", *12th International Conference on Industrial Applications of Artificial Intelligence and Expert Systems, IEA/AIE-99*. El Cairo, Egipto 1999.

Cuena J., Molina M.: "The role of knowledge modelling techniques in software development: a general approach based on a knowledge management tool", *International Journal of Human-Computer Studies* No. 52. pp 385-421. Academic Press, 2000.

Domingue, J.: "Tadzebao and WebOnto: Discussing, Browsing and Editing Ontologies on the Web", Proc. of the *Eleventh Workshop on Knowledge Acquisition, Modeling and Management*. Banff, Canada, 1998.

Eshelman L., Ehret D., McDermott J., Tan M.: "MOLE: a Tenacious Knowledge-Acquisition Tool". Academic Press Inc., London, 1987.

Farquhar A., Fikes R., Rice J.: "The Ontolingua Server: a Tool for Collaborative Ontology Construction", *International Journal of Human-Computer Studies*, 46, 707-727, 1997.

Fensel, D.; Angele, J.; Landes D.: "KARL: A Knowledge Acquisition and Representation Language". Proc. Expert Systems and their Applications, Avignon 1991.

Grosso W.E., Eriksson H., Fergerson R.W., Gennari J.H., Tu S.W., Musen M.A.: "Knowledge Modeling at the Millennium (The Design and Evolution of Protege-2000)". *Twelfth Banff Workshop on Knowledge Acquisition, Modeling and Management*. Banff, Alberta, 1999.

Hendler J., Feigenbaum E.A.: "Knowledge Is Power: The Semantic Web Vision" in *Web Intelligence: Research and Development*, N.Zhong, Y.Yao, J.Liu, S.Ohsuga (eds.), Lecture Notes in Artificial Intelligence 2198, Springer, 2001.

ISX: http://www.isi.edu/isd/ontosaurus.html. ISX Corporation (1991). "LOOM Users Guide, Version 1.4".

Macintyre A.: "KREST User Manual 2.5". Wrije Universiteit Brussel, AI-lab. Brussels. 1993.

Marcus S., McDermott J.: "SALT: A Knowledge Acquisition Language for Propose-and-Revise Systems". *Artificial Intelligence*, Vol.39, No.1, 1989.

McGuinness D.L., Fikes R., Rice J., Wilder S.: "An Environment for Merging and Testing Large Ontologies". Proc. of the *Seventh International Conference on Principles of Knowledge Representation and Reasoning*, Breckenridge, Colorado, 2000.

Molina M., Cuena J.: "Using Knowledge Modelling Tools for Agent-based Systems: The Experience of KSM" in *Knowledge Engineering and Agents Technologies* Cuena J., Demazeau Y., García-Serrano A., Treur J. (eds.) IOS Press, (in press).

Molina M., Garrote L.: "Decision support for flash flood early warning using bayesian networks", Journal of Hydrology, (in press).

Newell A.: "The Knowledge Level" in *Artificial Intelligence*, vol. 18 pp. 87-127, 1982.

Pearl, J., "Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference", Morgan Kauffman, San Mateo, CA, 1988.

Schreiber G., Akkermans H., Anjewierden A., De Hoog R., Shadbolt N., Van de Velde W., Wielinga B.: "Knowledge engineering and management. The CommonKADS methodology" MIT Press, 2000.