

KATS: A Knowledge Acquisition Tool based on Electronic Document Processing

Martin Molina and Gemma Blasco

Department of Artificial Intelligence, Universidad Politécnica de Madrid,
Campus de Montegancedo S/N 28660 Boadilla del Monte, Madrid, Spain
mmolina@fi.upm.es, gblasco@isys.dia.fi.upm.es

Abstract. This paper describes a particular knowledge acquisition tool for the construction and maintenance of the knowledge model of an intelligent system for emergency management in the field of hydrology. This tool has been developed following an innovative approach directed to end-users non familiarized in computer oriented terminology. According to this approach, the tool is conceived as a document processor specialized in a particular domain (hydrology) in such a way that the whole knowledge model is viewed by the user as an electronic document. The paper first describes the characteristics of the knowledge model of the intelligent system and summarizes the problems that we found during the development and maintenance of such type of model. Then, the paper describes the KATS 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 comparison between KATS and other approaches for knowledge acquisition.

1 Introduction

Knowledge acquisition is a key issue in the development of intelligent systems. Despite that the advances in this field of knowledge acquisition the last years have significantly improved the way to perform this task with useful resources and tools, the current experience shows that there is still an important gap between the way end-users describe their expertise and the type of communication followed by existing tools. In particular, this issue has received recently special 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 underlined within the *semantic web* context as one of the challenges for the twenty-first century AI research [15].

In particular, this problem is especially significant when the complexity of the knowledge model of an intelligent system is high due to the presence of more than one complex tasks (e.g. diagnosis, prediction, configuration) each one with its particular knowledge organization. In this case, the difficulty of development and maintenance of the knowledge model increases because, among other reasons, it requires using different symbolic languages for representation of local knowledge bases and they present important dependences that need to be considered to guarantee the consistency of the global model.

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 called SAIDA in the domain of emergency management in hydrology. This paper first describes a summary of the SAIDA system and its knowledge model to show the problems that we found in the development and maintenance of SAIDA models. Then, the paper describes a particular software tool called KATS (*Knowledge Acquisition Tool for SAIDA-Models*) that we designed to facilitate the development of hydrologic models for SAIDA. This tool presents an innovative approach that considers the knowledge model as an electronic document. The paper describes the characteristics of such a document and how it is processed by KATS to serve as knowledge model of SAIDA. Finally, the paper presents a comparison between KATS and other types of existing knowledge acquisition tools.

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. 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 [8; 22; 20]. 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. SAIDA 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 the 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 a *model-based* approach with which different problem-solving methods were used as building blocks to perform the different tasks. To implement such a model, the KSM environment was used [7] following and a methodology that use and extend some concepts that are present in methodologies such as CommonKADS [23]. 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. Each task is performed by the corresponding problem-solving methods (details can be found in [20]).

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. 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.

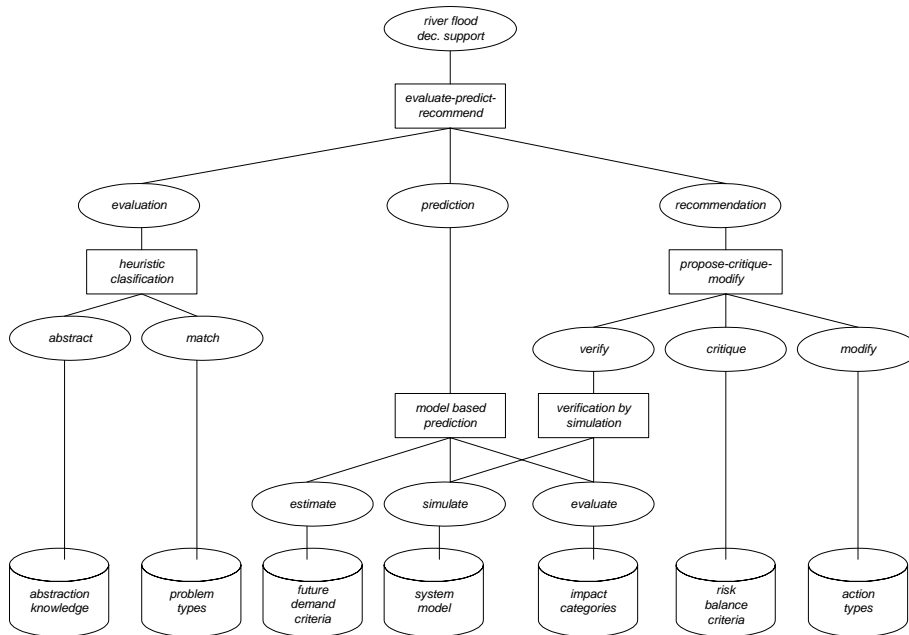


Fig. 1. The basic structure of the model for decision support in the field of emergency management during floods (circle = task, square = method, cylinder = type of knowledge base).

A model for a particular river basin is constructed formulating a set of knowledge bases. Figure 2 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.

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

Fig. 2. Summary of knowledge bases corresponding to a particular model for ariver basin in the case of the Júcar River Basin (East of Spain)

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 knowledge model 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 information processing, 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 to 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.

3 KATS: A Knowledge Acquisition Tool for SAIDA Models

In order to give an answer to the previous needs, we have designed and built a software tool called KATS that assists developers in the construction of hydrologic models for the SAIDA system. One of the most innovative characteristics of this tool is the user view of the knowledge model. We have applied the principles that we describe in [21]. According to them, the user perceives the whole knowledge model as an electronic document. This approach contrasts with the extended approach followed by knowledge acquisition tools that considers the user interface as a set of canned windows representing forms that must be filled by the user, together with a set of warning and help windows that supervise the process and guide the user. On the contrary, the basic idea in KATS is that when the user creates/maintains a knowledge model, she/he reads and writes a document in the same way that reads and writes a conventional document using standard text processors. A document is an entity very familiar to persons non specialized in computer science, so this a natural media to formulate the content of a model and it has been assumed and extended in the context of web-based applications with electronic features (hyper-links, multimedia, etc.).

The following sections analyze (1) the characteristics of the electronic document that describes to the complete knowledge model, and (2) the software tool that allows the user to consult and modify the electronic document.

3.1 The Electronic Document to Describe the Knowledge Model

The document is considered like a manual that describes problem solving methods for decision making procedures based on certain expertise. To satisfy the needs described in this paper, this document must accomplish the following three basic requirements:

- *Human understandability.* The content of the document must be totally comprehensible by users non-experts in computer science. A way to guarantee this assumption is that the complete set of procedures described by the document should be able to be performed manually by a user specialized in the domain of the problem.
- *Computer understandability.* The content of the document must be also comprehensible by computer programs that, based on this content, automatically perform the decision making procedures for which the knowledge based system has been designed.
- *Dynamic content.* The content of the document must be able to be modified by the user. For this purpose, in a document we consider three types of parts: (1) *static*, that describe prefixed areas assumed by computer programs that manipulate the document, (2) *user*, dynamic areas that can be edited and modified by users, and (3) *automatic*, dynamic areas that are automatically generated by programs that process the information provided by the user.

In order to accomplish the requirement for human understandability, it is important that the document presents a structure with the different issues about the problem that need to be understood by persons. For this purpose, we consider the categories of knowledge established by knowledge engineering methodologies and tools (for instance, CommonKADS [23], KSM [7], etc.). In addition to that, we distribute the domain knowledge and we add complementary views at different levels of abstraction to facilitate global views of the model (a description of the general format can be found at [21]). To be complete, the content of the document includes both inference (problem-solving) and domain knowledge. Normally, with a few exceptions, the static areas correspond to user-oriented descriptions about inference procedures, the user-dynamic areas correspond to domain knowledge and the automatic-dynamic areas correspond with complementary and global views of domain knowledge. Figure 3 shows a general view of the electronic document and its different parts.

An important factor to make the document understandable by humans is that the type of symbolic formalism used to represent domain knowledge must be familiar to persons non-expert in computers. Thus, instead of including in the text descriptions of algorithms and/or low level symbolic representations, the document must follow natural representations such as text (natural language) descriptions, tables, mathematical notation, graphics, etc. typically used in the professional domain of the problem. In the particular case of the KATS tool we used a set of what we call *document representation resources* that include common communication elements and others more familiar to professionals of hydrology (figure 4). This set of resources are presented uniformly in the electronic document together with the static information.

MANUAL DE SOPORTE A LA DECISION EN AVENIDAS

Este documento recoge el conocimiento relativo al dominio y a la resolución de la Toma de Decisiones en el caso de Avenidas de ríos.

Este documento muestra al usuario el conocimiento relativo al dominio contenido en el Sistema de Ayuda a la Decisión en Avenidas de ríos. La finalidad de este documento es adquirir todo el conocimiento relativo al dominio de aplicación de la herramienta final SAIDA, a partir del usuario final que interactúa con dicha herramienta.


Este documento está estructurado de acuerdo a los siguientes capítulos:

- Capítulo 1. Terminología del dominio.
 - Capítulo 1.1. Componentes topográficos de la cuenca.
 - Capítulo 1.2. Terminología de medición.
 - Capítulo 1.3. Topología de la cuenca.

Capítulo 1: Terminología del dominio

En este capítulo se define la terminología utilizada en el dominio hidrológico para modelar una cuenca, tanto desde el punto de vista de los elementos naturales o artificiales que componen su topología (subcuencas, embalses, etc) como de los elementos utilizados para obtener datos de control y medición del flujo de caudal del río, lluvia o niveles de embalse (alforjes, sensores de nivel, etc).


Los elementos identificados en la topología de una cuenca hidrográfica son los que se muestran en la ilustración 1.



Para poder relacionar los elementos definidos entre sí es necesario establecer el camino que sigue el flujo del río (el caudal) a través de ellos.

- La relación estación de alforje aguas abajo relaciona componente de cuenca con estación de alforje e identifica las estaciones de alforje que preceden de forma directa a éste. En líneas generales se utiliza para situar el elemento dentro de la topografía de la cuenca. La relación inversa se denomina componente aguas arriba.

En el Diagrama 1 se muestra la topología de la Cuenca del Sur.



Alforjones Guadalupe
R1 alforje virtual embalse Guadalupe
C1 alforjones Guadalupe
D1 alforje virtual embalse Guadalupe
E1 condenser Guadalupe Guadalupe Caudal de Guadalupe
F1 alforje virtual embalse Guadalupe Guadalupe

Q1 embalse Clotona
R1 alforje virtual embalse Clotona
R1 modificado embalse Clotona Invernal
Guadalupe
T1 embalse Casapala
E1 alforje Guadalupe

Capítulo 2: Fases de razonamiento

El objetivo de la tarea descrita en este documento consiste en la propuesta de planes de gestión para los embalses existentes en la cuenca, de tal forma que se reduzcan los posibles daños ocasionados por avenidas de ríos.

Para ello el razonamiento que se sigue consta principalmente de tres pasos: (1) evaluación de la situación, (2) predicción de la situación para las próximas horas y (3) recomendación de acciones de control que reduzcan los niveles de riesgo. La evaluación de la situación se lleva a cabo siempre que se toman o actualizan las medidas hidrológicas de las variables, ya sea por la toma directa de valores de un episodio hidrológico, por la predicción de la evolución de las variables o por la modificación de parámetros de gestión de los embalses, con el fin de interpretar y localizar situaciones de alerta.

Capítulo 2.3: Fase de interpretación de datos

Los datos proporcionados por el fichero del SARH se integran para obtener medidas con mayor significado hidrológico. Algunos de estos datos son interpretados de forma directa, como son los valores de las estaciones de alforje de salida de embalses y los sensores de nivel, mientras que otros como los valores de los pluviómetros, son distribuidos de forma ponderada de acuerdo a la influencia sobre unas cuencas u otras.

La lluvia media se obtiene de la siguiente forma:

$$M_i = \sum_{j=1}^{N_i} \alpha_j \cdot I_j \quad (1)$$

donde

- M_i es lluvia media de subcuenca i
- α_j es lluvia media de pluviómetro j de subcuenca i
- I_j es grado de influencia de pluviómetro j de subcuenca i

Capítulo 2.4.1: Identificación de escenarios de alerta

En el caso cuatralino, para cada elemento se establecen diferentes conjuntos de condiciones que caracterizan diferentes situaciones de gravedad en el mismo. La gravedad evaluada en cada elemento tiene diferente acepción de acuerdo al tipo de elemento del que se trate.

- La gravedad de un elemento de riesgo se define como el nivel de riesgo de desbordamiento o rotura en el caso de un embalse y de daños que puede sufrir a causa de una inundación. En líneas generales se utiliza para controlar la posibilidad de desastre en el embalse o área problema. Los valores posibles para este atributo son {nula, media, alta}

Se define un escenario de alarma como cada una de las combinaciones de características que identifican un estado de gravedad relevante para un elemento.

- Para el embalse Guadalupe Guadalupe

Nombre escenario	Lluvia últimas horas para Guadalupe Guadalupe	Lluvia últimas horas Guadalupe Guadalupe	Porcentaje de Embalse	Gravedad
Situación normal	<=50	<=50	<=50	nula
Situación normal	<=50	<=50	<=50	nula
Situación de alerta	<=50	<=50	<=50	alta
Situación de alerta	<=50	<=50	<=50	alta
Lluvia importantes	<=50	<=50	<=50	media
Lluvia importantes	<=50	<=50	<=50	media
Situación de vigilancia	<=50	<=50	<=50	media
Situación de vigilancia	<=50	<=50	<=50	media

Tabla 18

En la tabla 18 se expresa la forma en que se deduce el valor de gravedad de Embalse Guadalupe Guadalupe a partir de Nombre escenario (tipo de escenario de alarma), lluvia últimas horas Guadalupe (Lluvia media en las últimas 4 horas de subcuenca Guadalupe).

Anexo I. Nivel de terminación del documento

A continuación aparecen las características que deben ser cumplidas en el desarrollo del modelo de conocimiento para toma de decisiones en avenidas y el grado de cumplimiento de los mismos de acuerdo al estado de terminación del documento.

De entre ellos los requisitos generales deben ser cumplidos de forma estricta para tener la certeza de que el modelo de conocimiento construido es consistente. Los requisitos particulares, en cambio, aparecen como tareas planificadas por el usuario para ser desarrolladas de acuerdo a sus propios propósitos de creación o modificación del modelo de conocimiento.

Requisitos generales satisfechos:

- ✓ Existe al menos una subcuenca
- ✓ Existe al menos un embalse
- ✓ Existe al menos un área problema

Estado de los requisitos particulares:

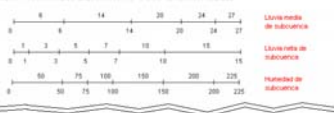
Requisito	Grado de cumplimiento
Decidir 0 subcuencas	75 %
Si hay número de cuenca entonces dentro la discretización de valores de cuenca para una subcuenca	75 %
Definir topología de la cuenca.	100 %
Definir fórmulas para el cálculo de lluvia media de subcuencas.	100 %
Definir fórmulas para el cálculo de humedad de subcuencas.	100 %
Definir temperatura para una área de cuenca	100 %
Definir velocidad de desborde de embalse	100 %

Anexo III. Modelo hidrológico causal

El caso de la predicción de comportamiento de una cuenca, requiere conocimiento basado en la heurística de la cuenca. Por la gran cantidad de datos que requieren ser manejados para ello, es necesaria la utilización de una herramienta que realice un estudio probabilístico de los casos recogidos en el pasado.

La herramienta utilizada para ello es GENEPROB. Esta herramienta se basa en relaciones causales entre variables hidrológicas y episodios hidrológicos pasados donde se refleja el comportamiento de la cuenca.

Las variables hidrológicas se expresan como variables cualitativas, es decir, variables que tienen un conjunto finito de valores que se corresponden con medidas de estados significativos del río para poner de manifiesto un rango suficientemente amplio de simulaciones a considerar en escenarios de avenidas. Para cada tipo de variable se considera un conjunto de valores que se obtienen mediante discretización de los valores numéricos medidos.



Anexo IV. Glosario de términos

En este anexo se recogen los conceptos específicos de hidrología utilizados en la descripción del modelo de conocimiento para la Cuenca del Guadalupe. Las características presentadas en la tabla indican:

Término	Definición	Fase	Naturaleza
DESCRIPCIÓN DEL DOMINIO	NUMERO DE TAREAS PRECISAS EN EL QUE SE DEBE DESARROLLAR LA PREPARACIÓN DENTRO DE UN PERIODO	PREPARACIÓN DEL PLAN DE PREVENCIÓN PARA PREVENCIÓN	Características numéricas de prevención
EFECTO INICIAL	EFECTO DE INFLUENCIA DE LA PREVENCIÓN	EFECTO DE INFLUENCIA DE LA PREVENCIÓN	EFECTO DE INFLUENCIA DE LA PREVENCIÓN
EFECTO DE Lluvia	EFECTO DE Lluvia CON UN TIPO DE TRAZADO	EFECTO DE Lluvia CON UN TIPO DE TRAZADO	EFECTO DE Lluvia CON UN TIPO DE TRAZADO

Introduction (static)

General terminology (static)

Domain terminology (user)

Model overview (static)

Task 1: Procedural view (static)

Task 1: Domain knowledge (user)

Task N: Proced. view (static)

Task N: Domain knowledge (user)

Task N: Complem. view (automatic)

Appendix I: completion degree (static)

Appendix I: completion degree (user, automatic)

Appendix 3: causal model (static)

Appendix 3: causal model (user)

Appendix N: Glossary of terms (automatic)

Fig. 3. Global view of the electronic document for knowledge acquisition

Representation resources		Descriptions
Paragraph	<i>Explicit term description</i>	Formatted text paragraph to describe explicitly a concept, an attribute of a concept or a relation between concepts.
	<i>Implicit term definition</i>	Formatted text paragraph to define implicitly a set of instances of a class, a set of attributes of a concept or a set of relations (for instance, there is a <i>path</i> for every pair <i>node-destination</i> connected through the relation <i>down-stream</i>).
Formula	<i>Simple formula</i>	Formula with arithmetic operators (+, -, /, *) and/or standard functions (sin(x), cos(x), etc.). The formula can be defined either for the attribute of a particular concept (specific) or for the attribute of a class (general).
	<i>Iterative formula</i>	Formula that is defined using an iteration (summatory Σ , productory Π , etc) on a set (or sets) of reference that is explicitly formulated in a table or implicitly defined in an attribute whose content is a list of values.
Table	<i>Table of instances</i>	Table to define particular instances of a class (e.g. the instances of reservoir in the domain of hydrology) with specific values for certain attributes (e.g., volume of each particular reservoir, etc.).
	<i>Decision table</i>	Table that includes conditional relations to be used in logical decisions. The decision table can adopt different shapes according to the amount of elements to be presented.
	<i>Table of causal relations</i>	Table that shows a set of types of relations between variables that are causes and variables that are effects. This is especially useful to formulate bayesian causal models in the domain of hydrology.
Graphic	<i>Qualitative number line</i>	A set of consecutive segments on a line with linguistic labels to describe a qualitative interpretation of a quantitative dimension.
	<i>Graph of relations</i>	Graph with nodes that represent concepts and arcs that represent relations. This is useful to present a global image of the relations between a set of instances.
	<i>Graph of processes</i>	Graph with nodes that represent components and measure points, and arcs that represent flows. This is useful to formulate a model of the structure and behavior of a river basin.

Fig. 4. Representation resources used in the electronic documents of KATS.

3.2 The Software Tool for Document Processing

The knowledge acquisition tool is considered as a specialized electronic document processor, i.e. a text processor specialized in the domain of SAIDA (hydrology) that allows the user to write the knowledge base as an electronic document (figure 5). The tool provides the following services: (1) *user interaction* to present the document to the user (with the corresponding representation resources) with facilities to modify the content, (2) *consistency checking* to guarantee the consistency of the model applying both syntax and semantic validation (the semantic validation is based on the processor specialization in the domain knowledge), (3) *development assistance*, to suggest what are the next steps to be done during the model construction, (4) *changes propagation*, to produce complementary views of the model, and (5) *document translation*, to generate operational versions of the document using symbolic languages (rules, frames, bayesian networks, logic clauses, etc.) to be processed by the intelligent system SAIDA.

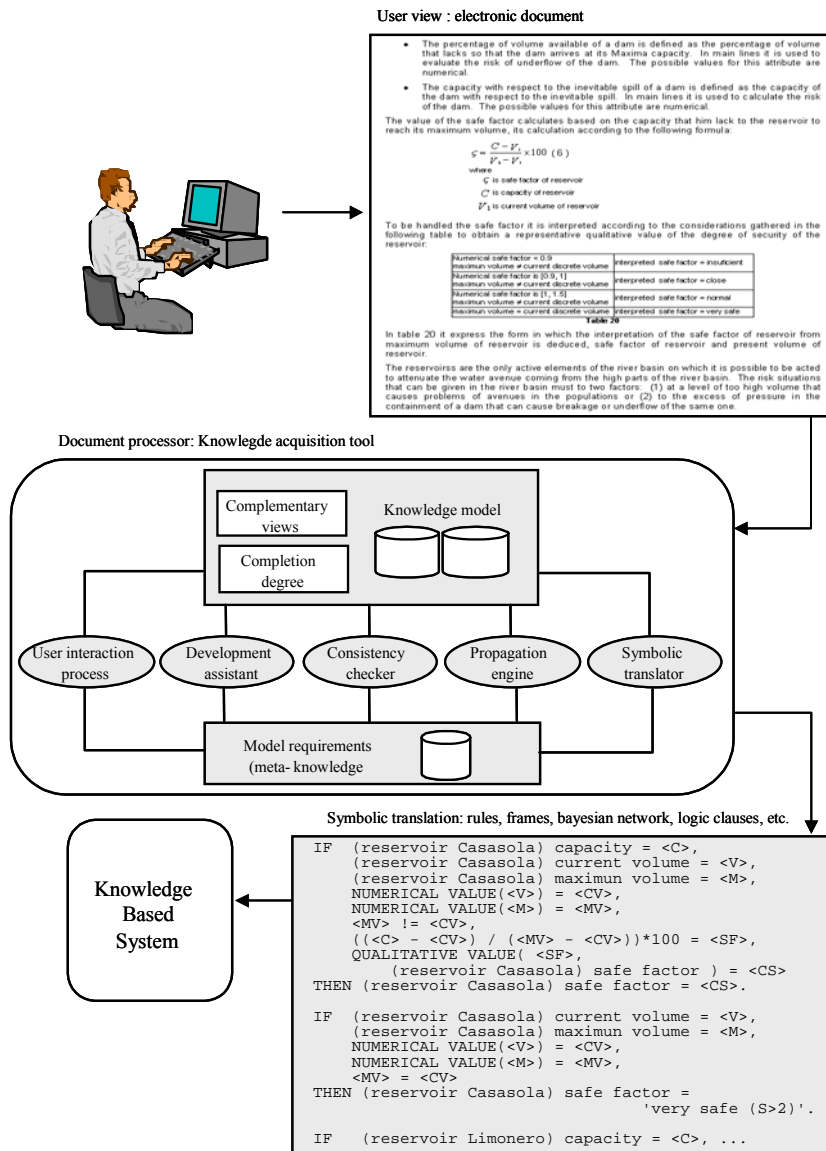


Fig. 5. Global view of the operation with the knowledge acquisition tool

Concerning the *user interaction* service supported by the tool, KATS provides a set of editing facilities to allow the user to modify each particular part of the document. The design of the user interface of KATS has been based on certain standards followed by the most extended text processors, so that it can be very familiar to persons and, consequently, easy to learn to use. The main window shows an image of the document as it is perceived by the user. Together with this, there is a set of editing facilities that are sensible to the part of the document to be modified. Thus when the user desires to change a dynamic part of the document, she or he clicks

directly on it and automatically a specialized editor window is presented. This particular window depends on the type of content in such a way that, for example, if a text paragraph for attributes is selected, an editor for attributes is presented. This editor allows to describe an attribute by selecting and completing prefixed natural language sentences. Similarly, if a decision table is selected, a specialized editor for decision tables is presented. Each particular type of presentation resource has its own editing facility. The window of each particular editing facility has the same appearance that the corresponding part presents in the document, together with certain buttons and resources for manipulation. Figure 6 shows examples of windows corresponding to editing facilities provided by KATS.

This idea of context-sensitive specialized editors is already present in conventional text processors, for example to modify the head of a document, a table or a formula. However, the KATS approach considers that these specialized editors are not only based on a syntax approach but their contents are interrelated to build local knowledge bases (e.g., a decision table only includes variables that were previously defined as attributes of concepts).

The services provided by KATS for *development assistance* and *consistency checking* are oriented to provide guidance to the user in order to know what is the next step (or remaining steps) to be done during the creation/modification of the knowledge model and guarantee that the model is complete and correct to be processed by SAIDA. For this purpose, we distinguish between local and global guidance. The local guidance is provided by each editing facility. For instance, the editor for a decision table helps the user to select appropriate elements for the table according certain construction requirements. Here, the editor uses the existing elements already present in the knowledge base to constraint the alternatives to be written by the user, which eliminates the possibility to write wrong information. For example, the editor for a decision table allows the user to write only existing concepts/attributes by selecting them in the corresponding menus, and also avoid to write autoreferenced sentences removing candidate attributes in the corresponding selection lists of menus. This editor also checks the logical expressions for attributes to be consistent with the allowed values.

The global guidance is especially useful for the construction of complex knowledge models. In KATS, the global guidance is based on what we call *model requirements*, i.e. affirmations about the form of the knowledge model that can be evaluated to check to what extent the model is complete. This is a kind of meta-knowledge and can be either prefixed by the knowledge based system (*hard model requirements*) or formulated by the user (*soft model requirements*). The hard requirements correspond to the minimum conditions that the model written by the user must satisfy to be operational (for instance, *there must be at least an instance of pluviometer* or *there must be a decision table for the attribute risk-level for each reservoir*).

The soft requirements depend on the particular model in a particular domain. We have designed a symbolic language for KATS with which the user may express in advance a set of global sentences about soft requirements. These sentences are referred to the presentation resources perceived by the user (e.g., concepts, decision tables, formula, diagrams, etc.). Figure 7 shows some examples of sentences (translated to English) that the user can write for this purpose. The KATS tool uses

these sentences to inform the user to what extend the model is complete. The sentences based on a quantitative reference (e.g. number of instances) allow to inform about the percentage of completion. The user may write this type of sentences with the help of a particular editor that presents the alternative options with the syntax details.

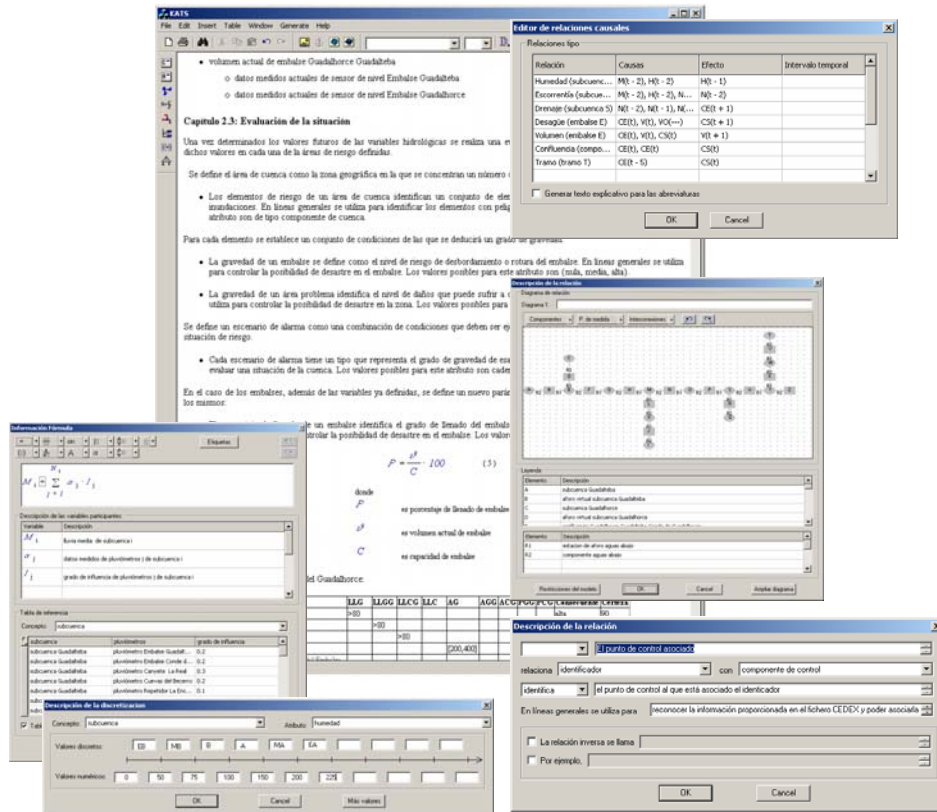


Fig. 6. Examples of editing facilities for different parts of the document

The service of *change propagation* provided by KATS is very useful for the construction and maintenance of complex knowledge models. With this service, when the user modifies certain part of the document, KATS automatically creates or changes other related parts of the document. Normally, these parts correspond to complementary views of the model that facilitate a better comprehension of the model. For example, the complementary views may correspond to a summary of certain elements that show a global view of certain part of the model. Other examples of complementary views are: cross references, glossaries of terms, etc.

Finally, the fifth service provided by KATS is *document translation*. In order to be operational (understandable by programs) the content of the document is translated by KATS to conventional symbolic structures to be manipulated by SAIDA. Thus, the changes in the document performed by the user with the help of KATS, modify the

knowledge base of SAIDA and, consequently, its problem solving behavior. Basically, KATS translates the representation resources of the document in the following way: (1) paragraphs for term description, graphs for relations, number lines, tables of instances and paragraphs for instances are translated to hierarchies of classes, subclasses with attributes and values and also the set of agents (2) formulas and iterative formulas are translated to functional expressions with temporal extensions, (3) decision tables are translated to rules and frame-like patterns, (4) tables of causal relations are translated to models based on bayesian networks, (5) graph of processes are translated to a temporal causal network.

Format of sentence	Examples
Number of instances of <class> [greater than, less than, equal to] <number>	Number of instances of reservoir equal to 10 Number of instances of pluviometer greater to 5
For every <class> where [<attribute> is <value>, <attribute> is known] a particular value for <attribute> must be known.	For every basin where number-of-curve is known a particular value for area must be known For every reservoir a particular value for objective-volume must be known
For every <class> there must be a qualitative interpretation for <attribute>	For every river-basin there must be a qualitative interpretation for average-rain
For every <class>, its <attribute> must be computed by [a formula, a decision table] [with the following attributes: <attribute-1>, ..., <attribute-n>]	For every reservoir its objective-volume must be computed by a formula For every problem-area its risk-level must be computed by a decision table with the following attributes: flow-upstream of problem-area
The relation <relation> [between <classes>, between <class-1> and <class-2>] must be defined by a graph	The relation downstream-section between river-basin-component and measure-point must be defined by a graph

Fig. 7. Some examples of sentences for model requirement

In certain cases the correspondence of the number of elements defined in the document in respect to the number of elements in the SAIDA knowledge base is not one-to-one but one-to-many. This means that the length of the document is normally significantly smaller compared to the length of the SAIDA knowledge base. This situation is present, for example, in the following cases: (1) a single paragraph that defines implicitly a set of instances is translated into the corresponding number of instances, (2) a general formula that is defined with an iterative definition is translated into a set of formulas with the corresponding particularization, (3) the definition of a particular attribute for a physical magnitude (e.g., flow) is translated into several specific attributes corresponding to different views (with different time steps and different levels of abstraction), etc.

4 Discussion

This section presents a comparison of the existing approaches of knowledge acquisition tools with KATS. 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) ontology management tools.

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 [10] for diagnosis systems with the cover-and-differentiate method, or SALT [18] for design systems with the propose-and-revise method. This category also includes other more specific tools in certain domains such as SIRAH [1] 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 significantly lower than other approaches. This category of tool is very similar to the case of KATS because KATS is a software tool oriented to help in building and maintaining the knowledge base of SAIDA which has a prefixed knowledge organization for a particular set of problem solving methods.

The main difference between KATS and the traditional approach of method-based knowledge acquisition tools is the user interface. With the second approach, normally, the user receives a kind of *data-base perception* of the knowledge model, i.e. a limited perception of the model mainly focused on domain knowledge that is presented as a kind of data base that is written with the help of prefixed windows. This approach is useful when the knowledge base presents low level of complexity. However when the complexity is higher (as it is the case of SAIDA that includes a number of problem-solving methods for different complex tasks, such as diagnosis, prediction and configuration, with different symbolic representations together with a multiagent organization) this approach does not guarantee that the user has an adequate uniform and global understanding and consequently an appropriate guidance to successfully create and maintain the models. On the contrary, the approach followed by KATS provides a *document perception* of the knowledge model, i.e. the model is presented as a document that brings together both inference and domain knowledge and could serve as complete documentation to perform manually the different tasks. This uniform and familiar way of presenting a knowledge model, with the corresponding editing facilities provided by KATS based on standards of text processors, make easier a complete understanding of the model.

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 [2] follows the KADS methodology [23] and allows a partial validation of the knowledge model using a computational language, KARL [12]. Other approaches such as KREST [17] or KSM [7] 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 [4]. 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. Thus, the KATS tool is less general but is closer to the user language which is important when the complexity of the model is high. However, although the specific KATS tool is particular for hydrologic models,

the principles of design followed in the construction of KATS are general and it can be reused for the development of other tools in different domains. In addition to that, certain components (such as the editing facilities) can be supported by programs whose code can be reused for the development of other tools.

The category of ontology management tools has been developed within the field of knowledge sharing and reuse, especially in the context of Internet. Examples of these tools are Protégé-2000 [13], WebOnto [9], OntoSaurus [16], Ontolingua/Chimaera [11], [19]. 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 standard knowledge representations (frames, relations, etc.) that can be limited in certain complex domains such as the case of hydrology presented in this paper. KATS, which is less general than these tools, uses more specific knowledge representations and inference procedures more appropriate for the tasks performed by SAIDA. In addition to that, KATS also is able to show the role that the different parts of the knowledge model plays in the model, which can be useful to better understand the model.

Besides the previous comparison with different types of knowledge acquisition tools, it is interesting to relate KATS with the concept of *semantic web* [3]. The semantic web approach is oriented to have electronic web documents that are not only understandable by persons but also by programs using standard *ontologies* [14] that establish certain common meaning about the potential content of documents. This idea is similar to KATS in the sense that both use the concept of document that should be understandable by humans and machines. However, the idea in semantic web establishes that programs are able to understand the *kind* of information in the document but not the information itself, for any kind of document. On the contrary, the approach followed by KATS, that is less general because it is limited to documents that represent knowledge models, *understands the content* of the document in the sense that certain parts of it are translated to symbolic representations that serve as knowledge bases of intelligent systems.

5 Conclusions

In summary, this paper presents the knowledge acquisition tool KATS that shows an innovative approach to improve knowledge acquisition and model maintenance of intelligent systems. In contrast to the traditional knowledge acquisition tools based on conventional user interfaces, we propose the idea of electronic document as basic media for knowledge acquisition. Under this approach, the knowledge acquisition tool is like an electronic document processor that allows the end user to read/write large and complex knowledge bases with automatic assistance (consistency checking, guidance, etc.).

KATS has been developed our group to serve as a knowledge acquisition tool to help users in building and maintaining the complex knowledge model of the SAIDA system. SAIDA is a decision-making tool that operates in the domain of emergency management in hydrology. SAIDA follows a knowledge-based multiagent approach with several symbolic representations and with, for example, 143 knowledge bases in a particular case. The first version of the KATS tool is now under evaluation by end-users specialized in the domain of hydrology.

It is important to note that the approach followed by KATS can be generalized to the construction of other knowledge acquisition tools in different domains. In particular, we are now applying this approach for the development of a new knowledge acquisition tool in the domain of transport management. For this purpose, we have reused not only the design concept of KATS but also certain programming libraries (for example, editing facilities for document representation resources).

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 KATS tool was supported by the Ministry of Science and Technology of Spain.

References

1. Alonso M., Cuenca J., Molina M. "SIRAH: An Architecture for a Professional Intelligence", ECAI Conference 1990. (L.Carlucci Ed.) Pitman, 1990.
2. 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.Cuenca (ed.). Elsevier, 1993.
3. Barnes-Lee T, Hendler J. and Lassila O. "The Semantic Web." Scientific American, May 2001.
4. Blythe, J.; Kim J.; Ramachandran, S.; Gil Y.: "An Integrated Environment for Knowledge Acquisition". Proc. International Conference on Intelligent User Interfaces, 2001.
5. Brown D., Chandrasekaran B.: "Design Problem-solving: Knowledge Structures and Control Strategies", Morgan Kaufman, 1989.
6. Clancey W.J.: "Heuristic Classification". Artificial Intelligence, vol 27, pp. 289-350, 1985.
7. Cuenca 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.
8. Cuenca 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.

9. 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.
10. Eshelman L., Ehret D., McDermott J., Tan M.: "MOLE: a Tenacious Knowledge-Acquisition Tool". Academic Press Inc., London, 1987.
11. 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.
12. Fensel, D.; Angele, J.; Landes D.: "KARL: A Knowledge Acquisition and Representation Language". Proc. Expert Systems and their Applications, Avignon 1991.
13. Grosso W.E., Eriksson H., Ferguson 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.
14. T. R. Gruber. "A translation approach to portable ontology specifications". Knowledge Acquisition, 6(2): 199-221, 1993.
15. 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.
16. ISX: <http://www.isi.edu/isd/ontosaurus.html>. ISX Corporation (1991). "LOOM Users Guide, Version 1.4".
17. Macintyre A.: "KREST User Manual 2.5". Wrije Universiteit Brussel, AI-lab. Brussels. 1993.
18. Marcus S., McDermott J.: "SALT: A Knowledge Acquisition Language for Propose-and-Revise Systems". Artificial Intelligence, Vol.39, No.1, 1989.
19. 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.
20. Molina M., Blasco G: "A Multi-agent System for Emergency Decision Support". Proceedings of the Fourth International Conference Intelligent data Engineering and automated Learning IDEAL 2003. LNCS, Springer.
21. Molina M., Blasco G.. "Using Electronic Documents for Knowledge Acquisition and Model Maintenance" Knowledge-Based Intelligent Information and Engineering Systems. Springer-Verlag. Lecture Notes in Artificial Intelligence 2774. Vol. II, pp. 1357-1364. Oxford, September 2003.
22. Molina M., Garrote L.: "Decision support for flash flood early warning using bayesian networks", Journal of Hydrology, (in press).
23. 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.