Supporting newly-appointed judges: a legal knowledge management case study

Pompeu Casanovas, Marta Poblet, Núria Casellas, Jesus Contreras, V. Richard Benjamins and Mercedes Blazquez






Pompeu Casanovas is a Professor, Marta Poblet is a Researcher and Núria Casellas is an Assistant Lecturer, all at Barcelona Autonomous University (UAB), Barcelona, Spain. Jesus Contreras is in the Innovation Department, V. Richard Benjamins is Director of Research & Development and a board member, and Mercedes Blazquez is a Researcher, all with Intelligent Software Components SA, Madrid, Spain.

Funded by the National and European projects: SEC2001-2581-C02-0 and EU-IST (IP) IST-2003-506826 SEKT.

Abstract

Purpose – *In this paper we describe the process of developing and implementing a knowledge management system for the Spanish judicial domain. Spanish judges, especially newly-recruited ones, hold a solid background of theoretical legal knowledge, but are much less familiar with the judicial knowledge of the more senior judges acquired from everyday practice and case resolution. The aim of this development is to capture and model these two aspects of judicial knowledge – theoretical and practical – for knowledge browsing and retrieving.*

Design/methodology/approach – *Semantic web technologies are applied to feed a question-answering system based on ontologies of professional legal knowledge (OPLK).*

Findings – *There is a kind of specific legal knowledge, which belongs properly to the expert domain, not being captured by current legal core ontologies, i.e. Judges require clues, hints or well-grounded practical guidelines that refer to the problem they have before them when they put a question or start the query. A scalable and useful frequently-asked questions system should have a simple, natural language interface, work in a real time environment, and the questions included in the system should be of high quality and reflect the current situation.*

Originality/value – *The final system will enable the users to ask queries in natural language and obtain answers, which are supported by legal documents stored in specialized legal databases. Special care is taken regarding usability issues, in order to ensure the highest user satisfaction.*

Keywords *Judges, Legal profession, Legal theory, Laws and legislation, Online databases, Knowledge management*

Paper type *Conceptual paper*

1. Introduction

Judicial systems are knowledge intensive domains as, on the one hand, judges are required to apply a vast bulk of theoretical knowledge drawn from legal textbooks, statutes, and codes and, on the other hand, they also need to master the standards of practice at courts when dealing with daily caseloads. While the former knowledge is highly codified and broadly shared by all legal professionals, the latter is exclusively acquired during years of practice and stored within the personal experiences of senior judges. Traditionally, this practical judicial knowledge was transmitted through informal apprenticeship: in every court, newly-recruited judges were trained by their senior peers (Casanovas, 1994). However, the renewal of the Spanish judiciary in recent years has made maintenance of this traditional method extremely difficult (Poblet, 2001).

The main goal of the project is to develop a software capable of clearing up doubts concerning judicial practice (as a senior judge would do) by providing justified and uniform answers to the questions raised by newly-recruited judges, avoiding possible inconsistencies. As in any other knowledge management system, this project will identify how and when the information should be offered to the user. Some particularities of the legal domain have been taken into account when planning and developing the project, often

related to the nature of the information supplied. For example, it requires that the user could track the information source (legal document) of each result displayed by the system to establish its provenance and reputation.

This paper is organized as follows: Section 2 offers the profile of the users in the judicial domain in order to establish the requirements needed for the use cases. In section 3 we describe the work done and the methodology used. The main objective of this section is to show how and what knowledge has been selected to be part of the knowledge base of the system according to the user needs. In section 4 we describe how the selected knowledge has been formalized in a semantic web format, using an ontology, in order to allow the final application to achieve all the functional requirements. The usage of semantic web formalism (ontologies) and technology makes possible to offer a precise information extraction and retrieval functionalities, as well as friendly interfaces, through the use of natural language processing. The semantic model underlying the application is oriented to the interpretation of user's queries in terms of the domain model, and then to produce an answer to the user's question. The next section deals with application functionalities, architecture, user interface and usability. The last section refers to our experiences regarding the analysis, the design and the development of this system in a real environment.

2. Knowledge acquisition and user profile

If a knowledge management system is to be able to adapt to user requirements and provide them with an efficient support in a fast and reliable way, the accuracy and the validity of the knowledge repository is critical. To adapt to the requirements of the newly-recruited judges, two national surveys have been conducted as a primary source of data regarding both the context of use and the contents of the questions for which the system provides answers[1].

These surveys have offered interesting and important data to elaborate the user's profile. There are two aspects of the professional profile of judges that are relevant to our project. The frequency with which the new judge talks about the cases he is dealing with and their usage of information technologies. Only 4.71 per cent of the judges interviewed stated that they never exchange information concerning their cases with other people (see Figure 1).

On the contrary, the vast majority of them gave affirmative answers regarding information exchange. When asked with whom they talked about their cases (see Figure 2), judges typically listed members of the judiciary (53 per cent of the total), including: senior judges or magistrates; colleagues from other judicial units of their same building; their tutors; and peers of the same graduation year. They also mentioned prosecutors (usually the ones related with the cases they had to decide upon). These figures are consistent with one of the core hypotheses of the research: the use of the senior judges' professional knowledge as a parameter when dealing with difficult or unusual cases.

Regarding the procedure followed to comment judicial cases, judges use either the telephone or a personal interview – most usually, an informal interaction with fellow judges to discuss their cases. Significantly enough, none of them uses electronic mail or instant messenger systems to communicate with peers or senior fellows.

Figure 1 Frequency of information exchange regarding cases

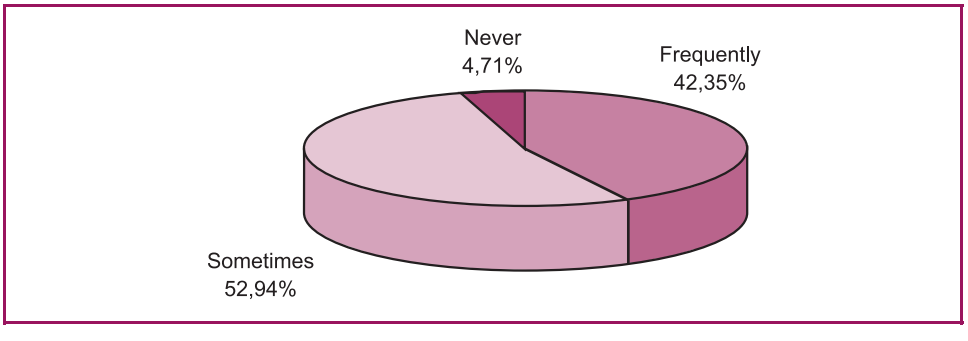
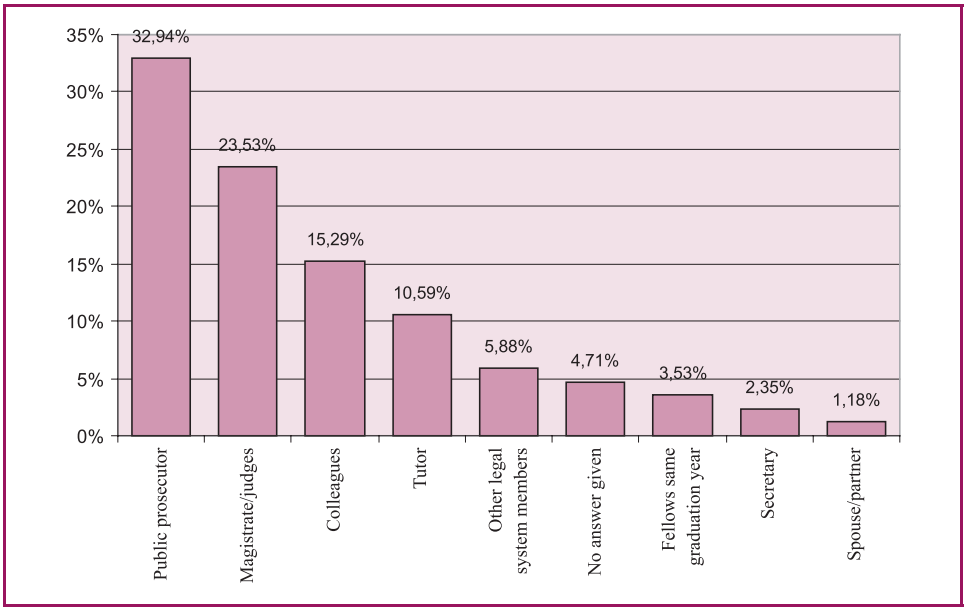


Figure 2 Who do judges discuss their cases with?



Therefore, it is necessary to insist on the fact that users of the system will be judges who have medium or low technological abilities, and are not used to new technologies. In relation to this, at the beginning of their training at the Judicial School, 47 per cent of members of the 52nd class declared to have no computer skills whatsoever, and only 35 per cent of the students declared to be internet users (Ayuso *et al.*, 2003) (see Figure 3). This is totally consistent with previous data showing a rather scarce use of information and communication technologies (ICTs) in court[2] (also see Table I).

Nevertheless, our survey shows that the use of internet among those same judges has improved: four years after accessing the School, the proportion is reversed: 84.7 per cent of

Figure 3 Internet use among judges

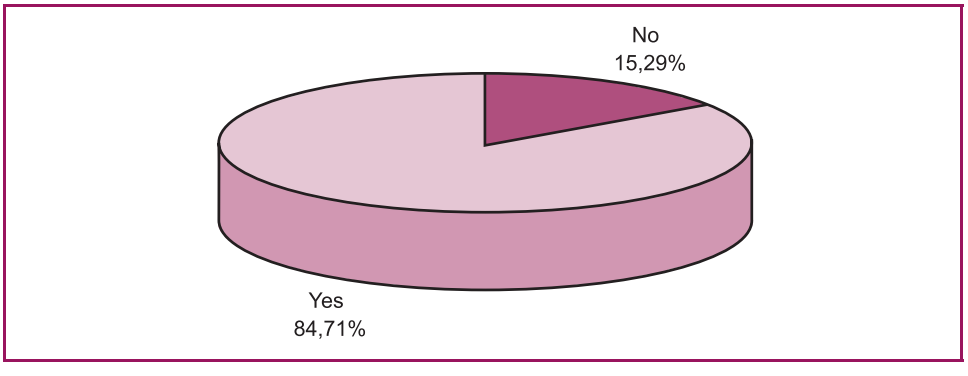


Table I Judges' internet use for professional purposes (2002)

Uses internet	*Inexperienced judges*	*Experienced judges*	*Total*
Yes	60.6	53.2	54.2
No	38.6	46.1	45.2
DK/DA	0.9	0.6	0.7
Total	100	100	100

the interviewed declare to use internet and only 15.29 per cent do not use it (see Figure 3). As for the type of information searched for through the internet, the page of the Official Bulletin of the State is the most accessed site, followed by legal information in general. Other types of searches are unusual, and judges will typically argue that they have no time to surf the internet.

Finally, judges offer an interesting answer to the question ''what would you like to find if judges were given a web service system?''. The majority of them proposed a site where doubts regarding professional cases could be put in common and discussed (see Figure 4).

3. Methodology and tasks performed for the knowledge description and definition

At the first stages of our work, the objective was to define the legal domains where the lack of professional knowledge was critical and what kind of information the newly appointed judges would require. We therefore focused on the identification and description of the desired knowledge areas.

Results from the two previous surveys allowed us to identify three main areas which presented some difficulties to new judges:

1. the organization and management of judicial staff (clerks working in judicial units);
2. the interpretation and implementation of new procedural statutes (e.g. *Ley de Enjuiciamiento Civil*); and
3. the ''on-duty'' period (or the *guardia*, that is, the week in which the entire court is on duty tackling the preliminary investigations of the criminal cases that keep entering the system).

We selected this ''on-duty'' domain to begin with and were provided by the Judicial School with rich material containing pairs of questions (competency questions) and answers regarding problems of practical procedural criminal law. Figure 5 shows a translated example and the way we reformulated the initial question.

These data provided by the judiciary were analyzed with textual statistics methods to extract the domains where the difficulties lay. In general, the problematic domains were mentioned in the answers by the use of substantives. Thus, the design of the ontology for the first Prototype of *Iuriservice* started from the analysis of these pairs of answers and questions, and using the ''competency approach'' (Grüninger and Fox, 1995) to identify relevant aspects and the coverage of the ontology.

Figure 4 Most needed information sources in a web service

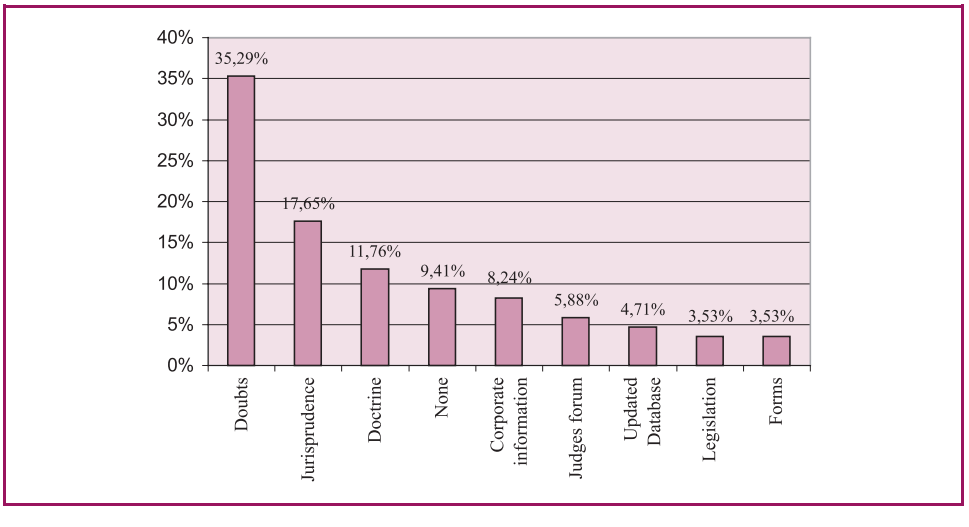


Figure 5 Example of an ''on duty'' question, rewriting and reply

(1) *Question*

– While on duty, an investigating magistrate receives a call from a hospital reporting a sexual assault. The victim has still not made an official report of the incident. Procedures to be followed. Which rules apply?

(2) *Rewriting*

– In a case where a medical center telephones to report a sexual assault, what must be done by the investigating magistrate who receives the call, and if the victim has not officially reported the incident, which procedure must be followed?

– If an investing magistrate is informed by a hospital that there has been a sexual assault, what procedures must he or she follow in order to ascertain the facts of the case, and which of the established official procedures must be followed if the victim has not officially reported the assault?

(3) *Reply*

– As for the procedures to be followed, a forensic scientist should be sent to the hospital in order to examine the victim and to take samples. If the crime has not yet been officially reported, the judge except in very exceptional circumstances may begin no procedures. Provided that it is clear from the telephone call alone that this is a case of sexual assault and that no other crime has been committed, then criminal proceedings must be initiated by the victim

The most general concept we found in the judicial criminal field was *proceso* (process, trial, procedure), the Spanish procedural notion that stands for all kinds of proceedings under the Spanish law. Thus, we reproduced the cognitive flow of a judge in an on-duty situation: he has to decide which procedure is the most suitable to the facts he has before him.

The second Prototype of *Iuriservice* is based on in-depth interviews with the judges, better focused on the difficulties due to the fact that the previous questionnaire had pointed out the areas where the difficulties were located. This new approach made possible the identification of more than 750 competency questions, in comparison to the 100 pairs of questions-answers that were analyzed for the first prototype.

Other changes have been introduced regarding ontology modeling issues. First, several KAON applications, provided by the SEKT Project partners, such as the ontology editor Oi-Modeler, have been used in order to model and visualize the domain ontology[3]. Second, we have also used two different software applications to analyze the competency questions and extract the relevant concepts: TextToOnto[4] and ALCESTE[5]. Finally, we have also followed the Distributed, Loosely-controlled and evolving Engineering of oNTologies (DILIGENT) methodology during the ontology engineering process (Pinto *et al.*, 2004). The visualization of the arguments took place on a wiki-based environment which allowed them to be traced.

Figure 6 and Figure 7 depict graphically some of the results concerning conceptual distribution. Figure 6 shows an extraction graph for the competency questions (obtained during the second survey) using TextToOnto's TermExtraction on KAON API. Its TaxoBuilder tool could not be used as the Spanish GATE components are currently being implemented[6].

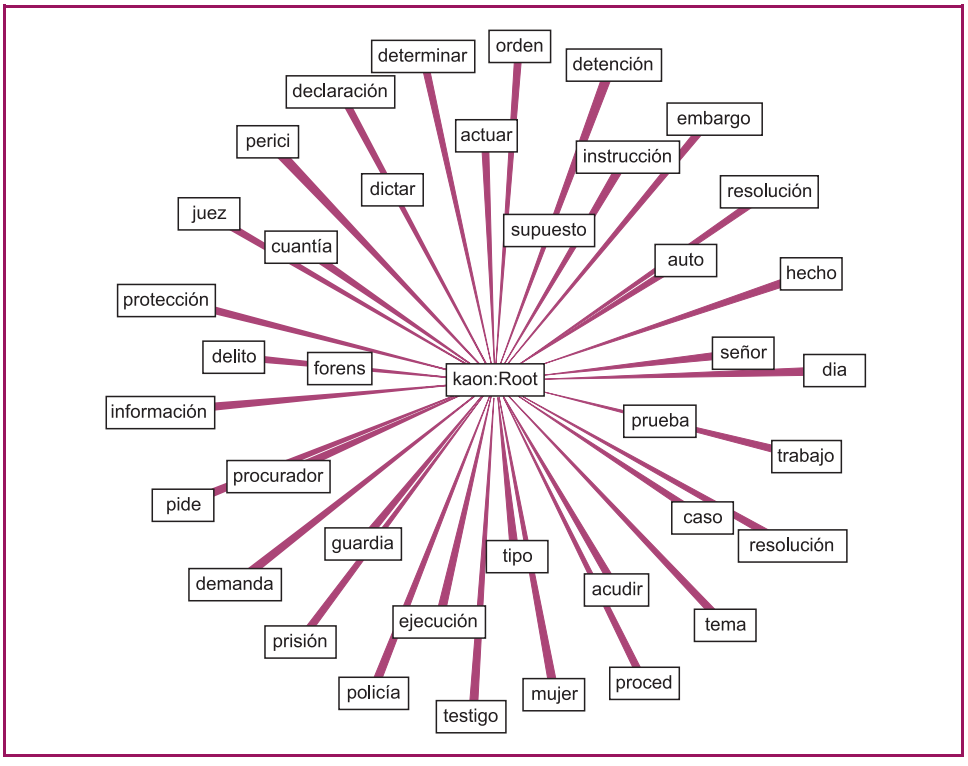
Figure 7 shows the visualization of the ALCESTE descending cluster algorithm applied to the competency questions obtained in the first survey. The program offers a list of most characteristic words assembling subsets of ''lexical worlds'' into classes according to a χ^2 metric. This has proved to be useful to flesh out the conceptual structure of the judicial competency questions.

4. Ontologies of professional legal knowledge (OPLK)

Legal ontologies

The legal domain has been of interest to artificial intelligence for many years. Gray (1997) has pointed out that the theory and the tools of artificial legal intelligence have developed in corresponding leaps, with the following progression of themes: legal language; deontic

Figure 6 Term extraction graph (TextToOnto)



logic; rule processing; case processing; stratification of reasoning; procedural reasoning; co-ordination of multiple tasks.

Legal ontologies have played a part in that process to help structure legal knowledge and create knowledge management tools. Many legal ontologies have been built so far. One way of describing the actual state of the art is by identifying the six main current legal ontologies[7]:

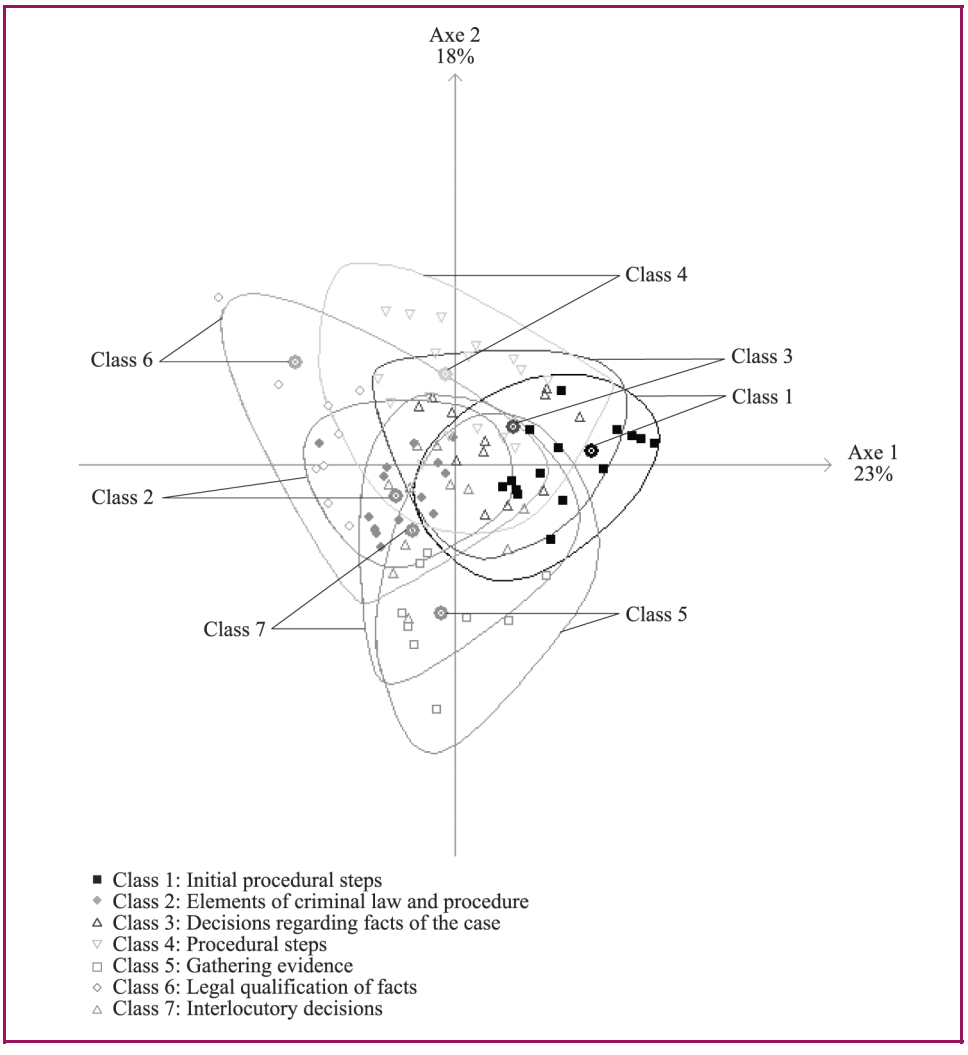
1. LLD (Language for Legal Discourse (McCarty, 1989)), based on atomic formula, rules and modalities;
2. NOR (Norma (Stamper, 1991, 1996)) based on agents behavioural invariants and realizations;
3. LFU (Functional Ontology for Law (Valente, 1995)) based on normative knowledge, world knowledge, responsibility knowledge, reactive knowledge and creative knowledge;
4. FBO (Frame-Based Ontology of Law (van Kralingen, 1995; Visser, 1995)), based on norms, acts and descriptions of concepts;
5. LRI (Core Legal Ontology (Breuker *et al.*, 2002)), based on objects, processes, physical entities, mental entities, agents and communicative acts;
6. IKF-IF-LEX (Ontology for Norm Comparison (Gangemi *et al.*, 2001)), based on agents, institutive norms, instrumental provisions, regulative norms, open-textured legal notions and norm dynamics.

All these ontologies have been discussed within the intellectual community of AI and Law, and presented at the annual meetings of JURIX and ICAIL.

Epistemology and legal ontologies

Legal ontologies differ from other types of domain ontologies in two special features. The first one is the bulk of common-sense notions that are carried out within the legal domain. Legal

Figure 7 Factorial plan of the 7 lexical worlds (classes) produced through its hierarchical descending clustering algorithm for the competency questions on on-duty problems (ALCESTE)



statutes and case law are written both in natural and in a more technical language. But practically all the common sense notions and connections among them that people use in their everyday life are embodied in the legal domain.

The second special feature has to do with the fact that the strategy of ontology building must take into account the particular model of law that has been chosen. This occurs in a middle-out level that it is possible to skip in other ontologies based in a more contextual or physical environment.

When the task to be done involves a whole process it is possible to figure out the ontology following some particular cognitive models that can be used as templates to build the basic ontological concepts and their relationships. In this case only two levels are really needed: the domain conceptual level and the upper ontological level.

For instance, recent conceptual work shows in which way a plant oil battery (Chan, 2003) and a smart home environment (Gu *et al.*, 2005) can be plotted into cognitive maps using these twofold ontological levels.

However, in the legal field, the modeling process usually requires an intermediate theoretical level in which several concepts are implicit or explicitly related to a set of decisions about the

nature of law, the kind of language used to represent legal knowledge, and the specific legal structure covered by the ontology. There is an interpretative level that is commonly linked to general theories of law. This intermediate level is a well-known layer between the upper-top and the domain-specific ontologies, especially in the so-called ''practical ontologies''[8].

Perhaps the most striking feature lies in the fact that in most of the legal ontologies built so far, the intermediate layer is explicitly occupied by high conceptual constructs suggested by general theories of law instead of a more flexible knowledge provided or induced from empirical or cognitive findings. Therefore, the link between epistemology and ontology is filled up with some intuitive or philosophical assumptions about the nature and function of law.

General theories of law seek to map into a single coherent body the most general concepts aiming to represent legal knowledge. In this way, several formulations have been approached. Law conceived as a set of:

- related static and dynamic norms;
- interrelated rights and duties;
- institutional rules and facts; or
- states of affairs, events and rules

has been the source of inspiration for legal ontology building[9].

This is not necessarily a critical issue. There is a free space for epistemological assumptions in ontology building and, besides this, each ontology might be constructed having in mind specific goals. But it seems to us that the lack of empirical knowledge in legal ontology engineering may broaden the gap between users' needs and expected solutions, as already noticed in the literature (Paliwala, 2001).

The interpretative middle level in which all fundamental concepts are defined is usually known as legal-core ontology. Breuker and Winkels (2003) have recently distinguished between legal ontologies originally based on normative knowledge (legal theory) and legal ontologies – or ''with an ontological flavour'' – in which modalities play the role of knowledge categories. This would be the case for McCarty's LDD or for deontic logic formulations applied to the legal domain (rethinking the Hohfeldian conceptions or based on modal linguistic functions: obligatory, forbidden, permitted ...). However, in both cases, the fundamental concepts are epistemologically set within a legal-core ontology, that is to say, an ontological representation of basic legal knowledge, in which the theoretical representation of abstract rights and duties count much more than the practical aim of a hypothetical user. Legal reasoning prevails over practical purposes.

Ontologies of professional legal knowledge (OPLK)

Professional knowledge (PK). PK, as encoding a specific kind of knowledge related to particular tasks, symbolisms and activities has been described many times in the law and society literature (Abel, 1997). Legal professions have changed dramatically under the impact of globalization.

Professional knowledge has been defined as possessed by professionals, which enables them to perform their work with quality (Eraut, 1992). Professional knowledge then includes:

- propositional knowledge (knowing that);
- procedural knowledge (knowing how);

- personal knowledge (intuitive, pre-propositional); and
- principles related to morals or some kind of deontological code.

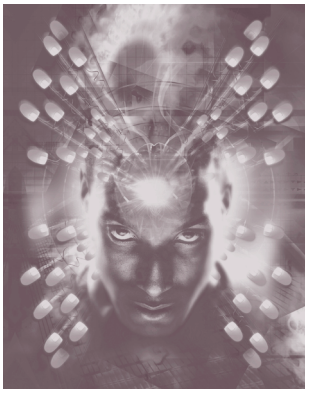
Such knowledge is not just represented mentally in the form of statements and rules, but also includes images, metaphors and attitudes to produce successful outcomes (Bromme and Tillema, 1995).

A distinction is generally made between knowledge base, which is being developed through practice, and experiential learning. But do notice that globalization produces pervasive and unexpected paradoxes. Moreover, in the legal field, there is a growing gap between the institutionalized professions closer to the state (judges, prosecutors) and those operating within the legal markets.

We could say that lawyers share with the judge, the prosecutor or other court staff only a portion of the legal knowledge (very likely the legal language and the most general acquaintance of statutes and previous judgments). But there is another kind of legal knowledge, the one having to do with personal behavior, practical rules, corporate beliefs, effect reckoning and perspective on similar cases, which remain implicit and tacit within the relations among judges, prosecutors, attorneys and lawyers.

Professional legal knowledge (PLK). The legal ontologies described above have been built up with several purposes:

- creation of regulatory metadata and content standardization (e.g. Legal-XML/LeXML/MetaLEX, ADR/ODR-XML, etc.
- information extraction from legal documents;
- regulatory compliance;
- statute-content harmonization;
- modelling of legal reasoning;
- support to decision making.



Although the legal domain remains very sensitive to the features of regional or national statutes and regulations, some of the legal-core ontologies (LCO) are intended to share a common kernel of legal notions. Therefore, LCO remain in the domain of a general knowledge shared by legal theorists, national or international jurists and comparative lawyers.

Our data indicate that there is a kind of specific legal knowledge, which belongs properly to the expert domain and that is not being captured by the current LCO.

Consider the following on-duty problems, extracted from different kinds of transcriptions of the research judicial protocols, below:

> Police are sending here all battered women to ask for a protection order without starting any proceedings or summoning anyone. What should I do?
>
> The doctor called explaining that he has a person who is insane and that should probably be confined in a hospital and he is asking me for a confinement order over the phone. Can I do this?
>
> Two women came today to court to tell me that their upstairs neighbor is a very old woman who lives alone and has Alzheimer's. They are afraid that she might forget to turn the gas off. What can I do?
>
> A body has to be moved. The body is in the public way and the forensic examiner is off duty and his substitute is not available and will not be back soon. The removal is urgent. Can I entitle a family doctor to do it? What should be done?

Technically speaking, these problems are not legally complex. However, they are difficult to solve. The judges' original questioning cannot be answered by simply pointing out a particular statute or legal doctrine. This is not only an issue of legal information retrieval.

Judges are experts; they take for granted the acquaintance with statutes, textbooks or former legal decisions. What is at stake here is a different kind of legal knowledge, a professional legal knowledge (PLK) (Benjamins *et al.*, 2004a, b; 2005a). What judges really

seek are some clues, some hints or well-grounded practical guidelines that refer to the problem they have before them when they put the question or start the query.

In this regard, the design of legal ontologies requires not only to represent the legal, normative language of written documents (decisions, judgments, rulings, partitions . . .) but also those chunks of professional knowledge that is used daily in courts.

From this point of view, professional knowledge of a legal topic (such as, e.g. gender violence, medical issues, police behavior . . .) involves a particular knowledge of: statutes, codes, and legal rules; professional training; legal procedures; public policies; everyday routinely cases; practical situations; people's most common reactions to previous decisions on similar subjects.

We may point out several properties of professional legal knowledge (PLK). PLK is:

- shared among members of a professional group (e.g. judges, attorneys, prosecutors . . .);
- learned and conveyed formally or most often informally in specific settings (e.g. the Judicial School, professional associations – the Bar, the Judiciary, etc.);
- expressible through a mixture of natural and technical language (legalese, legal slang);
- non-equally distributed among the professional group;
- non-homogeneous (elaborated on individual bases); and
- generally comprehensible by the members of the profession (there is a sort of implicit identification principle).

One of the main features of PLK is that it is context-sensitive, anchored in courses of action or practical ways of behaving. In this sense, it implies:

- the ability to discriminate among related but different situations (e.g. when is it really needed or required an injunction of protection to prevent a women of being injured or murdered by her husband?);
- the practical attitude or disposition to rule, judge or make a decision;
- the ability to relate new and past experiences of cases; and
- the ability to share and discuss these experiences with the peer group.

Especially in the judicial field, PLK presents two additional features: the attunement process produced in the everyday decision making with previous ''organizational memory'' of senior peers (institutional process); and the need to ground each new judicial decision on previous cases (legitimation process). The first process is almost completely tacit, but the second is totally explicit in the judicial decision: there is a substantial part within the written judicial decision devoted to it, named *fundamentos de derecho* (legal grounds). These two parallel information processes are required in order to pass judgment.

Ontologies of professional legal knowledge (OPLK). In order to build OPLK, we believe that we have to take into account the kind of situated knowledge that judges put into practice when they store, retrieve and use PLK to make their most common decisions. We use ''situated knowledge'' in a similar way in which Clancey *et al.* (1998) talk about ''situated cognition'': the concrete use of knowledge which is partially shared and unequally distributed through a certain ''community of practice'' who is able to use and reuse this same knowledge while transforming it[10].

"One of the main features of professional legal knowledge is that it is context-sensitive, anchored in courses of action or practical ways of behaving."

The main idea is that PLK is always situated in a particular context in which the judge or the lawyer needs to complete the information they possess about a particular case or problem to trigger or put into practice the basic knowledge that they already have. In this sense, they do not need to be provided with a complete legal reasoning, but only with some reliable information that they may use as a comparative parameter. They seek ''another opinion'', an external interlocutor to follow the full reasoning process that they build up any time they have to make a new decision. This is the reason why they discuss the case with their peers: not really to be helped in the decision making but to double check the decision they are going to take.

Other related concepts close to ''situated knowledge'' are the ideas of ''situated communities'' (Kavanagh and Kelly, 2002), situated meaning'' (D'Andrade, 1995), organizational memory'' (Vouros, 2003) and ''corporate ontologies'' (Roche, 2000) (Mentzas *et al.*, 2001).

Legal reasoning is eventually almost an automatic process in which lawyers or judges are involved without being fully aware of the devices they are using. But our data analysis makes clear that this is a collective and interactive process, even if it is usually performed on individual bases and remains tacit.

Tacit knowledge has been receiving an increasing attention from the knowledge management field. It has been argued that tacit knowledge:

- grows along with the complexity of situations at the organizational level (Nicolas, 2004);
- is linked to meta-abilities of the individuals (cognitive skills, self-knowledge, emotional resilience and personal drive) (Hysyam and Choudrie, 2004);
- involves emotional values that produce different identities, different levels of knowledge integration, and different degrees of attachment to the organization (van den Hooff and de Ridder, 2004);
- involves the articulation of the behaviors of different teams inside the organization as a sort of political dimension (Kim and King, 2004); and
- cannot be measured in simple quantitative terms (Koskinen, 2004).

Building ontologies means entering in a process in which this tacit knowledge is made conceptually explicit in a formal machine-readable language. But, because of its own nature, this is not made without some tensions.

On the one hand, for all practical purposes there is no such thing as absolute *meaning*: everything must ultimately be the result of agreements among human agents such as designers, domain experts and users (Jarrar and Meersman, 2001). On the other hand, in ontology knowledge modeling a concept is neither a class nor a set: the concepts which represent the term's meaning are structured into binary trees based on couples of opposite differences (Roche, 2000).

OPLK models the situated knowledge of professionals at work. In our particular case we have before us a particular subset of PLK belonging specifically to the judicial field. Therefore, we will term ontology of judicial professional knowledge (OJPK) the conceptual specification of knowledge contained in our empirical data.

SEKT ontology of professional judicial knowledge (OPJK)

The ontology of professional judicial knowledge developed by the SEKT legal case study team is learnt from scratch out of the competency questions posed by the judges during their interviews. Modeling this professional judicial knowledge demands the description of this knowledge as it is perceived by the judge and the abandonment of dogmatic legal categorizations.

The construction of the ontology is based on the term and relation extraction from the questions regarding practical judicial problems posed by the judges during their interviews. Due to the fact that at that time semi-automatic extraction software for Spanish was not

available, the extraction is performed manually, nevertheless, tools such as TextToOnto and ALCESTE were used to support manual term extraction and to visualize it. ALCESTE also showed that this knowledge is organized within distinct sub-domains (gender violence, procedural doubts, on-duty period, etc.).

As explained before, several modeling changes have been introduced towards the construction of the ontology used by the second *Iuriservice* prototype. The ''competency approach'' continues to be followed and the method used in building the ontology has now focused on the discussion within the UAB legal experts' team over the terms which appear on the competency questions. Thus, we have extracted significant terms for the ontology and their relations from the 756 competency questions. The ontology is still under construction, but more than 200 questions have already been discussed.

This method has several phases. First, it basically consists in selecting (underlying) all the nouns (usually concepts) and adjectives (usually properties) contained in the competency questions. The relevant relations between those terms also have to be identified (mainly is_a and instance_of). Once the terms had been identified, the team discussed the need to represent them within the ontology and their organization within taxonomies. Accordingly, we follow the middle-out strategy (Gómez-Pérez *et al.*, 2004). With this strategy, the basic terms are identified first and then they are specified and generalized if necessary.

As an example, and in relation to the competency questions analyzed, modelers considered that the concepts *auto* (interlocutory decision), *recurso* (appeal), *demanda* (private/civil complaint) and *querella* (public/criminal complaint) needed to be represented in the ontology. Moreover, a concept *documento* (document) had to be created as all terms: *auto, recurso, demanda* and *querella* describe documents. The result was the construction of a more general concept from those specific terms found in the competency questions.

However, the team also agreed that *demanda, auto, recurso* and *querella* were not only instances of *documento* but also constituted a specific class of documents used only within the judicial process. For that reason, *documento_procesal* (procedural document) had to be created as a subconcept of *documento*. At the same time, there are different types of appeals and court orders referred to in the questions, which have to be considered instances of *recurso* and *auto*. In this case, the terms where specified, not generalized.

However, the team faced difficulties in reaching consensual decisions. The lack of traceable lines of argumentation was also slowing down the construction of the ontology. The introduction of the Distributed, Loosely-controlled and evolving Engineering of oNTologies (DILIGENT), provided by the AIFB research team, offered a reliable basis for a controlled discussion of the arguments in favor and against modeling decision in this first building stage. A wiki-tool has been used in order to visualize and trace the arguments. The introduction of DILIGENT proved the need to rely on guidelines for the decision-making process within ontology design and the need of appropriate tools for discussion and argumentation.

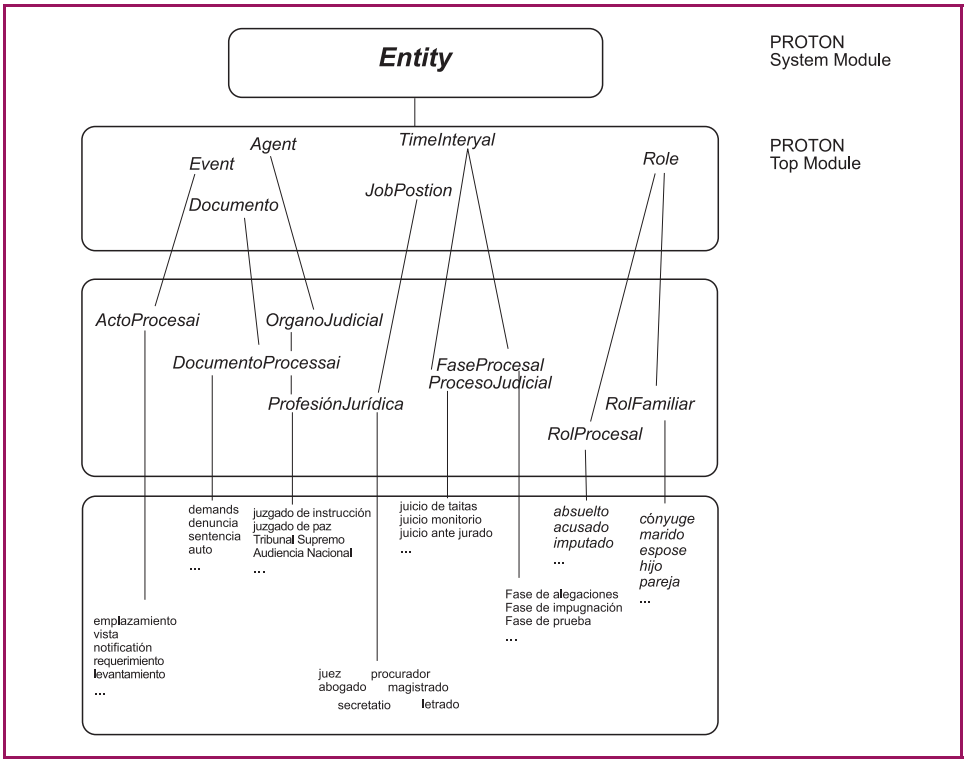
OPJK has, currently, nearly 50 concepts, 100 relations and more than 300 instances. This is result of a choice to minimize the concepts at the class level when possible in favor of creating instances and relations. The ontology is still under construction. There are still 600 questions to be discussed and also, the integration of the legal case study ontology into PROTON (Proto Ontology)[11], as part of the integration of SEKT technology, could generate some constrains towards the engineering process. Figure 8 shows the structure of the three-layer ontology at its present stage.

5. System functionalities and architecture

Functionalities

The software system offers two main functionalities. In the first interaction type the user – the judge – formulates a question using written natural language (Spanish) and retrieves an answer to his/her question. The answer is retrieved using semantic web technology from a

Figure 8 Structure of OPJK: domain, core and upper layers



''Frequently asked questions'' (FAQ) repository, where a set of question-answer pairs is stored (see Figure 9).

This answer is found using the underlying semantic model (OPJK) to compute the input question with natural language processing software (Tokenizer, Morphological Analyser, Basic Shallow Parser and Semantic Analysis). The last part, the semantic analysis, is performed on the bases of the domain ontology calculating the similarity between the user question and all the stored questions in the FAQ repository.

After displaying the computed answer, the system allows for its performance and accuracy feedback. Using a simple interface, the user is able to reject the answer. Since the semantic similarity for questions is calculated using a simple metrics, the system is able to offer the most probable alternative, in addition to generate the corrective actions.

The second interaction type deals with the answer explanation (see Figure 10). The user can ask for supporting documents for any answer the system offers. In this stage the semantic

Figure 9 UML use case for question-answering functionality

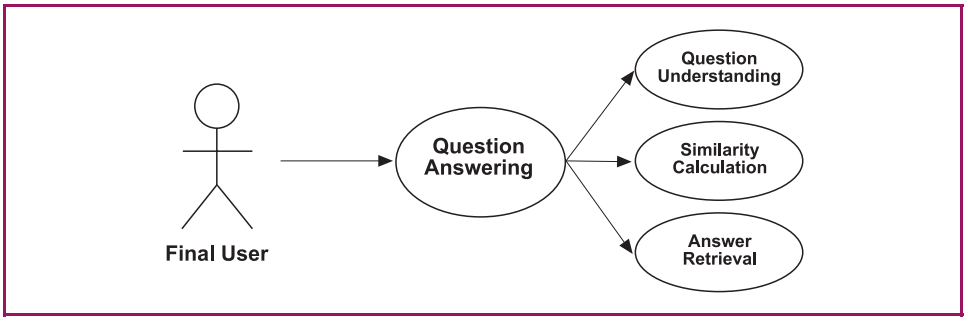
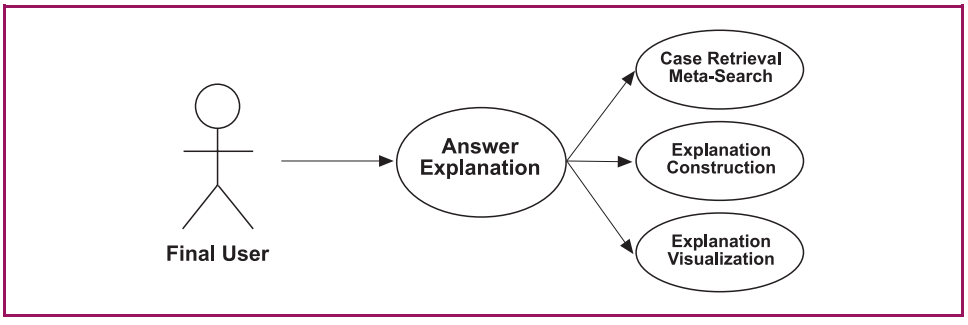


Figure 10 UML use case for answer explanation



search engine browses the case law databases and offers document references to relevant documents. This functionality allows the judge to learn from the cases that have originated the answer or precedent. This functionality can be used directly, as a meta-search engine working upon different case law databases, without the need to ask previously a concrete question.

Automatic document processing and understanding based on ontology mapping and alignment technology facilitate the introduction of case law databases into the system repository. In this way each format representing legal cases is translated into a common schema and processed to establish links to stored FAQ answers.

Conceptual architecture

The final software is a web-based application that retrieves answers to questions in the legal domain. It provides judges access to frequently-asked questions through a natural language interface. The system answers with an ordered list of similar question-answer pairs that might solve the problem of the judge. The advanced question understanding in the legal domain will be performed through the ontology previously described. The offered answers are supported and extended by existing cases extracted from local specific databases.

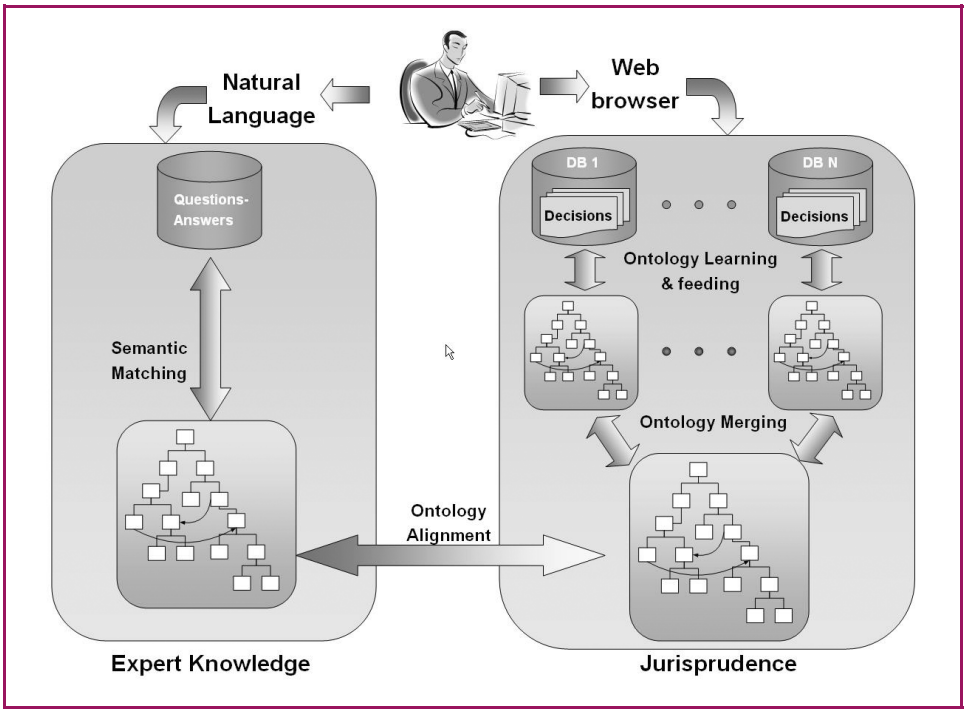
Figure 11 shows in more detail how the system that we are building manages two independent kinds of knowledge.

The system manages the expert knowledge related to judges' experience in the form of a repository of frequently-asked questions and an ontology representing this kind of knowledge (OPJK). This knowledge enables the system to answer the questions posed by judges. This is represented on the left-hand side of the figure. The user accesses the system using a natural language interface, and asks the question as he/she would do with an experienced judge. The question is analysed in order to detect the relevant concepts, using OPJK as background knowledge. The set of concepts obtained is matched against the questions in the repository, to check for the best answers available.

The right-hand side of the figure shows the other kind of knowledge considered in the system, the stored past judicial decisions (case law). As said earlier, legal justification is as important as the content of the ruling the judge is going to make. It is required to know how he/she can justify this decision, who took it before and why. This is exactly the kind of knowledge that is managed here. The application has access to a certain number of case law databases. Each database contains the cases produced by Spanish courts, at different levels. Each of these databases is modeled with different ontologies, that could be merged to obtain a case law ontology.

In order to connect the two kinds of knowledge contained in the judicial experience and the past judicial decisions, and to detect the useful cases to justify the answers in the FAQ repository, the concepts in the two main ontologies have to be aligned. Therefore, if a user selects a justification, the system will check the OPJK concepts appearing in the answer, will transform them into the corresponding set of case law ontological concepts, and eventually retrieve the appropriate cases containing those concepts (see Figure 12).

Figure 11 An overall architecture for the tool



Usability and interfaces

In order to build a scalable and useful FAQ system, the following requirements have been identified:

- Judges should not be bothered with a complex user interface. A simple natural language interface is probably appropriate.
- The system needs to work in a real-time environment. Each answer should be delivered in a very short time period.
- The questions included in the system should be of high quality, e.g. be rather exhaustive and reflect the present situation.

At the present moment, the application is entering the first testing phase, in which a selected group of evaluators from the potential user domain will check its usability. As shown in Figure 13, the web interface is very simple (Google-like). There is one single text box to formulate the questions, and answers are shown as a very simple document, including some control information (relevance, feedback, etc.) (see Figure 14).

Due to an open software design it is also possible to establish another way of accessing the application. In the next figure, an instant messaging interface for the question-answering functionality.

6. Conclusions

In this paper we have depicted the legal scenario in which to develop and apply the SEKT Project technologies. We have elaborated the following aspects:

- *user's profile and corpora*: we offered the preliminary results of the 2004 survey and fieldwork on the Spanish judicial units (focusing on young judges' technological skills) and an overview of the knowledge acquisition process (using ALCESTE, TextToOnto . . .);
- *legal ontology*: we presented an extensive state of the art, explained epistemological grounds, made the proposal for ontologies of professional legal knowledge (OPLK), and detailed the development of an ontology of professional judicial knowledge (OPJK);

Figure 12 Modular composition diagram

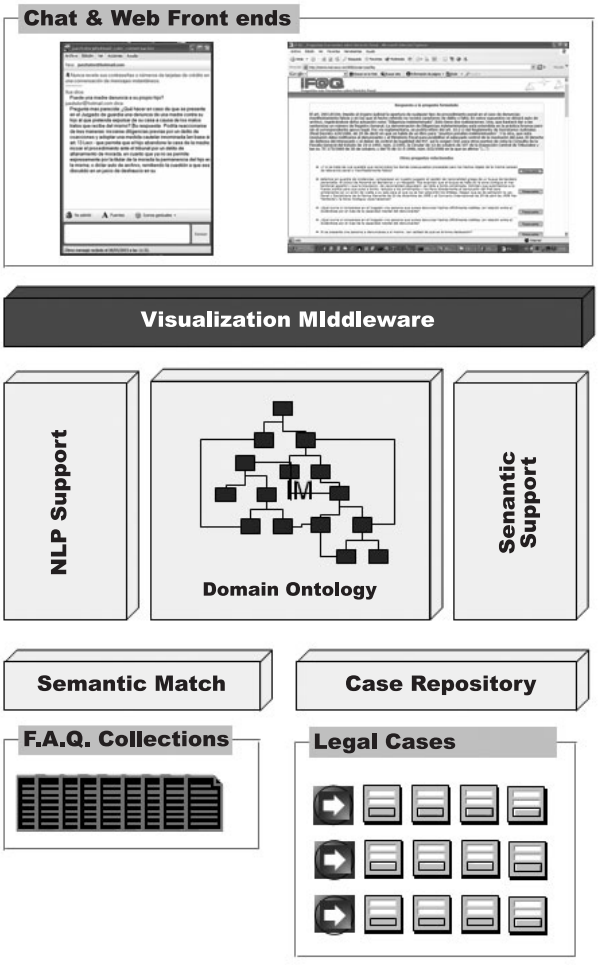
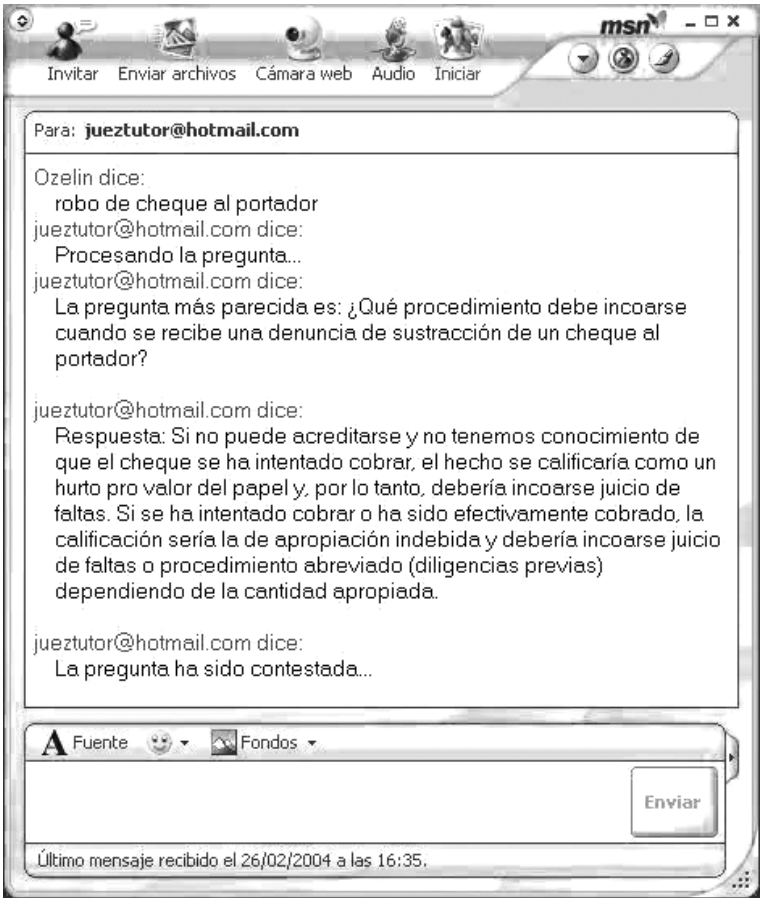


Figure 13 Web interface for question answering

Figure 14 Instance messaging interface for question answering



- *use cases* collection: relation of all use cases from question-answer to Jurisprudence Ontology and OPJK aligning;
- *use case* software design: presentation of the architecture for Iuriservice II.

This second prototype, due mainly to the implementation of SEKT technologies, has the following main features:

- it is designed to be accurate and technologically advanced by using NLP and semantic web techniques;
- it is designed to be efficient, extensible, customable and scalable;
- it makes use of incremental search as a process of narrowing the solicited FAQ set; and
- it uses a variety of pluggable searching algorithms.

Finally, incorporating practical professional knowledge, using OPKJ, not only will improve the performance of the prototype but also it will allow peer-to-peer knowledge sharing and will result in a faster start into judicial practice for judges in their first appointment.

Notes

1. The first survey was performed in 2002 by judges in their training period at the Judicial School – 52nd class – who interviewed their peers of the 49/50 class already at work. The second survey was done in 2004 by our research team and targeted the judges of the 52nd class. This batch of 248 judges had taken office by early 2002, so they had already spent two years at work. Consequently, the 52nd class fitted perfectly the two basic requirements of the ethnography: they were

newly-recruited judges who, at the same time, had spent time enough in office so as to provide researchers with a number of relevant questions regarding daily problems, on-duty periods, and legal procedures at large. To create a relevant sample, we randomly selected 150 judges scattered through the courts of 14 Autonomous Communities (out of 17). The method consisted semi-structured oral interviews with an estimated length of one hour. Judges were granted absolute confidentiality and anonymous treatment of personal data. For more information on Spanish legal culture consult (Escuela Judicial, 2001, 2002, 2003; General Council, 1999/2003; Poblet, 2001; Poblet and Casanovas, 2004).

2. Specific data regarding attitudes of Spanish judges towards ICTs can be drawn from two recent sources. On the one hand, the 2003 Barometer of the Higher Council shows that ''global computerization of the Administration of Justice'' ranks first among the most needed reforms mentioned by judges (81 per cent of them think that this is a very important or rather important issue). On the other hand, data from the 2002 survey to both inexperienced (less than three years in office) and experienced judges (more than four years in office) showed that judges' use of the internet for professional purposes was still low.

3. (see http://kaon.semanticweb.org)

4. TextToOnto is a tool embedded in the Oi-Modeler platform which supports the semi-automatic creation of ontologies by applying text mining algorithms. Although TextToOnto is not currently provided with textual analyzer components in the Spanish language, it is able to identify important concepts and instances and also relevant relations (or associations) between those concepts that the judicial domain ontology has to take into account (http://ontoware.org/projects/text2onto/)

5. ALCESTE is a software used to perform automatic analysis of textual data, developed by Max Reinert (2002, 2003) at the Centre Nationale de la Recherche Scientifique (CNRS). ALCESTE classifies different subsets of a given textual corpus based on a hierarchical descending clustering algorithm. Successive dichotomies are carried out along the first axis of a factor analysis. Therefore, for a relative semantically homogeneous corpus, the program seeks the list of most characteristic words assembling subsets of ''lexical worlds'' according to a chi^2 metric. This has proved to be useful to flesh out the conceptual structure of the judicial competency questions.

6. Although TextToOnto will not be further developed, the Spanish GATE components, result of the SEKT Project work, will be implemented into Text2Onto in a near future, a second version, designed by the same team, with improved features (e.g. will allow the identification of synonyms and meronyms). Then, the competency questions will be analyzed accordingly and more information will be retrieved to refine the existing ontology.

7. See Gangemi and Breuker (2002), Visser and Bench-Capon (1998a, b), van Kralingen *et al.* (1999). We offer a more detailed description of the content of these ontologies in Blázquez *et al.* (2004) and Benjamins *et al.* (2005b). Current legal ontologies have been constructed, among others, within the following EU project frameworks: CLIME (1998-2001), KDE (1999-2001), e-POWER (2001-2003), e-COURT (2001-2003).

8. ''An interpretation is the mapping (semantics) from one application instance (conceptual schema) syntactically described in some language into the ontology base, which is assumed to contain conceptualizations of all relevant elementary facts. [. . .]. The interpretation layer constitutes an intermediate level of abstraction through which ontology-based applications map their syntactical specification into an implementation of an ontology 'semantics''' (Jarrar and Meersman, 2001).

9. McCarty (2002) has formalized Hohfeld's legal fundamental conceptions (1919) to model property rights using deontic logic. Hage and Verheij (1999) have modeled a formal theory of law stemming from the causal or ruled link between events and state of affairs. Gangemi and Breuker (2002) have conceptually represented the law implementation or application process as a relationship between legal normative descriptions and cases,. Boella and van der Torre (2004) have constructed a normative multiagent system based on regulative and constitutive norms.

10. ''Situated cognition is an approach for understanding cognition that seeks to relate social, neural, and psychological views. From the social perspective, situated cognition provides insights about the content of knowledge, namely how people conceive of what they are doing in terms of their contribution to a community of practice and how this affects their attention and priorities over time. From the neural perspective, situated cognition provides insights about the physical structure of knowledge, namely how perception, conception and motor action are related through a

self-organizing coordination process with a memory. From a psychological perspective, situated cognition provides insights about how behaviour is improvised by resequencing and recomposing previous behaviours'' (Clancey *et al.*, 1998). See also Menzies and Clancey (1998), Menzies (1998) and Clancey (2002).

11. Available at: http://proton.semanticweb.org/

References

Abel, R.L. (Ed.) (1997), *Lawyers: A Critical Reader*, The New Press, New York, NY.

Ayuso, M. and Becue, M. *et al.* (2003), ''Jueces jóvenes en España, 2002. Análisis estadístico de las encuestas a los jueces en su primer destino (Promociones 48/49 y 50)'', Internal Report for the General Council of the Judiciary, within the framework of the Project ''Observatory of Judicial Culture'', SEC-2001-2581-C02-01/02.

Benjamins, V.R., Contreras, J., Casanovas, P., Ayuso, M., Bécue, M., Lemus, L. and Urios, C. (2004a), ''Ontologies of professional legal knowledge as the basis for intelligent IT support for judges'', *Artificial Intelligence and Law*, forthcoming.

Benjamins, V.R., Contreras, J., Blázquez, M., Rodrigo, L., Casanovas, P. and Poblet, M. (2004b), ''The SEKT use legal case components: ontology and architecture'', in Gordon, T.B. (Ed.), *Legal Knowledge and Information Systems: Proceedings of Jurix 2004*, IOS Press, Amsterdam, pp. 69-77.

Benjamins, V.R., Casanovas, P., Contreras, J., López-Cobo, J.M. and Lemus, L. (2005a), ''Iuriservice: an intelligent frequently asked questions system to assist newly appointed judges'', in Benjamins, V.R., Casanovas, P., Breuker, J. and Gangemi, A. (Eds), *Law and the Semantic Web*, Springer-Verlag, London and Berlin, pp. 205-22.

Benjamins, V.R., Casanovas, P., Breuker, J. and Gangemi, A. (Eds) (2005b), *Law and the Semantic Web*, Springer-Verlag, Berlin.

Blázquez, M., Rodrigo, L., Casanovas, P. and Poblet, M. (2004), ''Legal case study before analysis'', Deliverable no. 10.1.1, Semantically Enabled Knowledge Technologies, EU-IST Project IST-2003-506826 SEKT.

Boella, G. and van der Torre, L. (2004), ''Regulative and constitutive norms in normative multiagent systems'', available at: http://di.unito.it/PS/Boella-vanderTorre-kr04.pdf

Breuker, J., Elhag, A., Petkov, E. and Winkels, R. (2002), ''Ontologies for legal information serving and knowledge management'', *Proceedings of Legal Knowledge and Information Systems: Jurix 2002: The Fifteenth Annual Conference*, IOS Press, Amsterdam, pp. 73-82.

Breuker, J. and Winkels, R. (2003), ''Use and reuse of legal ontologies in knowledge engineering and information management'', paper presented at ICAIL 2003 workshop on legal ontologies and web-based information management, 24-28 June, Edinburgh, UK, available at: www.lri.jur.uva.nl/(winkels/legontICAIL2003.html

Bromme, R. and Tillema, H. (1995), ''Fusing experience and theory: the structure of professional knowledge'', *Learning and Instruction*, Vol. 5, pp. 261-7.

Casanovas, P. (1994), ''Towards a sociopragmatics of legal discourses. Discourse and decision making in an Aurelian court'', in Jackson, B.S. (Ed.), *Legal Semiotics and the Sociology of Law: Oñati Proceedings no. 16*, Oñati International Institute for the Sociology of Law, Gipuzkoa, Spain, pp. 151-225.

Chan, C.W. (2003), ''Cognitive modeling and representation of knowledge in ontological engineering'', *Brain and Mind*, Vol. 4, pp. 269-82.

Clancey, W.J., Sachs, P., Sierhus, M. and van Hoof, R. (1998), ''Brahms: simulating practice for work systems design'', *International Journal of Human-Computer Studies*, Vol. 49, pp. 831-65.

Clancey, W.J. (2002), ''Simulating activities: relating motives, deliberation, and attentive coordination'', *Cognitive Systems Research*, Vol. 3, pp. 471-99.

D'Andrade, R. (1995), *The Development of Cognitive Anthropology*, Cambridge University Press, Cambridge.

Eraut, M. (1992), ''Developing the knowledge base: a process perspective on professional education'', in Barnett, R. (Ed.), *Learning to Effect*, Open University Press, Buckingham, pp. 98-118.

Escuela Judicial (2001), *Datos Estadísticos de los Jueces en Prácticas (Curso 2000-2002)*, Escuela Judicial, Barcelona, Spain.

Escuela Judicial (2002), *Datos Estadísticos de los Jueces en Prácticas (Curso 2001-2003)*, Escuela Judicial, Barcelona, Spain.

Escuela Judicial (2003), *Datos Estadísticos de los Jueces en Prácticas (Curso 2002-2004)*, Escuela Judicial, Barcelona, Spain.

Gangemi, A., Pisanelli, D.M. and Steve, G. (2001), ''A formal ontology framework to represent norm dynamics'', *Proceedings of Second International Workshop on Legal Ontologies*, Amsterdam, The Netherlands.

Gangemi, A. and Breuker, J. (2002), ''Harmonizing legal ontologies'', in Gangemi, A., Guarino, M. and Doerr, M. (Eds), Deliverable 3.4: IST-Project-2000-29243, available at: http://ontoweb.aifbb.uni-karlsruhe.de/About/Deliverables

General Council of the Judiciary (2003), ''Opinion Barometers 1999 and 2003'', available at: www.poderjudicial.es

Gómez-Pérez, A., Corcho, O. and Fernández-López, M. (2004), *Ontological Engineering*, Springer-Verlag, London and Berlin.

Gray, P.N. (1997), *Artificial Legal Intelligence*, Dartmouth, Aldershot.

Grüninger, M. and Fox, M. (1995), ''Methodology for the design and evaluation of ontologies'', *Proceedings of the Workshop on Basic Ontological Issues in Knowledge Sharing: IJKAI-95*, Montreal, Canada.

Gu, T., Pung, H.K. and Zhang, D.Q. (2005), ''A service-oriented middleware for building context-aware services'', *Journal of Network and Computer Applications*, Vol. 28 No. 1, pp. 1-18.

Hage, J. and Verheij, B. (1999), ''The law as a dynamic interconnected system of states of affairs: a legal top ontology'', *International Journal of Human-Computer Studies*, Vol. 51 No. 6, pp. 1043-77.

Hysyam, M. and Choudrie, J. (2004), ''The diffusion of tacit knowledge and its implications on information systems: the role of meta-abilities'', *Journal of Knowledge Management*, Vol. 8 No. 2, pp. 128-39.

Jarrar, M. and Meersman, R. (2001), ''Practical ontologies and their interpretations in applications: the DOGMA experiment'', available at: http://starlab.vwb.ac.be/publications/STAR-2001-04.pdf

Kavanagh, D. and Kelly, S. (2002), ''Sensemaking, safety and situated communities in con(temporary) networks'', *Journal of Business Research*, Vol. 55, pp. 583-94.

Kim, J. and King, J. (2004), ''Managing knowledge work: specialization and collaboration of engineering problem solving'', *Journal of knowledge Management*, Vol. 8 No. 2, pp. 53-63.

Koskinen, K.U. (2004), ''Evaluation of tacit knowledge utilization in work units'', *Journal of Knowledge Management*, Vol. 7 No. 5, pp. 67-81.

McCarty, L.T. (1989), ''A language for legal discourse, I. Basic features'', *Proceedings of the Second International Conference on Artificial Intelligence and Law*, Vancouver, Canada, pp. 180-9.

McCarty, L.T. (2002), ''Ownership: a case study in the representation of legal concepts'', *Artificial Intelligence and Law*, Vol. 10, pp. 135-61.

Mentzas, G., Apostolou, D., Young, R. and Abecker, A. (2001), ''Knowledge networking: a holistic solution for leveraging corporate knowledge'', *Journal of Knowledge Management*, Vol. 5 No. 1, pp. 94-107.

Menzies, T. (1998), ''Towards situated knowledge acquisition'', *International Journal of Human-Computer Studies*, Vol. 49, pp. 867-93.

Menzies, T. and Clancey, W.J. (1998), ''Editorial: the challenge of situates cognition for symbolic knowledge-based systems'', *International Journal of Human-Computer Studies*, Vol. 49, pp. 767-9.

Nicolas, R. (2004), ''Knowledge management impacts on decision-making process'', *Journal of Knowledge Management*, Vol. 8 No. 1, pp. 20-31.

Paliwala, A. (2001), ''An intellectual celebration: a review of the Jurix legal knowledge based system scholarship'', *Artificial Intelligence and Law*, Vol. 8, pp. 317-35.

Pinto, S., Staab, S. and Tempich, C. (2004), ''DILIGENT: towards a fine-grained methodology for distributed loosely-controlled and evolvIng engineering ontologies'', *Proceedings of ECAI-2004*, Valencia, Spain.

Poblet, M. (2001), ''Spanish legal culture. Between the state and society'', doctoral dissertation. Stanford Law School, Stanford, CA.

Poblet, M. and Casanovas, P. (2004), ''Recruitment, professional evaluation and career of judges and prosecutors in Spain'', unpublished report within the framework of the project ''Recruitment, professional evaluation, and career of judges and prosecutors in Austria, Germany, France, Holland, Italy and Spain'', of IRSIG-CNR, Bologna, Italy.

Reinert, M. (2002), ''La tresse du sens et la méthode Alceste: application aux 'rêveries du promeneur solitaire'', *Proceedings of JADT 2000: 5es Journées Internationales d'Analyse Statistique des Données Textuelles*.

Reinert, M. (2003), ''Le role de la répéttion dans la representation du sens et son approche statistique par la méthode ALCESTE'', *Semiotica*, Vol. 147 No. 1, pp. 389-420.

Roche, Ch. (2000), ''Corporate ontologies and concurrent engineering'', *Journal of Materials Processing Technology*, Vol. 107, pp. 187-93.

Stamper, R.K. (1991), ''The role of semantics in legal expert reasoning and legal systems'', *Ratio Juris*, Vol. 2 No. 4, pp. 219-44.

Stamper, R.K. (1996), ''Signs, information, norms and systems'', in Holmqvist, B. and Andersen, P. (Eds), *Signs of Work*, De Gruyter, Berlin.

Valente, A. (1995), *A Modelling Approach to Legal Knowledge Engineering*, IOS Press, Amsterdam and Tokyo.

van den Hoof, B. and de Ridder, J.A. (2004), ''Knowledge sharing in context: the influence of organizational commitment, communication climate and CMS use on knowledge sharing'', *Journal of Knowledge Management*, Vol. 8 No. 6, p. 118.

van Kralingen, R.W. (1995), *Frame-based Conceptual Models of Statute Law*, Computer/Law Series, 16, Kluwer Law International, The Hague.

van Kralingen, R.W., Visser, P.R.S., Bench-Capon, T.J.M. and van Den Herik, J.H. (1999), ''A principled approach to developing legal knowledge systems'', *International Journal of Human-Computer Studies*, Vol. 51 No. 6, pp. 1127-54.

Visser, P.R.S. (1995), *Knowledge Specification for Multiple Legal Tasks*, Computer/Law Series, 17, Kluwer Law International, The Hague.

Visser, P.R.S. and Bench-Capon, T.J.M. (1998a), ''A comparison of two legal ontologies'', available at: http://webster.cs.uga.edu/ ~ maduko/SemWeb/comparisonOfTwoLO.pdf

Visser, P.R.S. and Bench-Capon, T.J.M. (1998b), ''A comparaison of four ontologies for the design of legal knowledge systems'', *Artificial Intelligence and Law*, Vol. 6, pp. 27-57.

Vouros, G.A. (2003), ''Technological issues towards knowledge-powered organizations'', *Journal of Knowledge Management*, Vol. 7 No. 2, pp. 114-27.

Further reading

Breuker, J. (2003), ''The development and use of ontologies of criminal law in the e-Court European project'', available at: http://cbke.prawo.uni.wroc.pl/PDF/wroclaw-2003-paper-effe.pdf