

LKMS – A Legal Knowledge Management System Exploiting Semantic Web Technologies

Luca Gilardoni, Chistian Biasuzzi, Massimo Ferraro, Roberto Fonti,
and Piercarlo Slavazza

Quinary - Via Pietrasanta 14 – 20141 Milan – Italy
{gil,bic,fem,for,slp}@quinary.com

Abstract. Semantic Web, using formal languages to represent document content and providing facilities for aggregating information spread around, can improve the functionalities provided nowadays by KM tools. This paper describes a Knowledge Management system, targeted at lawyers, which has been enhanced using Semantic Web technologies. The system assists lawyers during their everyday work, and allows them to manage their information and knowledge. A semantic layer has been added to the system, providing capabilities that make system usage easier and much more powerful, adding new and advanced means for create, share and access knowledge.

1 Introduction

After years of hype, there is clear evidence of an up-take of knowledge management in corporations. Today, knowledge is recognized as a strategic resource, with major key drivers being the need to cut time to market and the fear of missing business opportunities in a global market where companies have to cope with new products and services. At the same time, there is a general acknowledgement that existing technology behind most knowledge management products has somehow reached its limits. Current knowledge management systems are indeed still mostly built on top of conventional document management systems, without real 'understanding' layers. Tools are mostly designed as aids to human centered activities with a set of low level tools needing human guidance to deliver results [1]. Albeit relying on web technologies, built as intranet portal tools, current state of the art does not really leverage the expected potential of the semantic web. To go to the next step, we should move towards a architecture and an infrastructure providing a foundation for new generation services, semantically aware tools and proactive agents, able to better support human actors.

Technology born to support the development of the Semantic Web may be used to build such foundation. Moreover, the 'inside web', that is the web constituted by intranets, KM environments, portals, is worth to users as much – and often more – than the 'web out there'. If we consider the whole world of professional users within corporation – and to some extent even some virtual community build within closed spaces – we find out a huge amount of information available. The fact that such information is not available to the general public is scarcely relevant, as whenever we consider ourselves, we find that, in our space of accessible information, the outer web

and inside web often play an equally relevant role. What's even more notable is that the relevance is often in our capacity to connect internal information and external one. A collaborative environment, such as those currently found behind most intranets, could provide a natural place to add semantic capabilities, while the organization work which is behind most intranet initiatives provides the economic support and impulse to add what's needed – organized information and knowledge – to implement the semantic layer.

The system described in this paper is centered around this assumption. Based on a long experience in building advanced knowledge management systems, and derived from research made in the framework of the Dot.Kom project ([2]), we built an enhanced solution integrating such a semantic layer into an existent KM environment.

The semantic layer is founded over an ontology repository supporting knowledge integration and fusion and acting as the common glue for share and reuse services for knowledge management. Ontologies indeed play a key role in the context of the Semantic Web: they formalize the knowledge about the concepts related to the “world” of interest. Once the proper knowledge framework is defined through ontologies, one can identify in documents instances of the concepts described, and relations between them. Accordingly, they also play a key role in supporting knowledge management tools, making them a bit more ‘knowledge aware’. Semantic annotations, whether manually generated, or derived by information extraction techniques or other automatic processes, can provide a major framework for generating, preserving and sharing knowledge. Annotations provide the basis for advanced information retrieval, and for providing proactive services.

We will describe here a specific vertical solution targeting law firms. The described system has already been deployed in a major Italian law firm, and is currently a key component of our company offering for the legal market.

2 Knowledge in the Legal World

Law is a knowledge-based profession. Since law firms and law departments are knowledge-based organisations, knowledge management becomes critical to their continuing success. A knowledge management system enables lawyers to work more efficiently and to provide legal services quicker than ever before. By creating processes to support and facilitate the identification, capture and dissemination of a firm’s knowledge, knowledge management systems leverage a law firm collective wisdom.

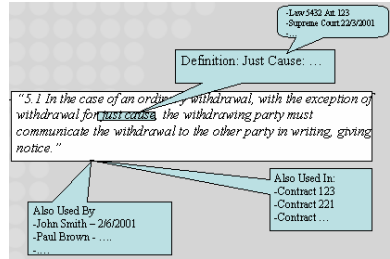
The legal industry has faced significant pressures in recent years, making knowledge management a business imperative. In the age of instant communication, lawyers have been forced to find quicker ways to deliver traditional legal services. Law firm clients have become very sophisticated buyers of legal services and therefore they expect a faster turnaround time.

Several components of knowledge management, such as precedent libraries or work product repositories, already exist in law firms. Innovative law firms however are already working to find a more efficient way to work, leveraging the knowledge of their experts by delegating work to more junior staff and hence looking for better ways to improve knowledge sharing and exploitation processes.

Work of professionals within a law firm – or a legal department in a corporation - ultimately leads to production of documents: acts, contracts or opinions. In this sense work processes are document centric. This is one of the reasons why most KM solutions targeting law professionals focuses on document management issues. However, from the point of view of knowledge building and sharing, what is really relevant is the intellectual process carried on to delivery the document.

The outcome of this process is constrained on one side by the task and the specific matter, from the other by contextual knowledge.

The context should be maintained, because it is this contextual knowledge that enables, for example, to maintain and revise documents (. this clause was made this way because of that law ...; if the law is later amended, or a different interpretation given by the supreme court, that clause has to be revised in future contracts and effects on old contracts has to be evaluated). Legal documents, moreover, are inherently interrelated; and so may be the process that leads to them. A contract – legal act binding two or more parties - may be designed taking legal opinions into account, and it may be in turn the source of a case which leads to production of legal acts discussing it, these in turns taking other opinions into account. Legal opinions - written by lawyers on request of customers who need advise on some topic - are also based in turn on decisions taken in courts on specific cases. Being able to keep track of context hence results to be of paramount importance.



The context is given as well by the work process. Lawyers, as many other professionals, are compelled by the need to share knowledge and competencies. Findings derived by analysis of a court sentence have to be somehow saved for usage by other members of the firm. Too many times people end up redoing the same work as the guy next door in the office. Even a search made against a database looking for specific cases may be reused in similar cases.

The more a law firm tends to specialize in specific sector, which is often the case, the more sharing this kind of knowledge gives the competitive advantage. Specific law firm knowledge is so relevant that the area of practice a law firm is specialized in highly qualifies the firm. A primary concern and a major activity directly functional to the primary process is therefore to keep this background knowledge up to date through knowledge maintenance processes.

The more a law firm tends to specialize in specific sector, which is often the case, the more sharing this kind of knowledge gives the competitive advantage. Specific law firm knowledge is so relevant that the area of practice a law firm is specialized in highly qualifies the firm. A primary concern and a major activity directly functional to the primary process is therefore to keep this background knowledge up to date through knowledge maintenance processes.

One way to share this knowledge is by similarity, which is often the approach taken by case based reasoning systems, which sometimes work rather reasonably. Point is that current tools reason by similarity at the textual level, which may work reasonably to find out similarity in matter (e.g. two employment contracts for managers with similar bonus plans) but can hardly support in linking at the clause level (to stay within contracts) or to maintain connections to supporting cases. Similarity is not taken to the conceptual level, and hidden links and background knowledge are ... just treated as hidden.

Hence knowledge management system to prove effective must support, other than ‘conventional’ search, a way to annotate and hyperlink elements to the surrounding

context, and be able to navigate and search hyperlinks. To be usable, however, the system must be able to support automatic (at least partially) hyper-linking, and make easy to manage annotations.

A rich knowledge layer and semantic web technologies provide the foundation to enhance existing knowledge sharing environments supporting these functionalities. In a sense, this is not surprising, as the same rationale (adding a semantic layer to enhance sharing providing a better user experience) is also behind the Semantic Web. Moreover, in a world where more and more information is going online, and where a number of public initiatives (e.g. NormeInRete, see [3]) are strongly driving to make available public legal information on the web, the more the technology supports integration of internal law firm material (the 'inside web') with external material, the more users are likely to take advantage of it.

3 System Description

LKMS (Legal Knowledge Management System) is a collaborative web-based platform for knowledge management, supporting law firms in managing a document base and the processes around it. LKMS is a vertical solution for the legal market built on top of “K@”, a generic KM system developed by Quinary since 2002.

With LKMS users can access and share a common repository of documents while the system keeps track of people interaction. Documents, including both physical documents residing inside the law firm, external URLs, notes and Wiki pages, may be organized according to one or more taxonomies, supporting multiple inheritance (DAGs): the environment provides a basic framework for sharing information by matching the way an organization is structuring its processes.

The core system supports browsing and searching using free text queries and provides a number of tools to track user behavior (who added a document or a node in a taxonomy, who added classification links between nodes and taxonomies, who visited nodes or read documents) to facilitate sharing and keeping track of workgroup activities.

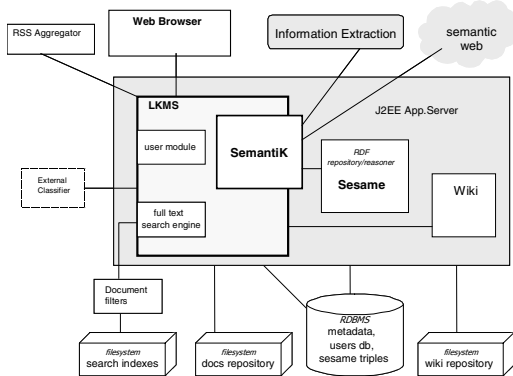


Fig. 1. LKMS Architecture

The system also includes a document drafting component, based on XML technology, enabling to build template documents as aggregate of clauses and composition logic. Drafts can then be built from templates by specifying constraints through a user friendly query answer interface.

The core version of LKMS has been enhanced in 2004, now it is able to maintain the association between documents and semantic annotations with respect to a formal ontology according to Semantic Web standards. Figure below outlines the overall system architecture.

Semantic Layer Overview. The core KM framework has been enriched with the SemantiK plugin to provide a semantic layer over documents repository. SemantiK is a platform featuring presentation, editing, integration, and searching of knowledge expressed through the RDF language. SemantiK has been integrated with LKMS as a plugin, allowing for connecting annotations to documents on evidence that, in most cases, annotations are motivated by or related to document content.

Storage and inferencing over annotations is given by an underlying RDF repository (Sesame, see [4]). The main purpose of SemantiK is to provide a middleware for high-level access to an RDF knowledge base, supported by a knowledge integration layer, and a web GUI for maintenance of RDF annotations tailored to end users habits. Main functionalities include:

- ontology supported GUI for viewing, browsing and editing annotations through web forms;
- support for semantic search;
- support for knowledge integration and resources disambiguation;
- interface and support for automatic annotation extraction from documents.

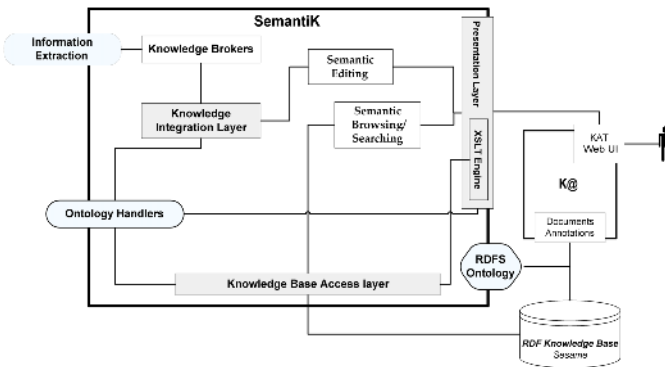


Fig. 2. SemantiK Internal Architecture

The whole architecture has been built to tackle flexibility in handling of specialized ontologies, with respect to both presentation issues and semantic integration. LKMS comes with a precompiled legal ontology, expressed in RDFS¹, which may however

¹ Enhanced with few custom meta-properties – porting to OWL is being considered.

be extended. The ontology is complemented by Ontology Handlers, a set of Java classes supporting specializations dependent on specific entities.

Moreover, Information Extraction services enable to extract annotations from documents or, more in general, fetch them from external data sources. Plugins has been developed for a number of information extraction systems.

The Ontology. The general legal ontology shipping with LKMS was created in cooperation with our first customer and has been only slightly modified since then (only manual adaptation is handled so far).

The ontology covers three main areas.

Laws. This area includes most of the subdivision of the Italian legislation. It also copes with law's articles and articles' subparts. Each concept is characterized by a minimal set of properties aiming at uniquely identifying each concept, such as law date and number. Given this information it is then also possible to compute URNs, that is a standard unique identifier used, in this case, in order to build links to the public site NormeInRete ([3]). Besides, name, description, source and references to other laws can be defined.

Legal Documents. This part of the ontology describes different kind of legal documents: contracts, legal Opinions, Sentences from the different kind of Italian or European Courts, Regulations, Decisions, legal doctrine, etc. Aside, we also include entities describing actors – i.e. organizations such as Tribunals and Judges, and other supporting entities (e.g. grades of legal cases, possible outcomes of “Supreme Court Decisions” etc).

Juridical Concepts. These kinds of concepts are expected to take over the glossary keywords normally used by lawyers, enabling annotating content on the basis of relevant matters. They have been derived transforming a digital glossary from a book about labor law into a structured and organized ontology. It resulted in a complex hierarchy of 1442 Legal Keywords, with references to each other and to more than 3500 laws or regulations automatically extracted from the same book too. An RDF representation of all instances was created and uploaded in the system, and at the same time patterns (a JAPE grammar) were automatically generated to support a Legal NEA IE tool – described later.

Knowledge Handlers. SemantiK has been tailored to the legal domain by providing a set of Ontology Handlers matching the legal concept classes described in the previous paragraph.

An Ontology Handler is a Java class that is bound to some RDFS class and that is in charge of handling a number of actions regarding instances of that RDFS class – such as rendering, searching, knowledge integration. The Java hierarchy must of course respect the RDFS hierarchy: this way, specialization of actions can be achieved straightforwardly. The root of the hierarchy is a Java class that by default is responsible for the instances of the RDFS class Resource – that is, of all resources in the KB.

Ontology Handlers may provide customization such as to find and match legal documents by number and date (as opposed for example to judges, matched primarily by name and surname), to render links to external resources for laws, and to generate automatic label for structured legal documents.

The Presentation Layer. Given a set of triples all having as subject a certain resource (in particular an annotation associated to a document in LKMS), the SemantiK presentation layer is able to render it (for viewing or editing) in a domain dependent way by means of XSL transformations, applied to XML-ization of the triples.

Using custom CSS we have been able to harmonize the output with web-application environment that constitutes the user interface of LKMS.

SemantiK uses custom meta-properties associated to RDF Resources and Properties for defining a number of presentation details like properties order, visibility and cardinality.

Moreover SemantiK can highlight annotated text in documents, given that the RDF resources annotated have an offset –automatic Information Extraction tools described later on provide such offsets. Annotations are given different colors to distinguish RDF Classes and are hyperlinked for fast querying.

Annotation Production and Knowledge Integration. Data can be inserted in SemantiK by manual editing or by means of external IE engines. In both cases, before instances are added to the Knowledge Base, they are passed through the SemantiK integration layer, which is in charge of detecting whether the intended resources exist already in the KB.

Manual editing is supported by the ontology management module: when creating new instances, the user is asked to choose the *type* of the new resource if more than one is possible: the RDFS classes are displayed in a tree-like manner in a listbox, and possibly some branches of the tree are collapsed (the behavior can be set using meta-properties). The wizard-like UI allows the user for expanding the class tree until the right type is found.

When the user is editing an annotation, given a certain property, he is asked to formulate a query in order to find in the Knowledge Base the intended instances. The query is dispatched to the proper Ontology Handler depending on the range of the property. In general, each Ontology Handler is responsible for parsing the query (for it could have a peculiar syntax) and then trying to use some specific method or heuristic in order to find some results, as already mentioned above.

In general, a fuzzy measure of closeness is computed between the candidate resource and the existing ones; then, the user is presented with the closest resources (if any) and is asked for disambiguate his intent. Note that the usage of a fuzzy match allows coping with misspelled words.

In the case of automatic annotation using a Knowledge Broker, the Ontology Handlers try to automatically defuzzify the closeness measures using a proper threshold.

Anyway, given some existing resource, at any moment the user is able to merge it with other existing resources, which can be selected from a list automatically generated of possibly similar resources, or manually searched.

Semantic Annotation Extraction. The Legal IE Application is the module in charge of automatically extracts annotations from documents. It is targeted at the legal domain and its task is to get as input a document, extract references to legal documents and juridical concepts, and get them back in RDF format (referring to the legal ontology).

SemantiK is responsible for calling the Legal IE Application passing a plain text version of the document (conversion from PDF, MS Word and RTF is supported by the base system) and for integrating the resulting RDF in the semantic repository. The Legal IE Application, wrapped in a web service based on AXIS [9], performs analysis and, using the legal ontology as reference ontology, returns RDF annotations. The analysis may be based on different engines. We integrated and tested a GATE [5] based NEA, Amilcare [6] and TIES [7].

The Legal IE Application uses GATE as processing framework and each external processing component is connected using a GATE component. There are four GATE components: Linguistic component, which uses Italian Linguistic tools to perform the basic linguistic analysis (tokenization, POS tagging, lemmatization); Legal NEA component, which uses JAPE grammar [8] and performs a NE analysis focused on legal entities; Amilcare component and TIES component, which use respectively Amilcare and TIES to perform IE processing.

The application is configurable to possibly use only some of the components (e.g. there can be different legal IE applications working concurrently, using different components).

Document Similarity Measures. An algorithm for computing documents similarity basing on RDF annotation has been developed in SemantiK. Roughly speaking, comparing annotations is accomplished by recursively following resources properties and values and counting matching values. The obtained measure is used to suggest similar documents motivated by semantic similarity. Experiments are ongoing to exploit the similarity measure for automatic document classification.

4 LKMS in Use

LKMS environment was designed in order to support a lawyer in his everyday work within a law firm. Accordingly, the system is used to maintain, in a centralised repository, and share between lawyers a number of documents, both coming from outside the law firm and produced inside.

By supporting indexing by semantic content, LKMS enables to perform better searches, based not simply on full text but on semantic content – or, better, on both. Moreover, it becomes possible to support better automatic classification, and to develop triggers based on content; hence people working on specific cases can be signaled about changes in relevant regulations or news about the specific matter.

Next picture shows the situation where a lawyer, browsing LKMS within a specific legal case, comes across a significant document. Starting from the SemantiK document info panel on top, the user can examine the annotations (in this example a number of references to laws). Then the lawyer can follow the hyperlinks leading to more specific information about a particular annotation (an article of law) and, possibly, to other useful information (juridical concepts) or related documents. Please note that in the example the relation between the source document and the one ‘discovered’ by following the semantic links is derived through the semantic layer through common references, while in a traditional KM system or DMS, such link should have been explicitly stated.

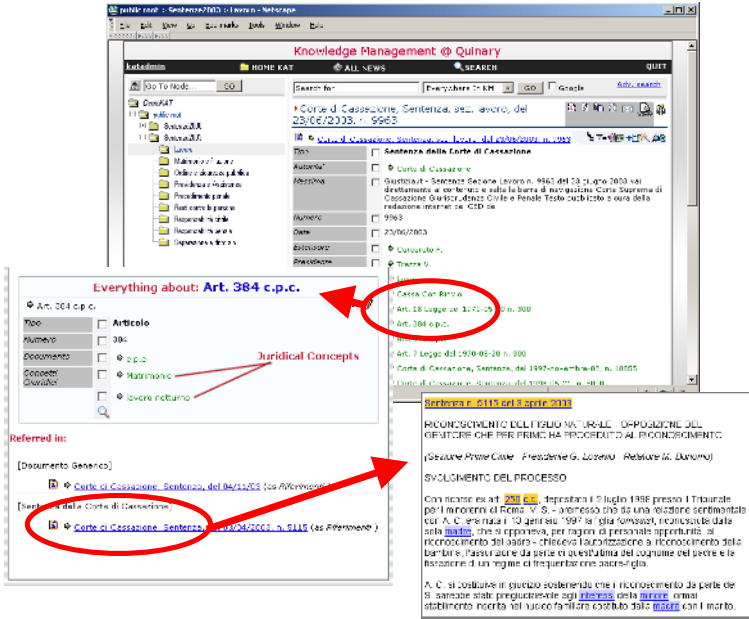


Fig. 3. SemantiK Browsing

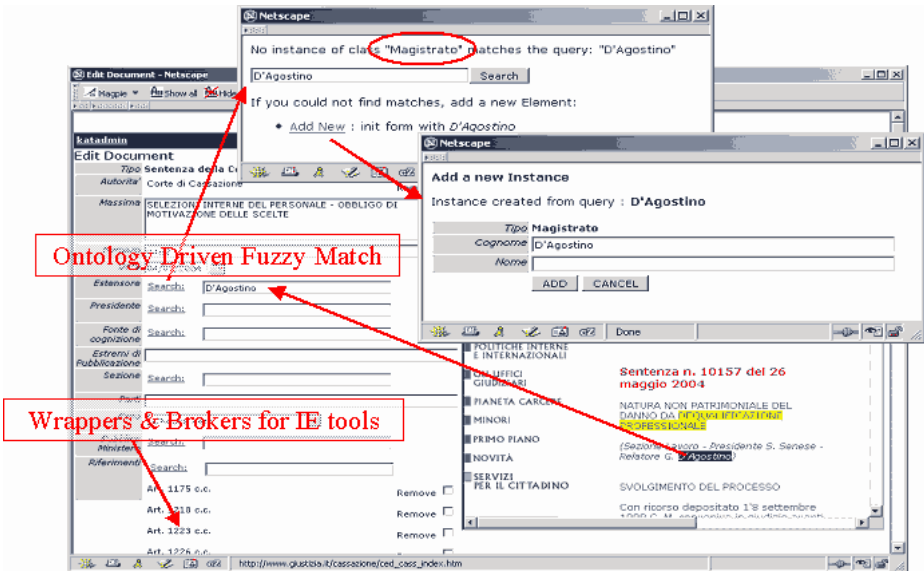


Fig. 4. Sentence Annotation

Annotation such as the ones just shown can derive from both manual annotation and semi-automatic information extraction processes. In this scenario a lawyer imports the sentence in the system, and gets a partial annotation of the new content, making use of some of the available IE modules. The lawyer may then further edit the annotation in SemantiK, removing mistakes and adding missed references.

An essential requirement to reach the LKMS goals introduced above is the development of tools supporting annotation of new documents inserted in the repository, keeping the complexity of the semantic representation behind the scenes.

Our feeling is that with SemantiK we are going in the right direction. The tool supports a lawyer with a point and click, workflow-based user interface, driving the user to a progressive refinement of the annotation with respect to our legal ontology.

In figure 4, editing of annotations is done within the browser having side by side the form based annotation panel on the left and the text with annotations highlighted in the context of the document on the right, rendered as hyperlink to the system knowledge base. Defining new annotations (e.g. adding a legal concept reference, or one of the expected attributes of the sentence) may be done very simply by dragging text elements from right to left. SemantiK features for fuzzy matching and the possibility to exploit ontological information to drive the GUI properly constraining input values, enable to minimize user burden in obtaining a rich and precise annotation.

Founding on a semantic layer also paves the way to development of new drafting tools, able to associate pertinent matter while editing legal documents. We are currently experimenting with tools enabling in-place annotation of text, such as AktiveDoc [26], developed during the course of Dot.Kom project by the University of Sheffield.

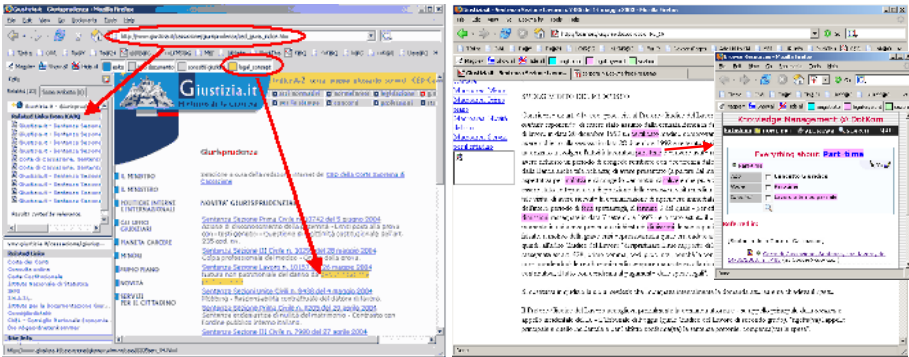


Fig. 5. Enhanced Semantic Web Surfing – Kzilla on the left, Magpie on the right

LKMS’s semantic knowledge base can be exploited proactively when the lawyer is browsing the web, either following links from online journals or institution web sites (e.g. during normal monitoring activities of selected sources) or actively searching for something relevant for a case he is working on. Tools like Magpie [24], a browser add-on that uses an ontology infrastructure to semantically markup web documents on-the-fly - which we tested integrated in SemantiK deriving automatically references for Juridical Concepts from our ontology - may support the user in making sense of

browsed pages on the web against internal knowledge, highlighting references to annotated material of LKMS stored documents (see **Fig. 5**). By clicking on the highlighted concept the user can follow the contextual link and browse the concept ontology in LKMS, reaching documents already in the system related to that specific concept and helping him to quickly make sense of new material. For the same purpose, we also developed Kzilla, a web browser plugin enabling to match browsed material against content of the LKMS internal repository, much alike what services like Alexa do for generic browsing.

LKMS Evaluation. Currently LKMS has been deployed in an Italian legal firm grouping about 50 lawyers. LKMS hosts now more than 30000 documents. Most material is in PDF - a large number however constituted by scanned images - and Word format, plus a number of XML documents generated by an automated drafting systems, a number of simple textual notes and a number of html documents or URLs (specialized publishers, newspapers, Italian and European public institutions).

Material includes legal documents produced within the firm or by other parts in cases, significant legal documents (cases, opinions etc) gathered from different sources, plus a number of general documents from newspapers. Material partly comes in from batch imports from older repositories, partly from daily work, with a minor part gathered by a specialized spider. Some types of documents (e.g. news from press and some legal related publication) is added by clerks and later on classified/annotated by lawyers, other are added by lawyers directly into the system or indirectly coming from the case management system (e.g. docs from corresponding parties).

The first LKMS version enhanced by the semantic layer SemantiK has been deployed during fall 2004. Currently an average of 50 docs are added every day, 25% of which are manually annotated. Most of them only hold generic metadata such as authors or sources, but there are also more than 400 references through laws or articles, about 150 sentences of various kinds and about 100 acts, laws, circulars etc, for a total of about 30.000 triples in the RDF repository.

The strong directives issued from the law firm management about having richly annotated material are by itself a clear sign of usefulness judged by end users eyes.

A recent experiment has been made in the legal firm: focusing on a specific topic ('non-competition pact'), 8 younger lawyers have been 'commanded' to collect selected material on the subject, add it to LKMS and properly annotate it. The experiment took 96 hours of work - including collecting and analyzing material - and resulted in 117 documents, properly annotated against the legal ontology, and enriched with references to laws, authors, keywords and so on. While manual annotation proved a daunting task, such work was judged useful and worth the effort from the senior partner's side.

Automatic Annotations. It is well evident that automatic IE support is what's needed to step up, as manual annotation may be accepted (and has been accepted - once suitable support for minimizing the burden has been put in place!) for documents where IE is not feasible (e.g. scanned documents), in case of major features (e.g. main metadata for a Supreme Court Sentence), full annotation on all material is out of question.

Usage of tools and IE techniques has however been explored and developed by the authors in the framework of the Dot.Kom project. An automatic information extraction application has been derived and already tested, and it is likely to be setup in production in short time. During the test, made on real data from the law firm, the system produced a number of automatically extracted annotations on a selected corpus of 1200 documents. 5800 references were extracted to a total number of about 250 laws, opinions or articles. 1150 articles to laws were identified. The repository was filled with 70000 triples.

The automatically extracted annotations have been evaluated qualitatively on random elements and judged of relevant quality.

A formal test has been done in parallel – in the framework of the Dot.Kom project - using a legal corpus composed of 197 Sentences from Corte Cassazione (the Italian High Court) in HTML format (2500 - 3000 words each). This corpus was fully annotated according to our legal ontology, starting from an automatic pass made using a NEA based IE module and further manually edited in order to remove mistakes and adding missing elements. While being very specific in coverage (i.e. it is based only on Supreme Courts Sentences) the corpora has the advantage of being based on public material, unlike most other legal documents, and sufficiently generic in both domain (the sentences are not only related to labor law but cover different topics) and in linguistic aspects. The structure of the documents is rather standardized, but the content vary; for example, considering legal references, a wide variety of forms is used, reasonably reflecting a much wider set of documents.

Results of the formal evaluation are listed in a forthcoming Dot.Kom deliverable, but results for all the system tested show average precision over .85 for recalls ranging between .70 and .80 depending on IE subsystem used, with better values for most frequent and useful entities.

Overall results from both tests has been judged more than adequate, and the manual work eventually needed to amend annotations has been judged feasible, given that the error rate is sufficiently low, and anyway worth while pursuing in the light of advantages in using annotated material.

5 Related Work

The idea of using Semantic Web technologies to enhance and facilitate the use of a Knowledge Management System is shared by several others systems such as, for example, KIM and Haystack. KIM [16] is a platform for semantic annotations of texts, supporting semantic indexing and retrieval, which also shares with our system the use of some underlying technologies (GATE, Sesame and Lucene). Haystack [17] as well aims at giving users a unified access to their own corpora of knowledge for organization, navigation, and search, enabling users to import a variety of information types (documents, email, calendar, web pages) into a single unified RDF repository. In either case the two platforms however focuses on general functionalities, while the system described here is strongly focused on the legal domain, making possible to tackle a number of specificities and to take into account peculiarities of legal processes. Another example of vertical solution – for the Environmental domain - is the Semantic Web Environmental Directory (SWED) [18].

Other systems, whose goal is to support the creation of semantic portals, have some similarities with our system. We can cite ODESeW [19], OntoWebber [20], SEAL [21], OntoWeb [22] and OntoView [23]. Other tools share with LKMS the semantic browsing approach, such as already mentioned Magpie [24], or Topicalla [25], a client application for the Semantic Web which allows one to view information using a UI that is generated based on the kind of data available. However these tools and platforms aim mainly to support accessing and retrieval of information, while the purpose of our system is to support the whole knowledge creation process.

The legal area is subject of a growing interest, and there is a lot of work ongoing - see e.g. the recent book on Law and the Semantic Web [10]. Moreover a number of EU funded projects, including e-Court [11], e-Power [12], CLIME [13], FF Poirot [14], and SEKT [15], coped to different extent with exploitation of Semantic Web technologies in the legal domain. To our knowledge however no system aims at supporting the whole knowledge lifecycle and none has yet reached the stage of deployment in a production environment.

6 Conclusions and Future Work

In this paper we have presented a knowledge management solution for lawyer enhanced by a number of semantic web technologies. The system has already reached the commercial stage, where most features has already been incorporated in a commercial solution, deployed in a main Italian law firm and likely to be deployed at other sites on finalization of ongoing deals.

A number of other features are currently in the research and development pipeline. We are currently doing preliminary experimentations on classifiers working on RDF expressions attached to documents – derived automatically from texts and enhanced by additional information derived from public web services based on partial data extracted. We also are experimenting Collaborative Filtering techniques to generate suggestion of interest, mixing data from user tracking – who read what where – with documents semantic features.

Work done, and feedback got, clearly showed usefulness of rich representation framework in knowledge intensive environments, and capacity of semantic web derived technologies and tools to effectively support end users in everyday work. It also showed feasibility of a number of features still too often confined to experimental labs. Work done also enabled to better assess and investigate a number of issues related to matching organizations knowledge resources against public web material, including expected role of end user's as active annotators vs. automatic information extraction, and influences of the new features on ways of working within legal organizations. Attention paid to GUI issues, to flexibility in handling different ontology entities, to tools supporting 'making sense' of raw material has been functional to take end users within the loop. Several issues have still to be solved, and most relate to exploitation of machine learning techniques [27] to ease building of the semantic web. However we feel that as semantic web technology is a key element to step up knowledge management systems, we also believe that knowledge sharing environments like the one described, able to really exploit users knowledge through the organization, will be strong drivers to support building of the Semantic Web itself.

Acknowledgements

Part of the R&D activities behind the work reported has been carried out within the IST-Dot.Kom project (<http://www.dot-kom.org>), sponsored by the European Commission as part of the framework V, (grant IST-2001-34038). Dot.Kom involves the University of Sheffield (UK), ITC-Irst (I), Ontoprise (D), the Open University (UK), Quinary (I) and the University of Karlsruhe (D) . Its objectives are to develop Knowledge Management and Semantic Web methodologies based on Adaptive Information Extraction from Text.

References

1. Salzburg Research and EC IST DG Unit E2, "The Future of electronic publishing towards 2010", <http://ep2010.salzburgresearch.at>, 2003
2. J. Iria, F. Ciravegna, P. Cimiano, A. Lavelli, E. Motta, L. Gilardoni and E. Mönch: "Integrating Information Extraction, Ontology Learning and Semantic Browsing into Organizational Knowledge Processes", Workshop on the application of Language and Semantic Technologies to support Knowledge Management Processes at the 14th International Conference on Knowledge Engineering and Knowledge Management EKAW 2004, 5-8th October 2004 - Whittlebury Hall, Northamptonshire, UK
3. NormeInRete project. <http://www.normeinrete.it>
4. J. Broekstra, A. Kampman, F. van Harmelen, "Sesame: A Generic Architecture for Storing and Querying RDF and RDF Schema", Lecture Notes in Computer Science, Volume 2342, Jan 2002
5. H. Cunningham, D. Maynard, K. Bontcheva, V. Tablan, "GATE: A Framework and Graphical Development Environment for Robust NLP Tools and Applications", Proceedings of the 40th Anniversary Meeting of the Association for Computational Linguistics (ACL'02), Philadelphia, July 2002.
6. F. Ciravegna, "Adaptive information extraction from text by rule induction and generalisation", Proceedings of 17th International Joint Conference on Artificial Intelligence (IJCAI), Seattle, 2001.
7. ITC-IRST "TIES. Trainable Information Extraction System". <http://tcc.itc.it/research/textec/projects/dotkom/ties.html>
8. H. Cunningham, D. Maynard, V. Tablan. "JAPE: a Java Annotation Patterns Engine (Second Edition)", Technical report CS--00--10, University of Sheffield, Department of Computer Science, 2000.
9. "Apache Axis, an implementation of the SOAP " <http://ws.apache.org/axis/>
10. V.R. Benjamins, P. Casanovas, J. Breuker, A. Gangemi"Law and the Semantic Web: Legal Ontologies, Methodologies, Legal Information Retrieval, and Applications", Vol. 3369, Springer, 2005.9.
11. "Electronic Court: Judicial IT-based management". <http://laplace.intrasoft-intl.com/e-court/>
12. "European Programme for an Ontology based Working Environment for Regulations and legislation", <http://www.belastingdienst.nl/epower/>
13. "Computerised Legal Information Management and Explanation", <http://www.bmtech.co.uk/clime/>
14. IST 2001-38248. "Financial Fraud Prevention-Oriented Information Resources using Ontology Technology", <http://www.starlab.vub.ac.be/research/projects/poirot/>
15. "Semantically-Enabled Knowledge Technologies", <http://sekt.ijs.si/>.

16. B. Popov, A. Kiryakov, D. Ognyanoff, D. Manov, A. Kirilov, "KIM - a semantic platform for information extraction and retrieval", *Journal of Natural Language Engineering*, Vol. 10, Issue 3-4, Sep 2004, pp. 375-392, Cambridge University Press.
17. D. Quan, D. Huynh, and D. R. Karger, "Haystack: A Platform for Authoring End User Semantic Web Applications", *Proceeding of the 2nd International Semantic Web Conference*, Florida, October 2003.
18. "The Semantic Web Environmental Directory" <http://www.swed.org.uk/swed/index.html>.
19. O. Corcho, A. Gomez-Perez, A. Lopez-Cima, V. Lopez-Garcia, and M. Suarez-Figueroa, "ODESeW. Automatic generation of knowledge portals for Intranets and Extranets", *The Semantic Web - ISWC 2003*, vol. LNCS 2870, pp. 802-817, 2003.
20. Y. Jin, S. Xu, S. Decker, and G. Wiederhold, "OntoWebber: a novel approach for managing data on the Web", *International Conference on Data Engineering*, 2002.
21. N. Stojanovic, A. Maedche, S. Staab, R. Studer, and Y. Sure, "SEAL - a framework for developing semantic portals", *Proceedings of the International Conference on Knowledge capture*, pp. 155-162, 2001.
22. P. Spyns, D. Oberle, R. Volz, J. Zheng, M. Jarrar, Y. Sure, R. Studer, and R. Meersman, "OntoWeb - a semantic Web community portal", *Fourth International Conference on Practical Aspects of Knowledge Management*, 2002.
23. E. Mäkelä, E. Hyvönen, S. Saarela, K. Viljanen, "OntoView - A Tool for Creating Semantic Web Portals", *Proceeding of the 3rd International Semantic Web Conference*, Hiroshima, November 2004.
24. M. Dzbor, J.B. Domingue, E. Motta, "Magpie - towards a semantic web browser", *Proceeding of the 2nd Intl. Semantic Web Conf.*, October 2003, Florida US.
25. "Topicalla", <http://topicalla.mozdev.org/index.html>.
26. V. Lanfranchi, F. Ciravegna, D. Petrelli: "Semantic Web-based Document: Editing and Browsing in AktiveDoc", *Proceedings of the 2nd European Semantic Web Conference*, Heraklion, Greece, May 29-June 1, 2005
27. F. Ciravegna, N. Kusmerick, S Staab, C Knoblock. "Machine learning for the Semantic Web" Dagstuhl workshop 13-18 February 2005, Dagstuhl, Germany. <http://www.smi.ucd.ie/Dagstuhl-MLSW/>