The E-Lex Project or How to Improve the Access to Law

Marie-Francine Moens

Interdisciplinary Centre for Law and Information Technology
Katholieke Universiteit Leuven
Tiensestraat 41
B-3000 Leuven
Belgium
marie-france.moens@law.kuleuven.ac.be

Key terms: Legal information retrieval, Legal drafting

Abstract: The E-Lex project investigates how the access to law can be improved by the use of systems that support the drafting of the documents or translate their content into forrmats that can be more easily read and processed by the computer. The article focusses on advanced information retrieval systems and on the requirements that they impose on the drafting of legal documents.

1. Introduction

Legal information can be found in case law, legislation and other documentary sources. Legal sources are increasingly available via Web portals that are maintained by public and private institutions. The legal information is usually searched by means of a full text search.

Governments currently take many initiatives in order to promote the electronic communication with citizens and companies. There are e-government programs that make information available via the World Wide Web and allow citizens to pose information questions to governmental institutions via e-mail. A citizen or a company might have a specific problem, for which the government is asked for advice. The problem and advice might be legal in nature and a solution might rely on information found in legal documents.

Institutions and companies also offer multiple Web services. For instance, insurance companies offer contracts via the World Wide Web and model transactions through the use of specific knowledge languages. Small expert systems are being developed as Web services that help the user to solve a specific problem. The modeling of transactions and the solutions to problems might require insight in legal knowledge extracted from legal documents.

In all of the above examples legal services are offered that are based on information found in legal documents. The tasks range from simple word searches in a document base to posing information questions (possible drafted in natural language) to the document base and to extracting problem-solving knowledge from the documents. Automation of these tasks becomes commonplace. But, as law is currently found in these documents, can we not call it an encrypted tangle of information, difficult to search by humans and very hard to effectively search by the machine? Law has become so complex that it becomes like encrypted with a secret key algorithm that borrows its power from its intricate complexity so it becomes very difficult to break the key. This complexity is the cause of many problems including difficult access to law by citizens, legal professionals and by lawmakers who should guard principles such as correctness, reliability, security and equality of the law.

Automating the tasks of information searching, answer finding and knowledge extraction require the analysis of the legal documents by humans and increasingly by the computer. The *E-Lex* project[1] that is currently under investigation at the Katholieke Universiteit Leuven studies the role of drafting for improving the access to law. We hypothesize that a better drafting of legal documents will improve both human and machine access to information in legal documents. We also assume that such a drafting can be supported by automated means. The goal of this short paper is to ascertain how drafting support tools can improve the access to law and advance the retrieval effectiveness of legal information sources.

2. Short State of the Art of Legal Information Retrieval

Automated retrieval from large document collections was one of the earliest applications of computer science to law. Being developed for a few decades now, large legal content providers such as LexisNexis[2] and WEST-LAW[3] are offering legal information services, though other private and many governmental initiatives take advantage of the current World Wide Web to promote access to legal information.

Legal information can be retrieved by means of a full text search, i.e., every word of a text can function as a search key. In addition the search can

---

[2] http://www.lexisnexis.com/ (13-4-2004)
[3] http://www.westlaw.com/ (13-4-2004)

be made more precise by searching based on descriptors attached to the documents called metadata that reflect, for instance, the domain of law, classification codes, titles, institutions that issued the document (e.g., court names), dates (e.g., date of enactment of a statute article), applicability designators (e.g., application area of a statute) and the function and role of certain components (i.e., identification of parties, of motivation and of conclusion in court decisions). A database or several databases are searched, in which the texts of the documents are stored. These texts are indexed with the terms that occur in the natural language texts of the documents and with extra descriptive data that are used in the information selection. Legal information is searched by lawyers, judges, legislators, civil servants, academics, students and ordinary citizens.

Legal information concerns the retrieval of structured content and unstructured content. Structured information regards information of which the semantics can be clearly, unambiguously determined and can be described with simple and clear concepts. These data are often added in the form of metadata (i.e., data that describe other data) to the documents (examples see above). Unstructured information regards the information that is communicated in natural language texts, or in other formats such audio and video of which the semantics are much more difficult to register in simple terms.

Most of legal information content is found in natural language texts. Users want flexible access to this information. Although a full text search offers such a possibility and retrieves quite useful information, especially when the search terms are automatically expanded with synonyms and related terms, users of retrieval systems that query large databases become less and less satisfied with the results of a full text search. Very often too many documents are returned by the system requiring a lot of the precious time of the searcher to consult. Users of retrieval systems want more precise answers to their information query without sacrificing the flexibility of requesting an answer for any possible information request.

To give a solution to the above problem, legal *question answering systems* are currently an important topic of research.[4] In a question answering system a searcher poses a question in natural language and the system does not retrieve the documents in which the answer might be found, but the answer to the question that is extracted from the documents (e.g., "When

---

[4] *Moens, M.-F.,* Interrogating Documents: The Future of Legal Information Systems, in: Proceedings of the JURIX 2003 International Workshop on Question Answering for Interrogating Legal Documents (2003), Utrecht University; *Queresma, P.,* Using Dialogues to Access Semantic Knowledge in a Web Legal IR System, in: Proceedings of the JURIX 2003 International Workshop on Question Answering for Interrogating Legal Documents (2003), Utrecht University.

is hunting with fire weapons on roe goats open?"). Single questions are automatically answered by using a collection of documents as the source of data for the production of the answer. In recent artificial intelligence research we see even research into systems that reason with the content of multiple sentences possibly from different documents in order to infer the answer to the question. Question answering integrates searching and inferencing and might become one of the leading future approaches to legal information systems as it combines traditional technologies of legal information systems, i.e., information retrieval and legal reasoning in knowledge based systems.

There is *query by example* retrieval. You input a case description and the system will find similar cases in the database. Finding similar cases is part of judicial and criminal research, here we see that police forces worldwide invest in simple statistical text mining software. This software uses statistical term correlations as found in large corpora in order to find similar content that is expressed in variant natural language expressions.

Information retieval lately has made great advances due to the progress in *natural language processing research* and the availability of software tools for part-of-speech (POS) tagging or syntactic word class tagging, and of parsing tools that analyze the syntactical structure of sentences.

With the start of the World Wide Web, retrieval algorithms that use the link structure of the Web in computing the importance or authority of a Web page have been developed. But, also law documents are full of implicit and explicit references. Currently this link *analysis* has not been explored in developing search algorithms for law documents – although *Turtle*[5] already noticed their potential –, but we are quite convinced that its value will be ascertained in future research.

An automated search for legal information should be efficient, reliable and flexible, and it should integrate different sources that might be available across different languages and cultures.

3. Short State of the Art of Legal Drafting Systems

From the very start of artificial intelligence and law research there have also been theories, models and systems that show how intelligent drafting can support the legal author in constructing law texts. In early years, the idea was that legal texts (e.g., legislation) should be drafted in an artificial logical language that allows for an automated reasoning with the legal

---

[5] *Turtle, H.,* Text Retrieval in the Legal World. Artificial Intelligence and Law (1995), 3, 5-54

sources.[6] The idea is still alive. But, instead of the model in which human drafters acquire such a language and commonly use it, we see gradually the growing idea that intelligent drafting systems assist the drafter towards an ideal legal language. We can define a drafting or drafting support system as a computer program that automates (an aspect of) the construction of a legal source. This is a rather wide definition, which we need in order to encompass all current trends.

Based on their functionality, we might distinguish four categories of legal drafting systems. *Informative systems* and help functions are computer programs that offer information about the drafting process to the user without any obligation, i.e. the user can decide for himself whether he will accept or ignore the hints generated by the system (e.g., LEDA[7]). The main advantage of informative systems is that they are very likely to be accepted by their potential users, because they do not enforce rules, but, the danger exists that their recommendations are simply ignored. *Text assembly* and *text generation systems* (e.g., certain functionality of SOLON[8] and DocuPlanner[9]) construct legislative texts by using information provided by the user and knowledge about the formal aspects and the content of legislative documents that is contained within the system. The drafter is forced into the system's rigid framework, which he will often perceive as an attack to his freedom of expression (even when the texts he produces turn out to be more correct and comprehensible). *Verification systems* (e.g., certain functionality of SOLON, Lexedit[10] and Enact[11]) give the user an unlimited

[6] *Allen, L.E. and Saxon, C.S.*, Better Language, Better Thought, Better Communication: The A-HOHFELD Language for Legal Analysis. In: Proceedings of the Fifth International Conference on Artificial Intelligence and Law (pp. 219-228) (1995), New York, ACM.

[7] *Voermans, W.*, Sturen in de mist ... Maar dan met radar (1995), PhD. Dissertation, Tilburg University.

[8] *Debaene, S., Van Kuyck, R. and Van Buggenhout, B.*, Solon. Een computersysteem ter ondersteuning van de wetgevingsactiviteit van de Vlaamse regering, in *S. Debaene and B. Van Buggenhout (Eds.), Informatietechnologie en de kwaliteit van wetgeving*, Antwerpen, Intersentia, (2000), 79-120.

[9] *Branting, K., Lester, J. and Callaway, C.,* Automated Drafting of Self-Explaining Documents, in: Proceedings of the Sixth International Conference on Artificial Intelligence and Law (1997), New York, ACM, 72-82; *Branting, K. , Callaway, C., Mott, B. and Lester, J.,* Integrating Discourse and Domain Knowledge for Document Drafting, in: Proceedings of the Seventh International Conference on Artificial Intelligence and Law (1999), New York, ACM, 214-220.

[10] *Mercatali, P.,* Computer-Aided Methods and Tools for Legislative Drafting, in: *S. Debaene and B. Van Buggenhout (Eds.), Informatietechnologie en de kwaliteit van wetgeving*, Antwerpen, Intersentia, (2000), 139-156.

[11] *Arnold-Moore, T.,* Information Systems for Legislation (1998), Ph.D. dissertation, Royal Melbourne Institute of Technology, Australia, 1998.

freedom. He can construct a text in any way he thinks appropriate and only afterwards the system checks compliance on a number of criteria (e.g., compliance with a predefined structure; compliance with the use of certain terminology) and suggest corrections, at which point the user is still free to dismiss or ignore the corrections. We might add a fourth category, which is currently gaining in importance and which we might tentatively call *translation systems* (e.g., Norme in Rete[12], POWER[13], REGNET[14]). They regard the automatic or semi-automatic translation of texts into concepts and logical statements that can be integrated in information systems for retrieval, decision support or compliance assistance. They are most useful for drafting legislation and regulations.

Because of the growing belief that legal drafting systems should support a facilitated processing of the documents by the computer, recent drafting systems offer the possibility to format certain information with mark-up tags in a *mark-up language* that form metadata of the texts. Metadata are used to structure (e.g., in mandatory and optional document components) and describe the actual content of a document. Metadata are usually present in the text in the form of tags and although in most representational formats it will be invisible to the user, computer programs can use them to correctly identify or classify documents or parts of documents. Above some examples of metadata are given. Metadata are usually encoded in a standardized markup language, the two most common ones being SGML (Standard Generalized Markup Language) and currently XML (Extensible Markup Language).[15] XML marked documents are in ASCII format and can thus be interpreted on virtually all platforms. XML is suitable for labeling the metadata of a particular document and at the same time it ensures maximal exchange possibilities given that standard labels are equally interpreted across the different institutions that exchange information or that labels are used that can be easily and unambiguously translated. In addition the mark-up information can be used for database search.

---

[12] *Biagioli, C., Francesconi, E., Spinosa, P. & Taddei, M.,* NIREditor, A XML Specific Environment for Legislative Drafting (Norme in Rete project), in: Proceedings of the JURIX 2003 International Workshop on the Development of Standards for Describing Legal Documents (2003). http://www.lri.jur.uva.nl/standards2003/

[13] *Van Engers, T. and Glassée, E.,* Facilitating the Legislation Process Using a Shared Conceptual Model, IEEE Intelligent Systems (January 2001), 50-57.

[14] *Kerrigan, S. and Law K.H.,* Logic-based regulation compliance assistance, in: Proceedings of the 9th International Conference on Artificial Intelligence and Law (pp. 126-135) (2003), New York, ACM.

[15] For information on XML and its specifications, the website of the World Wide Web Consortium (http://www.w3c.org) (11-4-2004) is a primary resource. XML is a subset of SGML.

For a specific document type a grammar that exhaustively describes how a document can be constructed in terms of possible configurations of its metadata and their possible values can be constructed. Such a grammar is called a Document Type Definition (DTD) or XML schema depending on the syntax used for describing the grammar. Based on such a DTD or XML schema the conformity of a document instance to the document grammar can automatically be verified.

4. The E-Lex project

The E-Lex project investigates how drafting can improve access to legislation. We have investigated metadata that are essential for the database management of legislation (e.g., metadata for identification, version management)[16]. We currently investigate how the drafting of the legislative texts can be improved in order to comply with the demands of advanced retrieval systems such as a question answering system. We are also implementing and testing a simple question answering system that interrogates legislation. Because legal documents often have a strict formal organization and are subject to a number of stylish conventions, their formal characteristics can be exploited to structure certain information when drafting the documents. This structured information can easily be searched. In addition, the free text of the documents could be improved in such a way their automated processing and structuring into representations that can be used in advanced retrieval models. As a result retrieval of legal information could be much more efficient and reliable.

Whereas early drafting support systems put the emphasis on producing qualitatively better texts and more uniform documents in order to increase their convenience of their manual use (such as their readability by humans), the focus of current drafting support technology is on producing digital documents that can be read and interpreted by computer, so that current information systems can offer advanced information services based on the document contents. This evolution is reflected by the functionality of the systems: early systems only provided informative help tools or offered some templates for text assemblage; the newer systems focus on verification and translation of the content into formats readable by machines. Correct drafting has become more important than ever, because computer readings of the documents are more sensible to different types of error. However many of the concerns of well-formedness by early drafting systems are still valid.

---

[16] *Moens, M.-F. (Ed.),* Digitale wetgeving, Digital Legislation (2003), Brugge, Die Keure.

Mark-up languages such a XML are tailored to encode information in texts that are understandable for humans as well as computers. XML allows for customization of markup languages with application specific tags and can encode information about document type specific structure and content and convenient editors can be used for drafting.

We can not ignore that we are still confronted with a large amount of existing legal documents that should be converted. In addition, a too large emphasis on structuring information at the time of drafting, puts a large burden on the drafter and might give him the impression that he is forced into a straitjacket which does not allow him to express freely all aspects of the content that he wants to communicate. The user might be reluctant to use specifically designed editors that make document drafting quite similar to filling out forms. As a result, we still need translation tools that automatically convert certain legal document content into structured information and mark up the information with the right metadata labels.

For accomplishing these tasks: artificial intelligence technologies play an important role. For instance, one can use technologies for information extraction and text categorization to automatically mark-up text with metadata.[17]

All of the above technologies allow for essential metadata of the legal documents to be tagged automatically. However, the use of drafting support technology can also improve the retrieval of information found in natural language texts, especially in advanced retrieval systems such as question answering, query by example systems and retrieval systems that exploit link analysis.

How can a drafting system help in order that the access to the legal documents is improved? The most simple approach regards the use of spelling and grammar control mechanisms that are now standard integrated in authoring tools. Secondly, the consistent use of the same term with the same meaning can avoid many problems of word sense disambiguation when treating the texts. Moreover, definitions and their scope should be clearly stated. References to other documents or document parts should be uniformly recognized. Future drafting tools might be designed to check the use of rhetorical relationships or compliance to certain text grammatical rules. Also the drafting system should make more visible the logical relationships between clauses and sentences, as well as exceptions to norms and regulations should be stated clearly. Future drafting systems could be very sophisticated and checking the logical consistency of legislation. In

---

[17] *Moens, M.-F.,* Innovative Techniques for Legal Text Retrieval. Artificial Intelligence and Law (2001), 9, 29-57.

such an evolution, we see that verification and translation systems are the most convenient drafting support tools, because drafters will not give up easily the use of natural language.

5. Conclusions

Quality drafting helps in improving access to legal documents. First of all, information that is structured in nature (e.g., identification information) can be easily structured and tagged with markup languages by means of text editors that have the functionality of text assembly and verification. This allows for an automated and reliable input in the information system. Secondly, in advanced retrieval systems the information in the free natural language text can also be structured for convenient use in advanced retrieval systems. This is not an easy job to do, but current drafting technologies might help in using the right terminology, and might already label definitions, references, rules and exceptions. More advanced technologies that translate legal information into knowledge rules are being researched.

The success of future information systems will thus largely depend on two factors: on how well certain elements of the information in the legal documents are structured at the time of drafting; and on the quality of the automated content analysis, which is somehow supposed to "order" the remaining unstructured information. Development of rigorous drafting and advanced content analysis might lead to some point in the future when all legal information can be correctly and unambiguously retrieved and when legal questions may be automatically answered based on document content.