BILETA
British & Irish Law, Education and Technology Association
2007 Annual Conference
Hertfordshire 16-17 April
UH

Paper and stone: How technology has not changed the retrieval of legal information, yet

Christine Kirchberger
Law and Informatics Research Institute,
Stockholm University
Email: christine.kirchberger@juridicum.su.se

1. Introduction

Lawyers and legal professionals depend on their subjective legal knowledge as well as on external legal sources. This has been the case for centuries. With the dawn of technology easier methods to retrieve legal information appeared, facilitating the search for relevant material in the growing amount of statutes, cases, preparatory works and literature.

Despite several research projects on information retrieval since the beginning of the 1960s the search process of a lawyer has not changed significantly. The question remains in what way it will change in the future and whether or not it will change at all. Legal texts are based on abstract and technical language. This means that the retrieval has to be done with both everyday and legal terms. The choice of these search words has not been a large focus in academic research yet. Rather it has been on technical retrieval methods behind the different legal databases and search techniques and how the system interprets different search queries and not so much how to increase the awareness of lawyers in their search technique.

This has led to the development of highly sophisticated systems in legal databases. However, this development is not always reflected in the user behaviour. More and more is happening technology-wise behind the scenes, which is not visible for the user, however. This means there is a lack of awareness.

The lack of trust that existed in the beginning of the digitalisation of legal material has reversed, which means an immense trust in relevance ranking and in search results. This does not mean that the algorithms of legal databases should not be trusted. Lawyers should be aware, however, of what is going on behind the curtains, as it is she or he who in the end is responsible for finding relevant legal material to support her or his decision.

In addition to including methods of retrieval in the legal education, systems should make the techniques used more visible and thereby allowing the legal user to adjust her or his search method or at least be more critical when searching. An effective search process requires the evaluation of retrieved documents and an adequate adaptation of the continuing search.

Starting with issues concerning legal information that have not changed, this article will then deal with issues that have changed and conclude with some remarks on possible affects of these observation. The taken approach will focus on the past 10 years, corresponding to the time information technology gained ground and influenced society in general. The legal profession is no exception to this development and offers a field that could gain immense benefits from technology. As the following section will show this has not been successful in all areas within law.

2. What has not changed?

2.1. Production of legal information

Legal information is produced both during legislative procedure as well as decision-making by courts and administrations. The role of legal academics and legal professionals in supplying written articles and text books on specific subjects should not be forgotten here either.

Legislative procedures are mostly based on the democratic foundations of the state and its constitution. This has, for obvious reasons, not undergone extreme changes in the past decades, not counting the

development in the Eastern European countries, but even there, law-making and judicial state affairs have kept the status-quo to a large extent. Politics might change, but the law stays the same.

Publication of legal information has not changed very much either. In most European countries statutes or amendments thereof are published in a printed gazette, case decisions are issued by the respective court and in some cases also promulgated in (private or official) publications. Interestingly, this tradition is kept even when official journals are published solely electronically, as has been the case in Austria since 2003.

The advantages of electronic publication are immense, not only concerning costs but also improved availability of public legal information as well as quicker and better access for the general public. When comparing the Austrian official journals from 1849, 2003 and 2007, the changes in format or structure are rather insignificant. [RIS/ALEX] There are legitimate reasons for that, mostly concerning the integrity of the legal texts and the principle of legal certainty. As in other areas of law and society in general, technology is used to perform manual, paper-based processes electronically. The processes as such are, however, not challenged or discussed.

The increase in technology has therefore not affected the publication process as such. In most cases bills of parliament, statutes, cases and especially legal text books and law journals are still mainly printed on paper, even if their digital equivalent is available in electronic databases. While searching is easier in the electronic versions, users might nonetheless prefer printed copies when reading and adding personal comments.

To some extent technology is nowadays utilized during the legislative process, e.g. preparatory works are sent electronically between the responsible civil servants and the documents are structured in a way that allows automatic processing at a later stage. [Eriksson 2005] Even in these cases legal texts are, principally, not prepared or made accessible for their retrieval at a later stage. [Saarenpää 2006]

2.2. Structure of legal information

Related to the question of publication is the matter of the structure of legal information. As previously mentioned, electronic publication of statutes, preparatory works and case law does not necessarily mean a change in format or structure. Their configuration and formation is still very much controlled by their printed form. In other words, the typical structure of legal material has been the same for decades. For example, a statute is divided into chapters and sections, a judgement into the presentation of the facts, the finding arguments, the conclusions, etc. and a legal textbook into chapters and sub-chapters.

This offers advantages, as a formalisation of this kind essentially facilitates information retrieval. If the conclusions of the court, for example, consistently are placed at the end of the decision, one can easily retrieve this information without having to scan or read the complete document. This, obviously, enhances the information retrieval, making it faster and more efficient. In addition, a certain tradition of formalisation gives legal information stability and ensures that searches are possible in material from different periods of time. Law does not change quickly, so the possible amount of relevant legal material is not limited to a few years, but actually stretches over decades.

The non-existing development in this regard has, however, not only positive effects. Information technology allows other possible setups of legal acts, preparatory works or case law. Texts do no necessarily have to be linear, but could be divided by chapters and sub-chapters, linked to each other and therefore enabling a complete picture of a certain legal rule. A legislative proposal could, for example, be split up in different parts discussing certain topics within the proposal. These parts could then be linked to the different sections in the statute which in turn are linked to case law concerning a particular rule as well as literature written on the same topic. The user could therefore get the complete picture of a particular regulation. This idea has already been implemented when it comes to commentaries. Commentaries, however, have to be updated manually by legal academics and professionals. Technology could contribute a great deal in this regard. [Schweighofer 2006]

2.3. Language of legal information

Naturally, language changes even less than law does, with primarily positive effects. The point to be made here is that the language used in constitutions, statutes, preparatory works, case law and legal doctrine commonly varies between the different types of legal sources. The dissimilarity concerns the type of words that are used. The language used in statutes and other laws is typically very abstract, while judgements include more everyday words.

For example, the sentence “The act of referring, on an internet page, to various persons and identifying them

by name or by other means, for instance by giving their telephone number or information regarding their working conditions and hobbies [...]" will be hard to find in a statute or directive, but might be included in a court decision [ECJ Case C-101/01]. In a statute, on the other hand, the same meaning might be expressed by the terms "processing of personal data". [Data Protection Directive]

As a rather apparent fact, statutes are designed for being applicable in numerous occasions wherefore the wording may comprise diverse circumstances. Decisions by the court, on the other hand, aim at solving a dispute between two parties in a particular situation, so references to facts are inevitable. The implications on information retrieval are, however, obvious.

For example, a search for "internet page, telephone number, hobbies" in a statute might not necessarily produce a result, but the same search terms used in a case law database, will probably lead to the particular case. The words "web site, mobile, free time occupation", however, might, depending on the database, not produce any results, as synonyms, principally, are not included automatically in the retrieval process.

Information retrieval brings the meaning of language to a head. Systems are not able to give meaning to language, at least no yet, but can simply interpret the syntax of a word or sentence, not its semantics. [Schweighofer 1999] Therefore the importance of the right use of search terms is not to be underestimated.

2.4. Search algorithms

Systems today basically use the same search techniques – algorithms – as 50 years ago. This does not mean that legal databases have not improved since the 1950s. The enhancements have, however, focused on performance and the user interface, not particularly on processing and search techniques. [Karlgren 2005]

A good example are Boolean operators (OR, AND as well as NOT). They have been in use for more than 40 years. Again, a certain tradition has its advantages, as users get acquainted with the functions. Boolean search holds, however, some constraints. One of them being that searching is completely dependent on the words contained in a document, which means that the user is searching for legal (or non-legal) terms and not legal concepts. If seen from a language point of view, as mentioned above, this means that the system does not necessarily look for synonyms of the same word, and even if this is the case, it is not obvious that synonyms are enough to express a certain need for specific information. Typically a lawyer knows what a legal document she/he is looking for should be about. In order to successfully search for these documents she/he has to guess, however, which words might be contained in these particular documents. At the same time she/he should avoid too common words, otherwise the result list will be too long to cope with. In addition, relevance rankings used in databases do not commonly indicate the criteria used when determining a document's relevance. [van Noortwijk 2006] Possible decisive factors include the frequency of the search word in relation to the total amount of words in a specific document, the location of the search term (higher ranking when the search term is mentioned in the beginning or in the end), and so forth. [Karlgren 2005]

In the past years, several projects have dealt with new alternatives for search methods. Examples for existing projects and suggested ideas include conceptual ranking and retrieval [van Noortwijk 2006] and dynamic electronic commentaries [Schweighofer 2006]. Theses ideas have, however, not found their way into the legal database systems used on a large scale.

2.5. Search skills of lawyers

Legal information, like any other information, is becoming increasingly remote from its users. [Saarenpää 2006] The amount of legal sources (legislation, preparatory works, cases, legal doctrine) is constantly increasing, and technology allows for this information to be accessed quickly. Technology does not show lawyers, however, how to retrieve information efficiently. This task is still for the lawyer to perform. In spite of technological advances, the development of technology has, unfortunately, not been reflected in a development of legal culture in this respect.

A progress seen within the last years is that commercial legal databases move towards simpler search functions, often limited to free text search and possibly including field search for words in headings. Focus is put on browsing, users can choose by topic or year and click until they reach the designated document. Database service providers are, understandably, striving to make searching easier for their users, but this development leads to a further decline in search skills, as browsing implies that the lawyer knows what she/he is looking for. In a situation where it is completely unclear if a statute or a case exists for this particular legal question, users have to rely on searching. Information retrieval is therefore still a skill, contrary to general believe.

While searching on the web is still considered an everyday task - Google has shown the way - lawyers and law students alike face the difficulty in putting effort in something that everybody is expected to be able to do. The risk therefore emerges in growing trust to databases and their relevance ranking. A user might be satisfied with the first retrieved document possessing some relevance and not consider the possibility of more relevant documents as continuing the search would be time-consuming and involve re-consideration of her/his search methods and the question to what extent the search terms should be adapted or replaced.

3. What has changed?

3.1. Availability of and access to legal information

With the development of information technology and the World Wide Web, companies, authorities and organisations are expected to have their own web site. This has also led to an increase in official information published on the Internet. Authorities with close citizen contact use their web sites to publish latest news, frequently asked questions (FAQ) and guidelines about how to apply for grants or social benefits, or how to declare taxes. At the same time, almost every household today possesses Internet access and a computer and can therefore theoretically receive this information.

Better availability, however, does not necessarily mean better access. Publishing reports, recommendations and FAQs on web sites does not equal users actually finding what they are looking for. If a person is inquiring about whether she/he can deduct subscriptions to newspapers from her/his tax, the fact that the national tax authority has published this information is not enough. It also has to be possible to search on the web site and/or the navigation on the web site should not be cumbersome, but clear and understandable. It is essential to secure quality and not merely quantity of information.

EUR-Lex is a good example, the official web site of the European union providing direct free access to EU law (http://eur-lex.europa.eu/en/index.htm). The search functions have been improved recently, so searching for specific EU directives has become less complicated. In terms of browsing facilities, the listed categories and sub-categories (http://eur-lex.europa.eu/en/repert/index.htm) might not necessarily correspond to the user's idea of the different topics regulated by the EU.

1996 a rather extensive public-opinion poll took place in Sweden, inquiring about the working conditions of lawyers, e.g. which legal sources were being used, which sources were available to them and how much of their daily tasks actually involved pure legal work such as researching for a specific legal question. [Sandgren 1999] One of the outcomes of the survey was that availability played a crucial role for the usage of legal sources. Peter Seipel called this the arm's length law. [Seipel 2004]

For example, the majority of lawyers had easy access to the code of laws, official journals, case law from the Supreme Court, law journals and legal textbooks. At the same time, when asked about which legal sources they use the code of laws lead the list, followed by textbooks, official journals and case law from the Supreme Court. One could therefore argue that availability is linked to the usage. [Sandgren 1999]

The survey did, however, not take fully into consideration why certain legal sources are used more than others. One factor mentioned in the report was the legal status of a certain source, i.e. laws having a prior status when it comes to legal work. Legal textbooks, however, do not have a recognised high rank within the doctrine of legal sources. I would like to claim, without being able to refer to statistical data yet, that a reason for their importance in legal life lies in the fact that legal textbooks provide an overview over a specific field of law, which is not the case with any of the other legal sources.

3.2. Google

While the reasons might not be very obvious, Google is probably the most widely used search engine at the moment and even gave cause to a new expression: “to google”.

One of the drawbacks of Google is that users are getting used to retrieving documents and results no matter which search terms they use, due to the theoretical high coverage of the database which contains several billions web pages. In most cases Google as a simple to use tool provides you with results. Users are, therefore, lead to believe that their search queries are effective.

On the other hand, Google offers many ways to retrieve information, due to its global character and its popularity, as well as its neutrality when it comes to ranking and result lists. One example, is Google Books (http://books.google.com/) that despite its legal implications concerning intellectual property rights, offers a way to search in printed literature, a feature that has not been offered by many commercial service providers yet.

Another example of a positive development is the discussion about the importance of references. The main idea of Google's search engine is to rank the results based on the amount of web pages linking to a particular site. In other words if web site A has 10 pages linking to it and web site B 7 pages, then site A will end up higher in the result list than site B. If, however, one of the sites linking to B has 5 other pages linking to it, site B will end up higher in the result list. (http://www.google.com/technology/index.html) This feature has been known and used in science for quite some time now. If many scientific articles or textbooks refer to a particular article in a law journal, that article receives higher credit than others within the same field. This feature, however, has not been successfully implemented yet in the digital world.

Again, Google has started such a project, though not limited to the field of law. It is called Google Scholars (http://scholar.google.com/), and takes advantage of exactly that, references from scholar articles. The biggest advantage being that the database is not static, but dynamic. This means that a certain article can link to other articles that have referred back to it, a function not possible in the offline word. As many commercial databases and also public ones for that matter do not contain legal literature to a large extent, this information is not searchable at the moment. With Google Scholars, you cannot only see how many times a specific article has been cited, but also go directly to the article containing the citation. This offers immense opportunities, also in light of the web 2.0 developments.

3.3. Types of legal information

As indicated above, a new openness and proactive approach within public administration has lead to more and more recommendations, guidelines, studies and FAQs being published online. This leads, however, to the question of the status of these documents within the doctrine of legal sources. Their status is not completely statutory, but not pure legal literature either as the author quite often is the responsible authority that interprets the existing rules and also applies them daily. This leads to the question to what extent the recommendation of e.g. the tax authority has legislative characteristics or is simply a certain point of view not having binding effect.

The doctrine of legal sources should therefore be revaluated and inquiries are necessary into the question to what extent the doctrine of legal sources is applied in daily legal life. One attempt has been made in the above-mentioned public survey. Already this study showed that the embedded ranking within the doctrine of legal sources was not completely followed by lawyers, which partly had to do with less access to certain material such as preparatory works and bills of parliament.

3.4. New tools for lawyers to collaborate

Knowledge management in law firms has been around for some time, and comprises creation, storage, organising and sharing of knowledge within a law firm or any other organisation. The main purpose is to avoid reinventing the wheel constantly and re-using once created documents in other projects and cases. This had already been done pre-IT, with lawyers re-using contract documents, but IT increases the efficiency immensely, as documents can be edited directly and more persons can work together with the same document.

Lately in the light of web 2.0 sharing of knowledge and information in general has become increasingly important, also for lawyers. Interesting projects include e.g. Law Underground (http://www.lawunderground.org/), web law 2.0 (http://weblaw20.wikispaces.com/) and law 2.0 (http://del.icio.us/tag/law2.0). Many legal academics and professionals have their own blogs or podcasts, though as of now it is mainly seen in the U.S. and the U.K. Also collaboration of lawyers in general, beyond the borders of a particular law firm, is receiving more and more attention. [Kennedy 2007]

These tools have a lot of advantages and allow lawyers to share ideas and information across borders and cultures. The question remains, however, whether to decrease or increase the information overflow. If the latter is the case, other tools are necessary in order to sort out the relevant information, which is why information retrieval is still vital.

4. Conclusions

4.1. Information technology and information literacy

One should not forget that information communication technology has had a positive impact on society. Especially when it comes to information retrieval more legal information is available electronically nowadays which increases the coverage of databases. In other words a larger amount of different legal sources can be searched at the same time, and systems might also allow referencing and linking between the different

sources, with direct access to a particular source document.

An increase in information, however, also implies an increased importance of retrieval of relevant documents, which leads to a need for effective retrieval tools and for an awareness within the legal professionals that searching, is a skill and not a simple task that does not have to be acquired. Lawyers should be aware of the techniques used in databases and be familiar with how relevance ranking is done, at least on a basic level.

4.2. New layers on old layers

Regarding publication and structure of legal information, information technology simply put another layer on the already existing layer of legal documents. One can compare this with public authorities introducing IT in their administration, but not seeing over the existing procedures. Progress does not mean doing the same things with new technology, but also seeing over how things are done and trying to make the process as such more efficient.

In the same light the publication of legal information should be evaluated and research initiated on how to effectively produce legal information and thereby facilitate its usage and retrieval.

4.3. Improved techniques for information retrieval

Despite efforts within research, the techniques used in commercial databases have been principally been based on the same algorithms for the past 40-50 years. The techniques work well, but the more the amount of legal information increases the more important it becomes to sort and retrieve the relevant "needle in the hey sack". Syntax search may soon become insufficient. Therefore research results should be tested in real-life circumstances as well and tried on a large scale in order to create opportunities for the future.

Development during the last 10 years has been significant, but we still have a long way to go, which will involve changes not only in technology but also in management of information and the legal culture.

Bibliography

[Data Protection Directive] Article 2 point b of the Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data

[ECJ Case C-101/01] Judgment of 06/11/2003 in Case C-101/01 (Reference for a preliminary ruling from the Göta hovrätt): Bodil Lindqvist

[Eriksson 2005] Bengt Eriksson, Maintaining information quality – Experiences from the Swedish Parliament – Sveriges riksdag, in Cecilia Magnusson Sjöberg (ed), Legal Management of Information Systems - incorporating law in e-solutions, Studentlitteratur, 2005, p 235-260

[Karlgren 2005] Jussi Karlgren, Information Retrieval System: Statistics and linguistics, in Cecilia Magnusson Sjöberg (ed), Legal Management of Information Systems - incorporating law in e-solutions, Studentlitteratur, 2005, p 295-336

[Kennedy 2007] Dennis Kennedy's blog, *http://www.denniskennedy.com/blog/2007/03/collaboration_tools_for_lawyers_the_book.html*)

[RIS/ALEX] The Legal Information System of the Republic of Austria (RIS), http://www.ris.bka.gv.at/; ALEX – Austrian historical official journals and legal texts online, http://alex.onb.ac.at/

[Saarenpää 2006] Ahti Saarenpää, Towards Legal Information and Legal Knowledge, in Cecilia Magnusson-Sjöberg and Peter Wahlgren (eds), Festskrift Peter Seipel, Nordsteds Juridik, 2006, p 509-525

[Sandgren 1999] Claes Sandgren, Vad gör Juristen? Och hur? - del I, JT 1999-2000 p 59-611, del II p 867-893

[Seipel 2004] Peter Seipel, Juridik och IT – Introduktion till rättsinformatiken, 8 ed, Norstedts Juridik, 2004

[Schweighofer 1999] Erich Schweighofer, The Revolution in Legal Information Retrieval or: The Empire Strikes Back, 1999 (1)The Journal of Information, Law and Technology (JILT) http://www2.warwick.ac.uk/fac/soc/law/elj/jilt/1999_1/schweighofer/

[Schweighofer 2006] Erich Schweighofer, Computing Law: From Legal Information Systems to Dynamic Legal Electronic Commentaries, in Cecilia Magnusson-Sjöberg and Peter Wahlgren (eds), Festskrift Peter Seipel, Nordsteds Juridik, 2006, p 569-588

[van Noortwijk 2006] Kees van Noortwijk, Ranking and Classifying Legal Documents using Conceptual Information, 2006 (1) The Journal of Information Law and Technology, http://www2.warwick.ac.uk/fac/soc/law/elj/jilt/2006_1/noortwijk/