A THESAURUS FOR IMPROVING INFORMATION RETRIEVAL IN AN INTEGRATED LEGAL EXPERT SYSTEM

ANTONIO CAMMELLI - FIORENZA SOCCI
ISTITUTO PER LA DOCUMENTAZIONE GIURIDICA - Italian National Research Council, Florence - Italy

1. Decision making Legal Informatics today: Evaluation and Prospects

The impact of computer technology on the legal world has brought about a fundamental change and this is widely recognized but a more analytical examination of the changes that have been brought about is still today an area that deserves greater attention.

Two disciplines that are conceptually so far apart have, in fact, proved useful to each other in evolving in such an interaction that they create the pre-conditions for a new discipline - legal informatics.

It is now established that documentary legal information technology has reached a point that it provides such a useful support in the legal world that is beyond discussion and, indeed, cannot be substituted, even if increasingly sophisticated information research and retrieval technology continues to be the subject of research and is susceptible to evolution and improvement. It would already be interesting in this domain to bring to the fore all the consequences that such information technology has produced in lawyers' minds and in their working life.

The biggest question marks about the positive results and substantial changes are, however, those originating from the use of the other emerging discipline in this sector, so called artificial intelligence, or better still, some artificial intelligence techniques that the lawyer may use for reproducing or simulating his reasoning.

In this sphere, the concept of the expert system or knowledge-based system or, in other words, a system capable of replacing the lawyer in some of his basic tasks or, at least, of offering him an “expert” aid similar to that which a human expert could provide, has become increasingly popular and all the sectors in which it can possibly be utilized are being explored. In fact, there are many problems of different kinds and dimensions. They are linked with the law-making phase, with the need for norms to be compressed and interpreted in an unambiguous manner, with their logical analysis, namely with the examination of their consistency and completeness within a normative system and, finally, with the creation of the individual norm or, in other words, with the decision of the judge who applies the general rule to the actual case: judicial law-making. Judicial law-making, due to its nature, is certainly fraught with considerable problems and is,
therefore, heavily weighed down by legislation, like that in Italy, which is chaotic and disorderly, with laws that are not easy to read and interpret.

Faced with a similar framework of reference, the advent of artificial intelligence was perhaps greeted with excessive optimism and enthusiasm, but advances in this sector have largely been disappointing because they have been both slow and laborious and because, in practice, they have not gone beyond examples and prototype systems.

The initial objective from which, as a research team, we started was to examine the norm, that is, to formally verify its consistency. The representation of knowledge in formalized classifications has achieved some results but in an instrumental and indirect way in the sense that it has demonstrated to us the extent of the deficiency of logical and linguistic rigour in many normative sentences.

From a less ambitious point of view, that of the decision-making function, better results can be said to have been reached, and, therefore, more ideas have emerged. The experimentation carried out over recent years has covered a reasonably wide area. We have built expert systems in the family law domain, relating to some particularly complex laws like that on divorce and the voluntary termination of pregnancy; we have experimented with some tools like simpler (Xi plus) and more complex shells (Flex), and, therefore, with different types of knowledge representation systems, such as production rules and frames. Today, our systems can be consulted fairly easily.

At this point, we are, therefore, now able to make some comments.

Above all, as we have already seen, whether or not we are dealing with a clear normative corpus or not emerges in the knowledge representation phase; but over and above this kind of assessment, there is no tool which automatically tells us where there is a gap or a contradiction.

Furthermore, leaving aside more theoretical questions for the moment like, for example, whether legal reasoning is deductive and, therefore, whether a tool based on this type of reasoning is suitable for reproducing it, let us now look at more practical issues.

When the judge carries out his institutional duties he has, on the one hand, the facts of the case before him and the norms, on the other. All he has to do is to give a description of the concrete case and identify the applicable norms in the legal order. The initial phase involves all the knowledge that goes beyond a mere legal evaluation but individual kinds of knowledge, evaluations of the circumstances relating to the case and so on come into play. These elements will unquestionably also weigh on the choice of the norm. In this sense, it is often said that it is the judge who creates the law in the sense that he establishes the individual norm.

It is now reasonably intuitive how intelligence attributed to a machine is a fictitious extension of the concept of the capacity to make the same kind of abstraction that a human being is able to make. The limits of expert systems must, therefore, be carefully assessed in both emphasizing their function in aiding the judge in his decision-making and in identifying the normative sectors in which such tools can effectively be employed. So while the application of such systems
in contexts where the possibility of solving a series of cases with common or more strictly procedural normative parts that require repetivity seems very useful, standard features or the application is, instead, more debatable and probably riskier to extend it to the facts of the case which require stricter interpretation by the court.

Naturally, the assumption for having written norms available with such a logical rigour that they can subsequently be formalized without strongly resorting to forcing them, could be considered one of the positive results of the interaction between law and informatics and, probably, resort to methods for formalizing even more complex contexts would not be so risky in producing misleading solutions.

In our paper, we plan to illustrate some of the features of the system we are building and, in particular, the construction of a thesaurus that is suitable for becoming a valid aid for the user for the purposes of using the knowledge available in the expert system better: a new generation expert system that, as a result of the experience we have gained, will not act as an alternative to the lawyer but should be considered by him as an advanced tool for information retrieval in a domain.

Practical considerations relating to the actuality, on the one hand, and to the flexibility of the system that should have the possibility of being implemented or corrected according to the consequent up-datings of the legislation being referred to, on the other, have also been influential in this evolution.

2. Thesauri for an Integrated Expert System

From all our previous experiments, we have come to the conclusion that there is a need to have complete, accessible and easily manageable systems, that is, systems, substantially, that are capable of following the development of a piece of legislation from the time of its enactment until it is repealed.

For this reason, we can see today that, along with the concept of expert system, there is the spread of a concept of advanced retrieval systems using, in practice, knowledge processing techniques based on artificial intelligence and on specific techniques for advanced retrieval. The two historical sectors of legal informatics - the documentary and the decision-making sector - after years of running in parallel come, therefore, together in constructing integrated expert systems where together with the classical presumption of decision support there is the equally important aspect of the full information retrieval.

Thesauri, in this framework, are not only a valid support for legal information retrieval systems but, in our opinion, can also have a central role in improving the fruition of articulated knowledge in an integrated legal expert system. Their main role is that of facilitating the acquisition of additional knowledge required for better utilizing an advisory legal expert system like the one we have been designing and experimenting with for several years.
Our experience with designing systems and the current literature available on expert systems in the legal domain have confirmed the need for enriching the system with additional information and lawyers are welcoming the novelty of these systems in the sense that, in the legislative sphere, for example, expert systems have, without doubt, made a contribution in the phase of legal decision-making and, at the same time, can provide useful suggestions in the direction of improving legislative drafting.

We believe this changing interest in decision-making legal informatics was made possible not merely because of the technological progress that has made sophisticated and flexible technology available today at reasonable prices but mainly due to the fact than many lawyers have understood that they themselves could become central figures in the building of an expert system, above all in the delicate knowledge representation phase.

The formalisation of the knowledge base not only represents an important step in expert system building but is also a central phase in a complex procedure of text and logical analysis of the normative sentence. For all these reasons, we are attempting, especially in the legislative sphere, to broaden the available knowledge base in order to see whether the ambiguities or gaps that have been found have, subsequently, been examined and in some way regulated. It means, in effect, updating the normative text so that the search becomes exhaustive, valuable and consistent.

After careful thought, we have attempted to make the system precise, correct and, in the last analysis, able to meet the objectives we have set. For this purpose, we have developed supplementary files containing non formalized knowledge taken from notes to decisions, case law and legal authority so that the represented domain is more exhaustive, so that the main user of our system (the man in the street) is provided with a more thorough and up to date search.

In our experience of the Italian legislative domain, we have often felt the need to resolve the gaps and ambiguities in ways other than those that the legislative text in itself is able to do. In these cases, it is advisable to present the user with all the updates of the text (if they exist) that can be queried during the search phase in the expert system with normal advanced retrieval procedures.

![Figure 1: Scheme of an integrated legal expert system](image-url)
Given this configuration, there is still the problem of making all the knowledge in the system easily utilizable: namely, of joining, in a harmonious and consistent manner, the system’s formalized knowledge and the knowledge that can be utilized from the files mentioned earlier. We believe that, if the shift between the two phases is to be facilitated, it is useful to have a support thesaurus, that is, a grid of key words that act as a filter directed towards the choice, within the dictionary of the supplementary files, of key words or, in other words, descriptors and, consequently, of contextual documents that are believed to be useful for going further into or filling in the search phase.

3. THES GIUR: A Thesaurus for improved Information Retrieval

In order to meet the objectives described above, we have created and are experimenting a thesaurus of legal terms called THES GIUR. It has constituted a very difficult task because the choice of the basic descriptors has had to contend with the ambiguity and difficulties of legal language compared with natural or literary language. A fairly strict criterion was used in choosing the descriptors so that the relations that are assigned do not expand too much in the automated generation phase resulting in the production of misleading and, therefore, noisy, links. Therefore, those terms, among the lexical relations, that derive from the same morphological root as the descriptor and whose use can change the syntax of the phrase, but not the sense, were defined as Equivalents (EQ). This relation appeared to be useful not only in broadening the user’s field of research, but mainly in enabling him to employ even richer linguistic material especially in dealing with legal language where, although it is not infrequent to find lexically equivalent terms, they are to be found in articulated semantic levels: one example of this is the chain of equivalents: adottare (to adopt), adottato (adopted), adozioni (adoptions), adottante (adopter), adottabilità (adoptability).

With the relations defined as Homography (HM), our aim was to attempt, in some way, to contain all the ambiguity of the polisemy of the homographic terms, where the homographic terms imply meanings or, legally speaking, very distant legal institutions. In this case, the Homography relation gives the user only the equivalent terms, restricting itself to pointing out two or more semantic paths that the term suggests. For example, we have two paths to choose from for the word appalto (contracting): either the implemented relations beginning with the term accollo (taking over), or the relations deriving from the term rivendita (resale), shown to the user in advance, by displaying appalto1, appalto2, respectively.

As far as semantic relations are concerned, we were very parsimonious in our use of the Synonymy, (SN) relation. We have seen how legal language continually interweaves, collides and interferes with natural language, but, in this phase of selection, we mainly contemplated the technical aspect and, therefore, only used this relation in the few cases in which the synonymy was considered to be complete.
In relations of superiority, *Broader Terms (BT)*, and of inferiority, *Narrower Terms (NT)*, we scrupulously complied with the hierarchical relations laid down by the law. As we shall see later, these three semantic relations constitute the central core of our thesaurus.

With the *Related Terms (RT)* relation, we selected those terms that can be identified as partial synonyms of the descriptor or that have close conceptual links with the same descriptor. For example, *contratto (contract) RT disdetta (rescission), liquidazione (winding-up)*.

Following this, we made choices or laid down selection priorities aimed at not only aiding the user in his use of the system but also in making the approach more consistent. The first screen of the thesaurus displays the descriptor (DES) and the three semantic relations synonyms (SN), hyponyms (NT) and hyperonyms (BT). As an option, the relations in the second screen can be displayed, namely, Related Terms (RT) and Equivalent Terms (EQ) whose aim is to conceptually broaden what has already been expressed by the descriptor. The initial screen displays the adopted semantic relations. In this regard, see figure 2 where the initial display relating to the term *società (company)* is illustrated.

![Figure 2: Initial display relating to the term "Società"](image)

The Thesaurus was built in DBase III for facilitating cross-searching of data banks, even by making significant changes without altering existing data, thereby giving the system an open configuration with the possibility of inserting new relations. Figures 3 and 4 illustrate the different files of the thesaurus and the various alternatives available to the user after gaining access to the data bank.

<table>
<thead>
<tr>
<th>name</th>
<th>identifier</th>
</tr>
</thead>
<tbody>
<tr>
<td>Multiple Word</td>
<td>MW</td>
</tr>
<tr>
<td>Equivalents</td>
<td>EQ</td>
</tr>
<tr>
<td>Synonyms</td>
<td>SN</td>
</tr>
<tr>
<td>Related Terms</td>
<td>RT</td>
</tr>
<tr>
<td>Homograph</td>
<td>OM</td>
</tr>
<tr>
<td>Narrower Terms</td>
<td>NT</td>
</tr>
<tr>
<td>Broader Terms</td>
<td>BT</td>
</tr>
</tbody>
</table>

*Figure 3: Relations and identifiers*
<table>
<thead>
<tr>
<th></th>
<th></th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Multiple word</td>
<td>MW</td>
<td>N</td>
<td></td>
<td>N</td>
<td></td>
</tr>
<tr>
<td>Equivalents</td>
<td>EQ</td>
<td>Y</td>
<td></td>
<td>Y</td>
<td></td>
</tr>
<tr>
<td>Synonyms</td>
<td>SN</td>
<td>Y</td>
<td></td>
<td>Y</td>
<td></td>
</tr>
<tr>
<td>Related Terms</td>
<td>RT</td>
<td>Y</td>
<td></td>
<td>N</td>
<td></td>
</tr>
<tr>
<td>Homograph</td>
<td>OM</td>
<td>Y</td>
<td></td>
<td></td>
<td>N</td>
</tr>
<tr>
<td>Narrower Terms</td>
<td>NT</td>
<td>N</td>
<td>BT</td>
<td>N</td>
<td></td>
</tr>
<tr>
<td>Broader Terms</td>
<td>BT</td>
<td>N</td>
<td>NT</td>
<td>N</td>
<td></td>
</tr>
</tbody>
</table>

1. Relation Description
2. Relator
3. Symmetric Relation? Y/N
4. Relator of Converse Relation
5. Transitivity Relation? Y/N

Figure 4: Relation and assigned properties to each relation

In figure 5, a general scheme of the thesaurus system is presented, underlining, once again, its open structure, namely, the ability to insert, at any time, further relations that are thought to be necessary like, for example, opposite or antonymous terms.

Figure 5: General scheme of the system. Example of open structured with the possibility to insert, at any time, further relation
For this thesaurus, we have taken advantage of the years of advanced documentary experimentation carried out for the very large legal data bank of the Institute which also has a thesaurus as an aid to better retrieval in the information retrieval system. This approach was assessed and discussed for the purpose of evaluating legal harmonization and pertinence in the automated generations of the thesaurus. Once the structure of the thesaurus was defined, we evaluated the configuration that could best make the links between the various files and, in particular, move between the thesaurus environment and the non formalized supplementary knowledge files. We developed an easy and reasonably complete system called the TextSearch system, that allows the stored documents, which are fairly different with regard, for example, to their format, to be retrieved. This means (which is common enough in legal documentation) legislation, court decisions and decrees that go from a few lines to many pages. Therefore, it was necessary to structure every document in such a way that every file became a collection of documents that could be perfectly identified and found: the identification details are the usual elements found in legal documents (kind of document, date, number, title, etc.). All the documentation stored in this way has its own dictionary that also sets out the frequency of the entry in the documentary universe. The scope of THES GIUR, built and implemented with the descriptors found in these files, is to provide an initial framework of reference by subject by suggesting, with its links starting off from the descriptor base, possible paths for textual research. It is up to the user to leave his navigation within the thesaurus environment and move on to the analysis of documents and to display all the documents in which the term he is looking for is to be found. In the initial phase, having decided on a reasonably rigid structure, all the basic descriptors and codecriptors in THES GIUR can be found at least once within the stored documents.

The results of this new phase in legal expert system building are encouraging and more or less confirm our initial scientific prediction, that of linking the two main elements of legal informatics by using a single advanced form of documentary expert system.

An integrated expert system is, therefore, more functional for achieving initial objectives: in fact, an expert system based only on formalized knowledge risks becoming a system that is not suitable for providing advice. One example from the many that we count amongst our experience is enough to illustrate this. The Italian law on divorce (No. 898 of 1970) lays down that, after a legal separation, the spouse who has been held responsible for supporting the other spouse should provide a sum of money that the law defines as “adequate”. It is understandable that the interpretation may be very varied and, consequently, controversial. Ambiguity is the exact opposite of certainty in the law which, somewhat ambitiously, is one of the foundations of our work. Well, the Constitutional Court, after various petitions, handed down two decisions, one of which in 1990 that clarifies the ambiguous formulation of the legislative text. The Supreme Court, in fact, holds that the weaker spouse should maintain the same standard of living that he or she
had during the marriage. It is obvious that an expert system on divorce cannot do without informing users about this decision which solves the original ambiguity in the legislative text.

Experience in building expert systems in the legislative domain has enabled us to see other opportunities for using the systems we have designed providing, also, suggestions for better law making. It is necessary to repeat that these systems cannot, in any circumstances, substitute those working in the sector, and preference is always given to the aim of the search and not to the tool that has been used. Remaining within the application domain of expert systems in the legislative field, we can state that textual formalisation is one of the most secure methods today for reading a legislative text, by addressing the legislator’s attention towards drafting legislation more carefully. It is to be hoped that there will be a rationalization of the lexicon of the law so that a dangerous cause of ambiguity is eliminated. Within this framework, the importance of making legislation more comprehensible and plain is understandable and this lays the foundations for closer adhesion to the text and, consequently, for its more secure and effective application.

We should, also, add that a well written legal text is not only more legible but also more easily stored and certainly more capable of being used in a legal information retrieval system. On various other occasions, we have already stated that even a lexical rationalization would constitute a positive step forward: if all the Italian regions, which are important sources of normative instruments would, for example, call “venatorial activity” “hunting” or any other equivalent syntagm, this would be a good start towards improving the quality of information.

The present phase of experimentation involves initial critical verifications by those who work within the legal world, namely, all those potential users of our systems. However, we are not neglecting any verification by the community which, in the final analysis, has the duty to obey and conform to a legal order.

The results are encouraging and positive and we believe the configuration of the integrated legal expert system, which we are discussing here, is the most suitable and complete as both an advisory system and for the logical analysis of the legal sentences under examination.

