Knowledge-based Hypertext for Legal Documentation

ROBERTO COLOTTI, ROSA MARIA DI GIORGI,
BONA INGHIRAMI, ROBERTA NANNUCCI

1. INTRODUCTION

Legal documentation appears to be particularly suitable for hypertext applications because it deals with texts which, by their very nature, are highly correlated, both explicitly and implicitly: legislation refer to other legislation, court decisions interpret legislation or refer to other decisions, legal authority comments on both case law and legislation. Therefore, hypertext systems become precious tools for acquiring complete legal information [Wilson 1988] [Wilson 1992].

The reference sources for integrated legal documentation fall into the following categories:

- Legislation: international conventions, treaties and agreements; European Community legislation (directives, regulations, etc.); national legislation (central government, regional, provincial and municipal legislative measures) that specifically regulate the matter or a different matter which is, nevertheless, pertinent to the subject being dealt with (for example, in environmental law, regulations regarding town and country planning, public health and health and safety).
- Case law: decisions made by the Constitutional Court, the Council of State, Regional Administrative Tribunals, the Supreme Court, lower courts, etc.
- Legal authority: text books, treatises, monographs, commentaries, collective works and proceedings of conferences, encyclopaedias, journals. This type of documentation may also include the codes, repertories and case note reports and bibliographies, namely collections of legislation, case law and legal authority, respectively.

2. CROSS-REFERENCING LEGAL SOURCES

If the assistance offered by hypertext systems makes them the ideal tool for consulting legal documents in general, which are so rich in notes,
citations and references, this is even more so in the case of legislation, where this feature is so accentuated that a single section of a law often becomes incomprehensible unless it is read together with other sections and/or other texts. Sometimes, it is even impossible to understand the subject of the legislation we are reading about if we do not resort to this kind of cross-referencing [Wilson 1990].

In the legislative domain, references, also known as citations, are of particular importance since they determine whether the legislation is in force, its enforceability or integration. By legislative references, we mean all the referral between the text of one Act and another (external references) or referrals to another part of the same Act (internal references).

An amendment, instead, is defined as any provision bearing in any manner on a prior legislative text or which, in any way, affects the normative content of provisions previously in force. There is a division within the amendments that may be termed as: substitution, integration, abrogation, derogation, prorogation and suspension [Osservatorio legislativo interregionale 1991]. This typology may prove to be interesting for hypertext applications in that each type of reference or amendment could be associated with a particular type of link.

From this point of view, the hypertext links may become the links for providing users with an up-to-date version of the legislative document thanks to the processing of the references. In fact, something more than just pure and simple navigation can be carried out with regard to amendments. These references apply in a certain period whereby it is possible, at any time, to have a specially reconstructed virtual text by incorporating the various amendments containing only the provisions in force then [Maioli et al. 1993].

With regard to case law data, the most interesting decisions are those complete with the massima or, in other words, a brief abstract of the principle of law applied in the given case. Certain law journals specialize in the publication of the «massimas» alone; others also report, fully or in part, the complete decisions. While a useful information tool, the «massima» may, in fact, turn out to be incomplete, making it necessary to refer back to the original decision. Furthermore, there are often references to precedents in the body of the decision.

Published decisions are frequently accompanied by a more or less extensive comment written by scholars on the subject, either critical of the decision or supporting it, citing previous case law and legal authority, whether upholding the decision or contrary to it. But even if we leave aside this particular type of contribution, known as the case note, legal
authority abounds with this type of reference. Authors support, refute, integrate or simply refer to opinions expressed by other authors and to interpretations given by the courts.

Finally, it goes without saying that, for both the legal scholar and the judge, the necessary point of reference is, in all cases, the provision they are called upon to interpret or apply.

3. A HYPERTEXT APPLICATION IN ENVIRONMENTAL LAW

The hypertext model allows us to approach information through a process of comparison which is similar to that of the human mind: the user «navigates» through the information along a network of links laid out between the single pieces of information that constitute the nodes. In these systems, special attention must be paid to the user interface if the system is to be easy to learn and user friendly.

On the other hand, the freedom in management potentially offered by the hypertext model also introduces certain drawbacks in searching a documentary collection that is too broad, due to the limited power of the help functions provided. A «knowledge overload» may occur to the user’s detriment, given the active role he is required to play [Baird et al. 1990] [Sillitoe et al. 1990].

3.1. The Hyperlaw project

A system prototype, called Hyperlaw was designed and implemented using the Hypercard software in a Macintosh Apple environment. This was part of a joint research project between the Istituto per la Documentazione Giuridica of the Italian National Research Council and the Department of Computer Science of the University of Padua. The aim of the project was to assess how the potentialities of hypertext technology could be used to design a tool for integrated management of the wide and non-uniform body of legal information [Agosti et al. 1989a; 1989b] [Di Giorgi, Nannucci 1991].

1 The research began in 1989 as part of a joint project between the Istituto per la Documentazione Giuridica of CNR and the Department of Computer Science of the University of Padua within the «Observatory on Environmental Law» Strategic Project, promoted by the Environmental Committee of the Italian National Research Council. It was subsequently developed (1991-1992) within the Institute also as part of a research unit within the Italian National Research Council’s Strategic Project «Metropolitan Areas and
The collection of chosen documents was that relative to noise pollution. The choice was not fortuitous. First of all, there is much interest in this topic in a society like ours, where noise has reached levels which are to say the least worrisome. Secondly, this is a sector in which the relative legal documentation, at the moment, is vast enough to constitute a sufficiently reliable test for evaluating the performance of such a system and at the same time limited enough to be managed with some pretension of completeness. Beyond the scientific aims of the experiment, the intention was also to supply interested users with a tool they could actually use. And finally, due to its necessarily technical implications involving graphs, tables and drawings, the legislation regulating the subject area is well suited to exploiting what hypermedia technology has to offer.

Hyperlaw is based on an information retrieval system model called Explicit (Fig. 1). This model uses a two-level structure to make the different
parts constituting the body of data explicit (and hence the name) at the document level, called the hyperdocument, and at the auxiliary data level, that is, at the level of the semantic structure according to which the indexing terms are organized, called the hyperconcept [Agosti 1988] [Agosti et al. 1991].

The two levels are linked by the relations between the concepts and the documents the concepts describe. At the same time the single elements contained on each of the two levels are interconnected: the documents are linked by references or citations, while the links among the auxiliary data are made up of the semantic structure in which the terms are placed.

The EXPLICIT model makes it possible to display the conceptual structure of the indexing terms, so that the proper semantic context in which each term is placed becomes known according to the meaning it is given in the indexing phase. This semantic structure can, at all times, be actively searched by the user, who, therefore is able effectively and immediately to enhance his query. Likewise, the documents linked to each single term may be displayed at any time [Colotti 1991].

The model makes use of a conceptual framework tailored for a specific application domain and makes that scheme available for active utilization by the user during his interaction with the system, thus providing a reference structure for the process of query formulation. Another feature is concurrent use of different conceptual schemes for the same application domain. This function can prove to be very effective in providing access to the same document collection by user groups with different information needs: a lawyer who is handling a case, a professional whose aim is to identify the constraints to which he will be subject when carrying out his activities or, finally, an ordinary citizen who is searching for the regulation regarding a specific problem of his own - three different needs, quite probably expressed in different terms and differently worded, yet all three addressed to the same collection of documents [Di Giorgi et al. 1989].

The Hyperlaw system has, therefore, experimented with the use of hypertext as a support tool for the associative web provided within each level and between the two levels of the architecture set in the EXPLICIT model.

3.2. The evolution of the project: Hyperlaw2

Given the interest aroused in both the information science and legal domains by the hypertext prototype created in 1990, the experiment was continued as part of the «Metropolitan Areas and the Environment»
Strategic Project promoted by the Italian National Research Council in 1992. A new version of the system, Hyperlaw2, was, therefore, created.

Since the first version had been implemented using a product which was less readily accessible to the majority of potential users, the primary objective was to transfer the experiment to a software, Toolbook for Windows 3.1, compatible with the MS-DOS operating system. We also experimented with processing data that was not exclusively text-based but also involving graphics – like, for example, tables, tabulations, designs and diagrams. The legal domain also remained the same for this second version: «noise and vibrations» within the field of environmental law.

3.2.1. The document base

The document collection of Hyperlaw2 includes legislation, case law and legal authority (Fig. 2).

In the pursuit of completeness, at least at the national legislative level, all those articles of the Civil and Criminal Codes and of the special laws which bear more or less directly on the subject were stored as well as the decrees implementing the numerous European Community Directives existing in the domain. In actual fact, noise pollution is an area which has been sadly neglected by the Italian legislator. There is still no law regulating

Fig. 2. Main Menu of Databases in Hyperlaw2
the subject in a unified and complete manner. Instead, as far as case law is concerned, due in part to this very lack of legislation, this abounds.

**Fig. 3. Documentary Unit: Legislation**

**Fig. 4. Documentary Unit: Case Law**
«Massimas» relating to the decisions of the different courts, from the Constitutional Court to the civil and criminal divisions of the Supreme Court, to the Council of State and finally to the lower courts deciding on the merits (magistrate's courts, tribunals, courts of appeal) were selected. And then the contributions taken from legal authority: monographs, miscellaneous works, conference proceedings, encyclopaedia headings and articles published in legal journals, many of which are case notes were added (Figs. 3, 4, 5).

3.2.2. *The semantic relations*

The semantic structure used takes the form of a classification table specially created for the system prototype. It is made up of a collection of terms relevant to the subject domain, subdivided within a hierarchically-organized table (Fig. 6). Currently, therefore, only vertical relations are handled. Nonetheless, a more detailed table will be built, a truly domain-oriented thesaurus which, alongside the already established hierarchical relations, will introduce relations of association and preference. The terms used are both those pertaining to the subject domain (namely noise pollution) and more strictly legal terms (such as competence, liability, penalties, etc.). The next step would be the creation of autonomous semantic schemes.
that makes it possible to take different approaches to moving around within the web of the hypertext [Chen 1990] [Davis 1986].

3.2.3. Navigating in the hyperdocument and hyperconcept

Special attention has been paid in Hyperlaw2 to the user interface: the work environment is totally graphic and user-system interaction occurs almost exclusively through use of the mouse. The interaction comes by clicking an information object; the system responds by displaying its contents.

Access to documents may be obtained either directly or through the semantic reference structure. The two types of access relate to the two levels of the reference model used for the system development: hyperdocument and hyperconcept. From whatever level the user begins his search, he can by following the links move either toward elements on the same level or toward elements on the other level by simply instructing the system to activate a reference.

In order to restrict the effects of disorientation, as navigation support the system offers functions which allow the user to go back over the trail he has taken up to that point, either node by node or by jumps [Kornbrot Macleod 1990].
Hyperlaw2 manages, as we have already mentioned, three distinct sets of data: a collection of legislative documents, a collection of case law documents, and a collection of documents taken from legal authority. Within each of these sets a network of links has been built that correspond to references to the legal texts (citations, amendments, etc.). Furthermore, another associative network links the elements of the three sets.

For each node, the links are subdivided into two categories according to the direction of the reference: active links and passive links. The term ‘active link’ refers to an association that originates in the current node, while the term ‘passive link’ refers to a citation from another node. For example, a citation becomes an active link to the document it cites and a passive link to the document cited: the different semantic value of the links between the two nodes is evident.

The semantic structure used for describing the information content of the document collections in the database is directly accessible for identifying significant terms. Once a node-term is activated, the system displays the node-documents linked to that term, distinguished by their type (legislation, case law and legal authority) (Fig. 7).

**Fig. 7. A Dialogue Example**

*When activating the reference button from the database, strings of documents retrieved through active or passive links, distinguished per typology, appear*
Fig. 8a. Advanced Search
An example of advanced search through keywords connected by boolean operators

Fig. 8b. Advanced Search
Result of an advanced search with documents retrieved according to their typology; their text can be displayed on demand
A special function, *advanced search*, permits the user to construct a virtual hyperconcept node starting from one or more terms. This function makes use of Boolean operators to determine unions, intersections or disjunctions with the references that link the documents to each of the selected terms. The user may, whenever he wants, move up and down between the two levels that constitute the system architecture. If it is possible to start from any term in the hyperconcept and move toward the document nodes of which the terms are the semantic representations. It is, likewise, possible to start from a single document-node and move toward the semantic term-nodes linked to it. Through this movement, the semantic context in which the document is placed becomes explicit and interactively able to be explored (Figs. 8a, 8b).

### 3.2.4. Other functions

Besides navigation, the system also offers the user a series of other functions designed to aid him in his consultation. One of these enables him to *print* selected documents once the search has been completed. Another function makes it possible to *store* a previously-consulted node in a special section of the system’s memory. Given the high degree of freedom of movement during a search, it may be difficult for the user to remember which documents he has already examined and found interesting or worthy of further analysis. With this last function a point of reference is created for each single node. A specially-designed procedure recalls and reactivates these references and re-positions the user at the desired point within the hypertext network. The user can, therefore, re-examine the node that interests him or start a new search within the data base. The references may be collected in a single list or subdivided in separate lists, stored for further consultations. The lists may be completely or partially deleted, or integrated by adding new references.

### 3.2.5. Updating and integration

In hypertext systems two aspects are critical for information management: the addition of new documents and the creation of new links between the newly-acquired documents and those already in the database. Thanks to the object-oriented approach on which the system is based and using a special data-description language, new documents can be acquired automatically from external text files. The same procedure, using the indications contained in the files, will link these new documents to the pre-existing
reference network. Another function, associated with that just described, makes it possible to delete documents as they become obsolete or to delete links which are no longer significant. Thanks to these two functions maintaining and updating Hyperlaw2's information base is greatly facilitated.

4. FUTURE DEVELOPMENTS

The experiment clearly demonstrates how useful hypertext technology can be for organizing knowledge in the legal domain. These techniques can improve the quality of the legal information insofar as they can provide, on a single support, such different types of data as the text of a law, a court decision and a contribution to legal authority, traditionally stored in separate data banks and accessible with very different retrieval systems. On the contrary, in hypertext systems the user can choose among the alternatives the route that best meets his query needs, and navigate among legislation, case law and legal authority until at the end of a search he has assembled a complete dossier of documentation on the issues that interests him.

One further development might be to handle the network of the hypertext links using AI techniques. As Rada suggests, this would involve transforming hypertext into *expertext*. That is, if the nodes in the hypertext network are related to the predicates and the links to implications, the implication will be activated when the predicate to a node is true. When a text or other type of information is associated with the node, the hypertext system becomes an expertext. We thus have an intelligent hypertext [Diaper Rada 1991] [Rada 1991].

Procedures might be associated with the links in the hypertext to make it possible, for example, to conceive differentiated routes within the hypertext network. During consultation, the system would ask the user about his level of knowledge of the specific domain and on that basis address him to different points within the network.

It would also seem that AI techniques could be extremely useful during hypertext system building. Developing the network of hypertext links can, in fact, be very arduous, especially when a large amount of data is being handled. Even only partially automating the implementation phase would greatly reduce the time factor involved in system building.

Integrated models consisting of data bases, expert systems and hypertexts are more and more often proposed: the unified system can provide access
directly oriented to specific document collections [Nanard et al. 1993] [Greenleaf et al. 1991].

Research in this domain, therefore appears to be especially promising and offers many prospects for applications in the field of legal documentation. Integrated systems would seem to be the direction development of legal information systems will most naturally take, since they enable the «intuitive» characteristics of hypertext to be combined with the «formal» features of expert systems [Kaindl Snaprud 1991] [Busse et al. 1993].

Further advances may occur with the evolution of system models based on neural networks, which may provide the solution to problems related to the automated construction of data links.

References


R. Colotti, R.M. Di Giorgi, B. Inghirami, R. Nannucci / Knowledge-based Hypertext 209


