Hypertext in Law: A New Paradigm to Explore Legal Sources

JACQUES SAVOY

INTRODUCTION

Hypertext systems manage information using a network of information units or nodes that are interconnected according to different links or relationships [Halasz 1988]. In such an environment, users navigate through the information network by following existing links or creating new ones, and the semantics associated with the links can guide users to appropriate nodes in their information search.

In education, hypertext applications emphasize the linking capability of this approach. The links established by the author or by the readers depict interesting relationships between various sources of knowledge, may clarify some points of detail or some ambiguous concepts. Moreover, based on the principle of «discovery learning», these relationships allow different perspectives on a given subject and favour the integration and comparison with other areas of knowledge. However, the particular features of the underlying domain must also be taken into account. This paper illustrates two essential difficulties appearing in legal applications namely: (1) the huge volume of documents that must be included to obtain an interesting legal hypertext, and (2) the need for an autonomous query-based search strategy to find relevant information in a large corpus.

Currently proposed hypertext systems do not provide an appropriate answer for the solution of these problems. As a primary means of accessing information, hypertext systems suggest browsing. Such an approach may be valid when working with toy-sized hypertexts because we do not need additional help when searching information within a booklet of 20 pages.

«Even in a 500 node single-user network, navigational access can be difficult as the network changes and its structure becomes heterogeneous. In these cases, navigational access is problematic because users tend to get lost while wandering
around in the network looking for some target information. Often, these users can describe exactly what information they are looking for, but simply cannot find it in the network.” [Halasz 1988, p. 842]

Moreover, the construction of a small hypertext can be done manually (for example, see [Nielsen 1990]). However, when the database grows or when the content of the hypertext is subject to many modifications, an automatic construction mechanism must be provided. When faced with a large amount of information, browsing is far from satisfactory, even with the introduction of a table of contents, an index, etc. [Alschuler 1989]. To promote the usability of hypertext systems, experiments with larger information networks have to be considered [Halasz 1988], and our research represents a step in this direction.

This paper describes a hypertext system designed and implemented for students in law at the University of Montreal. The first section provides an overview current educational applications of hypertext systems and the features particular to the legal domain. In Section 2, we describe three characteristics of our prototype namely: (1) the automatic transformation of legal texts into hypertext; (2) the linking and browsing features; and (3) the principles of our autonomous search mechanism used to complement the navigational tools.

1. Educational Applications

The physical composition of a book, a primary means of teaching, embodies three distinctive characteristics: it is portable, static and linear. The human brain by contrast does not work in a linear way. Humans think through association, comparison, inference, and they see many points of view when studying a given subject. In order to break with the two-dimensional structure of the classical book, hypertext offers an external support storing our thoughts in a manner that follows more closely the principles of action of the human brain.

If, as shown in Section 1.1, related educational applications are based on multimedia technology (sounds, images, photographs, video sequences, etc.), and natural language constitutes the primary medium to transfer information in legal studies. In this context, Section 1.2 specifies that not all textual databases may form an interesting hypertext application and that the legal sphere respects the golden rules of hypertext.
1.1. Use of hypertext in education

The linking facility provided by hypertext systems enhances their appeal as a computer-based learning tool by presenting not only the content of nodes but also pertinent relationships that can be drawn from different sources of knowledge. The first reported experience based on this new paradigm was conducted by Catano [1979] using the FRESS system at Brown University. The purpose of this study was to study how one can teach poetry and how students may analyze poems on-line by establishing relationships between related text excerpts such as other poems or literature fragments, by introducing their own annotations or by following remarks written by other readers.

Hypertext systems can also be used to learn foreign languages by establishing implicit links from any text to dictionaries or glossaries, or by presenting an explanation of any word or idiomatic phrase (see examples based on the InterMedia system [Yankelovich 1988]). Based on the hypermedia technology provided by the ATHENA project, Hodges et al. [1989] demonstrate how textual passages may be linked to video sequences to illustrate non-verbal information such as speaker agitation and mannerism or hearing a conversation in a foreign language. The project PERSEUS [Crane 1987] provides hypertext support for the study of ancient Greek literature. Storing relevant information on a CD-ROM, this system gives access to large amounts of original source texts in Greek, including a parallel translation. The study of original Greek text fragments are reinforced through dictionary lookup and automatic morphology analysis tools. Various links provide information to other original Greek sources, to interpretations given by scholarly and encyclopedia articles or by access to additional background data by linking text excerpts to cultural or historical notes, photographs and drawings.

Of course, the utility of hypermedia technology can also affirmed in other educational domains such as archaeology, chemistry or physics. For example, the specialized hypermedia system PALENQUE [Wilson 1988] using DVI technology (digital video interactive), can be exploited to teach Mexican archaeology to children by letting them take a tour on the Palenque ruins. Through photographs, the user may see the site before its restoration or in the days of the Mayas. In the arts, the Shakespeare project [Friedlander 1988] proposes a hypertext containing film clips from plays like Hamlet and Macbeth. In this prototype, links establish relationships between corresponding scenes in different films to illustrate how the same play has been interpreted by various directors and actors. Moreover, a simulation tool
called TheaterGame gives students the opportunity to stage their own interpretations of the plays from a database of hundreds of costumed actors and props.

Other research has been done in other domains of knowledge to account for their distinctive features; such as the use of simulations or guided examples to learn and play with mathematical models (e.g., the Ebook3 project [Savoy 1989]), or the automatic transformation of a linear medical handbook into hypertext [Frisse 1988]. The HyperTIES system [Shneiderman 1987] was used in various museum applications to provide auxiliary background reports about the subjects of exhibitions. This project was especially designed to work with novice and casual users.

1.2. Particular problems related to the legal domain

The current trend in educational software seems to include various media to motivate the student and/or to illustrate technological advances and their applications. However, in the legal domain, text constitutes more or less the only source of information. In such a context, is the linear presentation obsolete? Does the hypertext paradigm represent a good solution for teaching? Can this new approach help students to learn, to think in another way and to augment their skills by showing pertinent relationships that can be drawn between different legal sources? Such questions require that we demonstrate the underlying criteria of hypertext are respected in the legal domain.

Not all textual databases have appropriate structures for applications in hypertext systems. Before converting texts from a given domain of knowledge into hypertext, the «golden rules of hypertext» must be respected. These criteria are the following:

- There is a large body of information organised into numerous fragments,
- These fragments relate to each other, and
- The user needs only a small fraction at any time». [Shneiderman 1989, p. 115]

These three characteristics are fully pertinent and respected in our legal context. Of course, when the database is limited to a few pages, an hypertext does not really represent an interesting application comparing to a printed document. Firstly, legal documentation forms a voluminous and dynamic set of information which upon inspection is composed of various parts, divisions, articles, paragraphs, etc. This documentation is composed of statute texts, court decisions, scholarly treatises, commentaries, guidelines issued to
adjudicators, and information supplied to members of the public. When considering that legal corpora are formed by structured documents, this feature can be used to automatically build an hypertext based on the logical structure of legal texts.

Secondly, to build a legal argument, students in law have to consult various related sources. For example, to clearly understand a legal concept occurring in a particular statute or to think about its implications, the reader has to retrieve court decisions – and perhaps scholarly treatises as well – to interpret a particularly fuzzy concept. Of course, court decisions on a given point of law may be contradictory and latter decisions may affirm, question or over-rule previous ones. In such cases, further interpretation is necessary to make the decisions fit a coherent pattern. Moreover, the evolution of social customs may impose a different reading of statute texts that have not been formally modified to take that evolution into account. Of course, where the evolution has been formally translated via a change in the statute, one would want to consult the most recent text as well the earlier one. Thus, to determine the answer to a legal question, a minimum requirement is to obtain all the versions of a given statute text together with the case law and scholarly writing pertaining to it. As one can see, legal documentation contains a rich amount of interconnecting relationships.

Thirdly, when searching relevant information pertinent to a given case, lawyers need only a small subset of all the available documentation. When students in law search for information related to a given question, they may know the pertinent article(s) of the law or some relevant cases law. If not, users may formulate a request to find good starting points for browsing. However, when reading a text passage, they are faced with the following questions:

1. Has this court decision been affirmed elsewhere?
2. Has this court decision been questioned?
3. Has this court decision been over-ruled?
4. Has this court decision been clarified elsewhere?
5. Has this court decision been confirmed?

To answer these questions, hypertext links may be very valuable because the related nodes may be reached very easily. Even though the underlying principles of hypertext are respected in the legal domain, we are faced with two remaining problems; namely: (1) how to transform legal documentation into hypertext, and (2) how to efficiently search relevant information from a large set of connected documents. Such problems were not explicitly considered in our previous learning system [Savoy 1989].
2. Our Prototype

The use of computers has a long tradition in the legal domain. In the sixties, commercial services offered lawyers and law students access to legal documents through an interface based on Boolean queries. In response to a user's request, the computer displays a list of references to the given Boolean query. With this list of these records, the student can then locate the desired documents in the library. Such principles are still in use in current commercial systems.

Nowadays, legal information systems have evolved from the classic online system based on Boolean operators to hypertext technology. Providing a more user-friendly interface, this paradigm may have a profound impact on the way lawyers work. Instead of only informing the user as to the existence of information, a hypertext allows the user to see the retrieved documents. From a given legal text, the student may follow hypertext links specifying relationships with other legal sources. This feature is especially useful in our context because a legal document can rarely be studied «in vacuo» or without consideration of other related legal sources. For example, links may be followed to explain and comment on ambiguous, vague, imprecise and fuzzy legal concepts or interpretations (e.g., what exactly does a «reasonable period» mean? Can we consider a «slimming center» to be a medical center?).

For our first application, we have built a legal hypertext system with Canada's unemployment insurance laws. In the future, we plan to develop another application dealing with legal questions relating to the consumer protection laws. No area of law however can be considered as an island; general principles, cases relevant to other areas of law, and other legislation may be needed for a true understanding. Such considerations lead to the conclusion that we have to incorporate a very large volume of information into our prototype. In spite of these limitations, our primary concern is to demonstrate that hypertext paradigm can be useful as an educational tool, even when limited to a particular legal domain.

2.1. From text to hypertext

The transformation of linear documents to hypertext is not an easy task. The main questions are to define the content of nodes and to establish pertinent links. To achieve this purpose, some research projects have adopted a manual approach within which the author had to (re)write the content of each document and to establish the appropriate links among nodes.
[Nielsen 1990]. This solution is time-consuming and boring when the electronic versions of all documents are already available.

Frisse [1988] and Rada [1992] suggest that an automatic transformation of text into hypertext can be achieved, based on the logical structure of a large document. Guided by this structure, the system may form nodes and establish hierarchical links between these entities. In our context, we also employ the division of a statute law into parts, divisions, and articles to generate the corresponding nodes of the hypertext together with their hierarchical links. The guidelines issued to adjudicators were also transformed, based on these same principles.

This approach is far from satisfactory when considering case law. Although each court decision has a logical structure, in our context we face with around 2,500 cases and we want to organize them into a meaningful structure. That is, we want to group law cases law together that relate to voluntary departure or court decisions concerning misconduct. In the current stage of our project, this classification is done manually by students. Even though automatic clustering algorithms do exist, as pointed out by Botafogo et al. [1992], such approaches are far from perfect.

«One caution is that while metrics can be useful in making precise and objective comparisons, they are only approximations to interesting concepts. Good judgment by authors is more important than staying within strict numerical limits». [Botafogo et al. 1992, p. 143]

Moreover, the content of a statute law cannot be viewed as static. Law paragraphs are changed according to political or economical reasons, and in such circumstances, the legislator produces a modifying law. Current studies in our research group tend to demonstrate that the automatic adaptation of our legal hypertext can be achieved at a satisfactory level because the syntax of a modifying law is strict and rigid often leading with very few ambiguous paragraphs (see [Kokoroko 1993]).

2.2. Linking and browsing mechanism

Based on the considerations stated in the previous section, the linking capability of our system is rather limited; a user may: (1) follow the hierarchic structure of a given statute law or browse around court decisions according to the given classification; and (2) obtain previous version(s) of a given paragraph of the statute text. Clearly, the linking capability of our system must be extended to take into account other pertinent inter-document relationships.
To achieve this, we have designed an automata [Choquette 1993], [Marcil 1994] capable when inspecting legal documents, to automatically establish relevant cross-references links. The elaboration of this approach is valid in our context because the writing of laws follows a strict syntax when inserting references to other legal sources. For example, to refer to a particular article or paragraph, the legislator uses the same pattern (e.g., «article 7» or «paragraphe 9(1)»). When inspecting the result of this first attempt in the Act Respecting Unemployment Insurance in Canada, we find 168 appropriate cross-reference links over a total of 210 (correctness of 80%). A similar pattern can be found in case law (e.g., «CUB 6666» in which «CUB» represents an anachronism for a court decision relating to unemployment insurance law and «6666» its number).

Why does this automatic scheme not work perfectly? Some written references do not refer to our particular domain but tend to establish links with other legal documents not contained in our hypertext system. For example, the reference «paragraphe 227(10.2) de la Loi de l’impôt sur le revenu» (income tax law) can be found in the Act Respecting Unemployment Insurance in Canada and refers to another statute. Even in a medium
size legal hypertext, relationships to legal sources outside the content of storing information can thus appear. When we modify our automata to ignore references leading outside the stored information (ignoring references which included «... de la Loi»), the system may establish 166 appropriate links over a total of 169 (correctness of 98.2%). To remove the remaining errors, a semantic interpretation of the intent of the reference is required, which is beyond the potential of our automata.

Finally, our system also includes glossary links establishing a relationship between a given word appearing in the law and its definition providing in the beginning of the statute. However, considering this relationship as an implicit link cannot be the right choice in our context because the definition of a given term or phrase may vary among paragraphs. For example, in a given section of a law, «Minister» should be interpreted as «the Minister of Employment and Immigration», and in the other parts of the same legal document, «Minister» means «the Minister of Finance».

2.3. Query based searching

The resulting system described previously may help students to answer questions addressed in Section 1.2 (see Ex. 1). When reading a case or a statute law, links are highlighted to indicate a relationship to a related document (see Fig. 1). Such an approach which puts emphasis on browsing can be useful when users do not known exactly what they are looking for or when they only want an overview of the content of a hypertext.

However, due to the huge volume of information contained in our hypertext, students may be quickly overwhelmed with information and they will not efficiently find relevant facts. In order to complement navigational tools, we have designed and implemented an autonomous query-based retrieval mechanism. This search strategy may be helpful when users may express their information need but do not know where to find the pertinent information. To achieve this, Salton & McGill [1983] suggest different retrieval schemes and some of them already have been used by other prototypes. However, the main question is: are these techniques really effective? To answer this question, we have to first evaluate the best known retrieval models based on two-medium size test-collections: namely the CACM collection containing 3,204 articles published in the journal «Communications of the ACM» and the CISTI corpus (1,460 articles extracted from the information science literature).

In order to affirm that a particular retrieval scheme is better than another, we compute its precision and recall for each query. Precision is defined as
the ratio between the number of retrieved and relevant documents, and the number of retrieved documents. Recall is defined as the proportion of retrieved documents that are relevant over the total number of relevant documents in the collection. These values are well-known as a measure of the capability of a retrieval system to select relevant information on the one hand, and, on the other hand, to reject non-relevant documents. Instead of reporting these values for each request and each retrieval strategy, Table 1 depicts an average precision at ten standard recall values for each alternative (for details about our evaluation methodology, see [Savoy 1993]).

Finally, to decide whether a search strategy is better than another, the following rule of thumb is used: a difference of at least 5% in average precision is generally considered significant and, a 10% difference is considered very significant. From the data shown in Table 1, we can conclude that both the retrieval strategy proposed by the HyperTIES system and the traditional Boolean model do not present satisfactory retrieval performance. The p-Norm model displays better retrieval effectiveness. However, all these solutions do not account for hypertext links which may represent additional information about document content because they establish a relationship between nodes containing similar concepts and dealing with analogous subjects.

In our retrieval strategy, we suggest enhancing the p-Norm model in

<table>
<thead>
<tr>
<th>Table 1. Evaluation of Existing Solutions</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Model \ Collection</strong></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Traditional Boolean Model Justus [Wilson 1990]</td>
</tr>
<tr>
<td>Signature File Hyperties [Shneiderman 1987]</td>
</tr>
<tr>
<td>Hybrid Boolean Model InterMedia [Yankelovith 1988]</td>
</tr>
<tr>
<td>Vector-Processing Scheme [Frisse 1988]</td>
</tr>
<tr>
<td>P-Norm Model [Salton 1983b] CACM: p-value = 5, CISI: p-value = 1</td>
</tr>
<tr>
<td>Our Retrieval Scheme Based on Bibliographic Reference fixed value for $\alpha_k$</td>
</tr>
</tbody>
</table>
order to account for hypertext links. This retrieval scheme can be interpreted by analogy with the way lawyers work. In order to prepare a speech for the defence, they justify their conclusions using precedential reasoning. In civil and common law, precedent is cited as authority in drawing legal conclusions in favor of a legal argument and also in determining which other arguments might favour that of the defendant.

Our retrieval scheme works in two stages. In the first, the p-Norm model attributes to each retrieved document a retrieval status value (RSV) computed according to its similarity with the given request (see [Salton 1983b]). In the second phase, the retrieval status value of the s-best ranked documents is spread to their neighbors according to the following formulation:

$$RSV(D_i) = RSV_{p-Norm}(D_i) + \sum_{k=1}^{s} \alpha_{ik} \cdot RSV_{p-Norm}(D_k)$$ for $i = 1, 2, \ldots s$ (2)

in which $\alpha_{ik}$ reflects the strength of the link between Nodes $i$ and $k$ (for example a fixed value depending on link type), $r$ the number of links connecting Node $i$ to its neighbors, and $s$ the number of documents for which we modify the retrieval status value. The value assigned to $RSV_{p-Norm}(D_i)$ is given by the p-Norm model.

The aims of our ranking strategy are both to retrieve a larger number of relevant documents and to move them higher in the result list. As shown in the last line of Table 1, one can see that our retrieval scheme may

---

*Fig. 2. Retrieval of Information Using Bibliographic Reference Links*
significantly enhance the results obtained by the p-Norm model using the bibliographic reference links between articles appearing in the «Communications of the ACM». Because the CSCI corpus does not include links, this finding cannot be confirmed with another test-collection.

Nevertheless, our approach may also be confirmed by theoretical justifications. A hypertext link represents a relationship between papers expressing similar ideas (for example, Nodes 1572 and 2789 in Fig. 2) without using the same terms as those given in the request (or in Document 1410 in Fig. 2). In the same vein, Furnas et al. [1987] have demonstrated that the same idea or concept may be expressed by a large number of terms.

CONCLUSION

This study demonstrates that legal hypertext systems can be a useful tool for exploring legal sources because legal documentation respects the golden rules of this new paradigm. Moreover, the linking capability representing the cornerstone of this approach reveals a direct similarity with the way lawyers work, as mentioned by Mital & Johnson «It is often said that lawyers think in hypertext» [Mital 1992, p. xi].

Faced with a huge volume of information, our prototype incorporated an automatic transformation from text to hypertext by dividing large documents into nodes and by establishing links automatically between these elements. The large size of our hypertext requires, in addition to browsing, a mechanism such as an autonomous query-based retrieval system. In order to develop such a scheme, we have evaluated approaches suggested by other systems and have found that one can improve the retrieval effectiveness of a retrieval strategy by taking hypertext links into account. The current study presents interesting results than can be apply on other research projects using a large collection of documents (e.g., using a CD-ROM).

As an educational tool, our approach emphasizes case law and discovery learning. Cases represent experience and thus play a primary role in learning. From past cases people learn to avoid failures and to adopt successful plans as appropriate. A great deal of scholarship would be involved in selecting, organizing and inducing the lessons implicit in past cases in light of, or in contrast to, existing theories of the domain. This aspect is supported in our system by browsing through the hierarchical structure of court decisions. Moreover, in our hypertext system, a balance between directed search and browsing paradigms has been established. Although our current prototype
works on a SUN workstation, we also plan to develop a version for MAC-INTOSH, based on the HYPERCARD software.

Acknowledgments

This research was supported by the NSERC (Natural Sciences and Engineering Research Council of Canada) under grant OGP0090940, by the FCAR under grant 93-ER-1557, by the SSHRC (Social Sciences and Humanities Research Council of Canada) under grant 410-92-1858, and by the SNFSR (Swiss National Foundation for Scientific Research) under grant 21-37345.93.

References


