Semantic Models and Knowledge Representation

Document Assembly and Evidence Analysis: Two Approaches to Hypertext

Vijay Mital, Anthony D. Elliman

1. Introduction

Although the term Hypertext was coined as long ago as 1967 and Hypertext like systems has been widely studied in more recent years we still have much to learn about its appropriate use in information systems development. Rada (1991) describes it as a new technology for producing text and draws the analogy with the introduction of the printing press to explain the necessarily slow pace of development of hypertext based systems. Unfortunately it could be this very analogy with books (and libraries) which is restricting our thinking.

In discussing navigational problems Nielsen (1990) uses the words «loosely structured information» as if this is an implicit feature of all hypertexts. This may be true in some situations, but we believe this to be the exception rather than the rule. Among the most successful and widely used hypertexts are the telephone information systems such as British Telecom’s Prestel and on-line documentation systems such as Microsoft Windows Help. There is also a significant interest in the technology as a media for educational material. In all these cases readers are quite separate from the authors who take positive steps to produce a well defined structure. As with the printing press many users see hypertext as a publishing media.

Central to this issue is the extent to which a link is an unexplained association between two pieces of text. In Bush’s Memex (1945) links are just subjective associations left within the corpus of text by its readers. The hypertext software cannot «know» why the association is significant, and as Nielsen (1991) says, there is «no easy way to specify general actions or computations on the data». This clearly inhibits us in exploiting the information processing powers of the system being used to store and retrieve the text.

In his discussion of the need for different link types Murray (1993) identifies two types of knowledge or text base:
• «Everything» in a subject area;
• The knowledge associated with a complex tool.

The second of these identifies information provided to support a «mission critical» task and Murray argues that the process of producing highly structured links within the hypertext is particularly important in this latter case. Help like hypertexts extend to other complex activities besides software tools and it can be argued [Huang & Elliman 1991] that there are some common linking structures across this whole range of applications.

However, these views are still restricted by the «publishing» analogy and we would like to add at least one more type of text base:

• The local knowledge associated with the execution of a business process.

This is a hypertext acquired and structured by an organisation or work group in the course of some business activity. It is local in the sense that, although the business process may be common, the text represents specific information accumulated by a particular group. Examples of this type are the issue based discussion models [Rada 1991].

Since we begin with a good knowledge of the activity our hypertext can include a range of link types to represent meaningful semantics from within the application domain. Further depending on the strength (or objectivity) of the semantics we may construct the software in such a way as to react to the different semantics to support tasks in an «intelligent» way. Another aspect of these applications is that the roles of author and reader are not so clearly distinguished and since users add or modify certain types of links according to their role within the business process.

We will explore the power of this view of hypertexts by looking at two rather different legal tasks which are supported by domain specific hypertext systems:

• Authoring of a new document by assembling information from previously prepared documents;
• Organisation of «dead», evidential documents involved in a litigation.

The first of these deals with a corpus of text which represents the accumulated knowledge of a work group spanning several cases. It demonstrates that hypertext offers a powerful alternative to other implementation strategies when the links are given relatively strong semantics so that the system can take simple actions to advise the user during the task. In the second example we deal with a case specific hypertext representing the current perception of
an action in progress. This illustrates a system where the strength of the semantics assigned to groups of link types varies. The links in the strongest group actually constrain user actions to maintain the integrity of the text base. In contrast we shall see that those in the weakest group have a more significant semantic value to the user but cannot be interpreted by the system.

2. Document Assembly

Much of the work of lawyers involves creating complex legal documents — wills, leases, contracts, pleadings, etc. There are in fact three kinds of activities which fall under the umbrella term ‘document creation’. These are form-filling, template-based drafting and document assembly.

Form-filling is the process of creating simple and common documents by reworking pro formas. Template-based drafting describes the process of creating more complex documents which require relatively minor variations on a theme using a template and boilerplate text. When varied situations have to be catered for and complex documents are required, the construction involves the drawing upon and merging of fragments from previously created documents. The fragments are altered or added to suit the current situation. This is the process of document assembly.

Document assembly is qualitatively different from the first two types of document creation, requiring considerable application of legal knowledge throughout the process. In form-filling the assessment of the situation at the very start yields the appropriate pro forma document which can be instantiated in a routine manner. In template-based drafting, many of the decisions en route to the final document require assessment of the import of the various alternatives or options. However, the assessment is not open-ended and is confined to choosing from a small set of predefined elements.

Document assembly is time-consuming, involving significant effort and the exercise of considerable experience and discretion. This activity often requires discrete tasks by a number of workers:

- Lawyers specify the outline form of the document;
- Para-legals and legal secretaries take the outline, insert relevant clauses from sources indicated by the lawyer and make the document structurally consistent;
- Finally the lawyer rectifies omissions and verifies the finished product.

The feature of document assembly that is relevant here is the reliance on precedents. Precedents are previously created documents which deal
with similar or related situations. Clearly, when there is a choice between constructing a document de novo and adapting one or more existing documents to do the job, there is often little doubt that the latter will be preferred. This is not plagiarism, nor even an unwillingness to expend the requisite effort. Reuse of time honoured language is simply good legal sense.

Sole practitioners and those working in small groups usually have some sort of filing system which holds documents that they have personally created or received in the course of their work. Consequently, they have at least a vague recollection of what to look for, and where. This allows them to cope with the complex task of using the valuable information contained in precedents. In a law firm with more than a handful of fee earners, however, this ad hoc approach is not viable, since the volume of information processed is large and users may never have previously read material which they are to, or should, use and adapt. Consequently, law firms are increasingly turning to computerisation of the task of organising, maintaining and making information available to multiple users.

There have been a number of approaches to the task of providing computer-based support for legal document creation. We will first describe an approach that supports the first two of the activities we have described above – form-filling and template drafting. Following that, we will introduce hypertext-based work that is meant for the third kind of activity – assembly of a complex document by drawing upon multiple precedents.

2.1. Knowledge-based template systems

This essentially involves the creation of a template into which all the related precedents can be held in an amalgamated form and the relationships between clauses are captured as 'knowledge'. There is some fixed text which must be included, and interspersed with or embedded within the fixed text are markers or delimiters which indicate the places where some action, such as the following, is required:

- An optional piece of text must either be retained or omitted;
- One or more of some given text alternatives have to be selected;
- Some numerical value has to be calculated based on the state of facts and inserted;
- Some information in text form has to be inferred from supplied facts by the system and inserted;
- The user has to add some text.
Each of the actions implied by the delimiters is essentially an act of deriving or acquiring information. This may be done by executing a procedure which causes some calculations to be done or some rules to be invoked in order to derive the requisite information on the basis of the facts supplied to the system by the user. Alternatively, the information may be acquired directly from the user.

The seminal work in knowledge-based document drafting systems was done by Sprowl (1979) under the aegis of the American Bar Foundation. The system, and the representational language underlying it, are entitled ABF. Most commercially available document assembly systems which claim to employ knowledge-based techniques are based on Sprowl’s work to some degree [Evans, 1990; also see Wilson, 1990 which adds a hypertext interface to such a system].

2.2. Limitations of the knowledge-based approach

There is little doubt that any set of documents can potentially be reduced to a template in terms of the kinds of constructs provided by ABF and its successors. The effort involved in this reduction is likely to be proportional to the benefits only when the set has a sizeable substratum of commonality. The knowledge-based document template is useful in areas of legal practice involving documents which may be said to be relatively minor variations or adaptations of stereotypes. Where, however, there are large number of variations in the kinds of documents required within a particular domain, it may not be desirable, or feasible, to undertake the considerable programming effort required.

Clearly, the knowledge-based template approach is inappropriate for document assembly. It is not capable of guiding the user through complex and interrelated fragments of information in the form of multiple precedents, where the user – rather than the system – is to monitor and assess information and choose the form of the actions to be taken. Additionally, the actions which the user might take after reviewing particular information units are so varied that a comprehensive predictive statement, as is required in a knowledge-base, is either not possible or is commercially not viable.

3. Using Multiple, Independent Precedents in Document Assembly

An alternative to reducing a set of documents to a template is to leave the documents as they are and simply allow the user to adapt the informa-
tion contained in them to the current situation. But the user then has to be more experienced or skilled than the typical user of the template system. It is imperative that the system should allow the person to perform the task of locating and adapting the appropriate paragraphs, clauses and sub-clauses within the precedents, in a manner that conforms to existing manual methods, whilst taking advantage of the flexibility that computer representations potentially possess over paper representations.

3.1. The problem of distinguishing text units sharing a common vocabulary

Fulfilling the above requirements is not as simple as might be thought. It is unlikely that the user will know the detailed nature or character of each of the precedents, since the user is likely to be someone other than the original author or recipient. Additionally, the number of precedents relating even to an identifiable distinct area of law is likely to be too great to make going through each a worthwhile exercise.

In order for an alternative approach to work effectively, it must be possible for the user to locate in a precise manner the precedent(s) most apt for a particular situation. Here subject labels can help, but a librarian (or information system designer) cannot anticipate many of the possible headings that lawyers are likely to search for. In any case, it has been observed in the manual document assembly task that, with time, subject labels become general and their effectiveness degrades. This is the reason why the upkeep of manual filing and abstraction systems is usually allowed to lapse.

It is also, of course, possible to search for lexical patterns in the text of documents. However, even in simple applications of information retrieval in law, searching for lexical patterns has been shown to result in extremely poor performance. In one study which investigated information retrieval for litigation support, the recall was less than 20% while the users firmly believed that they were retrieving 75% of all relevant documents [Blair & Maron 1985].

In the case of document assembly the results are likely to be far worse, since it is often necessary to distinguish between two clauses on the same broad subject but with subtly different legal effect. As the subject is the same, the vocabulary will overlap significantly. Also, the legal draughtsman conveys much meaning, over and above that borne by the major words, by varying the use of function words – prepositions, conjunctions, articles and pronouns. These are usually discarded as being ‘noise’ words when indexing is carried out to enable full text retrieval. Searching is, therefore, problematic. For example, consider the following two possible clauses in a will:
I direct that during the lifetime of my wife no steps shall be taken by my Trustees to enforce the trust for sale applicable to the freehold property known as... or to realise my said interest or share therein... and that until sale my Trustees shall permit my wife to use and enjoy the same for so long as she shall so desire or to receive income therefrom at her option...
I declare that notwithstanding her appointment as an executor and trustee of this my Will my wife shall have the option to purchase my interest or share of whatever proportion nature or amount in the freehold property known as...
The purchase price shall be a sum equal to the value at the time of my death of the property in the open market and with vacant possession...

The former is a life time interest in the property; the latter is a far less valuable option to purchase at the market price. The situations where they might be used are likely to be entirely different. The former might be applicable where the spouse is advanced in years or where the property constitutes a relatively small proportion of the entire estate; the latter where there are other beneficiaries in need of support and/or if the spouse is capable of earning a living. A simple search mechanism would not necessarily be sensitive enough to distinguish between these clauses.

One might suggest that it is possible to differentiate between textual units if the relevant aspect of the information content were to be formalised using a representation schema such as frames, semantic networks or conceptual dependencies [Mital & Johnson 1992]. However, in the present application this is neither practical nor cost-effective. Another route to retrieval by semantics rather than form or manually appended labels is the connectionist paradigm [Belew 1987; Gedeon & Mital 1991]. There are difficulties with making connection system, e.g. neural networks practical. In particular, setting up a network and tuning it is still a black art. While certainly feasible in a laboratory situation, this extension is unlikely to feature in a commercial system in the near term.

A simpler approach to constructing such an information system can be achieved by:

- Using hypertext techniques to allow the user to navigate between text units in a manner which has semantic significance, while being guided by the system's knowledge of the interaction between information units gained from active hypertext links.
- Supplementing link-based navigation by a simple search facility that:
  - has a hashed access to any word pattern in the short names of text units; and
  - does a slower search through the main body of the text.
3.2. User-reliant informational retrieval

Hypertext technology is aimed at enabling users to make their way through complex information in a manner that is conducive to the ready appreciation of the contents. It is sometimes said that it is best suited to unhurried browsing. A printed book also has the same aim, but it is constrained into largely presenting its information in a serial manner. In a romantic novel, this may not be at all bad; in a thriller, only mildly irritating.

Legal treatises and reference volumes are a different matter and it is necessary to allow the reader to view information in a number of ways and to move from one idea to an associated one quickly, even though they may be in diverse physical locations. The tables of contents, book indices, footnotes, explicit references to other sections and, perhaps, a chapter/section dependency diagram are all steps in the right direction. But the needs of the reader go beyond this. No one seriously believes that it is possible to provide or specify links corresponding to all needs of the reader, even if the book is freed of the constraints imposed by its paper medium. Units of text contain simply too much information to be able to predict all the paths that the reader (or user of a computerised system) might wish to take.

The situation in legal document assembly is not dissimilar to that in the case of reference books and treatises. Tremendous amounts of information are contained in nodes – text units such as paragraphs, clauses and sub-clauses. The only connections or links that explicitly exist between such nodes are those that determine the sequential order. If the links showing sequential order were all that were to be embodied in such a hypertext system, we would be faced with what is sometimes called rich nodes and poor links. It is the aim to capture some of the other relationships between text units that are significant to the task.

We will now look at a hypertext-based system, HyperNotary, which departs entirely from the template approach and expects the user to play the major role. The increased reliance on the user is compensated for by way of requiring relatively little effort for setting up and maintaining the computerised document base.

4. HyperNotary: An Intelligent Brief-bank

The HyperNotary system [Southam, et al. 1991], developed by Mital V. and P. Thomas as an experimental model at the Centre for Computers in Law and Finance, Brunel University, was meant for situations where the precedents (i.e. model documents or selected, previously produced docu-
ments) in the possession of a firm are used as partial guides for the drafting of a new document. This is done where it is not easy to abstract from the precedents a manageable number of generic or skeletal forms. The situations to be dealt with are so varied that assessment and legal judgement is required at every stage of the process. So that maximum use may still be made of the knowledge accumulated in the precedents, it is necessary to organise the information and make it available in a manner that lends itself to the task of constructing a new document.

The user must be able to access the appropriate precedents and particular clauses in the precedents. Often, it is not possible to judge the appropriateness of a text unit simply from its name and without reading it; yet, it is not desirable that the user has to go through large amounts of text unnecessarily. Therefore, it is necessary to provide signposts throughout the collection of precedents which give the maximum possible guidance as to the optimal paths to follow, without rigidly predetermining them.

4.1. Controlling the complexity of linkages

Once the agenda is to provide maximum access to elements of the information contained in precedents, it may be thought that the designer will strive to provide links to cover all the paths that the user may conceivably take between text units. However desirable this may seem at first, such an approach is doomed to failure when the amount of the information is fairly large\(^1\). Most hypertext systems have been modelled, whether explicitly or not, on the concept of a personal information system. The methods employed do not scale up to applications in sizeable law firms. An idiosyncratic organisation of a personal library may be quite useful. Yet, it cannot possibly be allowed where there are both multiple contributors of information and users [Larson 1989]. It is absolutely essential to control the overall number of links and also to provide some meaningful way of deciding where a link should be inserted by the author or when it should be followed by the user.

If there is a proliferation of links, the user will find himself faced at any point with a large number of paths to supposedly related information,

\(^1\) A typical collection of precedents in one related area may consist of around 100 precedents, with an average of, say, 8 A4-sized pages per precedent. This is not a trivial amount of information because each document has scores of text units (clauses and subclauses) and there is much interaction between the units, both within a document and without. Still, it may be noted that we are not dealing with browsing through vast libraries where entirely different considerations would arise.
without much guidance as to which is the preferred route. Humans find it difficult to handle more than four or five choices at a time. Faced with more than that, the user will become disoriented and find it difficult to visualise the structure of the information [Parsaye et al., 1989]. Equally, it will be difficult for the system to direct the browsing when there are so many options and little possibility of differentiating between them without the user’s own appraisal of the contents at the respective destinations.

4.2. Basic information structures in HyperNotary

The HyperNotary system was built using HyperCard on an Apple Macintosh computer. Each precedent is stored as an individual stack object. The individual cards in the stack correspond to text units, such as paragraphs, clauses and subclauses. The sequence of the cards is the same as that of corresponding paragraphs, clauses and subclauses in the relevant precedent. The central feature of the HyperNotary system is that it employs named links which have semantic significance, to the user and, to an extent, to the system. Five types of links were provided:

- Usually follows
- Alternative
- Inconsistent
- Necessary part
- Refers

4.3. Example: the usually follows link

There is a source and a target. The semantics are that, normally, the source text unit is included in a document only after the target text unit has been inserted. To take a trivial example, leaving a sum of money in trust requires that the appointment of trustees already be provided for. A clause of a gift upon coming of age or at marriage usually, but not invariably, follows a clause appointing a guardian. This significance of the presence of such a link can be made clearer to the user by providing background legal information – e.g. the case law which shows the need for the extra clauses – contained in special «footnote» fields.

Where the user has imported a text unit from a precedent into the document-under-construction, and that unit is the source of one or more ‘usually follows’ links, the system would automatically display all targets of these links, i.e. text units that usually follow the imported unit.
Fig. 1 (originally produced for Mital & Johnson, 1992) shows the user viewing a clause labelled ‘Payment of executors’ (from a precedent, a will, entitled ‘In contemplation of marriage’). Links to this clause are shown under the heading ‘Links at Payment of executors’. It is seen that the clause is connected by a *usually follows* link to another clause, ‘appoint executors, spouse’, in the precedent entitled ‘Life interest to spouse’. It is also possible to see the second-level links, i.e. links attached to the latter clause, which are displayed under the heading ‘Further Links’.

5. HYPERTEXT LINKS WITH STRONG ACTIONS

We have now seen how hypertext is made usable for document authoring through assembly. We have seen the use of semantically differentiated links, the semantics being conveyed not only to the user, but also reflecting in actions by the system. However, it will have been noted that the actions taken by the system are relatively simple.
When, in the illustration given above, a text unit was imported and that unit had some "usually follows" links to other units, the HyperNotary system automatically highlighted the linked clauses, ready for importing. In other words, the system made the decision to follow certain links, rather than others, depending on the context.

Similarly, if a text unit is imported and there is an "inconsistent" link with a previously imported text unit, the system would take action to the extent of issuing a warning; if there were any footnotes giving reasons for the inconsistency, they would be displayed.

But the above are relatively low commitment actions: for instance, in the "usually follows" illustration, the system decided to follow a link only to the extent of highlighting the link and its targets, it did not definitively redirect the user in the indicated direction.

We will see next that a rather different type of situation pertains in the case of links between "dead", evidential documents in a litigation support system. Here, the task is not creating new documents, but analysing and grouping evidential documents, which, obviously, cannot themselves be changed. We will see that there it is necessary to have links with "strong" actions attached, actions which are actually determined by the system, with no expectation that the user will either confirm or override them. We will see that such it is feasible to give links such strong actions because the links are not so much between concepts per se, but between information structures.

6. Evidence Analysis in Litigation Support Systems

If a case involves a few hundred documents and a stable litigation team which has overseen the growth of the case file, then there may be little of relevance in the documents that the lawyers will not be able to recall. If they cannot recall it, their notes will tell them exactly where to look. Most cases are like this. Many are not.

Advances in information technology mean that it is easier than ever to generate a document, copy it and transmit it to multiple recipients. While this makes communications more efficient, its by-product is a great increase in the volume of information that is generated. When corporations are involved in complex litigation or arbitration, it is quite usual to find that 50,000 documents (with an average of, say, four pages per document) bear a greater or lesser relevance to the matters in contention. Mergers and acquisitions require lawyers to handle copious information generated by
diverse arms of conglomerates. Copyright actions need the tracing of entire recorded histories of designs and developments. It is not just corporate actions that are so affected. The defence in an action involving an industrial accident, products liability or labour relations often finds itself in a similar position. To the documents which may potentially be produced in evidence must be added those generated by the process of litigation itself: pleadings, interrogatories, requests for admissions, recorded testimony, etc.

As such, the use of computers for organised storage, management and retrieval of documents for the purpose of supporting litigation is growing. A study of large law firms showed that, of the one in three lawyers found personally to be using computers, a significant proportion did so because of the value of computers in litigation support [Wallwork 1989].

There is a considerable array of Litigation Support Systems (LSS) on the market. Essentially, they provide one or both of the following facilities:

- The user may search through library-catalogue type records containing salient subject headings and identification tags; possibly, the search may be for keywords in summaries created in respect of each document by an informed person.
- The user may locate documents by specifying any combination of (non-trivial) words and lexical items which occur in the original text.

The Empower system, developed within Esprit project «Document Logistics», extends the above capabilities in two relevant ways:

- In provides ways in which a document as a whole can be linked to one or more other documents;
- A document, or a given part of a document, can be linked to a concept, which in turn may be linked to other documents.

7. Inter-document Links in Empower

In evidential analysis several useful semantics for links can be identified. The sets of links available within Empower can be subdivided into different groups by looking at the degree of objectivity involved in assessing the links presence and interpreting its significance.

One group – the contextual links – identifies objective relationships between documents within the corpus of evidence. The existence of such relationships are well understood and rarely the subject of debate. However, the import of the relationship is a matter of subjective judgement and the system may only embody a simple interpretation of the semantics.
Another group – the process links – arise out of the litigation process itself. The existence of these links is again objective, but, because they represent part of the process Empower is designed to support it is beneficial to embody a strong interpretation of their semantics.

The third group – the conceptual links – relate directly to the subject matter of the documents and their semantics are defined largely by the issues, events and persona involved in the case. Although particular links of this type may not be contentious no such general assumption may be made. The semantics of such links are of necessity opaque to the system.

The first two groups are sufficiently similar that much of the implementation structure can be common. However the conceptual structure is quite distinct and we will discuss this in a separate section of the paper.

7.1. Contextual document links

A key issue in evidential analysis is the import of statements in the document when it was written and circulated. Note that in this activity documents are real physical entities each with its own history rather than the more usual conceptual entity which may be embodied in several printed copies. Thus the context within which a particular document was seen is a significant aspect of the evidence and the litigation team will wish to ask questions like:

- «What other documents were seen with the letter?»
- «Did John Doe see the draft or a final version?»
- «Did the copy sent to Jane Doe have John Doe’s hand-written note in the margin?»

We do not necessarily answer these question directly but provide a series of relationships which can assist in such a task. The existence of each relationships is rooted in the document’s history, however, it is usually possible to make an objective assessment from a study of the document itself. The three links which provide a contextual framework for a document are:

*Attached-to*
*Cites (Refers-to)*
*Duplicates*

Each of these links has slightly different properties which are recognised and maintained by the system.

Both *Attached-to* and *Cites* have an inverse relationship – *Attachments*
and Cited-by. In both cases the inverse has distinct semantics but is completely deducible from the primary relationship. We can also recognise that a document may only be Attached-to at most one other document but the other three relationships can link one document to many others. Our implementation presents these links as a series of global options which, on selection, retrieves the list of documents which stand in the selected relationship to the current item of interest.

The Duplicates link has no inverse as one document is simply a member of a set of duplicates. We specifically avoid giving the document which has been duplicated a special status for two reasons:

- We may not have that document within the corpus of evidential data;
- Which document «came first» involves a far more subjective judgement.

We also note that each of the three links yields a different global graph structure. Both Attached-to and Duplicates must partition the document space. Attached-to must yield a single tree structure in each partition but by implication each Duplicates partition is fully connected. In contrast Cites is simply a directed graph which may encompass the whole document space.

In summary the semantics related to document context links is used to control the generation of lists of related documents and the integrity of the database when the user declares that document X is Attached-to, Cites, or Duplicates document Y.

7.2. Document process links

The litigation process itself adds further structure to the context within which documents are seen. As one party to the process John Doe's representatives may regard his company ledger as a single document, however, another party might single out an account within the ledger and treat it as a document in its own right. Our respective pleadings will cite different documents yet we need to know that one is a subset of the other.

Another part of the process which needs to be recognised is the disclosure of evidence to other parties. In the course of responding to a disclosure request we may be forced to reveal part of a document while wishing to keep other statements within it confidential. For example in a patent action we would be required to disclose design documents dealing the component under dispute, but we should not reveal other aspects of the design to a competitor. The equivalent manual activity would be to
photocopy only some pages of the document possibly obscuring paragraphs or diagrams in the process.

Empower has two further inter-document links covering these issues:

*Sub-documents
Versions-of*

The semantics of both of these relationships have strong implications for the system, unlike contextual links, which are declared by the user, the existence of these relationships is deduced by the system. In each case they are alternative references to the same evidential material and are thus quite distinct from duplicates, which represent copies occurring within the body of evidence.

*Sub-documents* are identified when a document is introduced to the corpus of evidential material to be accessed by Empower. Access to the original form of documents is provided through the use of scanned images and machine readable versions as ASCII text. To achieve this the system needs each page to be clearly identified. The preferred method is through the use of a device such as Bates numbers which give each page a unique user identifiable reference number. When a document is introduced, or revised, to have some page identifiers in common with another known document Empower generates the appropriate *Sub-document* links, i.e. the documents are defined to relate to the same evidential material.

The creation of versions arises at a different stage in the process when using Empower to print documents and document bundles. The system allows the user to print single documents or compile bundles of documents to be sent to other parties which include both information about documents along with the text or images stored within the system. In the course of this activity the user may decide to only reveal part of an evidential document, thereby, creating a *Version*.

A significant amount of information about a *Sub-document* or *Version* must be inherited from the original document and may not be overridden by the user. For example, a *Sub-document* must be filed in the same place as its parent – it would have to be a *Duplicate* of a *Sub-document* to have a separate physical existence. However, we do allow *Sub-documents* to have such things as distinct authors because this may provide a useful mechanism to recognise such things as individual input to a corporate report.

Much tighter constraints are placed upon *Versions*. Since they are essentially an output filtering mechanism they inherit most properties of the original document. They also represent a record of action taken to reveal
information to other parties and once formal printing has occurred further editing of the pages is inhibited.

7.3. Presentation of inter-document links

Each type of inter-document link can describe one or more documents which stand in a particular relationship to the current document. Rather than provide visual clues or lists of these relationships within the presentation of the document we adopt an overall screen design which presents information along side the text of the document.

Fig. 2 shows a sketch of the screen layout.

![Fig. 2. Overall Arrangement of the Empower Screen](image)

The image (or text) of the selected document is displayed on the right of the screen with its associated bibliographic data in the form of a catalogue card on the top left. When the user selects a link type (on the speed bar or the Empower menu) the system displays a record card showing a list of the documents holding the desired relationship to the current document.

Fig. 3 shows an example of an attachments list record card. Any of these documents can be selected for study by using the mouse to point and

2 The system uses a large high definition screen to give good presentation of scanned images which does not reproduce well when reduced to an appropriate size for presentation within a document such as this.
click the desired entry causing Empower to display the relevant image and catalogue card.

Note the buttons at the foot of the card (in this case Attach and Detach) provide the mechanism for appropriately authorised users to declare and modify contextual links.

8. Links Between Concepts and Documents

Although the inter-document links provide some help with understanding of the evidence their semantics are not directed to the content of the investigation. We need to include something of the relevant conceptual framework within the system.

To achieve this we add a second set of records to the Empower catalogue each one identifying a concept relevant within the current case. Concepts are clustered in groups, such as Issues, Events, and Persona, which may be defined for each case or corpus of evidence. Within each group a concept is given a reference code and expanded definition as shown by the example concept record card in Fig. 4. For each concept type the
used may define a more complex (possibly multi-field) framework for the description.

The concept records provide both an alternative entry point and a means of linking documents which deal with the same fact or phenomenon or address the same issue. Wherever a document, or section of a document, deals with an issue the user creates an annotation record linking the document to the relevant concept.

Fig. 5 shows an annotation identifying the concept of a driver having been careless to some comments appearing in a letter. In addition to the references to the document and the concept other structured data such as a location within the document, a classification of the annotation (in this case probative value) and importance appear on the record card.

Although the concept links are the most powerful tool for the litigation team to express the significance of the evidential material they are the weakest in term of semantics which can be interpreted by the software. The only constraint on the graph generated by the annotation links is that one end of each link must be a concept and the other a document, it is even possible to have several links between the same document-concept pair because the concept may be expressed in several ways or places.
Some simple processing of these links is provided. To move between concepts and documents we again use intermediate lists of the conceptual links (or annotations) so that, for example, starting from a document one can ask for a list of relevant concepts (one can also list of all documents linked to a single concept). Since we anticipate some of these lists being lengthy facilities are needed to control the order or filter the items in the list, however, the system may only perceive a weak common semantics for all annotations and these processes can only work with the simple concept categories and qualifier labels declared by the user.

9. Issues for Discussion

The design and functionality of a hypertext system for lawyers must be placed in perspective of wider issues in hypertext. There are several matters of interest, of which the following points deserve more detailed consideration:

(a) When the information base consists of multiple documents, the level at which they are integrated.

(b) Minimising the disorientation during navigation through a large, complex information base.
c) When the information has to be available to multiple users, the degree of customisation of the structure which is to be allowed.
(d) The use that the system itself makes of the semantics of the links.

9.1. *Multiple document hypertext*

Many applications have focused on applications where the information they contain corresponds to a single document. These include conversion of reference works [Raymond & Tompa 1988] and manuals [Frisse 1988; Walker 1988] into hypertext format, as well as the presentation of new text by means of hypertext. Perhaps less common are applications which can be identified as *combining* information from several sources, such as the integration of documentation arising from the software development process [Bieg low 1988].

However, a common assumption amongst the hypertext community is that the ability to integrate documents is one of the major benefits which hypertext technology makes possible – such an assumption has been expressed by McLaren (1990). In the context of this belief, the dearth of applications which could be termed multi-document hypertext appears somewhat surprising. Whatever the reason for this, the fact remains that the issues which have to be considered when designing a multi-document hypertext have not fully been addressed. Glushko (1989) identifies four such issues, which relate to the suitability and usability of a multi-document hypertext. We will discuss each of these in the context of the HyperNotary application.

The first question is *which documents should be included* in the application? There would seem to be little justification in trying to integrate a telephone directory with the Encyclopaedia Britannica. This is not because it is not physically possible, it is just that the semantics of such an cluster would be suspect. We have to confront this issue constantly. Do we integrate all the wills with all precedents of commercial leases? And what about different kinds of wills? The answer to this generally follows from an analysis of the business process [Glushko 1989]. In legal document assembly, there is some customary aggregation – commercial leases relating to developments are considered together. Other than that, the degree of cross-referencing and comparison of precedents which would be required is our guiding heuristic.

The next question is *the extent of integration*. In other words, how tightly should the documents be linked. At the highest extent of integration, it would not be possible to identify the boundaries between documents. In our applications, it is not permissible to let the identity of separate prece-
dents or evidential documents disappear. This is because the legal significance of a clause or statement can only be fully appreciated in the precise original context. We keep each precedent as a separate entity—a HyperCard stack. Therefore, no matter how richly connected the information may become, it is still possible for the user to ignore the cross-linkages, to begin at the top of a stack and step sequentially through the clauses. The page reference codes used by Empower provide a similar standard of document integrity for the evidential data.

The third issue is the locus of integration—i.e. where within the integrated body of documents can the user enter and what further connections in the mass of information are available for traversal. There are several options open to a system designer. The simplest would be indexing the documents through a bibliography, as in the well known legal database LEXIS. Alternatively, the contents tables could be amalgamated. However, unless there were cross references between lower level text units, the user would be led into much serial browsing [Glushko 1989]. Consequently, both inter-document and intra-document links are provided in HyperNotary and Empower, though Empower does focus more on the former type of links. In addition, a number of entry points are available: e.g. direct access via free text search to one of the documents in a linked chain.

Lastly, the question is: what types of links are provided? This has already been dealt with extensively above. It is sufficient to note that we provide public links of types which minimise the risk of inconsistent insertion by different authors. This is because (a) the links have existing normative significance in the domain, and (b) the persons who insert them are domain experts, able to articulate these norms [Mital & Johnson 1991].

9.2. Minimising disorientation within hyperspace

The general problem of disorientation during navigation in hypertext is widely acknowledged and has been extensively commented upon [Conklin 1987]. A number of partial solutions have been proposed:

1. The user is alerted as to what might be found when arriving at the end-point of a link. Landow (1987) believes this to be a crucial aspect of hypertext research and has suggested the formulation of a «rhetoric» of hypertext which can be used to give impart knowledge about the structure of the information base. Currently, some systems follow this advice by making the significance of links apparent to the user by means such as icons and/or short description of the destination.
(2) A history mechanism is provided which displays trails of the users' movement in the information space [Nielsen 1990]. It has been suggested that such a record best makes sense when the goal or the purpose for which the user visited a particular text unit is annotated (in a volatile or non-persistent manner).

(3) Means are provided by which the user can recover to one of a number of known and familiar reference point in the information space. In a hierarchical arrangement, this might be the root or the apex of a branch.

(4) A very powerful aid to appreciation of the structure of the information is a graphical browser, such as is provided in the NoteCards system [Halasz 1988]. There is no one universally applicable type of browser and the manner of presentation is determined by the application.

Both of our applications use the first of these options by employing the semantics associated with links to generate lists showing the target nodes which can be visited. Indeed the browsing pattern in Empower will nearly always be alternately visiting documents and lists. Empower also uses the Windows, MDI architecture to give a brief analogy to the history mechanism. Rather than replacing the current node when a link is traversed the new record card is added to the ones on display, thus backtracking is achieved by simply a matter of pointing to the previous window and bringing it to the top of the stack.

The idea of having goals or purpose annotating the record of browsing path is interesting, though in our opinion, not practical for a generalised hypertext system. It is virtually impossible for the system to, by itself, gather the goal or purpose in constructing or traversing a link with any degree of certainty. The user could be asked to add some form of annotation but is likely to be resistant to the idea. In any case, a short statement of purpose might become meaningless after the user has lost track of an idea and has moved on to another. Consequently, what is needed is a history mechanism that is much leaner.

With an application specific hypertext, such as Empower or HyperNotary, it is possible to make crude inferences about the purpose of traversing a link and attach weights to the nodes visited. In litigation support the focus on documentary evidence gives the document catalogue cards a special significance and provides the user with a meaningful recovery point. The information carried by Empower can be perceived as clusters of annotations and lists «surrounding» a document (or concept) to give a two level structure to the graph (see Fig. 6). This means a more focused history mechanism
could be produced by just tracing the document, and possibly concept, references during a work session.

**Fig. 6. Empower Two Level Graph**

It helps when the hypertext model has a close and natural correspondence with the paper world of the lawyer as well as the structure of the document assembly or evidential analysis task as it would be manually carried out. The semantically differentiated links need to bear names and actions assignable in the paper world. If so, users can appreciate the significance of a link, not merely from the label, but also whence it originates and what it does. However, we do acknowledge that some disorientation is inevitable. The underlying problem is that information about a destination can rarely be fully conveyed – not just in hypertext but in other walks of life such as when directions are given to a motorist [Brown 1987].

9.3. Customisation in a multi-user environment

Bush (1945) envisaged the notional precursor of hypertext to be a library of information personal to an individual. Indeed, a large proportion of hypertext systems are treated as personal information managers by a single user. In such a situation, *ad hoc* linking of information units is not detrimental. However, this cannot be allowed when a number of persons
would access the information. As is the case with a second-hand text book, the markings by the previous owner can come in very handy; on the other hand, they can be most irksome. The matter of private versus public links is a live issue in hypertext. It has been addressed in context of systems to manage information related to the software engineering process.

One way in which software engineering systems allow customisation of information is by allowing annotations to be attached to information units such as pieces of code, design diagrams, and so on. These annotations may or may not be made available to others for reading and/or alteration. The multi-user hypertext tool, KMS, allows a frame-level security system to individual users [Akscyn et al. 1988]. However, it is rare for the structure of the information to be changed by individuals except in accordance with the strictest of guidelines set for a project.

In the legal document assembly application, we do not have the possibility of centralised direction and supervision of restructuring of information, yet we do not wish to curtail the users’ freedom unduly. A compromise solution is, therefore, provided by means of private and public links. The publicly available structure of the information – as constituted by the semantically differentiated links – cannot be changed. However, a user is permitted to add labelled links between any two points. These labels are not interpreted by the system. Also, the links of one user are not shown in the view of the information presented to another user. A similar dichotomy is present in Empower but in this case the conceptual links are visible to all users unless they select the author’s name as a filtering criteria for listing annotations.

9.4. Active links

Earlier hypertext systems did not entertain the idea of links which are semantically differentiated by the system – as opposed to the user alone. More recently, knowledge-based techniques have been used in the context of hypertext [Diaper & Rada 1989; Parsaye et al. 1989]. The aim is to provide the system with the ability to interpret the link in the context of the facts made known to it by the user, make suitable inferences and use that to drive the direction of browsing or provide guidance on what new links a user can add. As explained earlier in the context of knowledge-based template systems, strong predictions have to be made about the import of and relationships between text units. The two hypertext systems described above make very weak assumptions about the interaction between
text units. We believe that this is entirely appropriate in a fuzzy domain such as law. However, as has been seen in the case of Empower, when the linked units are relatively large and standard information structures, certain strong actions can follow the weak assumptions.

References


