Artificial Intelligence Applications Institute

The Application of Knowledge Based Techniques to Support Authors and Readers of Technical Documents

Andrew Casson

AIAI-TR-160

December 1994

This is a revised and expanded version of a paper published in *Proceedings of The Emerging Technology Showcase '95* (CCTA ET95) London School of Economics, January 1995.

Artificial Intelligence Applications Institute
The University of Edinburgh
80 South Bridge
Edinburgh EH1 1HN
Scotland

Abstract

The production and use of technical and legal documents is an important aspect of the day-to-day function of organisations, both in commerce and government. However, the size, content, structure and interactions of these documents can render them very complex, and as a result the processes of creating and using them are difficult and time-consuming.

This paper proposes a knowledge-based approach to providing intelligent support for the readers and authors of technical and legal documents. This approach, which is based on *rich hypertext* representations of documents with intelligent *navigation* facilities, concentrates on ensuring that documents are *used* correctly and efficiently, and leaves the *interpretation* of their contents to the reader.

Having outlined the technology, the paper goes on to discuss its applications. In particular, a case study involving building regulations is described. The application of intelligent knowledge-based hypertext to this domain has been successfully demonstrated using the PLINTH system designed and developed by AIAI and the Scottish Office. Other applications discussed here include *design rationale* and statute law.

Finally, the paper concludes with a summary and a discussion of why a hypertext-based approach is particularly appropriate at this time, given the recent massive expansion of the global hypertext network, the World Wide Web, and the exciting new opportunities for on-line publishing that it heralds.

Contents

The Application of Knowledge Based Techniques to Support Authors and Readers of Technical Documents

Andrew Casson
Artificial Intelligence Applications Institute
December 1994

1. Background

The production and use of technical and legal documents is an all-pervasive and vital component of the functioning of organisations both great and small. From Acts of Parliament, through regulations and standards, procedure and operations guidelines, and contracts, down to the manuals and on-line documentation necessary for the effective use and maintenance of office equipment, documents affect almost every aspect of the way organisations work. This is particularly the case in Government departments and offices. For example, it is estimated that up to 70% of Scottish Office work is involved with policy development and administration. A large number and variety of regulations and guidelines are produced by the policy makers, and must be used by those whose work the policies are intended to control and support.

These documents are often very complex as a result of their size, content, structure and interactions with each other:

- **size:** documents may have thousands of pages, and are often part of multi-volume publications.

- **content:** documents may include diagrams, tables and formulae, as well as text which may range from background discussion to precise instructions.

- **structure:** the way a document *works* may be quite complex, so that using it correctly and effectively requires a detailed knowledge of its internal structure, as well as an understanding of its contents.

- **interactions:** technical and legal documents do not exist in isolation, *e.g.*

    - the *Technical Standards* supporting the *Building Standards (Scotland) Regulations* refer to and therefore rely on the existence of over 160 British Standards; furthermore, because of the UK's membership of the European Community, each of these references must be construed as a reference to any equivalent standard of another Member State;
    - a legal document may directly modify the meaning of the whole or part of another, *e.g.* 'Regulation *x* shall be subject to the provisions made under section *y* of Act *z*' or 'The regulations in *x* shall apply, with *y* being construed as including *z*';

- documents may be subject to *harmonization* processes, *e.g.* to bring the regulations of different EC Member States closer together.

As a result of this complexity, producing, maintaining and using the documents may be difficult and time-consuming, even for those with long experience of doing so. Computer-based techniques can offer much-needed support for these activities. For example, 'expert systems' which represent the knowledge in documents (rather than the documents themselves) as logical facts and rules, and allow interactive consultation by the user, are well-established tools. Indeed, they can work very well for inherently logical documents, such as highly prescriptive standards. However, they have been found to be limited in expressiveness, *e.g.* in dealing with background texts and diagrams, and inflexible, *e.g.* in predetermining a single, rigid interpretation of legal texts, which is often inappropriate. Expert systems have also had the problem that, unlike the documents on which they are based, they effectively exist in isolation, *i.e.* they are closed, stand-alone systems. Consequently they only really suitable for handling relatively self-contained documents.

A more effective approach is to use knowledge-based techniques to provide flexible and structured access to the documents themselves, stored on-line as *rich hypertext*. The system's intelligence is then focused on ensuring the effective and efficient *use* of the documents, rather than the *interpretation* of their contents which is far better left to the user. Such systems can also be used to store a representation of the background research and rationale underlying documents, and help their authors to draft and revise them, drawing on this knowledge in a principled, logical fashion. Their open architecture also makes them much more suitable for coping with the size and complexity of the technical and legal document 'universe'.

The remainder of this report provides an overview of current knowledge-based techniques for providing intelligent access to and creation of on-line documents, tailored to the individual needs of the reader and author. We then discuss some specific potential applications, and end with a summary of the technology and some opinions on its maturity and the opportunities for exploitation. For a more detailed study of the techniques and applications, and references to related work, please refer to the accompanying *Technical Report.*

2. Knowledge based documentation systems

Conventional documents, whose structure is characterised by rigid hierarchy of sections in a fixed order, do not reflect or support the flexibility with which they are used in practice. As well as starting at the beginning and proceeding directly to the end, the reader may browse freely backwards and forwards looking for interesting sections, find specific information by looking in the index or the contents page, follow a cross-reference, and so on. *Hypertext* systems were designed to support this kind of document use, by simply representing the text of a document on the computer as a network of inter-linked sections, which the user can move around freely by activating the links in any order.

Hypertext offers a very natural and powerful interface to on-line documents, but experience has shown that ease of access does not necessarily equate to effective use. In large hypertext documents the lack of any obvious structure to constrain the reader's choice can lead to problems of disorientation and digression, *i.e.* readers lose track of their path through the document, and either don't know what to read next or get side-tracked into reading irrelevant material.

The knowledge-based solution is to supplement the textual contents of each section with high-level information representing, in a form the system itself can use, what the section is about and how it relates to the other sections in the document. This gives us a *rich hypertext* document. For example, in regulations and standards we typically find two types of clause (along with many others): *requirement* clauses which state the conditions that must be met and how to meet them, and *scope* clauses that specify when the *requirements* apply. A rich hypertext system can represent the type of each document (*e.g.* **standards**), and of each clause within it, and explicitly mark the internal relations between clauses (*e.g. scope S* specifies the *application* of *requirement R*), or the external interactions between different documents (as described above). Once this has been done, it is possible to use a *navigation* system that has been programmed with rules such as:

- check that the *scope* of a *requirement* is satisfied before checking the details of the requirement itself
- where a *scope* clause has several *sufficient* condition clauses for *exemption*, and one is satisfied, ignore the rest

to guide the reader around the document or a set of documents in an intelligent and structured manner, showing all relevant sections in a logical order. The system does not *interpret* the content of individual sections, but it does ensure that the documents are used properly. Furthermore, a single set of rules (which may well be only a few dozen), will support navigation of any document with the appropriate structure, so that if a new one is written or an existing one is revised no reprogramming is needed. In this respect the advantage over conventional expert systems can be huge.

Rich hypertext systems can also *classify* the subject and context of each document section, *e.g.* as being about *stairs* in *hospitals*, *lighting* in *theatres*, and so on. Using a fixed set of classifiers supports retrieval and indexing mechanisms that are much more reliable than free text searching, since the reader doesn't have to guess words that might be in target sections. The classification scheme may also be used by the navigation system to filter out irrelevant information.

3. Applications of the technology

The rich hypertext approach can be applied to almost any type of document in which there is a sufficient amount of identifiable structure to make intelligent navigation feasible, and which is complex enough to make it worthwhile. But whether it is on the scale of a computer manual or the whole of a country's statute law, the inherently scalable nature of hypertext makes the principles equally applicable.

3.1 Case study: Building regulations

We used the example of regulations and standards to illustrate the features of rich hypertext systems in the previous section, and mentioned the *Technical Standards for compliance with the Building Standards (Scotland) Regulations* in the Background. In fact, a prototype general purpose rich hypertext system, PLINTH, has been developed by AIAI in collaboration with the Building Directorate of the Scottish Office, with this document as its initial application.

Prior to 1990, the Building Directorate did a number of in-house experiments with computer-based authoring and information systems to support their work on standards. Two prototype systems resulted from this:

- an authoring system, providing facilities for input of text, tables and formulae, classification and indexing of sections and clauses, logical validation (*e.g.* checking completeness and consistency) and automatic generation of production rules for use in the second system;
- an end-user system, *i.e.* an interactive expert system for consulting the Standards, with dialogue, explanation and what-if queries.

Although the work demonstrated that computer-based support for standards processing was a useful pursuit, it ran into two main limitations. Firstly, as a result of the move from highly prescriptive standards to more 'functionally' based ones, much of the information in the Scottish building standards was at too high a level, too general, or too discursive to be readily and usefully converted into the formal representations of expert systems. For example, many standards now use phrases like *adequate provision*, *suitable devices*, *safely accommodate any likely maximum demand*, and so on. Although these may be backed up by more detailed and prescriptive 'provisions', they are still an essential part of the document and cannot be ignored in an expert system just because it cannot do anything useful with them. Secondly, the expert systems approach did not provide any support for representing the *design rationale*, i.e. the research, discussion and consultation on which the drafting and revision of standards is founded, and which can be very useful in actually interpreting the finished documents correctly.

In 1990 the Building Directorate began a four-year project in collaboration with UK university departments, including AIAI at Edinburgh, with the goal of developing a new approach to standards processing. The project set about rationalising, extending and developing the newly-emerging ideas on knowledge-based hypertext, and designed and implemented a system called PLINTH. The system will support all stages of the production, maintenance and use of the *Technical Standards*, from collaborative design rationale (see 3.2.1 below) to interactive consultation by end users. In addition, it has been designed as a general purpose shell that can handle, within a single integrated environment, the many other kinds of documents involved in the Building Directorate's work. PLINTH is currently being evaluated with a view to identifying requirements and strategies for future development.

3.2 Potential applications

3.2.1 Design rationale

A substantial legal or technical document such as a set of regulations represents only the final output of a lengthy process involving research, design, consultation and revision. The published text contains just a fraction of all the knowledge and data involved in this *design rationale* (DR) process. Consequently, a computer system intended to support the authors as well as the readers of legal documents would do well to support DR during the authoring process and provide structured access to on-line DR archives for reference when documents are revised later.

A rich hypertext system can do this easily. For example, the DR model IBIS categorises each bit of information used in the design process as one of the following:

- **issue:** a problem to be addressed or question to be answered;
- **position:** a potential solution or answer which *responds to* an issue;
- **argument:** evidence which *supports, objects to* or *questions* a position.

(where the terms in italics represent precise predefined relations between the bits of information). In addition, arguments and positions can *suggest* new issues, *contradict* each other, and so on. Given that each issue, position, or argument can (and probably will) be written down in some form, we can represent IBIS design rationale as a rich hypertext network, where each section is marked with its type and its relations to other sections. Each issue can then be linked directly to the document section it underlies. This could be used in two ways:

- after the issue is resolved and closed, to summarise the discussion that took place and archive the relevant documents in a structured form; or

- interactively, for on-line collaborative discussion, where new issues, positions and arguments are inserted directly into an active network by the various people involved in the design.

In the latter case the process of building up a DR network could be further aided by using navigation rules which not only traverse existing parts of the network but also allow the author to extend it in a principled manner, *e.g.* by ensuring that when a position is created it must be linked to an existing issue.

3.2.2 The Statute Law Database

The Statute Law Database Project, based at the Lord Chancellor's Department, aims to provide an on-line database of the whole of UK legislation, the main user and administrator being the Statutory Publications Office, who are obliged make statute law documents readily available to users in Government, Parliament and the public at large. Stated objectives of the project include:

- to provide facilities for enabling editorial references and comments to be appended to the database;
- to provide facilities for maintaining the database, including the addition of new legislation;
- to provide on-line enquiry facilities to end-users via their own PCs and printers.

The initial analysis of the system requirements identified the following features for the database:

- data with relational attributes;
- text retrieval on large data files of structured and unstructured text;

Although the project has already decided to use existing RDBMS and text retrieval software packages, these stated objectives and features indicate that a knowledge-based, rich hypertext approach would be appropriate and feasible, especially given relational attributes for representing the necessary supplementary information about the meaning and function of texts and the interactions between them. The sheer size of the database would seriously test the technology, but it has the flexibility and scalability required.

3.2.3 Other applications

Other potential applications of knowledge-based documentation systems include:

- procurement procedure manuals;
- authoring systems for contracts which select relevant fixed clauses and order them appropriately;
- instruction manuals and trouble-shooting guides, which lead the reader through a step-by-step process where each action may depend on the outcome of previous actions;
- computer-aided learning (CAL) material, where the information to be conveyed is predetermined, but the system has to present it in a way that responds intelligently to the reader's understanding and progress (this would be an ideal application for an adaptive expertext system);
- software engineering systems, in which code, documentation, development standards and process models can be fully integrated.

4. Summary and conclusions

Knowledge-based documentation systems, which use a rich hypertext model, provide a very flexible approach to providing computer-based support for authors and readers of technical and legal documents. The techniques used scale very well, being applicable to both very small and very large applications. Though the technology is relatively new, it is now mature and ready for exploitation. Its foundations on the well-established technologies of rule-based systems and hypertext, mean that take-up and development has every chance of proceeding quickly and painlessly.

The recent massive expansion of the World Wide Web, a global hypertext network which has the whole Internet as its repository and millions of users, is leading the way at what is probably the beginning of a fundamental transformation in the way information is produced, distributed, and used. Rather than producing thousands or even millions of copies of a publication, the technology is emerging which will support the ability to store just a single on-line copy, which can be retrieved and viewed on demand over international computer networks. The deployment of rich hypertext systems in business and government should therefore take into account not only the immediate improvements to existing technology as described in this report, but also the opportunities for beginning now to put in place the new information systems infrastructure that will be vital in the near future.