# Customizing the presentation of legal documents over the world wide web

## C.A. Royles & T.J.M. Bench-Capon

# Customizing the Presentation of Legal Documents over the World Wide Web

C.A. ROYLES & T.J.M. BENCH-CAPON

*LIAL—Legal Informaties at Liverpool, Department of Computer Science, The University of Liverpool, Liverpool L69 7ZF, UK. E-mail: Eroyles@csc.liv.ac.uk or T.J.M.Bench-Capon@csc.liv.ac.uk*

ABSTRACT *Large legal texts are often difficult to understand, and to extract the required information from. One particular problem is that the user typically requires only a specific, limited part of the content. In this paper we discuss how such documents can be customized on the basis of user supplied information. Customizations include the suppression of text elements, the organization of text, highlighting certain text elements, and the provision of individual links to other parts of the text, or to external texts. The result is a system which starts from a single complete representation and generates a representation appropriate to the needs of the specific user on the fly. In this paper we describe both an initial prototype, and a design for a more refined system which incorporates lessons learnt from the prototype.*

## 1. Introduction

For much of law to be effective it is important that those it concerns have a good understanding of the law as it applies to them. The Social Security system, for example, requires potential beneficiaries to be aware of their rights, so that they can receive what they are entitled to, and their obligations, so that they do not unwittingly defraud the system. Tax law provides another—currently topical—example of an area of law where it is important that the public are aware of their rights and duties if the system is to work effectively. Government Departments responsible for such law put considerable effort into attempting to raise public awareness. Currently the major medium for this is a collection of advice leaflets, some general, others specific to particular benefits or circumstances, although at particular times specific advertising will be produced. As well as such official sources, claimant support groups, such as Citizens' Advice Bureaux and the Child Poverty Action Group produce their own versions of such information. Similarly for tax, the Inland Revenue produce explanatory material, and many accountancy firms also produce their own advice leaflets. The need for advice is further evidenced by the existence of numerous 'phone-in' advice slots on radio.

There are a number of deficiencies with information leaflets, (see, for example Soper & Bench-Capon, 1994) and it is always worth considering whether new technology can offer a better solution. The Department of Social Security has always been active in investigating possibilities as they arise. Attempts were made as far back as Prestel, and now there is an extensive Web site at http://www.dss.gov.uk. This is actually a quite well produced and informative site, but is essentially just a transfer of advice leaflets into HTML. There are links

from references to other 'leaflets' in one 'leaflet' to pages containing these other leaflets, but otherwise the pages retain most of the faults of the paper versions. Some of the problems associated with paper advice leaflets, and web sites which are conceived of as no more than a window onto such leaflets, are discussed in section 2. In section 3 we discuss the possibilities that arise if we take the WWW seriously as new medium rather than simply as a means of delivering information already available in other forms. In section 4 we describe an initial prototype implementation. In section 5 we discuss the issues that arose from the prototype, and in section 6 we describe a more refined system that addresses these issues. Section 7 is a conclusion.

## 2. Problems with existing advice leaflets

Most of the problems with advice leaflets arise from the fact that such leaflets are necessarily few, whereas their readers are many. Every reader will have particular circumstances which determine the particular information that is required. As a result such leaflets attempt to please all their potential readers, and too often end up pleasing none of them. Some problems are addressed immediately by transferring the leaflet to the WWW:

*Distribution*

Currently leaflets are obtained by going to a Post Office or a Benefit Office. This creates problems of the availability of the most suitable leaflet, since stocks of all leaflets are difficult to maintain. This problem can be solved if we make the leaflets available over the WWW—assuming that interested parties have access to the WWW. But with ever increasing initiatives to make such facilities available in libraries, schools and the like, it could be expected that this will be a very accessible medium in the close future.

*Up-dating*

Laws change rather frequently, and with paper leaflets there is always the danger that out of date leaflets will be distributed, and there will be a significant lag in issuing an up-dated version. This problem is solved by WWW distribution: as soon as the version is up-dated, that version, and that version alone, is available.

Other problems, however, are intrinsic to the paper-like presentation, and remain even when the leaflet is distributed electronically. Appendix I gives the current page relating to housing benefit, and we will discuss the problems with reference to this page.

*Problems of details*

There is a trade off between the amount of detail that can be included in a leaflet and the length of that leaflet. In general it is important the leaflet does not become too big, as unnecessary detail obfuscates. The problem is that what is unnecessary detail to one section of the intended readership may prove necessary to others. For example, if a provision applies to EU countries, many people may be aware whether the country they have in mind is, or is not, a member of

the EU, and would find a list of EU countries unhelpful. Others, however, would be lost without such a list. The page we are considering offers an example with its frequent references to *Income Support/Jobseeker's Allowance (Income based)*. These could be replaced by a phrase like a qualifying benefit, since many of the readers will be aware of what these are. But clearly the writers of the leaflet thought it sufficiently important to make this clear that it is repeatedly spelt out in full, with consequent loss of readability.

*Thesaurus problems*

Different people will often associate different words with some of the key concepts to be found in a leaflet. For example the page in Appendix I speaks of *income related, dwelling, and habitually resident*: most customers would more naturally use *means tested, building* and *normally living in the UK*, and might well welcome some assistance in understanding the chosen terms. Of course, when writing the document one term or another has to be chosen, and used consistently, but very often any given choice will make difficulties for some readers.

*Overlap problems*

Breaking the subject down into discrete, self-contained leaflets is rather difficult. Several leaflets may need to refer to the same information. Duplicating the information in each leaflet makes the leaflets unwieldy, but a simple reference means that extra leaflets need to be obtained. Presentation through the WWW solves this problem to a certain extent: the example page contains links to other pages. However, when these links are followed, there is little guidance as to which section of the new page is relevant to the reason why the link was followed.

*Problems of personal circumstances*

Leaflets are typically written to cover large groups of people, and therefore contain a lot of conditional statements and sections that are intended to apply only to part of their readership. Our example in Appendix I applies to those renting from private landlords and local authorities, those in work and those on a qualifying benefit. As well as increasing the length of the material, wherever not relevant information is present, there is an increased possibility of mistakes being made.

*Use of the information*

Once readers have obtained the information they need, they still have to apply it successfully. Where—as is often the case—the rules are complicated, this process may itself be fraught with difficulty. Where the leaflet contains a form to be completed by the claimant, there may be problems in relating particular parts of the form to particular sections of the leaflet text. A static leaflet can, of course, provide no help in applying the information it contains.

## 3. Possibilities for presentation over the internet

At the root of the problems described in section 2 is the need for the author of a static document to make choices. These choices will be made in the light of expectations about the readers, but will, since targeted at a variety of different readership, involve compromises which make the document useful to more readers, but at the cost of making it less useful to any particular reader than it might be if targeted more precisely. The static document is like a blunderbuss: it can hit a lot of people, but its impact is often only superficial.

The choices are necessary if the document is to be typeset and printed. But an electronic document, such as can be delivered over the WWW, is not cast in stone in this way. What we wish to do is to involve the readers of the document in the process of choice and selection, so that their individual preferences can be accommodated, without compromising the general nature of the document. Thus we see the electronic document as a means of generating a variety of instantiations, which are what the readers will see, the particular instantiation shown being determined by information supplied by the reader.

Given a generic document there are a number of operations that can be performed on that document to tailor it to a particular user:

- Text can be *suppressed*: ideally readers should see only those parts of the document that are relevant to their particular circumstances, and their particular information needs at the time they access the document;
- Text can be *re-organised*: Different readers will start with different background knowledge, and may find different orderings of the material presented helpful:
- Text can be *highlighted*: Devices such as bold font, smaller or larger font sizes, use of white space and the like can greatly help readers to understand a document by drawing their attention to the critical parts of a document: what needs to be emphasised can again vary:
- Text can be *included*: Sometimes it may be helpful for text from another document to be imported into a document to supply critical background information, or to expand certain cryptic references;
- Text can contain *links*: If we don't want to overload the readers with background information in the document itself, it may still be helpful to indicate that background material is available, and to give easy access to it. What background material is best made available will again depend on the particular reader;
- Text can be *substituted*: Different readers may have preferences for different vocabulary and phraseology. Readers understanding will be enhanced if the vocabulary and phraseology is appropriate to them.

We can, therefore, see that given an initial starting document we can apply some or all of the above operations to transform the document into a version. The purpose of our system is to represent the document in such a way that we can, on the basis of information supplied by the user, select and apply a set of these transformation to produce a document tailored to the user's needs.

## 4. An initial experiment

In order to get a feel for what was involved in producing such a system we first implemented a small prototype. Our choice of source was an advice leaflet, *A*

*Young People's Guide to Social Security*, targeted at school leavers and students. It is a public information booklet published by the benefits agency (FB23 from October 1996). The leaflet has several interesting properties with respect to our goals: in particular there is a lot of information in the leaflet but only small amounts are usually relevant to any given individual. There is a distinction between the employed, the unemployed and students. There are particular sections for married and single people. Taking each of these sections and reading through the information, it is clear that large volumes of the information could be masked from particular users. This approach would provide readers with a quick and effective method of finding information, and avoid confusion caused by burying the relevant in the irrelevant.

The separate sections of the booklet were converted into individual HTML documents. These documents were then organized by using an index of topics. This is how most information of this type is made available by electronic means. By then examining each section, unique characteristics of the paragraphs were identified. A particular page may contain several paragraphs but only one or more may be specific to a given reader. By asking the user a set of simple questions, the relevance of the various paragraphs can be identified.

The resulting system is a set of documents that can be viewed with no modifications. Users are directed to different areas of the documents by a topic list, certain sections of which are indicated as relevant to them by a graphical symbol. Users can also run a questionnaire that alters the graphical symbols to point to information more relevant to them depending on their answers. This idea is extended further by altering the enhancements to the textual content. Different sentences can be emphasised to direct the attention of the user to the most relevant information on the page. We do not in this prototype alter the content or structure of the document, but rather give the user a set of enhanced bookmarks and guides to browsing the information. JavaScript is used to implement the dynamic alteration of the documents and to receive the information from the questionnaires. There is a standard set of functions that control:

- *Reading and writing cookie information*—Cookies are small amounts of information that can be stored on the client machine. They are a method of alleviating the state problems associated with inference. It is possible to save the state of the cookies so when the user returns to the documents the old state is restored. The JavaScript for this section of processing is stored in a JavaScript include file, not in the HTML.
- *Carrying out the logic of rules*—A small amount of code is required to carry out the logic by first extracting cookie information and then determining which sections will need modifying by a set of rules within a document.
- *Entries for each section that required alteration*—Each section of document, paragraph, sentence or word that requires alteration needs an opening and a closing section of JavaScript code. These sections of code determine how the document is altered within the browser.

We have also applied the prototype method to the Housing benefit example, and the results are shown in screen shots in Appendix 2. There we give the questionnaire used to get personalised information, the index page, the *who can claim* and *how to claim* pages customized in accordance with the questionnaire answers.

## 5. Problems with the prototype

The prototype was only a rapid implementation to enable us to get a feel for the issues, *en route* to a proper design. In this section we will discuss the issues that arose.

- *Hand coded control structure*: The underlying control code that has been written in JavaScript requires a good understanding of the JavaScript language. There is currently no automated method of generating a control structure within an electronic document. All development is currently implemented by hand. This approach is valid for a one off system that may require specialized code to improve the functionality of particular sections. As a generalized method however, ease of use and implementation is lacking.
- *Requires human expertise*: The technique requires a human expert to determine which sections of the text to enhance, modify or control. The document requires analysis to determine which sections can be adapted to take advantage of dynamic delivery. Once the sections have been identified, a questionnaire or information gathering form is designed. This form is the channel through which the user of the system can communicate their different circumstances. The information is gathered and must then be associated with JavaScript contained within the document. The placing of the JavaScript and the method in which the control is implemented can produce a very wide number of effects.
- *Non-standard technique*: The use of JavaScript is not compatible with some internet browsers. As new browsers are developed, the incorporation of JavaScript has become more widespread. The biggest problem is the use of cookies to store the state information between the questionnaire and the resulting document. Many users are concerned at the use of cookies for security reasons and it is possible to deselect the option within some browsers.
- *Small scale*: The prototype is currently only implemented on the small scale, although it is to be hoped that multiple pages can be interconnected to provide a larger resource. Because of the development time a large system has not yet been implemented. The WWW philosophy of breaking documents into pages does, however, lend itself readily to modularization, and there is no reason to doubt that a larger system can be constructed by combining smaller modules. The modules already developed (Housing Benefit and Income Support) have been combined in this way.
- *Development restrictions imposed by HTML and the Internet*: There are some restrictions imposed by HTML that make it hard to implement some desirable effects. Currently there is no method for incorporating another document into the text as this function is not available in JavaScript. There is no simple method for removing sections from the text, although recent experiments have shown cascading style sheets (CSS) can be used to do this. They are, however, as yet implemented in very few browsers.
- *Requires thorough testing*: During the construction of the prototype documents, it was obvious that the need for effective testing would be required. There are a large number of combinations of documents that can be generated, all combinations should be tested to ensure that they provide the correct information as originally intended.
- *Producing hardcopy*: Hardcopies of pages can be produced but they are little more than snapshots of the document as seen by the user. This project aims

at providing methods of viewing information on-line, it is also important that effective methods of generating hardcopies of dynamic content are considered.

- *Limited functionality*: Aspects of document presentation such as thesaurus and scope were not addressed in the prototype.

These problems can be categorised as being of several types:

(1) *Software engineering problems*: Issues such as testing, scalability and the need for expertise, only suggest that producing such a system is not easy, and the prospect of an automated add-on to existing WWW pages is a long way off. If you want this kind of facility, investment will be necessary. But this investment may well be worthwhile in particular cases: a standard cost-benefit analysis will be required, as is the case for any software system. One thing is, however, suggested: that it may well be worthwhile to develop general tools to support the development of this kind of application (cf. expert system shells).

(2) *Technology problems*: Some of the problems identified above relate to the (current) state of WWW browsers and the platforms used to access WWW information. This is something which changes almost moment by moment, and should not be allowed to drive the design to too great an extent.

(3) *The ad hoc nature of the prototype*: The prototype was proposed only as an exploratory vehicle. We need to generalize what was done there, and add functionality in a way which will progress towards the toolkit idea referred to in (1).

In the next section we will give our ideas about how we wish to generalize our ideas, and the kind of tools we wish to provide.

## 6. Generalisation from the prototype

The central idea is that we will mark up the document using a set of tags of the form:

```
< RULE ID = rulename STATE = rulestate ACTION = action >
```

where `rulename` refers to a rule in a rule base associated with the document, `rulestate` is a boolean indicating whether the conditions of the rule are satisfied, and `action` relates to the customization desired in virtue of the rule, and is one of the actions described in section 2 above. Each action is associated with a set of HTML tags which will have the effect of carrying out the given customization. Each rule is a set of conditions built up from terms the values of which are set by running the questionnaire.

When the document is loaded, it is first passed through a filter; this filter applies the rules and replaces our tags with the appropriate HTML tags. Thus when the document reaches the browser it contains only HTML tags, and so can be displayed by the browser, which will interpret these tags in its usual way. By removing the necessity for the document sent to the browser to contain JavaScript we avoid the compatibility problems noted in the prototype.

The process is shown diagramatically in Figure 1.

The process of creating the marked up document requires the following stages. First the document must be analysed, in conjunction with a domain expert. This analysis will produce:
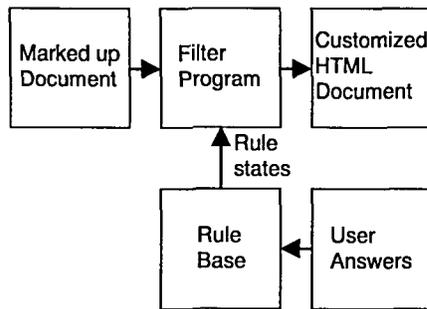
**Figure 1.**   System architecture

(1) a set of user attributes that will determine customisation:
(2) a set of customizations appropriate to each attribute:
(3) a set of questions to ask the user to determine the attributes.

Using this information the 'document engineer' will mark up the document with tags as described above. The process suggest a number of tools that can be used to support it:

(1) a questionnaire generator. This tool will also produce an attribute dictionary;
(2) a rule editor: this will make use of the attribute dictionary, and produce the rulebase;
(3) a document editor which will make use of the rule base to supply the tags.

These tools will both simplify the mark-up task, and enforce consistency between the various components.

## 7. Conclusions

In this paper we have described the requirement for customization of documents providing legal information over the WWW to tailor them to the particular needs of different categories of user. We have described how these facilities can be provided, and a simple prototype which implements a subset of them. We have discussed the issues that arose from the prototype, and advanced a modified system design which addresses them.

## References

Benefits Agency (1996) *A Young Persons Guide to Social Security*. Leaflet FB23.
Benefits Agency (1997) *Housing Benefit Page*. http://www.dss.gov.uk/ba/GBI/5a58abd.htm.
Soper, P. & Bench-Capon, T. (1994) Coupling hypertext and knowledge based systems: two applications in the legal domain, *Artificial Intellegence and Law*, 2, pp. 293–314.

**Appendix 1**

Current Housing Benefit page (http://www.dss.gov.uk/ba/GBl/5a58abd.htm)

Type: Non-contributory. Income related.

Taxable: No.

Who can claim? There are four basic conditions:-

   the customer must be habitually resident, and be permitted access to public funds:

   the customer must occupy the dwelling as their home:

   the customer must be liable to pay rent in respect of the dwelling: and

   the customer must either be entitled to Income Support. Jobseeker's Allowance (Income based) or have a low income.

How to claim: There are two methods:-

(1) If a customer makes a claim for Income Support/Jobseeker's Allowance (Income based). They can claim Housing Benefit at the same time.

Both forms should be sent to the customer's local Benefits Agency/Employment Services office:

(2) If the customer has not claimed Income Support/Jobseeker's Allowance (Income based) or is already receiving it, the Housing Benefit claim form should be sent direct to the local authority.

They can get the Housing Benefit claim form from the local authority.

How is it paid? In cases where the customer is liable to pay rent to the local authority, benefit is paid by reducing the amount payable.

If rent is paid to a private landlord, then benefit is in the form of a cheque or equivalent.

How long is it paid for ? Awarded for a period of up to 60 weeks at a time.

The award may be for a shorter period, depending on the local authority's administrative arrangements or if there is likely to be a change in the customer's circumstances.

Rate: Depends on four things:-

(1) amount of money coming in:

(2) amount of savings:

(3) personal circumstances (such as having children, being disabled. who else lives in the household): and

(4) amount of rent charged.

Maximum entitlement is 100% of the eligible rent. The amount may be less if:-

   non-dependents are in the household; or

   the rent officer determines that the rent is high or the accommodation is overlarge: or

   the rent officer determines that the rent is higher than the general level of local rents for similar property; or

   the rent payable includes charges for non-accommodation items; or

   the customer is single and under 25 years of age. The rent officer will make another decision based on the cost of a non self contained room (where the customer has shared use of kitchen and toilet facilities) but without being in the same area.

   Decisions made by a rent officer are valid for 12 months from the date they are made.

Other information

If a customer is thinking of renting a property form a private landlord, and intends to claim Housing Benefit to help pay the rent, they can apply for a Pre-Tenancy Determination. This will give them an idea of the amount of rent on which Housing Benefit is likely to be calculated.

People in work may have up to £60 of their Child Care Charges offset against their earnings as an allowable expense when calculating entitlement to benefit.

Housing Benefit is a qualifying benefit for funeral payments.

For further information and advice, contact the Local Authority Housing Benefit Section.

Main leaflets:
   RR1 Help with your rent
   RR2 A guide to Housing Benefit and Council Tax Benefit.
The Local Authority may also produce its own leaflets.

See also:

Council Tax Benefit Extended Payments

Housing Benefit Extended Payments

## Appendix 2

This Appendix contains four screen shots from the prototype application: the initial menu, the questionnaire with a typical response and two of the pages modified to take account of those responses: