INTELLIGENT PRESENTATION AND TAILORING OF ONLINE LEGAL INFORMATION

CHRISTOPHER ANDREW ROYLES

THESIS SUBMITTED IN ACCORDANCE WITH THE REQUIREMENTS OF THE UNIVERSITY OF LIVERPOOL FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

The University of Liverpool

2000

Acknowledgements

There have been many people that have provided support, guidance and advice during the time that this work has been undertaken.

First and foremost, I would like to thank my supervisor Trevor Bench-Capon who has been with me every step along the rocky road. I am very grateful for the confidence, feedback and guidance that I have received over the duration of this project. The effort and support Trevor gave, especially in the closing weeks has proved invaluable. Thanks for having faith in me and helping me fulfil my aims.

I am indebted to the members of two organisations that have helped with funding and experience during the period of my research. Particular thanks go to Colin Charlton of CONNECT for providing continued financial support throughout the first 2½ years of the project. I would also like to thank Charlie Whieldon and Nicolas Buckle of Vicorp.com for their support and patience. I am also grateful to my colleagues at Vicorp.com who have shown interest and support during the final 18 months.

I am very grateful to two close friends who have been with me during this project, Vicky Pye for her support and encouragement and Nora Hornyak for her patience and understanding.

Finally, I am eternally grateful to my parents and sister for their love, help and support.

Abstract

The Internet provides a vast source of information to a very broad and varied audience. For many years information has been presented online that is presented no better (and in many cases worse) than if it was published via traditional media such as paper. This is particularly evident of legal texts that are difficult for readers to interpret and put to use without seeking further advice. This research takes legal information sources available over the Internet and proposes a means of presenting the information. The presentation and structure of the documents can be tailored using Artificial Intelligence techniques in a manner more focused and relevant to the end reader.

This thesis provides a discussion of the current trends in online consultation and Artificial Intelligence applications. Many examples are either small scale or suffer from consistent limitations of the distributed architecture of the Internet. A number of questions are posed:

- Can information available over the Internet be presented in a form specific to the reader and the characteristics of the reader, their environment and context?
- Can a better architecture be developed for integrating Artificial Intelligence into the Internet that would provide control and intelligence in the way documents are presented?
- Is it possible to provide interactive online advice, developing the concept that an author can integrate their knowledge of a subject with the document that describes the topic?

The thesis provides a number of concepts and an example of an implementation, QERIS, which can be used to tailor the presentation of Internet content as a reader accesses it. The concept of embedding knowledge within the document and using this knowledge to invoke a consultation between the document and the reader provides the core of this work. This thesis develops the QERIS implementation by discussing two examples:

- The presentation of online Benefit Advice specific to the reader and their personal characteristics.
- The presentation of articles and content from the Convention for the International Sale of Goods based on the goals of the reader at various stages of the contracts life cycle.

The work undertaken provides further scope for research in the integration of AI and Internet technologies and also drive for the further development of new Internet applications in AI and Law.

Contents

List of Figures

CHAPTER 1

INTRODUCTION

1 Introduction

1.1 Presentation of online legal information

The Internet is a powerful medium for the publication of information, and is rapidly growing into the backbone of broadcasting, commerce, and communication. Legal information has been published on the Internet in many different styles, domains, countries, languages, and legal jurisdictions. The information is accessible to the public as a resource to inform, typically providing texts relating to queries of a legal nature. There exists a large amount of online legislation, cases, contractual agreements, and legal support tools. Many of these resources are available to practitioners of law and the public alike, without regard for expertise, language, or level of understanding. For legal information to be useful, it must be well targeted and accessible by those it concerns. As legal information is published online, the task of presenting the information in a clear, concise, relevant and useful form is becoming more important.

Practitioners of law have benefited from legal support tools for a number of years, providing support in areas such as document drafting, information searching and case retrieval. The Internet although providing large quantities of legal source material currently lacks tools for simplifying access to and interpretation of the material. More importantly, the public has access to government guidelines, advice, and legal information with little support in understanding or in the use of such material. The

Internet has provided the possibility of presenting legal information to a far broader audience, providing scope for legal publication on a broader scale. Because the Internet caters for different social, racial, ethical and jurisdictional boundaries large volumes of information available do not relate to many of its potential readers. It is time to address issues of presenting legal information to the *individual* user in a form that provides quicker access, easier interpretation, and in a form that can quickly be put to work to address the users needs. It is clear that legal information presentation using the medium of the Internet could benefit from new advances in hardware, software and networking technologies that have developed in recent years. The Internet is rapidly growing and evolving new and exciting opportunities for the communication and presentation of information. Originally the majority of legal information was designed for printing to paper and not for presentation over the Internet. Most online legal material is a simple conversion from this original form to an online format. The next stage has arrived where legal information is no longer acceptable simply as marked-up documents converted from paper representations but requiring development into more interactive and useful resources that take into account the new technologies available.

The large volume of legal information available online is aimed at a very general audience, and so in many cases not all sections of the information presented will apply to a particular user at any one time. In the case of Benefit Advice, a user is 16 or over or they are under 16. Knowing this small amount of information about a user can provide a mechanism for drastically reducing the amount of information that is of direct relevance; the reader then has less material to read and interpret. By identifying the characteristics of a user, it is possible to better guide users toward their goals; support them in their interpretation of information, and tailor the presentation of information to better suit their needs. As well as providing information more relevant to the users characteristics and situation, it is important to suppress information that could prove misleading or appear contradictory.

This thesis examines legal online presentation, discussing two problem domains, *Government Social Security and Benefit Advice* and access to articles from *The Convention for the International Sale of Goods (CISG)*. It will examine some of the

problems with current online legal information presentation, methods of reducing these problems and offer a solution that provides a new system for legal information presentation online.

1.2 Information and Advice

There has always been a demand for advice that can provide a quick method of identifying the information that will solve a particular problem. Users rely on experts to provide explanation and detail about a subject without the need to become totally familiar with a domain themselves. The Internet currently provides a mechanism for distributing information to the public and it is developing into a medium that can provide information in more interactive ways. There are a growing number of online advice solutions available over the Internet such as news groups, that provide interaction between human beings. A query can be posted and a useful response may be provided to that query. Automated methods of providing advice are also becoming more widespread. The Internet is well suited for interaction with an end user; information is accessible 24 hours a day, on demand in a very rich and flexible format. Once an advice system has been deployed it does not require extensive intervention by support staff and provides a good means of distributing and publishing information cheaply.

The Internet is designed to serve information at the request of the user. At a fundamental level the interaction is based on a one way discussion with the resource:

The User: "Please send me document X"

The System: "I have found it, here it is."

The extension of a Common Gateway Interface (CGI) protocol provided a more two directional discussion:

The User: "Please send me document Y"

The System: "I have found it, it contains a form I need you to fill in."

The User: "I have answered the questions you asked of me, here are the answers"

The System: "Here is the result of processing the answers you supplied, Thanks"

This technique is suitable to single interactions between the user and the system but if a new interaction is required then the process must start from the beginning. This method provides a means of document and domain selection but can only apply information collected during a single transaction.

A third and more natural form of interaction can be described as:

The User: "Please send me document Y"

The System: "I have found it, here it is."

The System: "By providing answers to these questions, the document can be tailored to your answers"

The User: "Here are the answers"

The System: "Here is the document tailored to your individual characteristics, Thanks.

In this model, an interaction is started every time the user requests a new resource from the system. The additional element is the loop that allows multiple questions and corresponding answers between requests. It is possible with this model to provide several interactions that can be combined into a final resource. The inside transaction provides a mechanism for changing more that a single component of a document to provide a single representation.

CGI provides a flexible approach of establishing a dialogue between the user and the remote system forming a core for online interactive systems. Different approaches currently used to provide online advice will be discussed in more detail in Chapter 2. The third model is the basis of this project and will be examined in more detail in Chapter 3.

As the number of users of the Internet increases, more people will resort to using the Internet as a primary source of information and means of communication. Information will be available from anywhere at any time through new devices such as TV, or mobile phones. Online navigation and support tools are already replacing the interactions usually provided by automated telephone switchboards. Modern network peripherals like printers; file servers and databases provide access for both user and support staff to the inner most workings of the physical device. Faultfinding tools, diagnostic applications and frequently asked questions (FAQ) databases are becoming more commonplace. The provision for enhancing the interaction with a user by providing guidance and targeting of information based on their individual characteristics can add considerable value to a site. But as sites become more integrated with devices then interactive and intelligent communication could be used to provide many new solutions.

The Internet technology currently being used is poorly suited to interaction in this style. Even well thought out implementations can be bogged down by the basic limitations of the Internet, its network speed, maintaining sessions, security and privacy. These and other problems will be examined and a possible solution, **QERIS** (**Q**uery; **E**ffect, **R**ule **I**nternet proxy **S**erver) will be explored. The implementation of **QERIS** as part of this project will be explained as a novel approach to expert system deployment, user support and online information tailoring. Examples of how this system can be used to improve current applications evident on the Internet will also be discussed.

QERIS is an amalgamation of three domains:

- The Internet acts as the presentation medium for information providing a rich networked multimedia platform.

- Law is the source material being used to represent information; two sub domains of law will be examined, Social Security and benefits advice, and the Convention for the International Sale of Goods (CISG).

- Expert system techniques are used to provide an intelligent query and processing engine, tying together the presentation of the material with the interpretation and semantics of the information. The result of the expert system processing is a single document tailored to each individual user.

1.3 Legal domains

There are many different legal domains evident on the Internet; two domains of particular interest will be examined in detail and a QERIS solution will be applied to them.

1.3.1 Social Security and Benefit Advice

The government provides information booklets to the public that cover many different Social Security and Benefit Advice subjects providing guidelines and instructions for claimants. An example section from the housing Benefit booklet is given below (Figure 1.1).

Housing Benefit

Who can claim? There are four basic conditions: -

- the customer must be habitually resident, and have recourse 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.

Figure 1.1: *Housing Benefit Section*

Simply converting the material available in the booklets straight into HTML currently creates the online version of these booklets. The online version includes navigational aids for the user providing access based on an index or specific events. Benefit information is very well suited to better presentation and support in its interpretation. The information is targeted at a very large and diverse population, the claimant, the advisor and third party support groups. Benefit advice is available in a number of formats including:

- Benefit advice documents - available to the general public from post office counters and benefit centres. Advice documents have been written from underlying Benefit legislation and fall into broad categories such as housing, child or invalidity benefit.
- WWW based HTML documents - many of the benefit advice documents are available online in HTML format. Documents are usually associated via hyperlinks and can be accessed via indexes in a number of different ways.
- Benefit advisor consultations - the general public can request an appointment with a benefit advisor in order to discuss particular points, fill in forms or seek advice. In many cases access to advisors is limited and usually involves a long waiting period. Benefit advisors usually work from books and other resources that outline the most recent interpretation of the underlying legislation. Benefit advisors can also provide information and guidance while filling in forms and documentation for claiming.
- There are local and national telephone helplines that can be used to get initial information about what benefits are available and can usually direct a claimant to the next step in the path when claiming a benefit. Helplines are also established where a benefit has changed or a new benefit becomes available in order to provide quick feedback to the public.
- A number of National media adverts during key topic periods - During certain key periods such as the change of a benefit or the introduction of a new benefit the media are used to provide advertising so that the public are more aware of what they may claim.

- Benefit Legislation - All benefit advice relies on the underlying legislation. Legislation is rarely used by a member of the public in favour of advice documents and benefit advisors. Legislation provides the most accurate representation of the law but requires an understanding of how to research all additions and amendments.

There are several problems in presenting advice to the public, which can be compounded by WWW presentation. A large volume of the information is formulated as questions and responses to those questions. Many of the sections replicate requirements and re-enforce previous statements. The target audience consists of a large number of individuals not familiar with domain specific terms or the interpretation of legislation and legal advice. It is clear that although benefit advice is available to the public it is in a format suited to a general population rather than the individual. Benefit advice will be examined as an example of tailoring general information to a more specific representation based on a reader's context and personal characteristics.

This project will investigate methods of using user characteristics to improve the presentation of advice and information. A mechanism for employing intelligent tailoring of WWW presented benefit advice will be investigated in detail in Chapter 3.

1.3.2 Convention for the International Sale of Goods

The Convention for the International Sale of goods (CISG) is a large document containing 101 articles. It is available online in a variety of languages and at a number of different locations. The advantage of the CISG as a source example is that it is standard across different legal jurisdictions, international boundaries, and languages. The CISG is also a document that is well suited to the Internets capacity for international publication.

Below is the first article from the convention

(Figure 1.2).

Article1

This Convention applies to contracts of sale of goods between parties whose places of business are in different States:

when the States are Contracting States; or

(b) when the rules of private international law lead to the application of the law of a Contracting State.

(2) The fact that the parties have their places of business in different States is to be disregarded whenever this fact does not appear either from the contract or from any dealings between, or from information disclosed by, the parties at any time before or at the conclusion of the contract.

(3) Neither the nationality of the parties nor the civil or commercial character of the parties or of the contract is to be taken into consideration in determining the application of this Convention.

Figure 1.2 *CISG Article 1*

At the core of the CISG are sets of articles similar to that above, that are related to different users, cases and stages of the life cycle of a contract. By extending the idea of interactive advice, the CISG document as a collection of articles can be reorganised and presented in a different form for each user, taking into account the different stages of their contract. The presentation of CISG components can be linked to the contractual obligations of the user and the state of their particular rights and obligations. By identifying attributes of an interested party and the current events in the contracts life cycle only the relevant sections and articles need presenting. There is currently a very active research interest in intelligent interpretation and presentation of the CISG (Yoshino 1997). The CISG will be examined as an example of tailoring a large legal document and its related resources to a more specific representation based on a reader's contractual obligation and the lifecycle of the

contract in relation to the CISG. The CISG will be investigated in more detail in Chapter 5.

1.4 Information presentation

1.4.1 Legal Information presentation

There are three key points in legal information presentation.

1. Legal information is currently available online and this availability will develop into a key resource as more people gain access and more systems integrate with those that already exist.
2. Legal information available on the Internet has lost elements of standardisation that exist in the printed version. Methods of indexing, highlighting revisions, making citations have been merged into computerised systems that can provide history and revision control but result in well established standards being diluted and ultimately lost.
3. Legal information online is just as difficult for a layman to understand but more worrying is that complex legal information is in some cases easier and quicker to find online than the simpler information aimed at the broader populace.

Legal information is presented in the form of documents containing formulations of statutes or cases. Material is presented in a way that maximises its usefulness by standardising headers, titles, sectioning, and indexes. Underlying the actual information is document structure and representation. In a digital format information has to be represented, the simplest form of representation is plain ASCII text, which can be extended by the use of a mark-up language. A mark-up language distinguishes strings of characters from each other so that different visual and processing effects can be applied to those characters.

There are four key layers in online information delivery that are important when considering the presentation and the tailoring of information, see Figure 1.3.

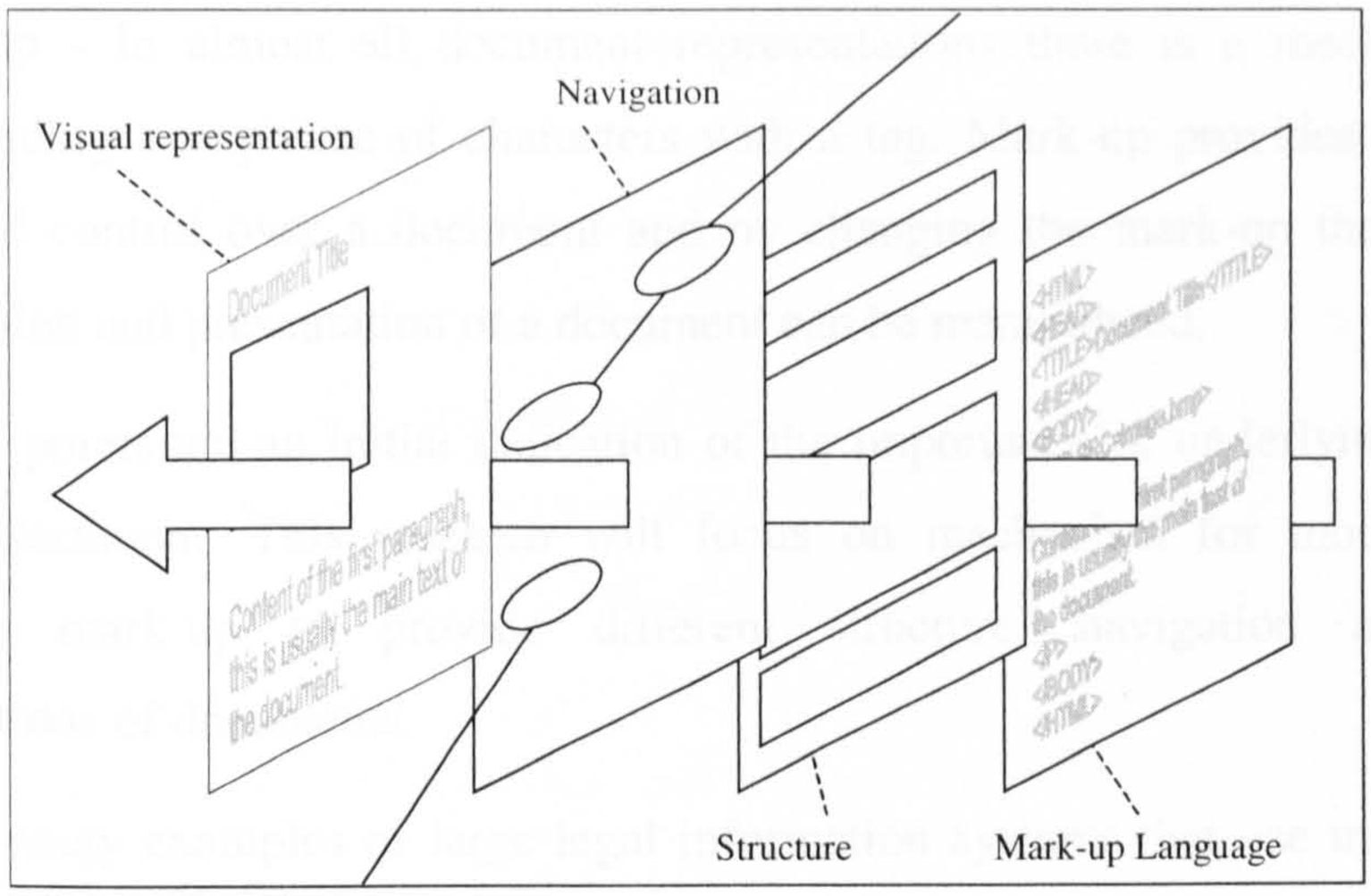


Figure 1.3 *Document Layers and structure*

1. Visual representation (style) - This is the final rendering of a document and is very specific to the client software used to read the document, the settings the user has made, the system the document is loaded on and the hardware being used to view the image of the document. The visual representation is the most dynamic element of the document and can be controlled by manipulating the three layers below it.

2. Navigation (hypertext) - Navigation provides association between different elements within a document providing a quick and efficient mechanism of moving between related subjects. Good navigation does depend on a mechanism for identifying which elements should link to which locations within the document or to external resources. In many documents, navigation is overlooked and left to just the contents and index pages. Navigation manipulates the visual layer in order to indicate which elements in the document can be manipulated by the reader.

3. Structure - The structure of a document provides standardisation and mechanisms for computer based processing. Structure within a document can provide far more predictable results across a collection of documents in the visual and navigational layers. Structural elements include paragraph separation, titles and mechanisms for compartmentalising the content of a document.

4. Mark-up - In almost all document representations there is a mechanism for surrounding a sequence of characters with a tag. Mark-up provides the lowest level of control over a document and by changing the mark-up the structure, navigation and presentation of a document can be manipulated.

The above points are an initial indication of the importance of underlying mark-up within a document. This research will focus on mechanism for modifying the underlying mark-up to provide different structure, navigation and visual representations of documents.

There are many examples of large legal information systems that use mark-up and hypertext as their main presentation mechanism such as DataLex, Greenleaf et al (1991), Widdison et al (1992). Linking this fine control over presentation to retrieval systems provides far richer and less cumbersome information interaction than reading linear, static paper-based materials.

1.4.2 The contribution of the Internet

Since the inception of the WWW in 1991, the number of attached machines has grown exponentially to the staggering figures known today. The popularity of the Internet fuels its growth, the volume of information, and the number of viewers on the Internet attracting more people to link in and participate. This growth also maintains a high level of research interest in the field of communication, distribution, and sharing over the infrastructure of today's Internet. All forms of information are available over the Internet and increasingly more legal resources are being placed online. There are requirements that certain legal information is published and made available to the public. Larger volumes of this form of legal information are being made available each day.

Legal material for practitioners of law has already undergone the transition to the digital format. Information is stored in vast databases, and can be searched remotely, cross-referenced, and manipulated with ease (LEXIS-NEXIS (1998)). The area that is lacking is providing legal information to the public over the Internet. Until recently, the vast majority of online legal information was produced by conversion to HTML from a different document format. As part of the conversion process

navigational and linking support can be added. An online resource also has the benefit of reaching a far wider audience. Conversions of documentation to the Internet do not explore all of the benefits of the Internet as a new legal information medium. In most cases the only benefits provided are the extended distribution and improved navigation. Understanding and taking into account the capabilities of the user in identifying sections of relevance are still essential to gain the most from the material being read. Not considering the capabilities of the reader may be acceptable in the case of a practitioner but not for the wider audience. Practitioners of a subject can usually make informed decisions and bring an element of experience to bear when accessing a document, they are less likely to take a document at face value.

The WWW heralded the introduction of a hypertext system on a grand scale. Although not as rich as some mark-up languages such as SGML (ISO 8879), Hypertext Mark-up Language HTML used as the underlying WWW mark-up language provides a mechanism for presenting different information formats, linking mechanisms, document structure and document appearance. The distributed nature, ease of use and linking and the location independence of the Internet provides a very powerful presentation medium with vast potential. The Internet has everything that is desired of a presentation medium, as discussed in Chapter 2, but it does have shortcomings although these are far outweighed by its benefits. Considering this new medium has such potential, it is regrettable that very little legal information is presented in innovative ways.

1.4.3 Intelligent presentation

The Internet provides a window onto a rapidly evolving space of information and media. Currently the most common interface to this information is via Internet WWW browsers. All a web browser does is download and represent a page on the screen. No attempt is made to determine if the document will mean anything to the reader, if the formatting is legible or if the reader is at all interested in the very large document that is currently being displayed.

One application area still to be successfully addressed online is the capability to present information to a reader that is relevant, in a format they can understand and

in a form they can make use of. Although information can be accessed and read there is still a stage where the user will need help in identifying or interpreting the information they are accessing. It is at the point where the user reaches for the telephone to contact customer support, a help line, or an expert in the domain that the Internet no longer proves sufficient. It is at this stage where artificial intelligence can be seen as a means of extending the support and augmenting the information delivered to the end user.

Artificial intelligence (AI) is a large domain with many practical online applications. The most obvious online applications currently using AI techniques are search tools such as AltaVista (http://www.altavista.com/) or Lycos (http://www.lycos.com/). Recently a new form of online system has developed. Online interactive databases are providing support for users alongside traditional customer support services, examples are usually based on software product support sites such as Microsoft's online support database (http://search.support.microsoft.com/). Server side solutions generate documents dynamically based on the user's demands, browser or profile. More accomplished examples of advice systems employ artificial intelligence techniques to provide a higher degree of interaction. Usually this form of application is based around a traditional expert system. Their use is becoming more common as Internet based interactive help desks or tutorial applications Ashley et al (1991), . Relying on CGI and server based implementation; these online expert system solutions have however shown little innovation in tackling the shortcomings of CGI implementation. Developers of such systems have identified problems such as their resource requirements, poor support for multiple users and their limited flexibility. The lack of successful systems available over the Internet is possibly influenced by implementation and deployment issues. These issues will be covered in more detail in Chapter 3.

1.4.4 The Internet in legal information presentation

There is historically a connection between legal software tools and hypertext based presentation. Large early legal systems such as DataLex used mark-up and hypertext techniques as modules for presenting textual information to the end user. Since the

early 1990s, intelligent legal information presentation via hypertext has undergone extensive research and development becoming an important element of several legal applications Greenleaf et al (1993). A good example of the transfer from a hypertext to WWW enabled application of an intelligent legal domain is that developed by the Australian Legal Institute AUSTLI. The wysh project (Greenleaf et al 1997) is a good example of the natural evolution of legal systems into intelligent online tools.

Because legal information can be presented effectively using a hypertext model, it is assumed that the Internet is also well suited to presenting legal information. Currently there is little evidence other than a handful of token examples of legal applications that address issues of presentation to the typical Internet audience. Current solutions tend to overlook important benefits of online presentation creating systems that although having potential actually emphasis the deficiencies of the Internet. This thesis will examine an approach for using the Internet for intelligent legal presentation focused more on the document being accessed, simple knowledge representation and practical integration rather than the reliance on large and complex knowledge bases.

1.4.5 Authors and users

Textual online information is written for a target audience. Authors must be skilled in writing material for that audience; the wider the audience the more general the description, the narrower the audience the more specialised the description.

The author can usually be considered an expert in interpreting the information. There are currently no means for an author to write a document that can adapt to different audiences as and when the need arises. There is a risk that expertise and mechanisms used to aid understanding can be lost in the translation of information between mind, page, Internet and reader and the translation mechanisms used in-between. Paper based mediums do not allow any adaptation of the original style, structure and meaning of information. The expertise of the author must be represented in the semantics of the document. Guides and indexes can be created to aid the user in the navigation of larger texts but each page, paragraph and sentence is fixed. There is little evidence of user support systems that allow guidance through and interpretation

of information. There should exist a mechanism for influencing a user beyond the actual writing and eventual reading of the text.

People (users) who access a document containing information are seeking to learn or answer questions specific to their circumstances. Users are all individuals; although they fall into broad groups, they have unique attributes that apply when seeking answers to their specific problems and circumstances. It can be assumed in the majority of cases that a married user will be more interested in marriage benefits than single parent benefits. If these attributes can be identified before information is presented to the user then they can be used to better represent the information. Conventionally, documents are written with at least a target audience in mind, for example this thesis is targeted at users with a working knowledge of computing.

> *"This thesis has been written with a particular audience in mind. It would be easier to target the information at a particular user or group if their interests could be identified before they read this document. An interest in Internet applications would allow me, the author to restructure the document and generate an index related to your interest thus simplifying your task as a user to identify points that are more relevant. All the source material would be available but in addition, the relevant sections about Internet applications would be highlighted and related links to additional information could be provided."*

This customisation goes beyond restructuring documents; whole contracts can be generated based on user attributes; application forms can be constructed as a user browses benefit advice information. Combining knowledge about the subject domain and tying that in with knowledge about the user provides a powerful foundation for controlling the presentation of information for the user.

Of the many users that will have access to a document, some of them will not be interested in the content, others will. Even if a user is interested in a document, there will only be certain sections that are specifically relevant to their interests or requirements. Users fall into three main groups when viewing a document instance. (see Figure 1.4)

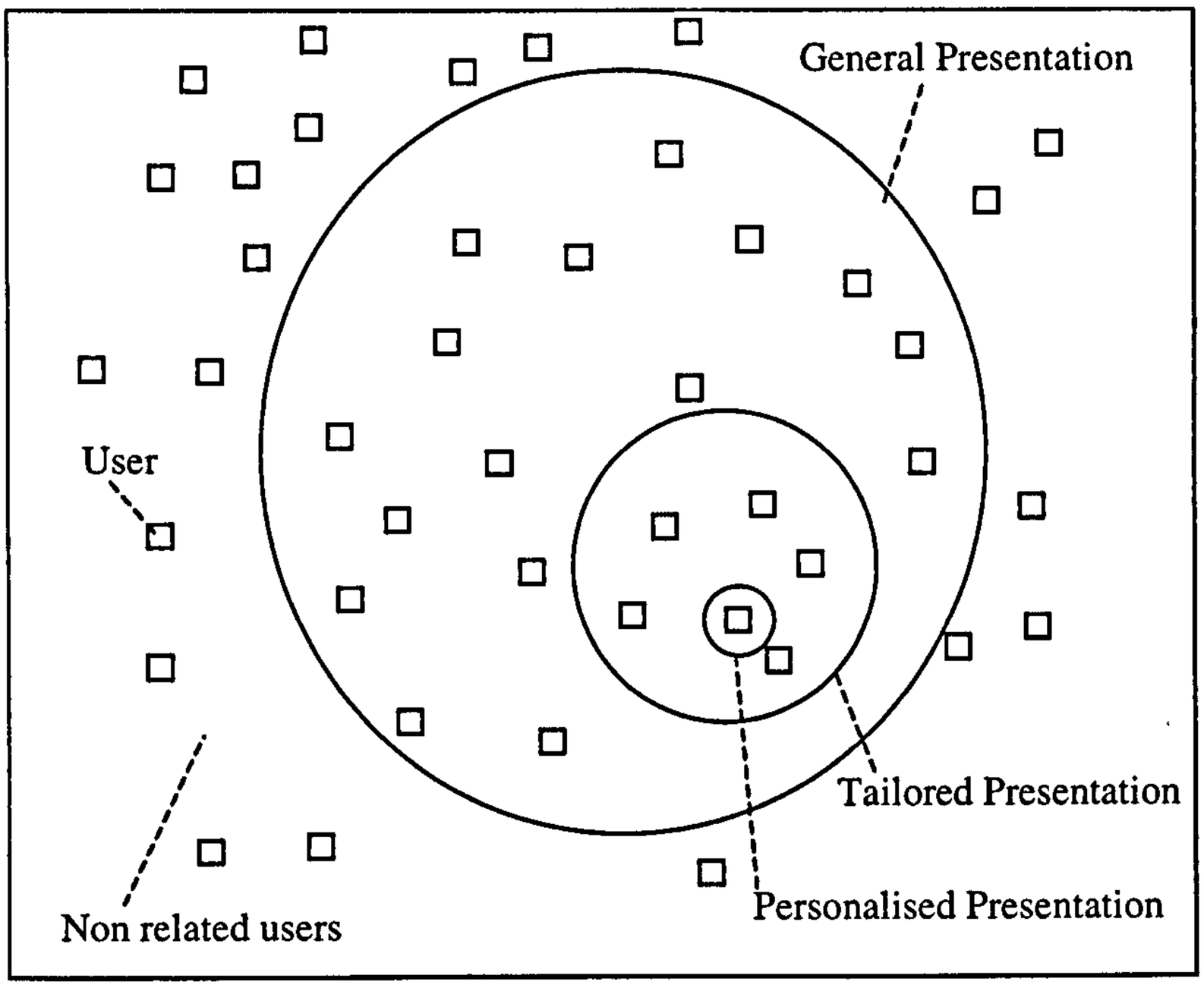


Figure 1.4 *User sets*

General Presentation – No specific user model, no user characteristics or context.

Tailored Presentation – User tailored information, the users characteristics and context are known but are based on groups of individuals. (e.g. User is older than 30 years.)

Personalised Presentation – Users unique characteristics and context are known. (e.g. User is 42 years old.)

A general presentation contains sections of information that may be relevant to a user and sections that are not relevant. By *tailoring* the presentation, a more specific version of the information is produced providing relevant sections to a subset of the users. By *personalising* the presentation further, the subset is reduced to a set containing a single user. The final stage is an ideal to be aspired too and is unlikely to be achieved in real domains due to the number of factors involved with such a process. Although this thesis discusses tailoring of a document to an individual user, tailoring actually reduces the presentation to a group of users that have similar

characteristics. The balance between tailoring and personalising is left to the developer of the document tailoring knowledge base.

1.5 Implementation

During the development of this project, a prototype implementation was carried out. The implementation provides a proof of concept and tool for explanation and realisation of the major concepts and actions of the system.

1.5.1 QERIS, a proposed model for document tailoring

This research proposes a new mechanism for delivery of information across the Internet. Documents are not only described by their content and their mark-up but can also include a knowledge base that is closely integrated with the mark-up. By using the knowledge base to drive an inference engine the mark-up can be modified based on the conclusions reached during inference.

The intention is not to create a mark-up for representing knowledge bases but rather a means of linking a knowledge representation with existing mark-up such as HTML. Mechanisms exist for providing more information about a document than is possible using the standard HTML tags and this is where meta-data is used. Meta-data are small amounts of information related to the current document but used to explain to underlying client applications special properties about the document. Meta-information can be used to represent the documents history, expiry date, origination and authoring tools used in its construction. In most situations meta-data is transparent to the user who is not aware of the information contained and in many cases not aware of processes that make use of the data.

Meta-data is usually known as data about data and it is this area that the knowledge representation attempts to bridge the gap. A document can be changed by manipulating the HTML tags that underlie its representation. By changing and switching tags on and off the entire representation of the document can be modified. The conclusions of inference are used to determine if tags should be switched and which state they should be left in. The inference is guided via a consultation with the

user so facts about the user can be used to ultimately effect the final representation of the document.

This project proposes a new mechanism for delivering information across the medium of the Internet. Content marked up with HTML can contain other information about its structure, semantics, and style using meta-data. QERIS provides a mechanism for combining a knowledge base with the HTML that underpins a document. By dynamically modifying the HTML using a one to one consultation with the reader controlled by the knowledge base, different representations of the document can be created. The QERIS application described later in this document is responsible for extracting the knowledge base from a document, carrying out inference based on the knowledge base, altering the underlying HTML and then delivering the resulting document to the reader.

There are a number of advantages to providing a knowledge base within a document. The document can be written in a number of different ways depending on implementation. It is possible to create documents that can be viewed without processing them, but during processing they are tailored and specialised to the user adding considerable value to their content. The use of a knowledge base and general mark-up concepts leads to a very general solution that can be used for many different documents and in many different domains.

1.5.2 System Architecture

There are three key points when considering the architecture of QERIS:

- Placement
- Integration
- Interaction

1.5.2.1 Placement

There are a number of methods that can be employed to create the control required over the source document. It is required that the source document be interrogated,

processed and then rebuilt based on certain criteria internal and external to the content. There are four main locations where the alteration of the source material can take place.

1. The server where the information is stored and where it is delivered across the network provides the developer with an environment for controlling documents when they are requested. Most modern WWW servers can provide some means of control over the documents being accessed by adding information to the documents as they are requested. The use of server side processing is almost exclusively the area that has undergone the most development related to online expert systems, extending already established expert system shells by adding a CGI or server integrated module allowing remote interaction.

2. The document can be processed at the user side, or the browser. The browser usually has a one to one relationship with the user; the information being accessed and viewed is tailored to that user. Currently browsers are used to process and generate the final graphical representation of the page, animate elements, and provide multimedia support.

3. Documents can be processed while they are in transit between the server and the user which provides control over both the information requests and the information source material. The use of a proxy server provides many more avenues for development and extension of a system while providing the greatest operating system, browser, and server independence.

4. Internet based languages such as JAVA and ActiveX allow the application to reside on the remote machine but to execute as a local application with restricted access. These languages are designed to animate and improve the functionality of WWW documents.

It was evident early on in the development of this project that the positioning of the system would be critical in maximising the advantages of the Internet over the disadvantages. It was evident that many implementations of Internet expert systems focus on server side processing and cross Internet interaction. The most obvious flaw

with these systems is the considerable lag between each successive query. There are also issues related to multiple user interactions with CGI programs and the control afforded over the processing of the source documents. This project proposes a different strategy for expert system deployment over the Internet providing a faster, robust and secure method of interacting with the user. The process uses a proxy server to provide the Internet support and integrates and expert system to provide the intelligent processing of documents.

1.5.2.2 Integration

The Internet although young is well established and built upon standards. It is the case that although standards exist there is incompatibility between different software platforms that must be overcome when developing a new Internet based application. Many applications available provide foundations that can be built upon. This project uses standards such as HTTP and CGI to overcome incompatibilities between Server and Browser software platforms. There are three areas of integration that will be considered:

- The integration between software components such as web servers and browsers that are well established and understood.
- The use of standards such as HTTP and HTML upon which much of the Internet is already built and highly dependant on.
- The integration with solutions already available on the Internet such as databases and search engines.

1.5.2.3 Interaction

As with many other Internet solutions, there is a degree of interaction with the user. There are some key elements that should be considered when developing an interactive solution for delivery over the Internet:

- The Internet is largely a transaction-based media and there are few well-established means of maintaining session and state between different requests between client and host machines.

- If users provide information then this information requires securing and the user must have a degree of trust in the groups that may have access to their information.
- Information should be re-used as much as possible to limit the amount a user has to enter during and between sessions.
- The method and representation of queries presented to the user should be controllable so that a service provider can create a consistent interface between the user and their service.
- The interaction with the user should return more than the user puts in so that there is a reason and advantage in using such a system.

The design and development of the QERIS solution will be discussed in detail in Chapter 3. To summarise, the solution provides an OS independent, well-integrated solution for interacting with and tailoring documents for a broad range of possible users.

1.6 Summary

Although attractive, eye catching, interesting and functional presentation of documentation via the Internet can be achieved, there is a lack of research into the basics. Many developments improve the aesthetic and visual elements of an Internet site but do not necessarily make it easier to access or to understand. A successful presentation of information allows the user to identify, understand and apply the information that they are seeking at any given time. There is a large volume of information available on the Internet that cannot be found or understood, and so cannot be applied.

From the user's point of view, the Internet should provide the answers to their questions. Can the volume of information be reduced, the representation improved and interpretation provided to aid the user in applying the information they are accessing? A solution will require integration with existing technologies, proliferation to reach a broad audience, be easy to author and provide sufficient advantages to the end user to be useful.

The acquisition of understanding and the seeking of information to provide that understanding are the major applications of the WWW. There are answers to every potential question available via the WWW if the user can find, interpret and use the information available to formulate a corresponding answer. Examining these three stages, *find*, *interpret* and *use*, there are many different systems available for searching and indexing content on the Internet, and although there is much work still to be undertaken, the hunt for information is now better addressed. The actual use of the information is beyond the capabilities of the WWW; it is a task that in most instances is carried out independent of the computer. The area that is of greater interest is that of interpretation, currently there are few systems available via the WWW that tackle this problem. Some of the questions that will be addressed include:

The aim of this thesis is to evaluate the techniques employed in intelligent presentation of information over the Internet. It is proposed that a lack of Internet and AI presentation solutions, highlights a difficulty in bringing together these two fields. Once existing techniques have been identified and evaluated then a more appropriate solution that attempts to address those difficulties will be investigated. It is further proposed that the knowledge an author has of a subject domain can be used when presenting documents about the same subject domain.

Three key questions that will be raised in this thesis include:

- To what extent are Intelligent-processing techniques currently used in conjunction with Internet information delivery?
  - The identification and analysis of existing systems.
  - The extent to which the identified systems exploit the Internet as a presentation medium.
  - Review the capabilities of existing solutions with the aim of identifying a more appropriate solution to intelligent online presentation.

- How can Internet distributed information be better presented to the end user? Can information available over the Internet be presented in a form specific to the reader and the characteristics of the reader, their environment and context?

- Is it possible to provide interactive online advice developing the concept that, an author can integrate their knowledge and experience of a subject domain with the document that describes that domain? This involves the:

  - Identification of suitable domains

  - A Representation for integrating an author's knowledge with the document that has been created using that knowledge.

  - A means of combining the document content and the authors knowledge into a technique for improving the presentation of the document to the end user.

- To what extent can the presentation of legal content be improved, by employing AI techniques in its presentation over the Internet? Several associated topics include:

  - The suitability of the legal domain for investigation into the techniques described and proposed in this thesis.

  - The identification of relevant sub domains within law for providing example scenarios to develop and demonstrate improved presentation techniques.

This thesis will examine techniques for presenting information via the WWW; how new techniques can be employed to extend the current capabilities of the medium to improve the presentation capabilities beyond those currently employed and how these presentation techniques can be applied to legal information. It will examine information sources from the domains of contract law and government benefit advice, using these sources as examples of presentation of legal material over the medium of the Internet. The use of artificial intelligence techniques to reduce the input required by the authors and recipients of legal information will also be examined as a mechanism for improving information presentation, system interaction and user guidance.

1.7 Thesis contents

Chapter 2 examines the background of the three contributing domains:

Legal information – Legal information is used as the source material for this research, and there is a large volume of legal information available on the Internet but there are very few methods for providing access and support in understanding such material.

Artificial intelligence – There are many different techniques available in AI for helping users interpret information, generate documentation, simplify task and for finding information. There is little evidence of AI techniques that can be used to alter currently existing documentation into a format that can be read and better understood by smaller groups of people within a large more general audience.

Internet technologies – Technologies associated with the Internet continue to be developed at a fast pace, there are new concepts, standards, and utilities available, and it is important to push their capabilities and investigate new techniques not yet employed. By integrating AI and Internet technology, methods for presenting legal information can be investigated and new techniques developed.

Chapter 3 describes in detail the underlying concepts and physical components of a prototypical system, QERIS. The integration of AI and Internet technologies to provide a general tool for representing knowledge within documents, extracting that knowledge and then rebuilding the document based on the characteristics of the reader.

Chapter 4 examines a real solution for presenting benefit advice over the Internet using QERIS. The structure and current representation of advice documents will be discussed. An example of how the document may be represented at a knowledge level and how this knowledge will effect the final document presentation will be described.

Chapter 5 examines the Convention for the International sale of goods, a much larger and more involved document currently available over the Internet. Methods for representing this document and the different approaches that may be taken in representing it within QERIS will be investigated. The CISG differs from benefit

advice due to the involvement of time. The documents presentation will be affected not only by the attributes of the reader but also the characteristics of the contractual agreement they are involved with.

Chapter 6 discusses the overall findings of this thesis and summarises the work undertaken.

CHAPTER 2

RELATED WORK AND LITERATURE

2 Literature Review

2.1 Overview

This chapter reviews the separate strands of previous work, which have been brought together in this research. The aim of the research is to investigate and extend the presentation techniques of legal information over the Internet tailored to the needs of a particular reader. The separate strands relevant to this work are found in current systems that make legal information available, expert systems; the application of novel techniques to the legal domain; and the explosion in communication and Internet developments. The research in each of these areas will be reviewed in turn.

The Internet provides an enabling technology that extends the types of interface available between the author, information sources and the end users.

- There are many tasks carried out in the legal domain that can benefit from the fusion between Internet technologies and intelligent systems.
- Currently there is little evidence of approaches that exploit the benefits of combining these three domains, and exploit the new capabilities of the Internet as a legal information-publishing tool.

2.2 Legal Information

Unlike other sources of information, legal information is prevalent throughout society and applies to a very large audience. The audience is not only the people who

work in the field of law, the clerks, solicitors and judges but also the people affected by the law, the general population. In the case of European and International Law the "general population" includes groups from different countries, religions and languages.

An action by an individual or group will usually involve some degree of legal responsibility; tasks such as buying a house, getting married or driving a car each have a corresponding source of legal information that will apply to the circumstance, and which needs to be observed and taken into account. Some actions require the generation of new legal resources such as contracts, licenses or case reports.

It is important that when carrying out an action, a person is aware of their rights and duties and is properly informed of the legal implications. They must therefore have access to the resources in a format that can be understood and put into practice. The person should also be provided sufficient time and access to resources in order to carry out this task. Since practitioners and the public have different needs and capabilities we will consider them separately.

2.2.1 Legal information and practitioners.

The volume of legal information available to practitioners of law is vast and constantly growing.

A solicitor will usually allocate a large portion of their time to research when undertaking a new case. It is essential that to provide good support to their client they are up-to date on the relevant fields of law. Many lawyers will also specialise in certain areas of law such as, contract, criminal, or divorce law so that they may focus their effort and provide more specialised advice. Because of the volume of information available, identifying the most recent and relevant sources is becoming a far harder task. This task has been eased by the introduction of computerised legal research tools. Non-computing techniques of indexing and storage have become over stretched while the volume of material being added to such systems or requiring maintenance has continued to grow. The vast amounts of legal information and resources are the perfect subject for computerisation. Since the introduction of computer systems in the legal domain, computer based information systems have

been in great demand. Computers play a very important role in the modern law firm as described by Widdison (1993) who catalogued the use of computer databases within law firms identifying the tasks they are put to and the growing use of computer systems in everyday legal tasks. His observations showed that database solutions and integration between different legal IT systems would continue to expand within legal firms with expected expansion of around 15% over three years. These observations have been borne out by subsequent developments Brooke (1998).

2.2.1.1 Legal Resources

The key resources in the legal domain are *Legislation* and *Cases*. Legislation represents the rules and procedures by which a certain action is governed. Once passed by Parliament a bill becomes an Act and thus law. A case is a document that concisely records the elements of and the decisions reached in a trial. Cases are a very important mechanism in law as they can be cited as support or counter examples in future judgements. Usually cases made in a higher court set precedents, which are binding on lower courts. The successful identification of cases and the correct interpretation of legislation are important for a task in the legal domain to be carried out.

2.2.1.2 Introducing the CISG

The Convention for the International Sale of goods (CISG) is a good example of a legal instrument that applies across geographical boundaries. Because of its International nature, the CISG is available in several languages and is available on the WWW in many formats. The CISG has a large collection of associated cases and represents a good example of a legal document and related resources. There has been considerable interest in developing intelligent systems that use the CISG as a source. A series of International workshops (Amsterdam 1993, Washington 1995 and Melbourne 1997) are devoted to the CISG and the journal of Advanced Computational Intelligence produced two special issues largely related to this topic in 1997. This research builds on the work carried out in by Yoshino (1998) in defining the logical structure of contract law.

Due to the interest in presenting the CISG to such a broad international audience across the medium of the Internet the CISG has been selected as a subject domain that will be studied in more detail. The CISG will be discussed further in Chapter 5. A demonstration of how a proposed document tailoring system can be applied to improve the presentation of the CISG to a wider group of readers using the medium of the WWW will be described in that chapter.

2.2.1.3 WYSH: a WWW based AI legal tool

The most influential Internet based AI solution that is grounded in the domain of Law has been developed by a group led by Graham Greenleaf and Andrew Mowbray at the Australian Legal Information Institute (AustII). The institute is prominent in AI and Law research, continuing to push the limits of legal systems, investigating novel ways of combining AI technologies with legal retrieval to further aid the practitioner in their task. One of their accomplished systems is an expert system shell that was used in the DataLex Workstation project Greenleaf et al (1991), ysh. ysh has more recently been developed into wysh, Greenleaf et al (1997), a WWW based adaptation of their previous work. The work carried out by AustII is a good contrast to the different source materials and techniques discussed in this thesis, although in recent years the institute has tended to focuses more on database representation and publishing of material than the use of intelligent techniques.

2.2.2 Legal information and the public

There are many sources of legal information that are published as advice to the general public. Issues relating to areas such as tax law and social benefits must be available to people so that they are not ignorant of the law as it applies to them.

Susskind (1989) in his description of the future of law identified the implications of the current and future volume of legal information.

> *"By coining the term hyperregulated I have in mind that we are all subject, in our social and working lives, to a body of legal rules and principles that is so vast, diverse, and complicated that no one can understand their full applicability and impact."*

Hyperregulation is only part of the problem. The lack of understanding of the law that regulates the public by the public itself can compound the problem. Providing information is not sufficient; if information cannot be accessed, in a language and format that provides quick and correct interpretation and application then it loses its immediacy and its usefulness is reduced.

2.2.2.1 The Department of Social Services

Many government groups such as Tax, Inland Revenue or benefits have a duty to inform the public of their obligations and rights. In the UK, this form of information is widely available as a collection of advice booklets broken down into broad categories such as Housing Benefit, Child Benefit or Income support. Attempts are made to provide the public with the information they need, and media such as the national press and TV are used to provide advertisement of important issues. The information provided by these media is limited and usually designed to attract the attention and guide the user toward further support or advice. The forms in which the information is presented are constrained and therefore limited. The people that the information is aimed at are not always receptive to the information, having limited interest in information that does not appear immediately relevant. Another aspect to information availability is the rapidity of status changes; actions such as starting or leaving a job, bearing a child or changing house will each make different topic more or less relevant. Having quick access to the relevant information can smooth difficult transitions. Some topic domains can also have very serious implications if the person is ignorant in situations such as signing off the dole and starting a new job. Not abiding by such requirements can give rise to liability and possible prosecution, as well as the loss of potential benefits.

Gilbert (1991) reviews two systems developed for the DSS as attempts to provide different methods of interacting with users seeking benefit advice.

- The Forms Helper was designed as a demonstration of how computer applications could help in supporting a member of the public in completing application or information forms. It is important that when a person identifies their right to claim a benefit or apply for a service, they can successfully provide the details that are required to complete the claim. There are many instances of

claims being delayed or by incorrectly completed or spoiled claim forms. Another problem is people incorrectly filling a form in such that they do not receive their entitlement or the opposite, which could result in a fraudulent claim. The Forms Helper was designed to alleviate some of these issues and guide the claimant through a questionnaire presenting only the relevant sections and providing explanation where required.

- The Advice System is designed to provide a wide range of information sources. The information is made available in a similar dialogue that may be undertaken by a human advisor. The particular model for the dialogue was based on radio phone-in advice sessions.
- Both of these systems used sophisticated computer systems, and were designed to be run as stand alone systems in Social Security offices, where support in their use would be at hand if required.

2.2.2.2 High Tech Public Advice

Historically the DSS has been interested in exploring new means of interaction with end users. The limitations of technology have limited the success of initial approaches. For example PRESTEL (a predecessor of teletext offered by British Telecom) was explored as a means of presenting advice to the public, but the technology proved inadequate at that time. (Bench-Capon T. J. M, personal communication (1998)). Methods of reaching a wide public audience were limited to non-interactive distribution mediums such as TV and telephony.

The Internet has removed several limitations in distribution by providing a rich and easily accessible resource. The Department of Social Security now provides information online at its WWW site http://www.dss.gov.uk. All of the paper based advice leaflets have been converted into an online version. This migration to the WWW does not address the initial issues but distribution and maintenance is now improved. As yet the potential that the WWW offers for interaction has not been exploited.

Other groups who also provide public information include the Citizens Advice Bureau, Local Authorities and independent groups such as Accountants, Lawyers and interest groups.

Their principles are

> *"to ensure that individuals do not suffer through ignorance of their responsibilities or of the services available; or through an inability to express their needs effectively."*

This statement sums up the capabilities of providing advice by trained advisors to the people that really need support or legal advice. Even though the CAB maintains such strong principles, their access to the "individual" is limited to very conventional channels. People must approach the CAB, book appointments and wait on lists before they can successfully "express their needs effectively". The lack of channels to the individual hampers the immediacy by which issues can be addressed. By opening more channels and providing a larger volume of advice without the need for human advisors many minor issues could be resolved far quicker without the need to resort to the already over-stretched resources of the expert advisors. The Internet and intelligent presentation of information and advice could provide the means to lift constraints and lessen loads. Public advice and documentation will be discussed further in Chapter 4.

2.3 Expert Systems

By encoding a human expert's knowledge within a knowledge base a system that emulates the task of the expert can be created. Such systems are called "Expert systems" Jackson (1990). Current expert system technology is far from recreating the full skills of a human expert due to the inability of encoding experience, common knowledge, ingenuity and other elements humans take for granted. Even with these limitations, however expert systems are used in many subject domains applied to many different tasks (e.g. the Retirement Pensions Forecast Advisor (RPFA) which provides written information in the form of letters on likely pension entitlement (Springel-Sinclair 1988)).

Expert systems can provide an automated solution to a number of mundane tasks and there are key advantages in using expert systems over human experts.

- Expert systems are available on demand, 24 hours per day. Because expert systems are highly automated, there is little need to take them out of service as is the case with support and help telephone lines.
- Expert system advice is consistent and the advice can be validated and verified for correctness. This is in some cases better than advice provided by a human which can vary greatly between advisors based on experience and training. This was seen as the chief benefit of the RPFA.
- Expert systems can be extended and maintained providing immediate advances in its capability. If designed well, the knowledge base can be continually refined and updated to provide new services or better advice for current services, and as the law changes.
- Expert system can be scaled to deal with a large number of enquiries easily and quickly. Human advisors need to be trained, and there can be a very difficult period between heavy demand and the training of new advisors where active advisors will be stretched to their limits. Expert systems can simply be duplicated, limited only by hardware and network constraints.
- Experts are *very* valuable and their knowledge represents an important business asset, encapsulating their knowledge ensures that if the expert becomes unavailable another option for maintaining a consistent level of service exists.
- Expert systems can interact with many different users about many different issues simultaneously.
- Expert systems can act as a pool of knowledge from a number of different experts.

There are, however, problems with expert systems that still limit their use:

- Knowledge acquisition is a complex and time-consuming task. In a legal context it is important to collect correct and complete information from the domain expert.

- Knowledge bases can be very difficult to maintain and extend. Rules are abstractions of information and understanding, in many cases it requires a skilled knowledge engineer to understand and provide the abstraction. Every time the knowledge is modified it is essential that it is tested as fully as possible.
- Expert systems have a very clinical interface with the end user that in some cases is not desirable. Asking many questions without establishing a conversation with the user can be distracting and tedious.
- There are psychological issues of expert systems providing adverse decisions, it is easier to accept bad news if explained by another person who can provide emotional feedback and adaptive support, see Rainer Born (1987).

2.3.1 Architecture

Expert systems consist of two main elements, an inference engine and a knowledge base. The knowledge base contains the domain knowledge in an encoded format, the knowledge representation. An inference engine, the heart of the expert system, processes the knowledge base, applying the knowledge to a task and interacting with external elements such as an operator. In addition to these two key components, there is usually a front end, the user interface, now typically a GUI and additional interfaces for interacting with external resources such as databases. See Figure 2.1.

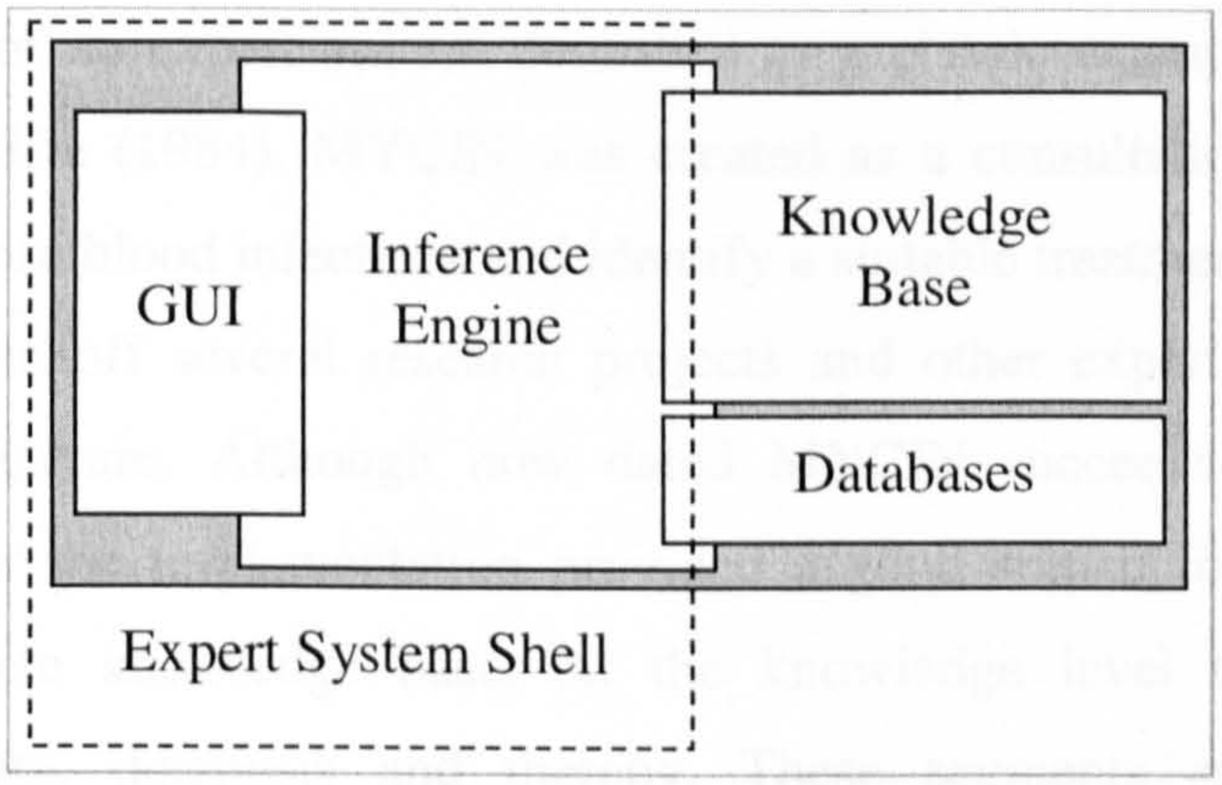


Figure 2.1 The key components of an Expert System.

In some cases, simply providing a different knowledge base allows the adaptation of the expert system to a different domain. In this situation the inference engine and related elements are known as an expert system shell. The task of the inference engine is to search for elements in the rule base that will help to progress the current thread of investigation. The inference engine must also overcome conflicts between rules and prioritise them and their sequence of application. The knowledge base can be built using a number of techniques such as production rules, frames or semantic graphs. Expert system architecture is well know and has been well developed, there are a number of good references and examples such as Jackson (1990) and Buchanan et al (1984).

2.3.2 Rule based expert systems

The term 'rule based' relates to the way knowledge within the knowledge base of the expert system is represented. Rule based systems have proved the most popular way of building expert systems and are well understood and widely documented (Stafik (1995), Luger & Stubblefield (1997)) There are many methods for extending the simple rule based model and some of these approaches are discussed in the next section.

2.3.3 Classic Examples

An early example of an expert system described as a classic example is MYCIN, Buchanan & Shortliffe (1984). MYCIN was created as a consultation tool to help medical staff diagnose blood infections and identify a suitable treatment. The success of MYCIN has spun off several research projects and other expert system shells based on its architecture. Although now dated MYCIN succeeded because the domain selected for the implementation provided a good source for generating a limited but complete knowledge base. At the knowledge level the domain is segmented into data, diagnosis and therapy. These segments are represented internally as a backward chaining rule base with the representation shown in Figure 2.2.

RULE177

If 1) the infection is primary-bacteremia, and

2) the site of the culture is one of the sterile sites, and

3) the suspected portal of entry of the organism is the gastrointestinal tract.

Then there is suggestive evidence (.7) that the identity of the organism is bateroides.

RULE178

If 1) the identity of the organism id bacteroides,

Then I recommend therapy chosen from among the following drugs:

1 – clindamycin (.99)
2 – chloramphenicol (.99)
3 – erythromycin (.57)
4 – tetracycline (.28)
5 – carbenicillin (.27)

Figure 2.2 MYCIN Rules

MYCIN employs certainty factors to represent uncertain data, which is shown by the use of numbers in the rules above. So for rule 177, the value .7 indicates that the system would have a reasonable degree of Belief in the result. CF are combined to form a CF for the conclusion. By using CF, the degree of confidence in results can be provided to the user.

MYCIN is aimed at a user that is familiar with medical terms, as a tool in aiding a doctor in making a treatment decision. Much of MYCIN's success is the manner in which it interacts with the doctor using familiar (to doctors) terms and following logical deduction as used by a human carrying out such a task.

It is the case that the design of expert system will not always reflect the methods and terms employed by their users, which is where the development of expert systems proves difficult. The group of users that provides the greatest challenge is the general public due to the diversity of skills, expectations and goals. The major reason for considering MYCIN is its influence on expert system design, as it represents the prototype from which all subsequent systems of this type derive. Other leading examples of expert systems include:

- DENDRAL which was one of the first early examples (1965) of an expert system, using a rule based approach. DENDRAL was used for the analysis of molecular structures. DENDRAL is no longer an active program and no longer a subject of academic research. Popular references to DENDRAL include Buchanan et al (1978), Gray et al (1980) and Lindsay et al (1980).
- PROSPECTOR was an early and successful example of an expert system used for geology and mineral analysis. (Duda et al 1979)
- XCON, used for configuring DEC minicomputer systems McDermott (1981),(1983).

2.4 AI and Law

In the legal domain there are many tasks that are considered mundane, repetitive and time consuming. Many of the tasks require processing, cross-referencing and correlation of vast amounts of information and data. There already exist successful systems for storing and retrieving legal resources such systems include Lexis (http://www.lexis.com/) LEXIS-NEXIS (1998) and WestLaw (http://www.westlaw.com/). More recently interest has grown in the development of intelligent support systems that can help interpret legal resources. This section will discuss several AI techniques and how these techniques can be used in tasks common in the legal domain.

2.4.1 Techniques applied to the legal domain

AI has been applied to law in many ways and the interest in research and developing AI applications is considerable. There are many AI techniques that can be applied to

the legal domain. Expert systems can be applied to many different subjects but have unique qualities in the legal domain. An area such as advice for legal practitioners has seen extensive research and development in recent years. The vast quantity of legal information has demanded the development of retrieval and research tools for the practitioner. Some legal applications are taught to lawyers at a very early stage in their career and are a key element in a lawyer's armoury and it is common for experience with Lexis or WestLaw to be an important requirement. Conversely, since the vast amount of information needs to be created easily, practitioners often use automated tools to support the creation of legal documentation. Many legal applications also employ presentation tools and hypertext has been very successful in this area and is integrated with many AI tools as part of the interface between the computer and the user. Many of these techniques are described in further detail in the survey paper of Sergot (1990).

2.4.1.1 Information Retrieval

The first developments in legal tools were large databases designed to make the storage and retrieval of legal resources easier for practitioners of law. One of the earliest examples of a database retrieval system was developed by WESTLAW. Initially WESTLAW developed an abstract-text retrieval system that allowed users to query a database and retrieve abstracts of legal material. In contrast a competitive group LEXIS developed a full text retrieval system that like later generations of WESTLAW is still in use by professionals today.

Information retrieval available through the use of search terms to a database has limitations. Such systems because of their reliance on Boolean combinations of keywords are not capable of retrieving information that is similar conceptually to the goals of the querying party, unless the same terms are used. The use of conceptual retrieval attempts to blur the black and white constraints of a straight matching between search criteria and the documents returned to the user. Conceptual retrieval provides a means of supplying examples to the system rather than search terms. An initial query may begin by providing a search term and an initial set of documents would be presented as likely candidates. The user may then indicate which of these documents matches what they consider to be their goal. By providing this feedback

method the system then begins to search not on fixed search criteria but based on the 'relevance' of documents in the search space with the documents indicated as similar to the users goal. De Mulder et al. (1993) consider a concept in legal document retrieval as "a set of documents", and that furthermore

> *"A concept is an ordered pair of sets of documents consisting of exemplars (of relevant documents) and counter examples (a set of non-relevant documents that are as similar as possible to the relevant documents)."*

Concept retrieval uses examples of supportive and non-supportive cases selected by the user to determine the most likely candidate set of documents. This methodology is well suited to the nature of legal information.

The key components within conceptual retrieval systems are:

- The Search mechanism employed.
- The mechanism that constructs the dimensions and measures a documents relevancy.
- The method of providing search criteria and representing selected documents (the GUI).
- The mechanism for determining the relevancy of a document is the unique element of conceptual retrieval and contains the AI elements.

2.4.1.2 Legal Expert Systems

Artificial Intelligence and Law is an active research area represented well with an annual International conference, International Conference on Artificial Intelligence and Law (ICIAL). There are a number of groups involved in extensive research. The following sections will introduce a number of relevant examples.

Graham Greenleaf and Andrew Mowbray have emphasised a need for developing practical and useful legal AI solutions One of their major systems was the DataLex workstation designed to provide a practitioner with a general-purpose tool that combined hypertext, inference and retrieval into one system. DataLex combines a number of separate modules supporting document retrieval and generation; case

based inference, backward and forward chaining rule based inference and a hypertext GUI for interaction with an end user. The Hypertext front end acts as a means of interacting with the knowledge base and instantiating an inference session. Feedback and prompts are also provided in a hypertext format so links can be provided to relevant resources. DataLex is a general solution and could be described in each of the following sections due to its Case based support, document drafting and integration with hypertext.

The inference engine used as a module of the DataLex workstation is called *ysh*, a well developed inference engine capable of many different AI tasks. The engine supports both case and rule-based inference as well as limited support for document generation, the rule-based inference can be either forward or backward chaining. Ysh has a well-defined and flexible knowledge representation that supports a natural language format described as a "paraphrase of ordinary English" as discussed in Greenleaf & Mowbray (1993). Ysh also forms the foundation of a WWW based solution wysh (see section 2.2.1.3).

Other legal expert system examples include:

- TAXADVISOR: (Michaelson R 1982) - TAXADVISOR is a system aimed to help practitioners in tax planning by collecting information from a client and providing recommendations and suggestions for organising their estate.
- Latent Damage Advisor: (Capper P & Susskind R 1988) - The Latent Damage Advisor is an expert system that provides advice on the domain of the UK's Latent Damage Act 1986. (Capper P 1987)

2.4.1.3 Case Based Reasoning

An extension to document retrieval is case based reasoning. Like document retrieval case based reasoning relies on a large database from which examples can be selected, but this time the database contains cases. As well as storing the content of a case, factors offering an analysis of the case are also typically generated and stored.

Cases tend to be more highly structured than documents (a vector of factors rather than free text). These factors represent how closely a case fits within a certain concept. When a user wishes to retrieve a case from the database a matching between

the factors of a test case and the factors of the cases stored within the database is made. In a similar manner to conceptual retrieval, the returned cases can be accepted or discarded and a further query to the database can then extend its search to take into account those examples accepted and discarded by the user thus refining the search criteria. Case based reasoning is not unique to the legal domain but does lend itself well to the large volume of reports and judgements of the many cases made each year or in some systems, representation of the "landmark" cases which have clarified the law relating to a particular issue.

Case based reasoning provides a research tool for legal practitioners rather than providing conclusions or interpretation of a decision. An established example of case based reasoning is HYPO.

HYPO (Ashley & Rissland 1987, Ashley 1990) is a case based reasoning tool applied to the domain of Trade Secrets Law. As with the general view of a Case based reasoning system, HYPO maintains a collection of *dimensions* that represent the factors of a case. In addition, conditions are associated with each case to determine which dimensions apply. It is not required that dimensions match exactly for a case to be successfully retrieved, similar cases are also identified by conditions which define similar dimensions between cases. Cases are then arranged in a lattice representing the similarity to the current case. HYPO does not simply generate a list of similar cases, it attempts to generate similar cases, counter example cases and form a legal argument that applies to the case under enquiry.

The developers of HYPO have gone on to produce more sophisticated systems. CABERET, (Rissland & Skalak 1992), integrates the case based reasoning of HYPO with rule based elements and CATO (Aleven 1997) introduces a hierarchy amongst the case factors which allow for more elaborate argument moves. Interesting as these developments are, the basic mechanisms remain those of HYPO.

2.4.1.4 Document Drafting

In the legal domain the vast quantity of documentation gives considerable scope for tools to assist in document creation. Documents of a particular class must be

consistent, information about a case must be detailed and accurate. Many legal documents must conform to a standard structure and style.

Daskapopolu & Sergot (1995) explain that document drafting is a central feature of legal expertise. Document drafting systems must provide tools for the legal expert that can carry out many diverse tasks while providing feedback about the reasoning employed in creating a legal document. In that paper, they go on to describe a technique whereby a document can be created, the result of which can then be queried explaining the structure and content of the document to the user. Constraints are defined that allow the construction of documents based the relationships between document sections. Sections are included based on constraints within a document template and between different sections that may be incorporated within this template. A similar approach has been used in commercial systems such as Scrivener (Dianoetic Development Company) and HOTDOCS (Culley A 1995). Other systems are summarised in Soudakoff & Lauritsen (1997).

Document drafting techniques fall into three main categories:

- Procedural applications follow a script that determines which sections to include/exclude for the document being generated based on user responses. A good example of such early systems is described in Sprowle (1980).
- Template-Based applications where a document specific template defines what sections should be included, what requirements are imposed for inclusion/exclusion and where within the template the subsection applies. This is still the most common approach.

Discoursed-Based techniques identify relationships between separate sentences and statements in documents. Branting et al (1997, 1999) continue by defining a discourse approach as a structure comprising 'a number of rhetorical relationships among sentences, such as elaboration, exemplification, generalisation and sequence' into which the various illocutions required by the document are placed. These techniques are still very much at the research stage.

PLAID (Bench-Capon & Staniford 1995) is an example of a legal document-drafting tool which has a very focused goal, the generation of a brief concerning a legal

question. PLAID brings together many different sources of information to provide not a simple answer to a question but an argument and a collection of supportive materials that may help a practitioner reach the answer to a question. PLAID constructs documents from sections retrieved from a database that match clauses within the knowledge representation based on a technique described in earlier work by Bench-Capon et al (1991). The close association between the knowledge base and the source materials is known as 'isomorphism' Bench-Capon & Coenen (1992). Having a close mapping between the knowledge representation and the source material provides easier control over updating and maintaining the knowledge base. Changes to the source material would determine the changes required in the knowledge base.

2.4.1.5 Integration with hypertext

Knowledge based systems in law usually handle sources that are very large and predominantly textual. For this reason the representation of the legal sources as hypertext provides a natural progression in information presentation and knowledge based system integration. There are two main techniques for integrating with hypertext:

- The resource level where information is referenced and linked to from knowledge based consultations. These systems still separate the navigation through the consultation from the navigation through the legal resources.
- Full integration where the navigation through a consultation is carried out using the same hypertext mechanism used to navigating resources.

There are examples of integrating knowledge-based systems and hypertext. One example of this integration is Soper and Bench-Capon (1994) where animated advice documents are presented to the user as hypertext pages. The use of hypertext mechanism within knowledge based systems has been evident before hypertext became well known by exposure on the Internet. Many document presentation tools already possessed hypertext components such as HyperCard on the Apple Macintosh and the legal knowledge based example, DataLex, Widdison et al (1992).

2.4.2 Tasks in the Legal Domain

The legal domain like any other subject has tasks that are commonly carried out. A number of attempts have been made to classify the tasks with a goal of better understanding the thought processes and operations carried out by practitioners, so that these tasks could be better supported. It is believed that a clearer understanding of the interactions between a practitioner, legislation, cases and other legal resources would provide clearer distinction of the AI methods that could be employed for such tasks.

> *"These tasks are thought to be generic, and applicable in one way or another to any piece of legislation. For all laws are made, giving rise to the policy task, all laws are applied, and so require adjudicators, and it is desirable that all laws should be understood by those whom they apply, so that they can regulate their behaviour appropriately."*
>
> (Marek Sergot 1991)

Although the knowledge is arguably the same, what is done with the knowledge, and hence the appropriate tools, can be very different. Bench-Capon (1991) provides a classification of tasks placing them into three main classes: advice systems for the public, advice systems for lawyers and adjudication systems. There are a number of other miscellaneous tasks that fall outside of these three main classes, but these cover the clear majority of implemented systems.

Figure 2.3 below summarises the key tasks and their relation to legal sources.

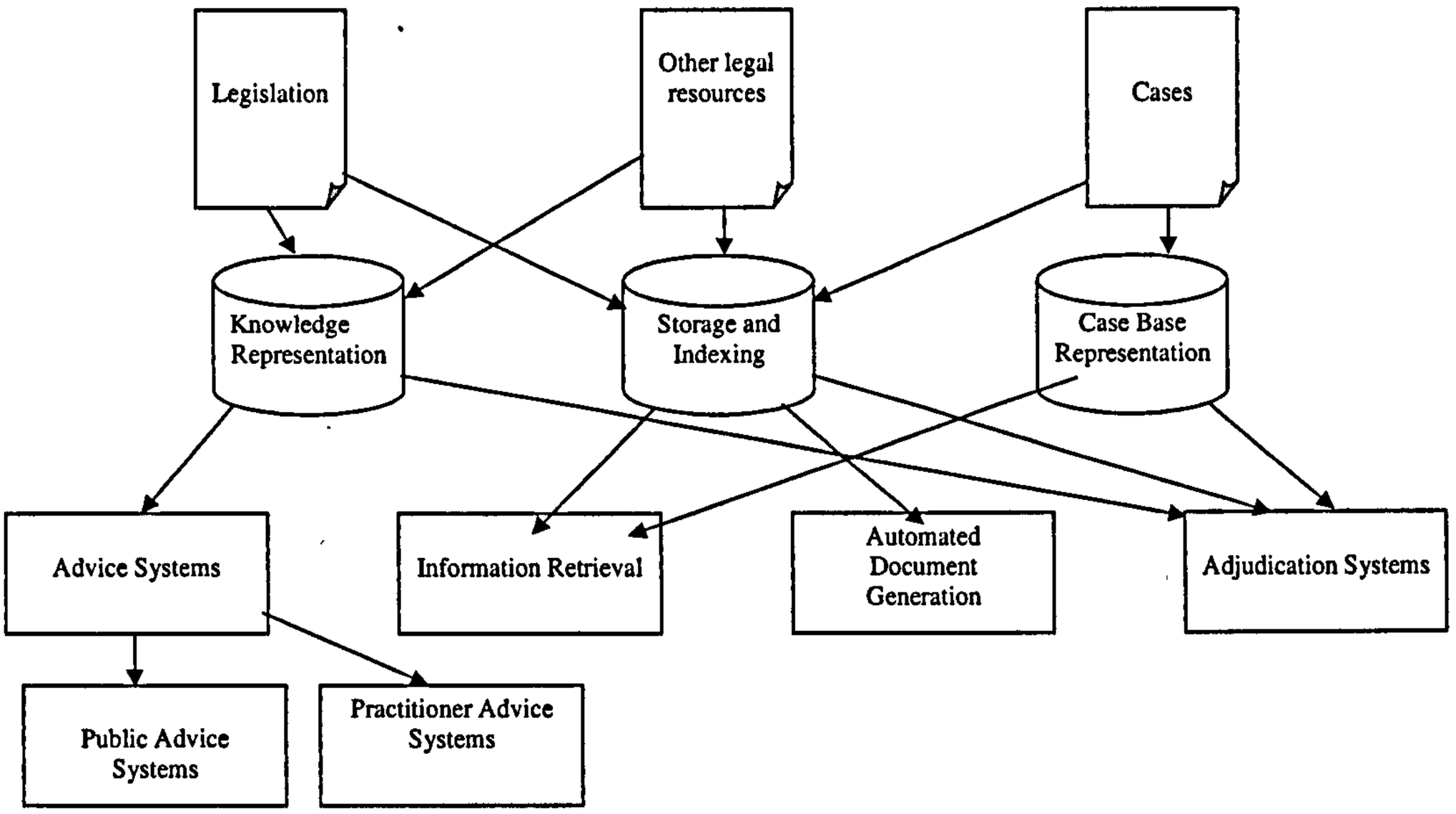


Figure 2.3 Classification of system types

An important task identified as legal service in Susskind R (1996) pp 42 is described as follows:

> *"Legal Service of today is predominantly advisory and consultative in nature. Generally, the lawyer is asked to provide advice in relation to the specific details of a client's case, problem or circumstances. The relationship is advisory in that the lawyer's response takes the form of recommendations for action (or inaction), focusing only on these particular details; and so too with any implementation of the advice (for example, by drafting a contract or commencing a court action).*
>
> *It is relatively rare for the lawyer to be invited to impart knowledge at a more generic level, in the form of reusable legal guidance or information; rather, the advice tends to be more disposable in nature, geared to specific circumstance, and to be relied upon only in the context of these specifics."*

Such observations emphasise the need to be able to provide easily constructed tailored representations.

2.4.2.1 The lawyer and the public

In developing a legal application, the end user must be considered carefully. The domain of law has a very distinct gap between people who are trained in the legal profession and those who are not. A lawyer can be trained to use a system such as LEXIS for information retrieval as it aids in a task that will be useful often. The lawyer would be familiar with the legal terms and devices used throughout such a system. In contrast, a non-legal member of the public would first need to understand how legal documents are constructed and some of the important terms used before useful tuition in LEXIS could be employed and infrequent use makes such training excessive. The results collected from LEXIS would also be easier for a practitioner to understand and put into use.

Applications aimed at public consumption need to be easy and quick to learn. The reliance on understanding law must be kept to a minimum and legal terms if required should be explained in a form that can be easily understood. The contrast between practitioner of law such as a lawyer or adjudicator, and the general public can be summarised as:

The Practitioner

- Understands key legal terms and is familiar with their use and meaning. A practitioner would become frustrated by detailed or simplistic descriptions of legal terms.
- Is directly experienced in key legal issues and may have access to resource collected in previous or similar situations.
- Is qualified to make interpretations and informed decisions based on legal terms. It is also likely that a practitioner would be liable if interpretations and decisions were poorly informed or incorrect.
- Understands the formal format of legal information. The practitioner can understand and make use of the format as intended and documentation can also be created in the required formats.
- Is inclined to undertake detailed research from many different and large sources of information. A practitioner can effectively put retrieved information to use.

- Uses legal applications sufficiently frequently to make training in these applications worthwhile.

The General Public

- Understands few legal terms and requires detailed information in order to make informed decisions.
- Has very limited and possibly indirect experience of legal issues and is therefore not qualified to make interpretations.
- Is less inclined to undertake detailed research due to a lack of resources and information to draw from.
- Due to a lack of experience and training in law, may miss important elements.
- The effort required to train in legal tools is not justified due to the infrequent use of such tools.

There are few intelligent legal systems available to the general public and developing support systems for the practitioner has proved far more popular, perhaps because it avoids some of the difficult issues above. There have, however, been several attempts to produce computer based intelligent systems that help to provide legal support for the general public.

There are several commercial drafting tools available to the practitioner and also the public. The most popular of these tools are aimed at people who require guidance in the creation of wills or other form documents. Such automated tools are also used to improve lawyer productivity and save time undertaking repetitive tasks. One commercial application employed in legal document assembly is HotDocs. The application is described as 'intelligent template software'. HotDocs provides integration of template document assembly with many commercial word processors. By including symbolic references within the templates used by HotDocs an interaction by use of user queries can be performed. HotDocs allows simple word replacement within a document right through to complex integration of document sections based on a hierarchical structure of related templates. Legal applications are usually not aimed at individual lawyers but as a general solution for a law firm.

2.5 Document Tailoring

This work focuses on the tailoring of online documentation, the following section describes the key issues with traditional paper based mediums along with observations of online documentation. Much of this section builds on early work carried out by Sopper P & Bench-Capon T. J. M that looks at coupling expert systems and hypermedia. Much of this work predates the introduction of the Internet yet hints at the possibilities of providing better hypermedia representations of paper publications.

2.5.1 Issues with Traditional media

Traditional publication of information on paper is still the most common way of providing information to the general public. There are a number of issues that can be identified that reduce the usefulness and immediacy of the media in getting the message across. An example domain studied as part of this thesis is the presentation of benefit advice. Below is a single page from the document for claiming child benefit (CH1). This is a good example of many of the shortcomings of traditional information presentation. This document is also available online (see section 2.5.3) in the same format without little attempt to exploit the benefits of a digital media.

Who is entitled ? A customer who is responsible for a child living with them; or

is contributing to the child's maintenance at a rate of at least equal to the Child Benefit payable for the child; and

the customer's right to enter or reside in Great Britain is not subject to a limit or condition, and the customer or the child have been resident in Great Britain for a total of more than 26 weeks in the preceding 52 weeks. A person coming to Great Britain from abroad can claim Child Benefit if he or she:

has been granted refugee status; or

has been given exceptional leave to remain in the United Kingdom; or

is a national, or member of the family of a national of an EEA country; or

is lawfully working in Great Britain and who is a national of Algeria, Morocco, Slovenia or Tunisia, or a member of the family of such a person living with them; or

is covered by one of Great Britain's agreements on Child Benefit with other countries; or

has leave to enter or remain in the United Kingdom which is not subject to any limitation, for example, has been given indefinite leave to enter or remain or the right of abode in the United Kingdom.

Asylum seekers are not entitled to Child Benefit.

An employed or self employed person, who comes under the United Kingdom National Insurance arrangements, can be paid Child Benefit during the first 26 weeks of residence in Great Britain if they intend to stay here for at least six months.

Figure 2.4 online *Child Benefit Claims (1998)*

2.5.1.1 Document Specific

It is important that a document gets the required message across to the reader. There are two issues, detail and overlap that are physical constraints on the layout and volume of information. Detail can cause a loss of interest, distraction and confusion in the reader. Overlap or dependencies on other resources are not always clear until the reader can no longer proceed with the current document.

2.5.1.1.1 Detail

A document is of a finite size and can only include a certain amount of information. Different user characteristics will lengthen the document due to the need to address each of them to a sufficient level. For example, does the reader own their property or are they renting it? If they are renting the property then do they rent from a private landlord or their local authority? If they are renting from a private landlord, do they pay over £X per month for the property?. It can be seen that the volume of information is dependent on the different characteristics that apply to the total audience of the document. An individual has to accept a document that is longer than necessary due to the 'completeness' of the document in addressing the characteristics of the whole intended audience. It is also the task of the reader to identify which sections of the document are relevant to them. The example in Figure 2.4 is an attempt to clearly identify reader's characteristics, a group of options provide a means to quickly determine validity to claim child benefit. Many documents do not provide such a clear distinction.

2.5.1.1.2 Overlap

Due to the size of most advice domains, the information is typically split at logical bounds into several documents. These separate documents can be combined in different combinations to provide the reader with their complete advice. Advice on tax depends on an understanding of benefits such as retirement pensions. Claiming income benefit will require some understanding of income support. These separate subjects are available from different information sources and should be available to the reader for them to have a complete understanding of the domain. The dependencies on other documents are not always clearly defined until the reader has reached a point where they require another resource. It will then take time to find and read these other documents to get a clear understanding of all the factors effecting their situation.

2.5.1.1.3 Media Specific

There are physical constraints on the media used to present the information to the reader. Distribution is important for advice documents because they have a very wide

and diverse audience. Updating information that has such a large circulation can also become an important and difficult issue.

2.5.1.1.4 Distribution

The current method of accessing benefit advice leaflets is to collect them from Post Office counters or from benefit centres. Sometimes information is sent through the post to the claimant. In most cases the claimant has to find all the necessary documents before they can begin to understand their situation. The recent publication of advice online has now provided a new means of getting the information to the people that need it. Internet access is available from home, libraries and social areas, and there are few reasons why such information is not made available through interactive television.

2.5.1.1.5 Up-Dating

Laws and benefit advice change frequently, with paper leaflets updating the information is difficult. Many advice leaflets have expiration dates but these can be misleading as changes may render them out of date prior to the anticipated documents expiry date. Also a leaflet may be correct after the expiration date yet the reader may lack confidence in its use. There is also an inherent lag in updating paper based resources. Old versions may still be available when they should be replaced by new versions and readers may not be aware that changes have been made.

2.5.2 Reader Specific

There are a number of issues that are specific to the users characteristics, their situation and their capacity to understand and interpret that document they are reading.

2.5.2.1 Personal Circumstance

Each person is unique but advice documents are written to accommodate a very large audience. This mismatch results in documents where large sections are not relevant to the reader, and the converse is also true in that there will always be individuals that are outside the scope of the documentation.

2.5.2.2 Use of information

Once the reader has collected all the relevant information it is important that they can successfully put that information to use in filling out forms, submitting claims or applying it to their query to reach a solution. Unlike a human advisor, static documents cannot provide adaptive advice; for example what to do if a person has a name longer than the cells provided to enter the information. Is it suitable to write over the end of the line or should their name be shortened? A reader may inadvertently enter incorrect information resulting in a delay before their situation is resolved, or may misrepresent their situation. In the best case they may have to fill in sections of a form again, in the worst case a mistake may not be immediately evident and the reader may incorrectly be denied benefits they have a right to claim.

2.5.2.3 Thesaurus

Words have a meaning and in a legal context it is more important that the reader is aware of the precise meaning of a term. For example the terms "income related", "dwelling" and "habitually resident" may be better explained using terms such as "means tested", "building" and "normally living in the UK". While providing advice, it is essential that the reader understands the correct meaning.

Converting these traditional formats into an electronic version does not immediately address all of these issues, it is required that other techniques are identified and employed to help improve the presentation of information in parallel to the developments a research into search and retrieval. These issues were further extended in Royles & Bench Capon (1999) when investigating presentation of advice over the medium of the Internet.

2.5.3 Issues with online information

The Department of Social Security now provides information online at its WWW site http://www.dss.gov.uk. All of the paper-based advice leaflets have been converted to an online version. This migration to the WWW does not address the initial issues but distribution and maintenance is now improved. Other groups also provide online information including the Citizens Advice Bureau, Local Authorities and independent groups such as Accountants, Lawyers and interest groups.

Each Internet site is likely to be authored and maintained by a different person. Because of the large number of formatting mechanisms available the structure and layout of sites can vary greatly. In the legal domain there are formatting standards for documents, they have a fixed structure and information specific to a document can be found where expected. Due to the physical layout differences between paper based formats and online formats and the use of graphical or aesthetic elements in sites the use of these formatting standards is being diluted. As such, there is little consistency between elements such as content pages, indexes or navigation.

In traditional media the search mechanisms such as citations, references, library catalogues work well and their use is well known. On the Internet each sites search mechanisms if available at all vary greatly in their reliability to identify the information a user is looking for. Referencing online information also provides other issues such as availability of the material due to network outages or performance, locations on the internet are rarely static over time so it is difficult to rely on information staying where you expect to find it. There are no standards in formatting the location of information, the location is usually constrained by the address of the host machine. Also the path to the files have no bearing on date and location of publication, author, or title of the document. These constraints add another layer of abstraction between the information and the user trying to find the information.

In many cases documents have simply been converted directly into an online equivalent such as a PDF file. These forms of information add little value other than the distribution mechanism to the information. The same problems as paper based documents still apply and is further emphasised if the document is printed. If the document has been generated in HTML then it is very unlikely that layout standards have been applied, the information is usually split over different pages, hyperlinked and shoehorned into the sites look and feel.

2.6 The World Wide Web

The WWW has seen massive development since its inception in 1986, and today the Internet provides a vast resource of information on every topic imaginable. The legal domain is a subject that has a very wide network of successful and established sites. The Internet simply provides a vehicle for publishing huge amounts of information in

a format that is easy to access, easy to browse and rich in its presentation capability. Unlike other resources WWW based information can combine verbal, visual and textual elements. As well as providing the information in a form where it can be accessed, and interpreted, the information is also in a means that can be easily reproduced or stored for later reference. All these advantages result in a publishing network that cannot be surpassed by traditional means.

The Internet relies on the storage, transfer of and presentation of information across a very large network of connected machines. It is important that each machine can understand the format of the information being transferred. If information is not in the correct format to process then an application must be installed for the user to make any use of the information. The key components of the Internet are the physical network, the Server and Client software, the information formats and the transfer protocols. This section will examine these separate components and how they are used to facilitate intelligent information access.

2.6.1 A channel of communica tion

There is a growing trend in IT to use publication technologies to provide many parallel means of interaction with an end user. Service groups such as banks have recently started to provide these different means of extending their services to the consumer. Several years ago the bank service was augmented by automatic teller machines so that the customer was then able to withdraw and deposit money outside office hours. More recently information about a users account has been available through phone services. These systems have developed to provide many more functions such as fund transfers, bill payments through to financial advice and the ability to purchase new services. Within the last year (1999) the phone channel has been extended to provide Internet banking. A user can now access their account details and carry out many tasks through the interface of the web. This interface has now become available through interactive television further extending the number of channels through which a user can interact with a service. These different enabling technologies provide the user with the ability of carrying out tasks whenever and however they wish.

2.6.2 Languages

The Internet has spun off several standards that provide a base for the development and integration of new technologies. Possibly the best known WWW language is HTML but there are several other mark-up languages and document formats for storing and transferring information online.

2.6.2.1 The WWW glue

If the network infrastructure is the medium to replace paper, what replaces the pen? During the early development of the Internet the language for creating content and developing pages, Hypertext Markup Language (HTML) was a very important element. The language at that stage was very simple, uncluttered and designed for placing text and pictures onto the screen. As more people created online content pages were constructed that pushed HTML capabilities and stretched its limitations. This approach to creating online content has changed dramatically, now very few people 'write' HTML it is now more commonly generated by WYSIWYG applications. These applications have evolved from initial experiments such as that in Punin et al (1994) through to more accomplished examples such as Macromedia's (http://www.macromedia.com/) DreamWeaver and Softquad's (http://www.macromedia.com/) HoTMetaL. The language has become more complicated and functions have been added to meet the expectations and reduce its limitations, HTML 4.01 Specification (1999). Using additional document formats such as Postscript or RTF can supplement HTML mark-up presentation, although such formats require the use of additional applications to format them for presentation and lack some advantages of HTML such as hypertext navigation and portability.

Developers of Internet Browsers have extended the original HTML. Netscape extended HTML to provide provision for dynamic visual elements and the inclusion of scripting languages such as JavaScript allowed the content developer to control interactive elements within the WWW page. This extension termed Dynamic HTML (DHTML) has grown and is now better defined and widely used.

2.6.2.2 Style Sheets and Dynamic content

Cascading Style Sheets (CSS) were introduced into WWW mark-up at a similar time to DHTML to provide standardisation over page properties such as fonts, margins, layout and positioning. A style sheet is a definition of how tag attributes should be represented visually by a browser. The intention behind style sheets is to provide a single style sheet representation for many HTML documents, in this way the look and feel can be standardised across an entire site. Cascading Style Sheets provide the ability to hierarchically inherit from different CSS files so a more and more refined representation can be provided for each subset of pages. The presentation of a page can be further controlled by the use of scripting languages such as Netscape's JavaScript (http://home.netscape.com/eng/mozilla/3.0/handbook/javascript/index.html) or Microsoft's Visual Basic Script. These scripting languages provide control over elements of the page so dynamic and interactive content can be created. As well as providing control of the browser using scripting languages, more functional systems are available such as:

Java – A complete programming language in its own right, Java (www.javasoft.com) provides a means of running a program within a web document.

ActiveX – ActiveX are small components that can run within Microsoft applications such as Windows and Internet Explorer.

Flash – Flash is a scripted graphical presentation system developed by Macromedia. (www.macromedia.com) Entire pages and sites can be constructed in Flash which when accessed are displayed within the browser in a form similar to an animation.

2.6.2.3 Protocols

The Internet is dependent on a protocol used for transporting WWW documents over the network. Hypertext transfer Protocol (HTTP) is the common language the WWW servers and browser talk in order to find, access, download and present HTML content.

HTTP is a stateless protocol, which means that it does not maintain a constant connection between the client application and the server. Each transaction uses a new connection, after each transaction the connection is closed. When downloading a single WWW page, several parallel connections may be opened and each image or object within the page will be downloaded separately in an individual transaction see Figure 2.5 below.

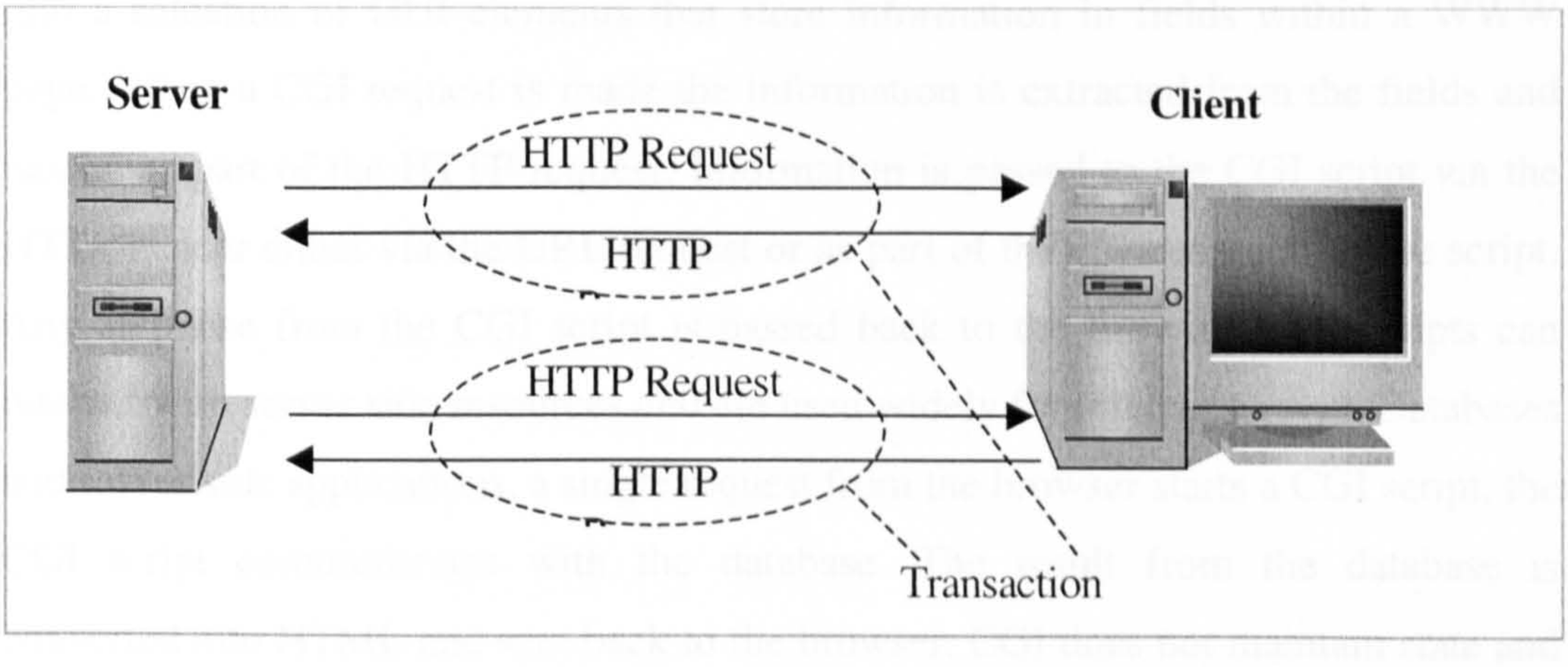


Figure 2.5 HTTP transactions

If a hypertext link is selected, this will invoke a new transaction independent from the last. This stateless interaction has important limitations in AI applications that will be discussed in section 2.7.1.

2.6.3 Applications

2.6.3.1 Servers

Servers are a very simple concept. They receive a request for information and then provide the information. A simple WWW server parses an incoming HTTP request header, extracts the path to a file, opens the file and then transmits it back through the network connection established for the request. The server then closes the connection and awaits another request. WWW servers become more complex and

interesting when they are used to process information and automatically generate documents.

The Common Gateway Interface is the means by which dynamic information can be provided to the server and a dynamic response can be created in reply. Information is passed to the server as part of the HTTP header, and within the header is the name of the CGI script to invoke on the server. HTML supports the provision for sending information to CGI scripts by the use of forms which allow information to be entered into a selection of GUI elements that store information in fields within a WWW page. When a CGI request is made the information is extracted from the fields and passed as part of the HTTP request. Information is passed to the CGI script via the HTTP header either via the URL request or as part of the environment for the script. Any response from the CGI script is passed back to the browser. CGI scripts can interact with server side resources and are used widely for interacting with Databases and server side applications, a single request from the browser starts a CGI script, the CGI script communicates with the database. The result from the database is converted into HTML and sent back to the browser. CGI does not maintain state and each query is handled independently. No information is available across requests other than that sent from and to the browser or stored on the server.

2.6.3.2 Clients

For most WWW tasks, a browser application is used such as the popular *Microsoft* Internet Explorer or *Netscape's* Navigator. Essentially a client such as a WWW browser reads a document format such as HTML or Acrobat and converts it into a visual representation. The current tasks of browsers are now far more demanding and can include support in playing multimedia such as movies sound or to animate web pages by providing scripted control of the content. Most common browsers also contain a Java Virtual machine to provide support for Java Applets. To provide flexibility for application developers, browsers support a number of different means of extending the browsers functionality.

Application Programming Interface (API) – this interface provides a means for external applications to interact and in some situations take control of the browser. The browser manufacturers provide a set of function calls that an external program

can make. These calls provide very close integration between the browser and the supporting application. A disadvantage of using the API is that it is unique between browsers and sometimes across hardware and OS platforms. The supporting application will be browser and platform dependent.

Helper Applications – A helper application is a program such as Adobe's (www.adobe.com) Acrobat that can be loaded to display a format not supported by the browser. If the user chooses to download an Acrobat format document then the content will be downloaded and then passed to the helper application for display. There is very little connection between the browser and the helper application. The document is usually passed to the supporting application through the command line and both applications then work independently from each other.

Plug-in – Most browsers support some form of plug-in, which is a small application that uses an API but once installed is considered part of the browser application. Plug-ins are usually employed in presenting content such as movies or sound that are displayed inline with the document content. A good example of a plug-in is Macromedia Flash which provides a window within the document; the plug-in downloads and plays a Flash format movie within the window. The plug-in thus extends the capability of the browser. Many plug-ins are available for download from the Internet and many browser companies have made alliances with plug-in developers such as Microsoft and Real Media so that plug-in technologies can be packaged with the main browser application.

Java Virtual Machine – The Virtual Machine allows Java applets to be downloaded and then executed within a secure area known as the sandbox. The sandbox protects the host system from malicious code that may be downloaded. Java applets usually run inline in a window that limits their area of effect. Different browsers can interact with Applets in different ways. Netscape provide a means for Java Applets to give and receive messages from JavaScript programs. There are also means of extending the capabilities of Applets by employing signed applets which carry the signature of their creator thus instilling a degree of trust between the creator and the user. Signed applets have extended capabilities allowing access outside of the sandbox which can include file IO, and enhanced network capabilities.

2.6.3.3 Intermediaries

Intermediaries are applications that sit between a browser and a server and are able to act as a server or client depending on which party they are communicating with. Every transaction carried out over the Internet travels through several different machines as it is routed to its destination. Most of the machines that the transaction passes through only forward the message on to another machine, but there are some machines that process the message. This processing can be used to carry out several tasks such as security by limiting access to certain services, caching of files to help decrease network latency, the generation of documents to provide help or warnings and possibly the most interesting, the dynamic alteration of documents. The advantage of an intermediary is they can be located on any machine within the network path. The use of intermediaries located on the user's machine to implement intelligent Internet interaction will be discussed further in Chapter 3.

2.7 Expert Systems on the WWW

Looking again at the capabilities and advantages of expert systems discussed in section 2.3 it can be seen that the Internet applications can benefit from the use of an expert system approach. Many different users each with their own goals and characteristics can access a site on the Internet at any time of day, multiple interactions can occur simultaneously. It is important that each person receives correct information that is consistent and immediate

2.7.1 The Main Issues

There are a number of issues and restrictions that apply to providing expert and knowledge based system applications over the Internet.

2.7.1.1 Maintaining State

Expert systems rely on maintaining an active session with a user, asking several questions before a conclusion can be reached. It is important to maintain state during an expert system interaction so that dynamic rules and facts about the user can be carried over to the next interaction. There are two main methods of maintaining state between a browser and server I have additional references to add here.

The first method uses hidden fields within a WWW page that can be used to store unique information about the current session. As information is passed backward and forward between the browser and the server this state information is kept up-to date. Information can also be stored in a server side database identified by unique indexes or username / password combinations. The problem with sending state identification through hidden forms is that the WWW page can be saved and state information examined, and it may then be possible to manipulate this information and gain access to data from other states. State information passed to the browser is usually encoded to reduce the risks of manipulation like this occurring.

The second method of maintaining state is to use cookies. A cookie is a very short message that can be stored on the client machine. The main purpose of cookies is to uniquely identify the client browser. In order to maintain state the cookie information can be queried and altered by the server. Cookies are linked to a particular URL and can only be accessed when that URL is identified. A cookie set by the server www.foo.com can only be accessed by that server or its sub domain. Like hidden fields cookie information can be examined and altered on the client machine, and so for secure applications the data should be encrypted.

2.7.1.2 Security and Trust

Many expert systems especially those in the legal domain require interactions that contain user confidential information. How can online expert systems convey a sense of confidentiality and provide security of the users information? It is not enough to protect information in transit, for example by encryption; the end user must also be confident that their personal information is not being collected by the remote party so that it could be used in the future against them. In the domain of Social Security the information that would need to be acquired from the user can be very personal. The initial concern is the protection of the personal information, the second concern is the perception the user has of where their information is going and who would then have access to it. If the user thought that their information was being used to confirm their entitlement to claim then it would be possible that the user may provide false or incomplete information to provide a better set of characteristics. Such a response would be misleading to an AI system resulting in incorrect and poorly informed

advice. This lack of user confidence in Internet transactions must be overcome in order to provide sound advice to a user.

Providing correct advice is also an issue if the source material is not reliable. There have been a number of high profile cases where false reports on Internet sites have had an impact on company share prices. The spread of such advice is very fast and many people respond to the information without confirming that it is valid. If advice is provided it must be possible to authenticate the source of the material whether it is the DSS or an independent party. Currently all legal resources and examples of WWW based expert systems carry warnings explaining that the information they supply should not be relied upon and the conclusions reached should not be trusted. These declarations are there to provide protection from litigation if incorrect advice is provided by the AI system. These forms of declaration show a lack of trust in the advice provided from the point of view of the developers and does not send out a promising message to end users of such a system. Such service providers need to protect themselves if incorrect or misleading advice is provided but it is still necessary to confirm advice given with a human advisor or council before taking action on it. At what stage is advice considered final and where does the responsibility lie if things go wrong?

2.7.1.3 Presentation

The successful use of the Internet and hypertext media is very important when developing an intelligent interactive system. The WWW provides many rich forms of information presentation, so we must decide how these forms of presentation can best be exploited by intelligent WWW systems. Many legal applications successfully use hypertext as a means of presentation to the user. Linking to citations, related topics, definitions and other resources is well suited to the legal domain.

Internet mark-up provides a rich palette for presenting information in many different formats. It is also very easy to generate or tailor HTML dynamically to increase the flexibility of the presentation provided to the end user. Many information providers allow a user to select the services that they wish to receive, so those sources can be added, removed or tailored at will. In many cases, due to the volume of information it is not practical to represent every page as a separate HTML document while

storing it as such on the host machine. This would result in a large amount of duplication, as every file would include the same formatting and header/footer content. Many online resources that present large volumes of information rely on database technology to store and dynamically generate HTML documents when and in the format requested.

2.8 Tailoring of documents

2.8.1 General purpose systems

There are a few publications available about WWW based inference in the legal domain but there are a number of examples of more general solutions. A simple but effective WWW based expert system is the Whale Watcher expert system. This site provides the user with an interactive expert system that helps classify a whale by its individual characteristics. Although the inference and knowledge base are simple, the deployment of such a system as far back as 1996 demonstrated how such interactive systems could be deployed over the Internet. Many of the examples of expert systems available over the Internet are commercial products and there is limited published material about these products.

2.8.1.1 Exsys (Wren) (http://www.multilogic.com)

The group Multilogic provides a commercial development tool called Exsys (originally known as Wren) for developing WWW based expert systems. A number of legal examples of this expert system used to be available online but these are no longer accessible.

Exsys employs a simple IF-THEN-ELSE rule structure that can be developed using a visual editor. The inference engine in Exsys provided backward and forward chaining inference as well as support for certainty factoring and fuzzy logic. The applications developed for Exsys can then be deployed over the WWW providing all the functionality of networked access and HTML hypertext. Exsys relies on CGI interactions to communicate with the expert system shell. There is a large volume of development documentation and a number of different resources available for

developing knowledge based solutions. Exsys is possible the most promising of the WWW capable AI solutions current available.

Exsys is based on the classic CGI server, client model that can result in slow access times and broken sessions. The user has little control over the information they provide, they can either choose to send it to the expert system trusting that the information will be protected or they do not enter the information so may get to stages in a consultation where they are unable to proceed. Exsys is a well-developed expert system with many powerful options for knowledge representation and inference including fuzzy logic and certainty factors. Exsys also includes a full graphical interface for the creation of rule bases.

2.8.1.2 WebLS (http://www.amzi.com/WebLSManual/)

WebLS was created in 1996 Sehmi & Kroening (1996) as a system for providing Internet based problem solving and advice. The goal of the system was to create a tool for web developers without having to rely on the specialist skills of a knowledge engineer. WebLS uses CGI style forms to collect information about a user and then construct a HTML document based on this information. WebLS uses simple IF-THEN rules to determine how to generate the HTML representation of the document using an outline to provide the skeleton. As with many CGI based solutions, WebLS had to address issues such as maintaining state between interactions, providing a concise way of asking multiple or related questions, providing a natural integration with the current Internet technologies and provide a easy to learn and use knowledge representation language. All presentation to the user is achieved through HTML documents and forms, the inference engine was created using a backward chaining Prolog, Merritt (1996) implementation.

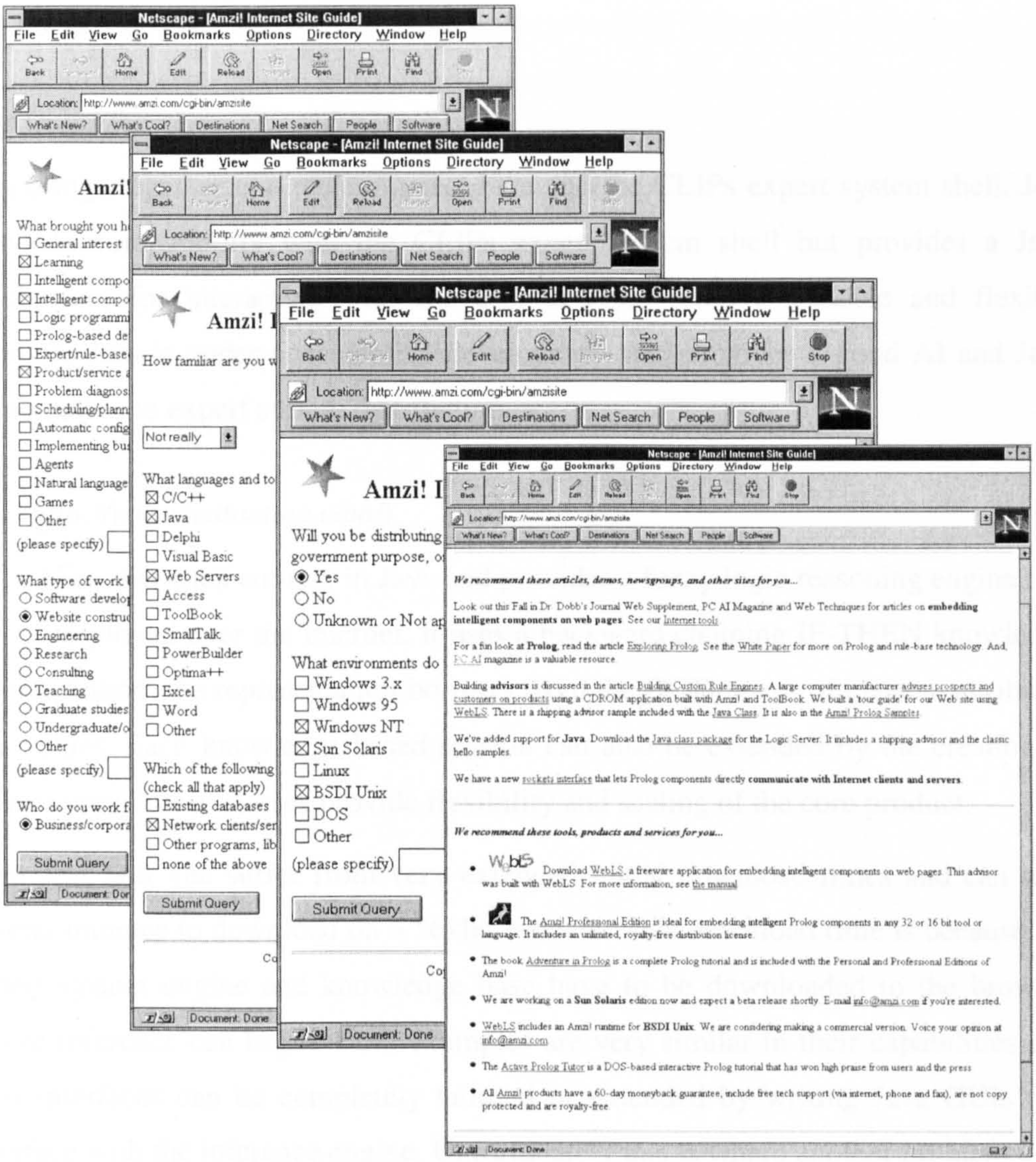


Figure 2.6: Several stages in a WebLS consultation

(images reproduced with permission)

WebLS is possibly the closest example of any solution on the Internet that provides a similar result to that proposed in this thesis. Yet WebLS relies heavily on the CGI mechanism that can result in issues similar to those for Exsys. Moreover the system we propose is much more directly related to the original information sources that provide the definitive versions of the information sought.

2.8.1.3 Jess & Tea

Jess (http://herzberg.ca.sandia.gov/jess/)

Jess brings together two technologies, Java and the CLIPs expert system shell. Jess provides compatibility with the CLIPs expert system shell but provides a Java mechanism for interacting with the user and providing a scalable and flexible solution. Jess is under considerable development and provides a good AI and Java combination to expert system distribution.

Tea (http://www.instanttea.com/)

InstantTea is written entirely in Java and provides a fuzzy logic reasoning engine that can be deployed over the Internet. It uses a backward chaining IF-THEN knowledge representation, the representation borrows from the OOD terms of classes, attributes and values. Each knowledge-based session can also be extended by the creation of Java classes. These classes provide flexibility and scaling of the core product.

Both Jess and Tea suffer from very extended initial download times and can take several minutes to download on a 56kbit modem. This download time is because the expert system engine and knowledge base have to be downloaded to the browser before inference can begin. Both examples are very similar in their capabilities and their interfaces can be completely tailored or extended by writing Java GUIs that interface with the inference engine. Unfortunately this is where another problem with Jess and Tea arise, in order to create a full interface a new Java front-end needs to be created. Interfaces are not generated using HTML and the browser is used simply as a run time environment rather than making use of its HTML and presentation capabilities.

2.8.1.4 Whale Watcher (http://www.aiinc.ca/demos/whale.html)

The Whale Watcher site is an entertaining and educational example of an online expert system. The Whale Watcher expert system uses ACQUIRE which is a web based expert system shell in order to provide a consultation aimed at identifying different types of whale.

Like the Exsys example, the whale watcher suffers from the limitations of CGI interactions. There also exists a Java version of this expert system implementation but this also suffers from poor download times. The Whale Watcher/ACQUIRE example are however very comprehensive and make good use of media components to improve the user to expert system interface.

2.8.1.5 ACQUIRE (http://www.aiinc.ca/products/acquire.html & /articles.html)

ACQUIRE is a full knowledge engineering and expert system delivery package that provides a suite of tools for constructing knowledge bases and delivering them in a number of different ways to a users browser.

> *"ACQUIRE® is a software package that can be used by non-programmers to build expert system applications. Its two major components are: first, a knowledge acquisition system which lets an expert construct knowledge bases for use in expert system applications; and second, the inference engine to run knowledge bases on particular cases (i.e., situations in which the captured knowledge is applicable). ACQUIRE® also provides facilities to customise data entry forms and session reports that provide details on case runs."* http://www.aiinc.ca/products

The ACQUIRE product is very well presented in a similar manner to Exsys and there are a number of demonstrations available over the Internet.

2.9 Summary

This chapter has provided an overview of all related technologies and has presented examples of similar work and systems that have provided inspiration and also shown where limitations of Internet presentation and Expert system control over the Internet lie.

Section 2.2 provides a discussion of the legal resources that will be investigated in more detail later in this thesis. Giving an introduction to the Benefit advice documentation and the Convention for the International Sale of Goods. Distinctions between the two groups who access legal information, the general public and practitioners of law are also made.

Sections 2.3 & 2.4 proceed to give and overview and then examples of Artificial Intelligence. A number of example systems are described that are used for general solutions and then a number of examples specific to the legal domain are described.

Sections 2.5 & 2.6 focuses on the Internet technologies available that provide access to information and the components and relevant standards. Issues with the technologies are also discussed to provide an indication of the complexity of successfully delivering expert system solutions via a WWW interface.

Section 2.7 provides an overview of existing examples of WWW based expert system giving a brief description of each system, its benefits and issues related to their design and implementation.

CHAPTER 3

SYSTEM ARCHITECTURE

3 Background and Implementation

This chapter discusses the implementation of an Internet based intelligent document tailoring system (QERIS). First related areas will be discussed to provide a context for the system within current mark-up, AI and Internet technologies. The chapter will then discuss the internal architecture of the system from its method of inference to the document tailoring approach.

- QERIS relies on three key components to provide a means of tailoring Internet documents in an intelligent manner:
- An extension of HTML that adds support for a knowledge base representation that defines how the document should be tailored, the effects of inference and the style of interaction with the user.
- A proxy server that provides a mechanism for interrogating Internet content for knowledge representations and executing the tailoring of the document. The proxy server will be located on the client machine and operate alongside the browser.
- An inference engine that interprets the knowledge representations retrieved, activates specific tailoring of the document, interacts with modules that control the type of tailoring and provides real-time interpretation and feedback.

3.1 Document Mark-up

A mark-up language is used to represent disparate media on a single display as a collection of associated elements. A mark-up consists of a set of tags that identify sections of a document and gives the sections specific attributes. This identification of document sections can instil underlying structure, meaning and representation to the document. Mark-up languages date back to typesetting applications such a TeX, (Knuth 1984), developed by Donald Knuth in the 1970's as a means of representing complex documents before the availability of WYSIWYG word processors. Mark-up languages have a number of advantages over binary files, in that they provide greater portability and interoperability as they can be defined using a standard character set. Mark-up tags can also be used to define associations and hyper-linking between relevant resources.

3.1.1 Hypertext Markup Language

HyperText Mark-up Language (HTML) W3C html4.01(1999) was developed to meet a need for a distributed information sharing and management system and was originally developed at CERN. In the eight years it has existed, HTML has seen major development, revision and growth. HTML has undergone four major revisions since its inception. There is a core set of tags that can be used to mark up a document and deliver the content as part of the collected hypertext of the Internet. A simple example of a HTML document follows:

```
<HTML>
  <HEAD>
    <TITLE>A Simple HTML document</TITLE>
  </HEAD>
  <BODY>
    <H1>Introduction to HTML</H1>
    <IMG SRC="pic.gif">
    <P>
```

```
        This is a demonstration of how
<EM>HTML</EM>
        documents are represented. This
        Mark-up language will be familiar to
users
          of the Internet.
        </P>
        <UL>
          <LI>HTML is simple to use.
          <LI>HTML is portable.
        </UL>
        <A
HREF="http://www.csc.liv.ac.uk/">home</A>
      </BODY>
</HTML>
```

The above example results in a graphical representation of the document, which can be viewed by using a browser. The result of the conversion from HTML into a graphical format is application dependant, and different results occur for different HTML browsers. An example representation can be seen in Figure 3.1 below.

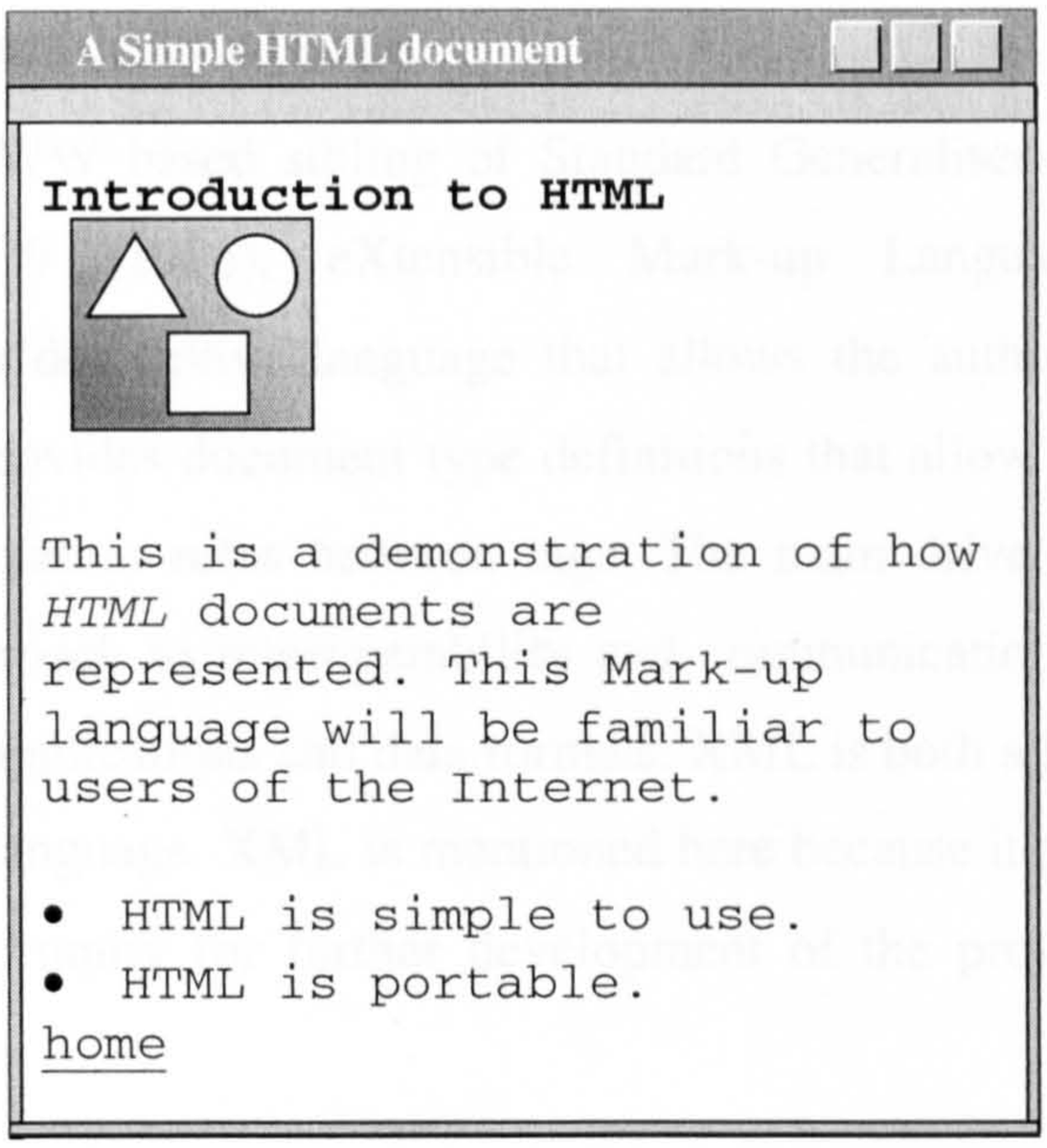


Figure 3.1: *HTML rendered in a browser*

In HTML tags are delimited by an opening character '**<**' and a closing character '**>**'. There are two types of tag, an inline tag and a block tag. An inline tag such as `<IMG SRC=pic.gif>` has no content and no closing tag. A block tag such as `<H1>Introduction to HTML</H>` has an opening and corresponding closing tag that surrounds a segment of content to be influenced. For example the <A HREF...>Link</A> tag allows the implementation of hyper-linking between documents since the enclosed section of text is displayed as an active area that can be selected by the user and is associated with a URL. HTML tags have a position and area of effect within the document, for example, an image will appear in the right relation to its surrounding document sections. Other tags such as <TITLE> affect external elements of the document such as the applications title bar. HTML is important for this project because it provides a flexible approach to representing information. HTML is also a standard used throughout the Internet providing an essential foundation for content storage and presentation. HTML tags can also be added, removed or altered to provide many different representations of a single document.

3.1.2 Extensible Markup Language

Developed as a WWW based sibling of Standard Generalised Mark-up language (SGML) ISO 8879 (1986), eXtensible Mark-up Language (XML) W3C XML1.0(1998) is a descriptive language that allows the author to structure their documents. XML provides document type definitions that allow the development of new tags and association rules between tags. The main drive behind XML is to develop a new approach to interoperability and communication between disparate operating systems, applications and data formats. XML is both a human readable and computer readable language. XML is mentioned here because it provides support for this work and opportunity for further development of the project as discussed in section 3.9.

3.2 Internet Connectivity

The Internet is now a well-established method of presenting information to a wide audience. There are standards that provide interoperability between applications and resources. (see Figure 3.2)

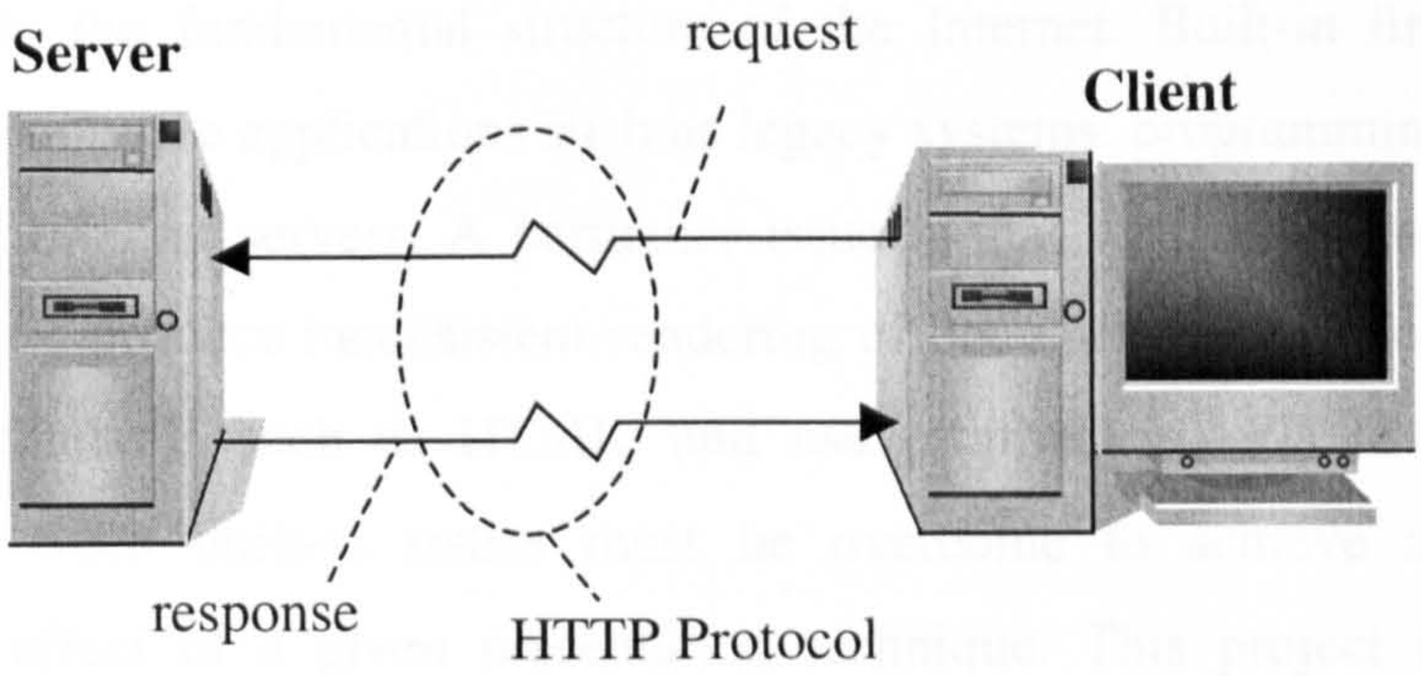


Figure 3.2 Client/Server architecture

The Internet relies on a network association between server and client applications. Information is transmitted using standard protocols such as HTTP and FTP. If a user wants to access an Internet resource they use their browser to request that resource from a remote server. The browser constructs a header that tells the server what information to send as a reply and the address to send the information to. The server then collects the information from its file store and sends the information over the

Internet to the client. The client then interprets this information and presents it to the user via a browser application.

3.2.1 Internet Limitations

3.2.2 Software Limitations

Standards such as HTTP and CGI, languages such as Java and applications such as Netscape and Internet Explorer have defined limits on the development of new tools. For a new concept to succeed it must integrate well with the current technology while also providing capabilities beyond what is currently available. The Internet is limited by a number of physical and also "built-in" limitations that reduces the ease of introducing a new approach to the field.

3.2.3 Physical Limitations

Limitations such as network latency, restricted bandwidth, variations in operating system and hardware capabilities introduce a physical limit to some designs. Many of these physical limitations will be reduced over time although they will always be evident due to the fundamental structure of the Internet. Built-in limitations are dependent on software applications such as legacy systems, programming languages, browsers and Internet servers. A particular issue is that well established and wide spread browsers produce inconsistent rendering of documents, introduce divergence of Internet standards such as HTML, and use proprietary APIs to extend their functionality. Such built-in issues must be overcome to achieve a stable and reproducible effect of a given presentation technique. This project identified an approach that limits the effect of these limitations and provides flexibility allowing adaptation to future standards and requirements.

3.2.4 Common Gateway Interface (CGI)

CGI provides a means for applications to be invoked from a browser that reside on an Internet server. A good example of a CGI application is an interaction with a database. A user can input information into a form and submit the information to the CGI script. Scripts can be written in a variety of languages and can interface with other server side applications such as legacy databases or expert systems. A script

usually takes its arguments from environmental variables or from the command line. The script returns an HTML document containing the additional resources provided by the server side applications. CGI is designed as a one-shot approach to client/server interaction; one form is submitted and one document is returned. This approach can be restrictive if an ongoing session needs to be maintained. In some applications such as expert systems, answers from earlier questions must be maintained and applied to future interactions. This problem can be overcome but imposes overheads in server side storage and client/server communication.

3.2.5 Proxy Servers

To augment content transmission over the Internet, proxy servers were introduced to provide access and control of the content between its source and final destination. (see Figure 3.3)

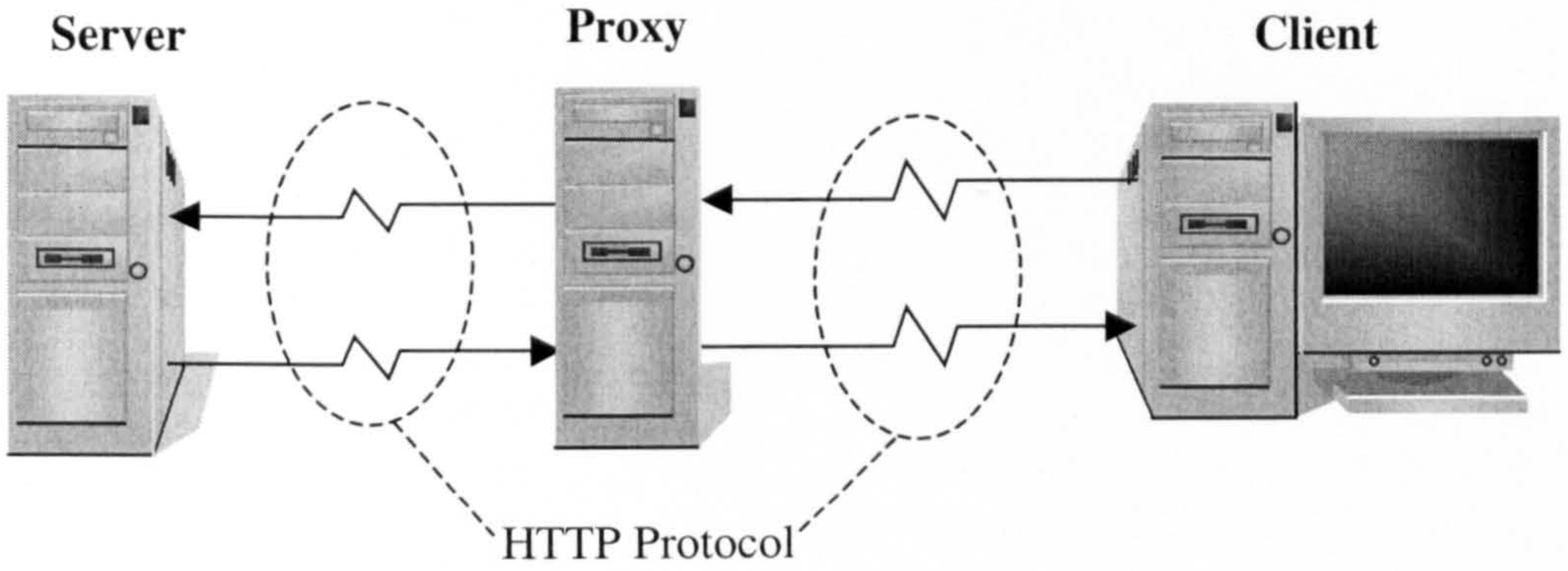


Figure 3.3 Client/Proxy/Server architecture

Proxy servers have the ability to communicate between both server and client applications and can be installed anywhere in the network path. A request for a resource is made to the proxy server including the address of the server where that resource can be located. The proxy then requests the resource from the server and transmits the result back to the browser. During the outgoing request and the returning reply, the HTTP headers and content can be examined and manipulated by the proxy.

Proxy servers are used for a variety of applications where access to data during transit is required. Examples of Internet proxy servers include:

- **Firewalls:** Firewalls are used to control transmission of data through authorised TCPIP ports and ensuring only recognised protocols are used. Firewalls are generally used to protect sensitive information on a LAN from malicious access by external groups. Communication through standard or high-risk ports such as telnet and ftp can be blocked completely by removing the ability of an outside agent to access machines on the network inside the firewall.

- **File Caching:** To reduce the latency in downloading Internet content regularly accessed elements can be stored in closer proximity to the destination of the content. Many browsers provide client caching, which can be supported by a central cache facility provided by a caching proxy server. Companies and academic establishments may employ a local cache to reduce the amount of requests external to the organisation and decrease the apparent latency of those connections. Cache proxies are also used throughout the Internet, employed as part of the Janet Cache (http://www.wwwcache.ja.net/). There is an annual International Workshop, WWW Caching Workshop that covers all aspects of Internet cache technologies. There are many other Internet based research projects that are exploiting the capabilities of proxy servers as an intermediate location on a network.

Using a proxy server helps to address a number of issues. Proxy servers are independent applications that communicate with servers and browsers using standard Internet protocols such as HTTP and CGI. The only requirement for successful operation is that the server generates standard HTML for the proxy to interpret and the browser is compatible with the HTML generated by the proxy. HTTP has maintained its standard; the protocol still supports the old HTTP/1.0 and now the more recent HTTP/1.1 recommendations providing a single and well established interface. There is no need to meet varying API requirements when communicating directly with Internet applications or designing under the constraints of security issues with Internet capable programming languages. By implementing a proxy server in a platform neutral language a solution can be achieved that is OS, browser and server independent, because the proxy is free to communicate with other resources it can query databases, push content to browsers, collect documents from Internet servers and deliver in-page networked applets.

There are, however, a number of disadvantages of using a proxy server when compared too more established approaches such as CGI. Proxy servers add an additional point of processing in the network pipeline. Each point of processing applies a network overhead in finding the additional machine, establishing socket connections and actually carrying out the forwarding of the data. The proxy server also applies processing in addition to the passing of data. If the proxy loads the entire page before it is forwarded then this will impose a lag in the network as the proxy downloads all of the document before it is passed on to the browser.

The application and any supporting tools must be installed on the users machine in order to benefit from client processing. For the proxy to work in parallel with a browser, options for the browser's proxy must be set. The options are currently set manually but there are automated methods that can be used to simplify the task. If the client owns the application then they have more control over its use and availability. Unlike network applications the proxy would not have to rely on a network being present in order to operate. Regular or dynamic updates to the software are harder to implement than an Internet Applet but can be achieved at the implementation stage. Useful updates may be the upload of new effect plugins, updates to the interface or even application dependent knowledge bases and expert system shells.

Because the proxy is a stand-alone application it duplicates tasks undertaken by the browser. Modern browsers have plug-in APIs that provide access to the HTML processing pipeline. A single HTML parsing engine would provide less programming overhead and might reduce the delay in processing as the expert system and document tailoring could be carried out within the browser rather than in two separate applications. There is a lack of integration with the only form of communication and control being the HTTP protocol and CGI used to communicate between the proxy and the browser.

3.3 Expert Systems

This project introduces a rule based expert system implementation, which, although simple, provides real inference and user interaction in a very wide range of applications. QERIS provides a high level of control over a source document aimed

specifically at the presentation of information rather simply as a consultative expert system.

This project introduces several new ideas in expert system design not addressed in other Internet based expert systems.

- *The Knowledge Base (KB) as a distributed Internet resource* - The system introduces incremental download of the knowledge base, since only the section of the KB associated with the rendered document needs to be downloaded. The questions about that KB are also defined as part of the document.

- *Driven by the source material* - Very close association between the information and the knowledge that controls how that information is presented. The source material and the KB representation are both stored within the same file. When the file is downloaded all the related information is available to the system for processing.

- *Document tailoring* - Direct and effective control of information and how that information is rendered to the user. The use of effects provides a flexible and extendable manner of controlling the document contents. Effects can be developed to provide added functionality and specialised applications.

- *Client based inference* - The ability for the user to control the functionality of the expert system. Due to the client side operation, a very powerful GUI can be developed that provides immediate updates and interaction between the application and the user. The inference engine also benefits form the power and resources of the client machine where the processor is less likely to be shared between other users.

- *User control over `their' information* - The application runs on the client machine providing the ability to store session information on a floppy disk or the machine's hard disk. Due to the client side processing no user information needs to be sent over the Internet. This may provide benefits where confidential information is being used.

- *Limited knowledge reuse and sharing* - The ability to apply knowledge from one knowledge base to problems stored within another knowledge base. Given the

manner in which the KB is defined and the manner in which documents can be downloaded across the Internet, the KB can be distributed over distance and across hardware platforms in the same manner as HTML documents.

3.3.1 Server based inference

Larger Internet based expert system solutions are server based using the CGI as a means of interfacing the Internet server with an expert system shell wysh discussed in Chapter 2 is a good example of such a system. The advantage of CGI is that it provides a simple mechanism for interfacing databases and well established expert system shells with the WWW. (see Figure 3.4)

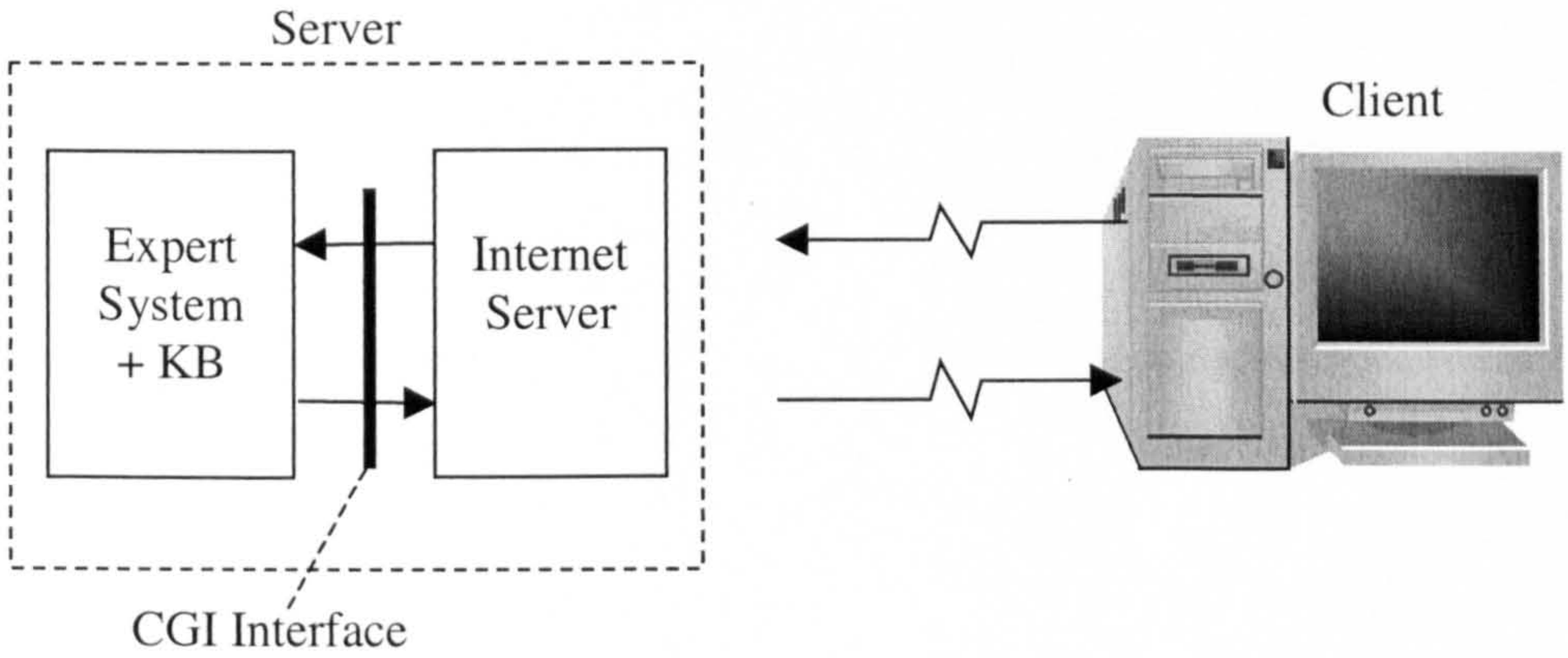


Figure 3.4 Expert System and CGI integration

A CGI based expert system uses arguments passed as elements of a WWW form as the responses to its queries. The developers have total control over the method of interaction, the interface style, the KB used and any information sent to the expert system while a consultation is undertaken. CGI although providing a solution imposes some restrictions on the interaction. Greenleaf et al (1997) state that the approach is poorly suited to multiple interactions requiring state to be maintained between those interactions. Servers generally require the capability of interacting with multiple clients simultaneously. There are methods of maintaining state but they require data to be sent backward and forward between the server and the client as "cookies" RFC2109 (1997) or for the server to dedicate resources to the storage of state for each user accessing the expert system. Another problem with server side inference is the delay between each query. There are delays both in network

communication when sending the query and receiving the response as well as the time it takes to process the knowledge base.

CGI interaction can be used to generate dynamic documents altering the presentation of the document before it is sent across the Internet. Documents can be constructed from templates and database interactions providing a very fluid representation of information. This control over the WWW documents is the strength of CGI, providing a degree of tailoring and personalised versions of documents, although restricted to instantiations of templates. The database and document tailoring elements of CGI will be discussed further in Chapter 5 as a means of integrating a client solution with server processed information. CGI should focus on the processing of the database queries, the construction of document sections and the tasks that it is better suited too. CGI is poorly suited to expert system consultations due to the requirement of immediate feedback and independence from network.

A question has to be built and sent to the client over the Internet as a CGI form.

The answer has to be sent back to the server as a CGI response.

The expert system processes the CGI data and initiates a `singe step' inference of the knowledge base.

If a conclusion is reached then the result is constructed and sent to the client as a HTML document.

Between each step of this sequence, a network connection must be established over the Internet. The connection depends of the network being available and the latency being low so that the response appears to happen in real-time.

It is not possible to save a consultation to disk, since the processing and the data required for the consultation is not available to the user. If the network connection is lost or slow then the user loses the ability to continue their consultation.

3.3.2 Client based inference

The development of Java applets, Cornell et al (1998) has provided a means of developing expert systems that although maintained and updated on the WWW

server, can be downloaded on demand and executed on the client machine. This form of implementation provides expert system functionality within documents

A WWW document and expert system applet is downloaded to the client.

The applet begins an expert system consultation.

Java provides an interesting solution to WWW based expert system consultation but there is one issue that reduces its usefulness in intelligent presentation. A Java applet has very limited control over the Internet documents that contain it and therefore not suited to document tailoring applications.

3.4 Introducing QERIS

QERIS (Query Effect Rule Internet Server) is the main implementation carried out for this project. QERIS has been developed as a solution to intelligent legal information presentation. QERIS can identify meta-information within HTML documents and provide a real-time interface between the user and an Inference engine. QERIS identifies special mark-up tags providing a knowledge representation that is used to describe a knowledge base specific to an HTML document; the mark-up will be discussed further in section 3.5.4. Once the tags have been extracted from the document they are used to build a knowledge base. The goals within the KB are identified and used to invoke an expert system consultation with the user. Information provided by the user is combined with the KB building up an internal model of the consultation and the user. The KB is used to create a tailored version of the document that can be displayed on the user's browser. QERIS uses a proxy server implementation to provide its Internet capabilities, which means that any requests for documents are forwarded through the proxy providing total control over the formatting sent through to the browser.

3.4.1 Implementation Requirements

The implementation of a prototype is designed to better visualise the operation of a concept, identify limitations imposed by current applications and standards, and identify further opportunities that may arise from such an approach. The implementation of the prototype is also guided by the following requirements:

- *Provide the author with a flexible development tool for intelligent presentation of information* – The author is the person who can provide the most insight into an information resource. It will ultimately be the task of the author to provide the knowledge used in presenting the information.

- *Interaction with the user must be responsive and its approach appear intelligent* – The interface with the user must provide a fluid interaction with the user for the minimum information required and following an intuitive path of questioning. The queries act both as a means of gaining information about the user and as guidance and prompts to the user to emphasis important elements of a document.

- *The system should provide the user with an individually tailored document that provides benefits over simply accessing the original material* – A tailored document must provide some benefit or value added to the client.

- *Rely only on well-established standards and not on application platforms that may incur incompatibility or restrict use* – There are many applications and languages that are restricted by hardware or software requirements. There are also many disparate methods, browser and server platforms that demand a certain approach. A solution should not restrict the distribution and use of the system.

- *Must integrate well with current popular technologies* – Many Internet technologies are already well established within the online community. These technologies should be used to their advantage.

3.4.2 Implementation language

The programming language used to implement the prototype is Java. Rather than using the applet functions of Java, a Java application has been implemented due to the comprehensive suit of Internet APIs to allow the development of client and server elements that can communicate in HTTP (HyperText Transfer Protocol). The Java language is portable without the need for recompilation, which allowed the development to be carried out on the University of Computer Science, department of Computer Science UNIX HP-UX systems and standard Win95 PCs. The final application is also compatible when distributed to a wide number of operating systems and hardware platforms. Due to the Object Orientated nature of Java, the

prototype has been developed as a selection of distinct modules with particular tasks. The potential of this modular construction allows for incremental development and updating of the core application. There are opportunities for replacing modules such as the Expert System when a new or better approach is developed. It is also possible to update modules of the system over the Internet when they are available for updated.

3.5 Prototype architecture

Traditional expert systems consist of two main elements, the inference engine, and the knowledge base. Elements such as the GUI are external to those at the core and are usually tailored for specific application domains. In the case of QERIS at the core a proxy server replaces the knowledge base component. The proxy server is responsible for providing a domain dependant knowledge base.

The prototype has a number of external influences that determine its:

- Network placement
- Server interaction
- Client interaction
- User interaction

3.5.1 Network placement

The Internet allows for a wide number of application locations that provide advantages and disadvantages that have to be balanced to achieve an optimal result.

Users make a request for information, which is then served to the user. Implementing a search engine requires a single request followed by a single response. The user receives what they expect and can then progress by making additional requests. In the case of an expert system, the user makes a request for information but a result is not returned, but an intermediate request is made instead. The user has no idea of the number of intermediate requests that may be made before a useful response is provided. If there is a delay between the intermediate requests then an additional

burden is placed on the user, who will be uncertain about the total time it will take to finish the consultation.

There are two main elements to consider when identifying a suitable placement:

- The latency of the Internet – Each call to an Internet resource consists of two components, the time it takes to establish a connection and the time it takes to transfer the data across an established connection. Internet latency is a very dynamic element dependant on the number of intermediate machines accessed, bandwidth available and the type of media being transmitted.

- The application latency – The application latency depends on the speed of the server the application is located on, the number of simultaneous users and the particular task been undertaken. The time it takes for the machine to finish the task is divided between the number of users accessing the machine.

The total latency of the system is the delay for each connection made and the delay for each processing period by the number of requests made for the duration of the session. A reduction in the number of users and a reduction in the number of requests would have the greatest impact in reducing the overall latency of a session.

Because of the volume of communication and style of interaction between users and online expert system solutions, the placement of the main processing element of the application is very important. Poor placement can result in sluggish performance, long delays and a reduction in flexibility. Good placement can reduce delays between interactions, provide faster processing and more flexible development opportunities. Many online expert systems rely on a CGI solution, which has the following qualities:

The application is located on the server, it is under the direct control of a development team and can be updated readily. A server-based solution can interact with other CGI and server resources such as Databases, object brokers and proprietary programs without communication with the Internet, see Appendix A.

3.5.2 Server Interaction

QERIS can communicate using several different techniques. The proxy can communicate using HTTP; QERIS can also understand CGI form submissions, make CGI requests to other online resources, server Java in-page applets and establish connections with the applets to provide real-time feedback.

- *HTTP server* – The HTTP server awaits a connection by a client and a request by that client for a resource. The server acts in a very similar manner to a WWW server checking the HTTP headers for information about the resource required.

- *HTTP client* – The HTTP client is the second half of the proxy function providing requests for resources from WWW servers. The HTTP header is rebuilt to include the correct machine addresses so that the information is delivered to the proxy for processing rather than directly to the original client.

- *CGI client* – The CGI client is similar to the HTTP client but it generates HTTP headers that include CGI POST requests. This allows the proxy server to query online databases for textual segments or additional user data. The CGI client then processes any responses, the results can then be used for tailoring or inclusion in the tailored documents.

- *Java Applet handler* – The Java Applet handler is a very prototypical implementation but provides a communication between the browser and the proxy server. The proxy can generate documents that originate as if sent from a WWW server. These documents can contain an applet located on the same machine as the proxy and then downloaded to run on the client machine. Initial experiments showed that statistical and graphical information could be updated in real time between the proxy and the client by establishing a TCP/IP port connection between the applet and the proxy server. It would also be possible for the expert system to interact with the user via a Java Applet front-end embedded in the browser document.

Like most standard WWW servers, the proxy has to interact with more than one resource at a time, downloading a HTML document along with all the graphics that go with it. Also the user may have more than one browser window in use. For this reason the proxy server is multi threaded providing the ability to carry out multiple

tasks simultaneously. The multithreading is extended in that the expert system and WWW based queries each run in their own thread also.

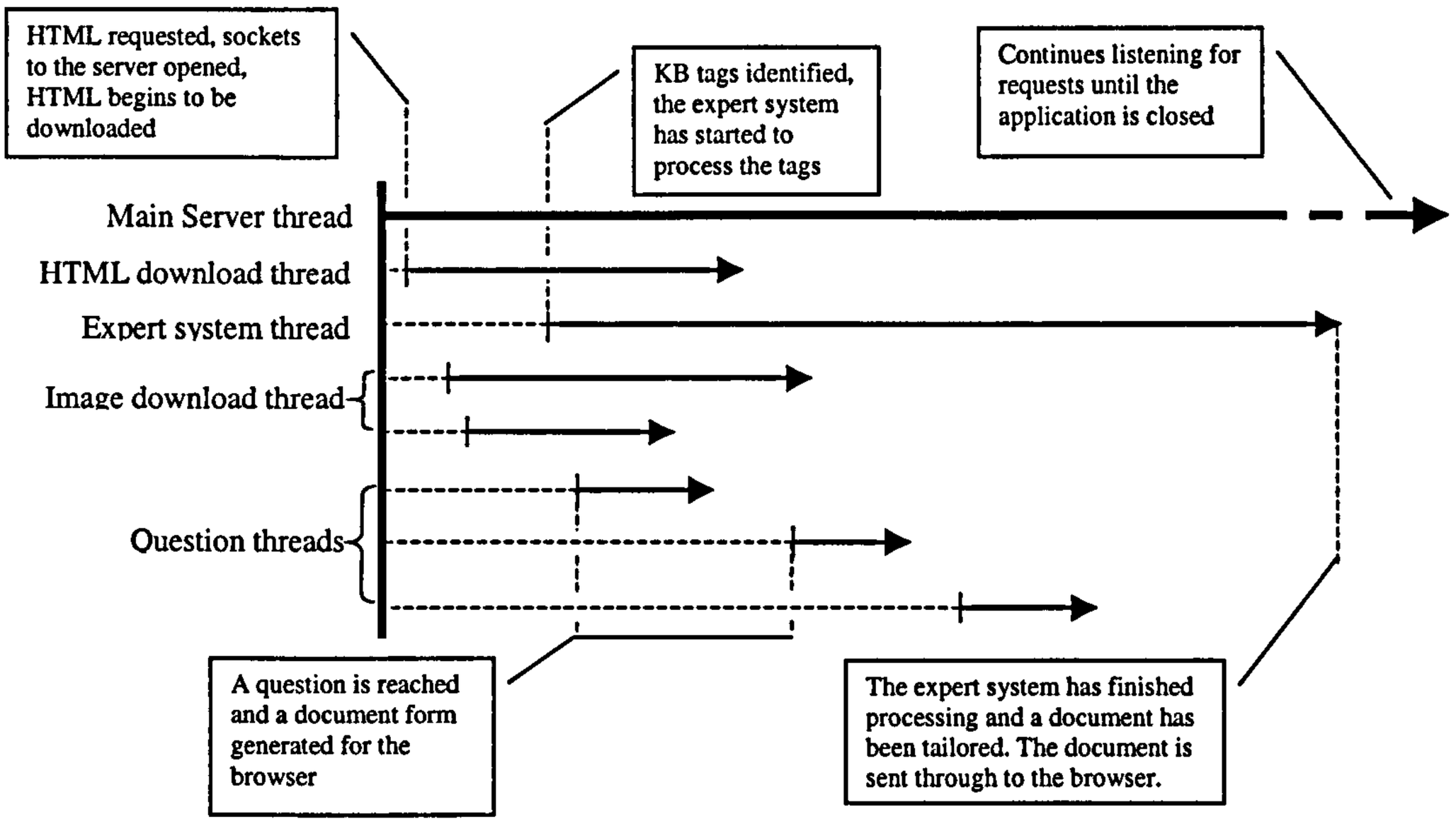


Figure 3.5 Execution threads in QERIS

Figure 3.5 shows a typical request for a document containing two images and a consultation consisting of three questions. The multithreading can become far more complex when multiple clients are involved. For this reason the earlier versions of QERIS only support a single client and are designed for client side operation.

3.5.3 Client Interaction

QERIS has two main modes, local or remote of communicating with a user accessing Internet content. The local presentation uses dialog boxes created in Java to prompt the user. Dialog boxes can be presented very quickly and do not need to establish a network connection to the browser. The implementation of the dialogs is also far simpler and provides a standard means of presenting queries to the user. In the remote mode the proxy "pushes" query pages containing forms to the browser. The implementation of this approach is more complicated and requires careful control of the networking elements of the proxy. The advantage of using the browser is that page templates can be loaded to tailor the appearance of the query forms. These two

separate approaches also provide a means of locating the proxy server in different places on the network. When using the remote method, the proxy does not have to reside on the user's machine.

In addition to the different presentation techniques, there are two different dialog styles, multi-choice or open. If the response from the user can be limited to a set of pre-defined answers then the query is defined as having a list of possible answers.

```
<QUERY ID="client_claiming"
QUESTION="Is the user claiming"
ANSWER="income_support,housing_benefit">
```

The answers "income support" and "housing benefit" would be displayed as a multi-choice list and the user would be expected to select an option. If no answers can be provided, then an open question is presented where the user can enter an answer string.

```
<QUERY ID="client_name"
QUESTION="Please enter your name">
```

The final presentation of the tailored document is the culmination of all the processing and user interaction. There are many different approaches to altering the presentation of the document controlled by the activation of effects. A number of approaches will be used to

3.5.4 QERIS tags and knowledge representation

QERIS uses a set of three tags that extend the set available in HTML. The tags are used to define a knowledge base, the actions that inference through the knowledge base will have on the document and the interaction that will be undertaken with the user. Each of these tasks is reflected in each of the three tags. Although the tags are current implemented as inline tags only, a block tag implementation would provide a

more flexible approach. Inline tags were selected to simplify the parsing of the HTML and the tailoring of the source document.

Effects – The location within the document and the type of effect applied indicates what results from the goal of the inference.

An effect does exactly what its name suggests, it applies an effect to a document at the current location determined by the chain of inference. There are different types of effect that could be applied to the document such as inserting strings or files; altering the HTML mark-up code; or removing a section of HTML from the document. Effects are the goals of the inference, the expert system attempts to prove, or disprove the effects condition. Effects have two possible cases, they either activate and the corresponding action takes place, or they do not activate in which case an alternative action occurs.

```
<EFFECT ID="can_claim"

CONDITION = "claim is valid"

PLUGIN="insertString"

TRUE="<B>A valid claim <A

HREF="make_claim.html">

  Continue?</A></B>"

FALSE="<I>The claim is invalid <A

HREF="moredetail.html">

  Why?</A></I>">
```

The above example requires further information; the information "claim is valid" can be determined by activating related rules or queries to form a chain of inference. If the result "claim is valid" is determined then the string

```
<B>A valid claim<A HREF= "make_claim.html">
Continue? </A></B>
```

Is placed into the document at the location of the effect, otherwise the string

```
<I>The claim is invalid <A HREF=
"moredetail.html"> Why? </A></I>
```

Is inserted. Note the use of HTML mark-up within the strings, which provides control over layout and final presentation of the document.

Rule – Rules act as the linkage between Effects and Queries providing simple backward chaining inference. Each rule has a unique identifier. If this identifier is used elsewhere then the result of the rule is used rather than the condition. The rules are designed to be simple to create and have the following structure:

```
<RULE ID="claim"

CONDITION="client_claiming is income_support or
client_claiming is housing_benefit or
client_income is low"

CONCLUSION="valid">
```

Each rule has a condition, which must be met for the rule to `fire'. If the rule fires then the conclusion of the rule is asserted into the knowledge base as a fact.

The above rule asserts "claim is valid" to the knowledge base as a fact if the client is claiming income support or housing benefit or has a low income. The effect shown in the example above may call this rule in order to determine if "claim is valid".

Query – Provide knowledge acquisition from the user. This is where the characteristics are identified and obtained.

```
<QUERY ID="client_claiming"

QUESTION="Is the user claiming"

ANSWER="income_support,housing_benefit">
```

The Query is constructed from two main parts, the question string that will be asked of the user and the possible responses that the user can make. The answer will be represented as a multiple-choice query of which the user selects a single response.

By combining the above examples a simple scenario can be constructed.

```
<EFFECT ID="can_claim"
CONDITION = "claim is valid"
PLUGIN="insertString"
TRUE="<B>A valid claim <A
HREF="make_claim.html">
  Continue?</A></B>"
FALSE="<I>The claim is invalid <A
HREF="moredetail.html">
  Why?</A></I>">

<RULE ID="rule1"
CONDITION="client_claiming is income_support or
client_claiming is housing_benefit or
client_income is low_income"
CONCLUSION="claim is valid">

<QUERY ID="client_claiming"
QUESTION="Is the user claiming"
ANSWER="income_support,housing_benefit,none">
```

The inference begins from the goal claim is valid, to determine if a client can claim or not. In order for a client to claim their claim must be valid. A rule with a

conclusion that matches valid is rule1. In order for rule1's conclusion to be determined, the rule must fire. What requirements does the rule impose to fire? By examining the condition,

```
client_claiming is income or client_claiming is
housing_benefit or client_claiming is low_income.
```

It can be seen that if any of these cases succeeds then the condition will be met. There are no other rules in the KB and the effect still is not satisfied so we can examine possible queries that we may ask the user. There is a query "client_claiming" which matches the requirements. The following question will be asked of the user:

```
Is the user claiming?

  income_support

  housing_benefit

  none
```

If the user selects either "income_support" or "housing_benefit" then rule1 succeeds and its conclusion claim is valid will be asserted into the KB. Once this information is known then the effects "can_claim" condition is met and the effect is activated. If this information is contradicted then the effect tag fails to activate.

Two examples of using these tags, a Benefit advice site and a Convention for the International Sale of goods will be discussed in chapters 4 and 5

3.5.5 Internal Architecture

The prototype consists of five elements that combine to provide a flexible approach to document tailoring. The prototype is designed around a modular approach so that each section can be updated independently without effecting the other elements. It is envisaged that the expert system/knowledge base will be "pluggable" allowing more specialised inference engines to be downloaded and integrated with the rest of the system as a particular subject domain requires. A good example of this is extending

the current expert system to provide support for uncertainty such as Certainty Factors or fuzzy inference.

Main Elements

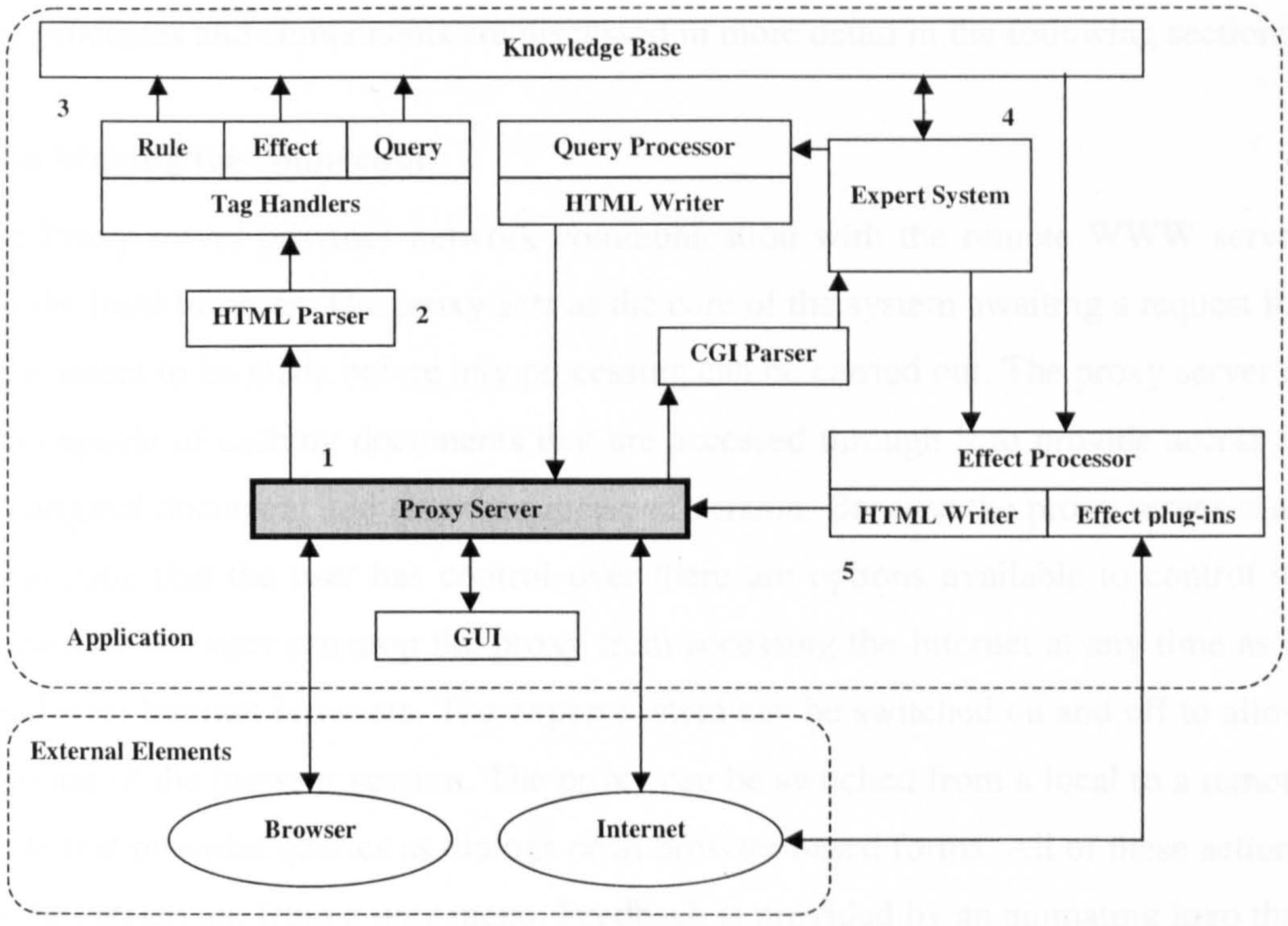


Figure 3.6 Internal Components of QERIS

The internal components of QERIS are shown in Figure 3.6. These components include:

1. The proxy server controls the HTTP messages and connections between the client and web server.
2. The HTML parser reads the HTML document and breaks it down into an internal data structure that can be processed by the expert system.
3. The tag handlers are used each time a rule, effect or query tag is identified. This could be expanded in future implementations by adding additional tag handlers.
4. The expert system is started once the last HTML tag has been processed and registered in the knowledge base. The expert system controls all inference, consultation and final document generation.

5. The effect processor takes the resulting document and applies any active effects to it, external content is downloaded, knowledge tags removed and the final document is produced.

The processes and components are discussed in more detail in the following sections.

3.5.6 Making the connection

The Proxy server provides network communication with the remote WWW server and the local browser. The proxy acts as the core of the system awaiting a request for a document to be made before any processing can be carried out. The proxy server is also capable of caching documents that are accessed through it to provide access to the original document and also the processed version. Because the proxy server is an application that the user has control over there are options available to control its operation. The user can stop the proxy from accessing the Internet at any time as is usual with Internet browsers. The expert system can be switched on and off to allow tailoring of the browser session. The proxy can be switched from a local to a remote mode that provides queries as dialogs or as browser based forms. All of these actions can be carried out from a user menu. Feedback is provided by an animating logo that indicates that the proxy is currently accessing the network or processing the knowledge base.

3.5.7 Parsing the HTML document

The HTML parser is used to identify, interpret and extract the knowledge meta-tags from the documents as they are accessed. If the documents contain meta-tags then this information is passed to the corresponding tag handler for processing and combination with the knowledge base. If no meta-tags are found then the HTML parser stops and the document is passed directly to the browser without modification.

3.5.8 Building the knowledge base

The knowledge base uses a production rule model for storing domain knowledge from the documents being accessed and attributes about the user provided during consultation. The knowledge representation has three main types that match the three main tags that can be processed. The effects are treated as goals for the inference

through the rule base. Each goal initiates a new consultation between the user and the expert system. The goals are processed in sequence, and once the result of a goal is known and a conclusion reached then a fact is asserted into the knowledge base to reflect this. The facts or conclusions to previous goals can then be used in subsequent inference through the knowledge base. Rules define the linkage between the goals and the possible queries posed to the user. The rules follow a simple production rule model of multiple antecedents and a single consequent:

```
Goal:Is_NI_number_issued

IF NInumber issued

THEN provide some result to the user


Rule:NI_number_issued

IF Age >=16 AND Nationality is British

THEN NInumber issued


Rule:Claimant

IF Ninumber issued

THEN may claim


Query:Age

Age "<16",">=16" ?


Query:Nationality

"Non British","British" ?
```

The rule base is backward chaining, so that it attempts to solve a set of goals by matching rules backward until a fact is reached. If the system does not find a

matching rule then it attempts to match with a query. Queries allow the system to interact with the user in a flexible manner. In this model questions are either multiple choice or open. A multiple-choice query provides a selection for the user from which a single option can be selected. If no query is matched then an open query is asked that prompts the user for a textual string. Although the knowledge base is constructed a page at a time, the rules are available not only to a single page but also to successive pages loaded into the system. The knowledge bridges across documents providing an expanding knowledge base. Parts of the knowledge base can be loaded in an index page, so that these segments can then be used by goals identified later in the consultation. During a browsing session the rule base can be saved by the client to freeze the current session and store any rules or facts that have been added to the knowledge base. Users can then reload the knowledge base to continue the consultation where they left off. They can also investigate different options and "what if?" Scenarios even after access to the network has ended.

3.5.9 Fact storage

QERIS has a mechanism for storing facts and rules that are collected while accessing a document. Because rules are downloaded and stored each time a document is accessed they can be used across several different documents. This allows sites to be constructed with common rule sets that can be available to the inference engine during the full sessions and any sessions that follow. Currently rules are only identified by their ID string, and if the ID string occurs and a rule exists in the knowledge base then it will be used. If a rule exists in the knowledge base already then it cannot be overwritten. It would be possible to extend the implementation to group rule sets based on host IP address or other session static values in the same ways as HTTP cookies. The storage of rules provides some interesting effects. Rules can be pre-loaded into the knowledge base by including them in a document that does not contain any effects. Facts are stored, a fact can be registered based on which document has been viewed, and such a mechanism can be used to direct the navigation of the user.

Text segments loaded as part of an effect mechanism (the insertFile effect plug-in) can also include rules and additional effects which provides a very dynamic way of generating an inference path. If an effect is triggered and a document loaded then the document may contain only knowledge base tags resulting in a completely new thread of inference.

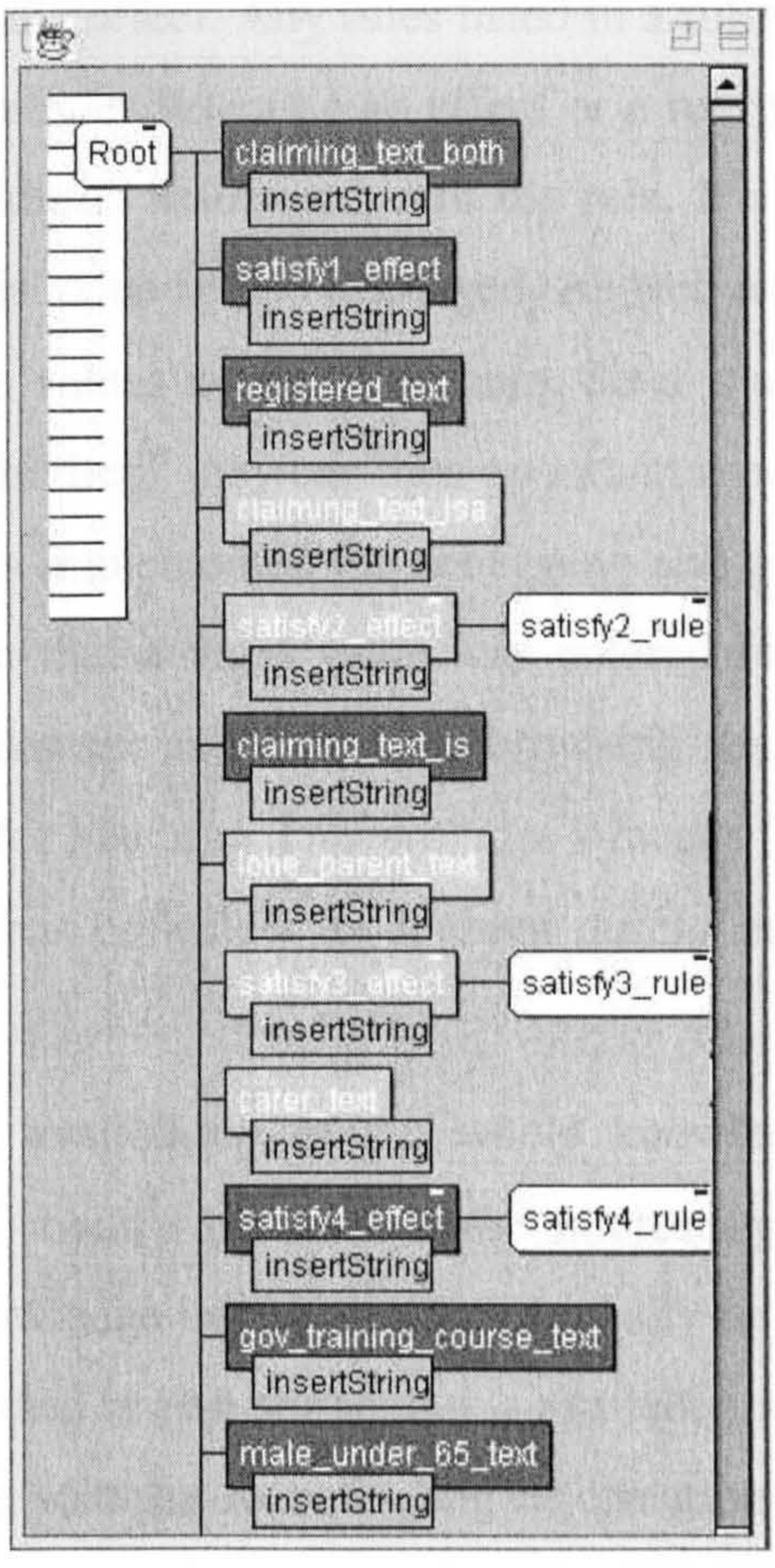


Figure 3.7 Rule and Effect display

The storage of effects, rules and facts is currently very simple: they are stored in a structure that indicates their relationship with other rules and effects. They can be viewed visually (see Figure 3.7) as a folding tree structure. The structure has a root that indicates the set of documents that has been loaded. Selecting the root will provide a list of URLs that have contributed to the knowledge base. Off the root are the effects that are displayed in the sequence they occur in the documents loaded. As new documents and new effects are identified, they are added to the list. Each effect can be folded or unfolded, and its icon (square purple box) indicates the name of the

effect and the plug-in that the effect will activate. Selecting the effect icon will display a dialog that indicates the arguments that will be sent to the plug-in. The darker colour of the effect also indicates if it has been activated. Each effect may have rules associated with it, a rule is added to the first effect in the list that has the rule listed as part of its condition. All rules listed in the effect's condition are placed as separate branches off the effect. Any rules listed in a rules condition are placed as separate branches off the rule. Selecting an effect or a rule will display the name of the element and the condition associated with the rule. If a rule has been triggered and a value assigned, this value is also displayed. As well as viewing the values it is also possible to alter the values assigned to them. So if a user answered "yes" to a question "do you have children" they can then go in and modify the value of the rule to "no". Currently this is implemented for debugging and prototyping purposes and there is no confirmation that a valid value was entered for the rule. Once a user wishes to close their browser and exit their browsing session, they can save the knowledge base to disk for later use. This provides a means for the user to "store" the knowledge the QERIS has collected about them during an interactive session for later use.

The flexibility and the availability of the whole knowledge base for a user or developer to view and modify provides some interesting capabilities of such a system. Because the knowledge base is generated locally on the user's machine then all the information collected is also stored and is available on the user machine. This provides a more private solution to collecting information about users in order to generate information more specific to them. It is not necessary to transmit any of the user's information across the Internet, the only way of finding information out about a user of the system is by monitoring their navigation patterns.

Although not implemented there are a number of areas that would benefit from further research:

- The reuse of facts between organisations: in many cases information such as date of birth, name and marital status may be useful in a wide range of resources. By maintaining a history of which rules and facts are more commonly used there presence may be maintained by moving them to a more static knowledge base.

By generating a core knowledge base that contains key information the user will be asked to enter the same information less often. The only concern with storing information for long periods is the easy access to the information for updating, deleting or modifying by the user.

- What level of detail should be associated with an effect, rule or query? It would be possible to register the document or documents an effect is located in. The frequency rules are activated, so the number of times they succeed or fail could be monitored. Rules that are infrequently used could be removed from the knowledge base, so that we could consider what impact would such a dynamic modification of the knowledge base have on its function? Questions could be monitored and the most common answers registered. The originating document of the query could also be stored so the user knows which document asked what question.

- Should sub-sections of the knowledge base be made sharable across groups of people or with the host that delivered the knowledge base to the user? Which organisations should have access to which facts and what mechanisms can be put in place if such sharing occurs for the user to control such access?

- Should rules and effects have an expiration date or period, where their state and even their inclusion in the knowledge base will be modified? If a rule is removed because it has "expired" what effect will this have on other rules in the knowledge base?

- How can the visual representation of the knowledge base be improved? The important goal is to allow the user to control their own information, bringing the knowledge base up-to date, modifying incorrect facts. The current implementation is very simple but still provides very powerful mechanisms for visualising the inference and controlling the future of the facts that are stored. What can be achieved even with this simple implementation will be demonstrated by the case studies in chapters 5 and 6.

3.5.10 Interface

The QERIS interface is very minimal so it does not detract from the browsing session. It is designed to provide access to important elements such as the loading and saving of knowledge representations, resetting of the inference and proxy engines and clearing of the rule base. These functions are all available off a small interface that also provides feedback about the number of threads being used and a graphical indication that the system is processing documents and rule bases.

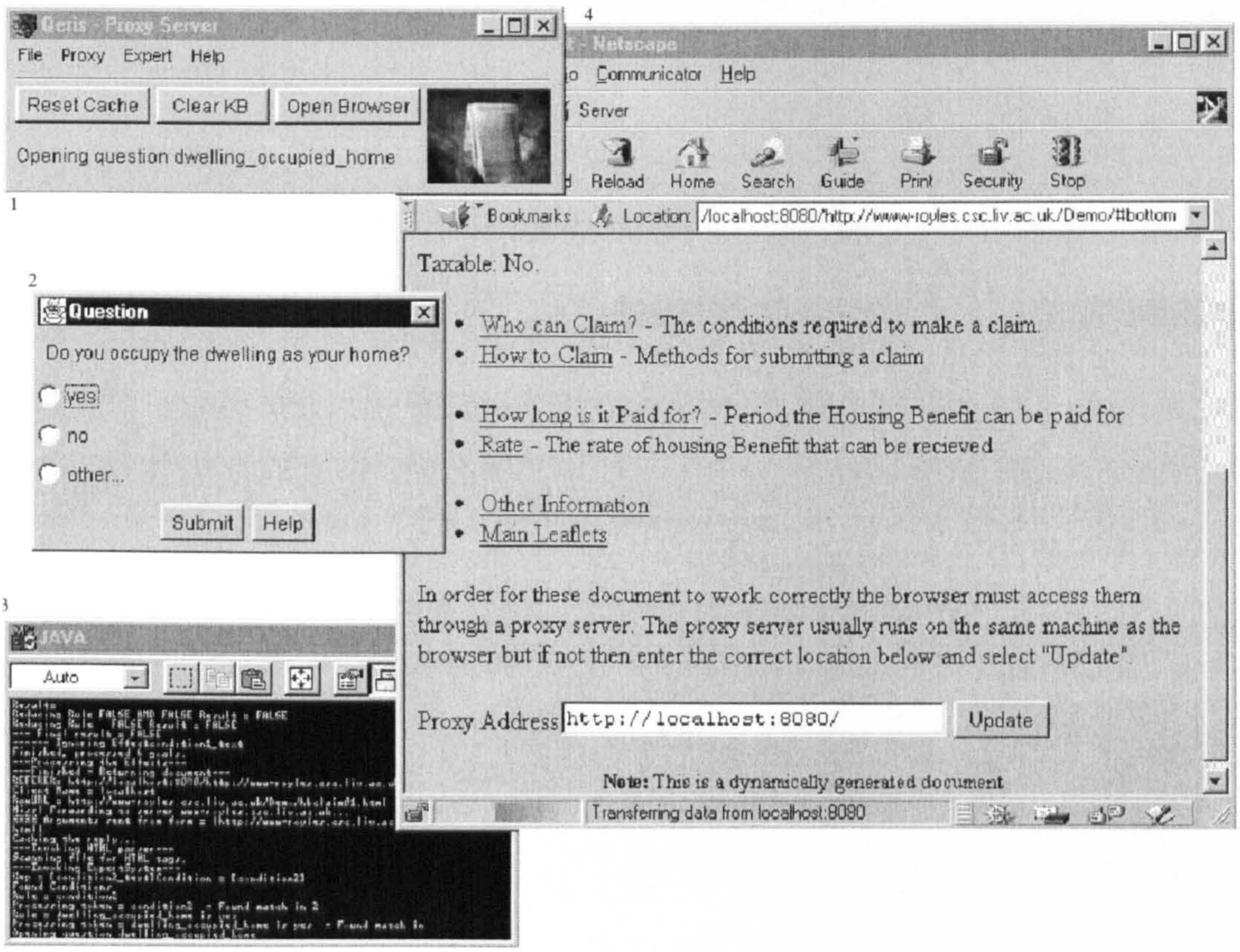


Figure 3.8 Main QERIS interface components

When using the QERIS system, there are 5 main windows (see Figure 3.8),

1. The proxy server interface used to display information about the state of the server, its active connections and a graphical indication that the service is processing a document.

2. A number of different question dialogs are displayed during as browsing session. Query dialogs can be displayed within the browser or as stand alone Java AWT components.

3. The initial command prompt used to start the proxy server can be used to gain more detailed information about the state of processing, currently debug and tracing information is displayed in this window.

4. The Web browser is the main interface and is used to access pages and display queries and final document representations to the user.

5. The 5th window is the rule structure window which is described in detail in section 3.5.9

3.5.11 Execute the inference

Once the knowledge base has been extracted from the source document and a list of goals has been identified then the expert system is invoked. The expert system operates on a changing knowledge base so its main task is to match rules that relate to the current inference. If a query is activated then the expert system formulates a response for the user as a dialog or a WWW document. The response is then processed by the expert system and asserted into the knowledge base. The expert system maintains a history of rules and queries fired for a particular session. This provides a backtracking facility and although not implemented in the prototype, the ability to give an interactive explanations.

3.5.12 Tailor the document

The effect processor is used to apply the tailoring to the source document. If a goal succeeds then its action is applied to the document at the location of the corresponding effect. Each effect has an associated plugin that determines what processing action is applied to the document. Applied effects include the insertion of textual elements such as strings, files or CGI database responses; insertion of HTML tags to alter the presentation and structure of the document; or textual strings provided by the user during the consultation. The prototype hints at a "pluggable"

architecture by providing a means of adding new effects to the Effect processor. The class name is provided in the Effect mark-up tag

```
<EFFECT ID=uniqueid
 CONDITION=condition
 PLUGIN="insertString"
 TRUE="string to insert if effect triggered"
 FALSE"string to insert if effect NOT
triggered">
```

This class is loaded and its task is carried out on the source document. The above example will insert a string at the position of the effect within the document. Although effects cannot currently be downloaded on demand it would be possible. This capability would provide document authors with a means of designing a knowledge base and also the effects that carry out the tailoring of the document, rather than relying on some standard set of effects.

3.6 Advantages of QERIS

QERIS varies from other expert system solutions in that it focuses on the tailoring of downloaded documents. The prototype contains a very simple implementation of an expert system shell, and a more complete solution would provide far more functionality and control for the user. QERIS relies on an expert system to undertake its task. The important concepts are the placement of the expert system within the network, the interaction between the expert system and the knowledge base and the tailoring of Internet documents.

QERIS provides an analogy with real life advisors where a user would communicate with an advisor, sometimes that advisor may resort to accessing documents for further information to help the interaction progress, and drawing his client's attention to particular sections of them.

QERIS provides a model that does not require confidential information to come into contact with machines where it may be accessed. No user information needs to be

transmitted across the Internet or stored on a remote machine. The only method of determining information form the client is by inferring from the documents or resources they select to download.

The knowledge bases and use of the expert system does not require alteration to and control over CGI elements of a hosting server. This provides a means for anybody to develop knowledge bases contained within their documents in a similar manner to Meta-data within HTML.

QERIS complements currently evolving technologies such as XML and Java, integrating well with current technologies.

3.7 QERIS as a Legal Solution

Legal information is hard to read and interpret especially by non-experts. For this reason any system that improves the readability and the ease of understanding will be a benefit in legal information presentation. Although none of the individual components of QERIS were specifically designed for legal material, its main ability of taking complex documents and providing a tailored representation is well suited to legal advice and legal online material. Areas where QERIS can help in legal information presentation include:

Interpretation - Legal information has to express sufficient information in a clear and concise manner as to provide a useful resource to the end user. In many situations the interpretation of legal information is a skill in its own right, practitioners of law require many years of experience to become proficient. Any support in interpretation will provide a member of the general public with a better means of interpreting raw legal information.

Format and structure – A large majority of legal information abides by official or agreed standards in format and structure. Citations are mad in a similar manner to aid in their understanding. Wording and terms are maintained across like documents to ensure consistency. QERIS can provide information tailored around templates providing a uniform and automated technique for generating standard structures.

Relevance and Interpretation - Heavy on text and terms, legal material contains large volumes of text that are not always relevant to the reader. The text does not

always contain familiar words and terms. By providing a support tool for such material, the volume and the wording of such text can be tailored to aid the user in its interpretation.

Prototype examples of QERIS and its use as a legal presentation tool will be discussed further in Chapters 4 and 5.

3.8 QERIS as a General Solution

QERIS was initially designed as a solution for presenting legal information. The result is s system that can also be used to tailor documents in other domains. The expert system shell can interpret knowledge bases from many different subject domains. The advice capabilities could be used to sell a product, guiding a user to the item that best fulfils their requirements. The fault finding elements of the expert system could then be used to provide online support and intelligent documentation. The QERIS system can even be used to provide interactive and dynamic navigation of an Internet site or narrative such as an interactive novel.

3.9 Further work with XML

This project began before XML developed as a mainstream Internet based mark-up language. The project consists to two main elements, the mark-up and the processing engine. Both of these systems can be developed independently from each other. The mark-up provides a mechanism for placing knowledge information within Internet content in a similar fashion to layout mark-up provided by DHTML. The processing engine provides a conversion from HTML containing this knowledge mark-up into plain HTML ready for processing by a browser. These two ideas are key in the development of XML. In a WWW context, an XML document defines a structure for the document, and applications are then used to process this structure. The processing of XML is left to third party applications which when identifying XML can process it into a browser ready format. The similarities between the original key concepts of this project and the development of concepts for XML provide validity to the techniques being used.

It is possible to develop this project so that it carries out the same task and has the same effects of a document if XML is used as the mark-up language. The only requirements of such a system would be a Document Type Definition (DTD) that defined the new tags developed as part of this project, see Appendix B.1.

This example demonstrates how the knowledge base mark-up language could be represented in XML. This representation would then follow the models proposed for the processing and development of XML applications. Any author wishing to use the knowledge base mark-up in their XML content would need only to link with and abide by the formatting constraints the above DTD in order to produce valid XML and knowledge base mark-up. Although the tags do not map very well into elements within XML, the use of attributes allows the tags to be represented. With careful design, it may be possible to develop an object-oriented approach to the structure of the mark-up allowing greater flexibility in the tags. To ease parsing the tags have been developed as single elements with no nesting, and have also only been implemented as inline tags not as block tags. This is an implementation restriction that has been imposed and not a concept or design restriction. Another way of representing the knowledge mark-up in XML is represented in Appendix B.2. This representation would be far harder to process but may provide far more flexibility. For example it would be possible to represent multiple conditions and conclusions within a rule and plugins could be handled better within the effect tag.

Another interesting parallel between the project and the development of XML is the use of the Java programming language in the development of processing engines. The XML model requires client side applications to apply processing to the XML content before the information is rendered to the browser. Currently the minimum requirement is the conversion of XML to HTML so that current browser technology can interpret the content. As technology develops, XML parsing engines will be integrated into the browser applications, and the developers of these applications will have to provide mechanisms for third party development and interaction with the XML engine allowing useful processing of the XML content to be achieved.

The mark-up used in QERIS, although currently resembling HTML, requires further development. The current mark-up could be used in many different content formats

including HTML, SGML, XML, or any other format that would allow the mark-up to be hidden from other processing engines. This degree of flexibility far outweighs the disadvantages of a proxy server especially during the prototype and development stages of the project.

3.10 Implementation Limitations

Although the concepts and prototype behind the QERIS application provide potential for knowledge based system design, the underlying concepts and prototype implementation are restricted by time and resources. The following points would be areas for further investigation if the time and resources were available.

3.10.1 Inference engine

The current implementation of the expert system and knowledge base is a simple proof of concept. There are no tools allowing explanation of conclusions reached or user control over the consultation such as undoing a previous answer. The internal structure of the knowledge base restricts the functionality of the inference engine. A better more structured internal representation would allow interaction with different inference mechanisms. Due to the modular nature of the current implementation it would be possible to develop a knowledge base API that would allow for the development of separate `pluggable' knowledge bases and inference engines. Extensions to the inference engine could include certainty factors, fuzzy logic or possibly management of the different knowledge base sources and better integration with the distributed nature of the source knowledge.

3.11 Summary

This chapter has provided an overview of related technologies including Internet mark-up languages, Internet proxy servers and networking and expert systems. The chapter then described the key elements of the QERIS system as an intelligent Internet based document tailoring system. The key elements include:

- A client side expert system that provides intelligent interaction with the user with the goal of tailoring the requested document better to meet a users unique characteristics.

- A knowledge base that is integrated with the requested document providing very close association between the knowledge and the information. The authoring of the knowledge base is carried out during the authoring of the document.
- A fast and responsive Internet based expert system approach than can provide applications from simple template construction of documents through to full consultation with the user.
- An online solution for presenting legal documentation or providing legal advice specifically tailored to an end user.
- A broader concept of providing client side knowledge based systems that can adapt to the domain that the user is interacting with at a particular time.

CHAPTER 4

SOCIAL SECURITY BENEFIT ADVICE

4 Benefit Advice

The Benefits Agency is the largest executive agency of the Department of Social Security. It is responsible for the administration and payment of 22 social security benefits to 20 million customers, amounting to £90billion in 1999. The agency has an obligation to inform the public about the benefits they may claim and to help an individual through every step of making that claim. The Agency provides a set of information leaflets on all aspects of claiming benefits in 7 languages. From the 23rd November 1998 the number of leaflets made available by the agency dropped from 140 to 50 as the advice was condensed and restructured.

4.1 Overview

4.1.1 Providing advice

Throughout everybody's life they go through changes in their personal circumstances, leaving school, starting work, getting married, and having children. There are also negative events that can affect a person such as losing their job, divorce or injury. In many cases when changes occur, the person has time to adjust and learn about their new situation, but in other cases very quick adjustments must be made. It is important that whatever the situation, the information regarding the benefits available can be identified, is accessible and provides is sufficient information for the person to be aware of their rights to claim, or not to claim, a particular benefit. Benefit leaflets are very dependent on the characteristics of a reader, their contexts, surroundings, level of education, job situation and the like. For

this reason benefit advice must cater for all possible aspects of a persons situation. If the advice does not provide information about a person's situation then this can delay or mislead that person in making a claim.

4.1.2 Traditional Advice

Advice is provided by the benefits agency in two ways, as paper based documents and as verbal expert advice. Paper advice documents are usually 10-20 pages thick and are available from many outlets, the most common being social security and post offices. The documents are all laid out in a similar style and the content is aimed at the full spectrum of potential readers. Each document covers a section of the available benefits such as Income Support, Child Benefit or Housing Benefit. In many cases several documents are required to get an overall picture of a particular benefit. For example, in the case of Child Benefits,

> *"The customer will be excluded from the right to Child Benefit if ... the child is receiving Income Support or Severe Disablement Allowance ..."*

The claimant must therefore be aware that the child is either claiming these described benefits already or has the potential to claim them. If the child has the potential to claim then the claimant needs to know which benefit takes precedence and which would be most beneficial to the claimant. Problems of this nature are discussed in detail in Chapter 3. In many cases the information in the advice documents is not sufficient and then the claimant may visit an advice centre, or a local benefit office, to speak with a Benefit advisor.

4.1.3 Internet based Advice

With the advent of the Internet a new method of providing information to the claimant became available. The Department of Social Security has a large collection of information available on its web site at http://www.dss.gov.uk/ and during the development of this project the site has been improved to help the user navigate to and then access the information relevant to them. The initial site available in early 1998 presented a large list of benefits from which the user could select. The information presented was simply an online version of the paper based advice leaflets.

Who is entitled ? A customer who is responsible for a child living with them; or

is contributing to the child's maintenance at a rate of at least equal to the Child Benefit payable for the child; and

the customer's right to enter or reside in Great Britain is not subject to a limit or condition, and the customer or the child have been resident in Great Britain for a total of more than 26 weeks in the preceding 52 weeks. A person coming to Great Britain from abroad can claim Child Benefit if he or she:

has been granted refugee status; or

has been given exceptional leave to remain in the United Kingdom; or

is a national, or member of the family of a national of an EEA country; or

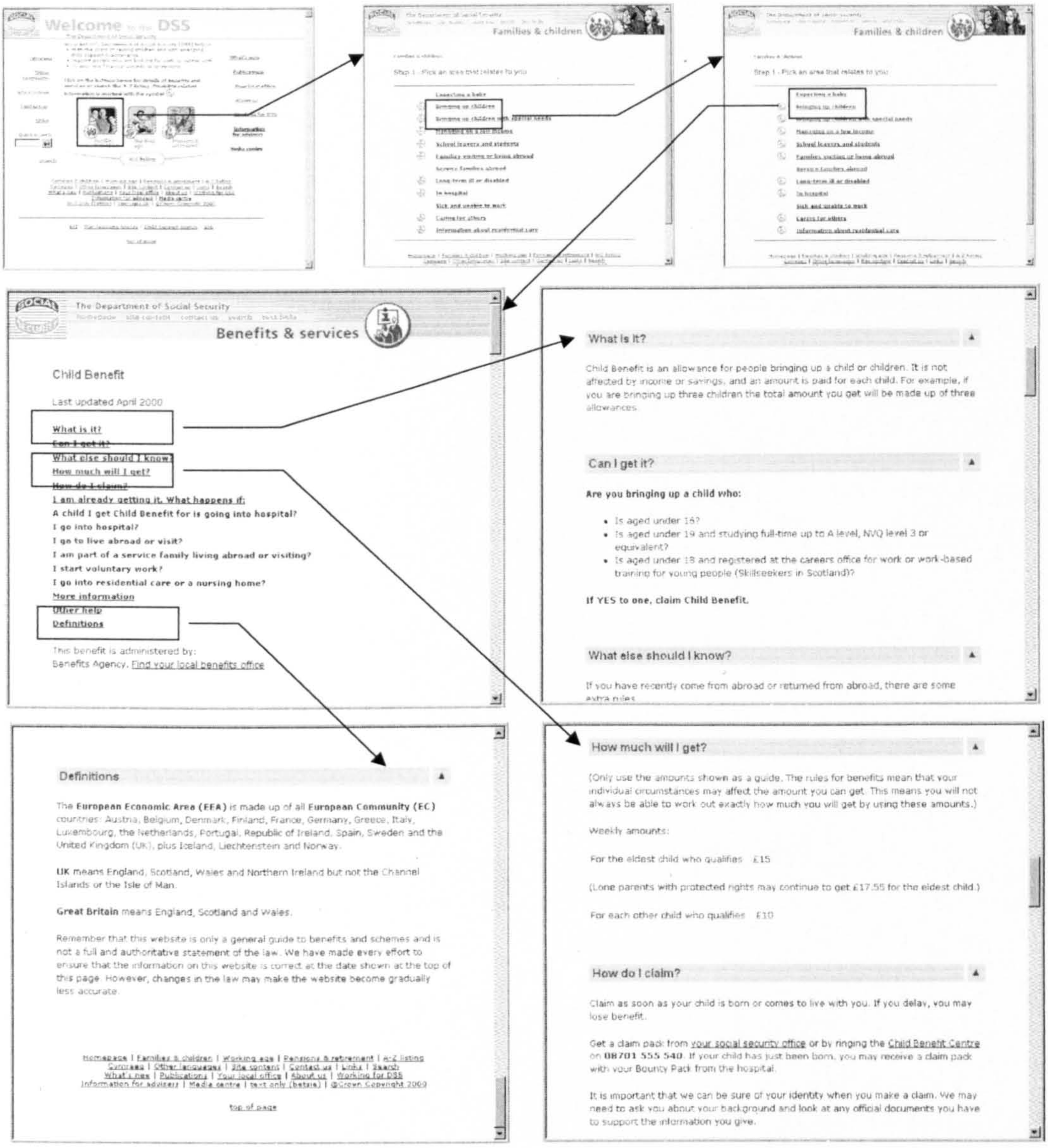
is lawfully working in Great Britain and who is a national of Algeria, Morocco, Slovenia or Tu or a member of the family of such a person living with them; or

is covered by one of Great Britain's agreements on Child Benefit with other countries; or

has leave to enter or remain in the United Kingdom which is not subject to any limitation, for example, has been given indefinite leave to enter or remain or the right of abode in the United Kingdom.

Figure 4.1: Original Online format of Benefit Advice

Since the availability of the benefit advice online, the site has undergone several changes resulting in its current format (September 2000) where navigation and presentation has been improved, although the content is the same. The benefit Information for Child Support that will be described later in this chapter can be reached through the following pages.



(Reproduced in appendix)

Figure 4.2: *Current Online Child Benefit Advice*

In relation to Benefit Advice there are very few additional resources to the official site available. The Social Security site is kept very up-to date with daily changes. Information such as rates of benefit are also provided with dates when they are applicable. Although the navigation through the pages helps to identify the information quickly, there is still no attempt to try and improve the interpretation of the information presented or to help the reader further to understand if they are entitled to claim. The other obvious limitation is the inability to submit information to the site, or to engage in any kind of interaction. There is no way of filling in forms

or updating the Social Security records with information using online mechanisms. None the less, the information presented online has a number of advantages over a paper-based documentation:

- The information is always the most recent and up-to-date version. Some of the values in the documentation had been updated within the first 10 days of the month of September. All users that access the site can benefit from the updated information. Many of the advice pages also had real figures in them that represent the rate of benefit for that month. With paper based media there is a high cost incurred in the collection of old documentation and the distribution of the new. It is also common to see references to a second document containing rates rather than placing the constantly changing rates within the larger benefit advice document. This can result in readers not having access to all of the information within a single document.

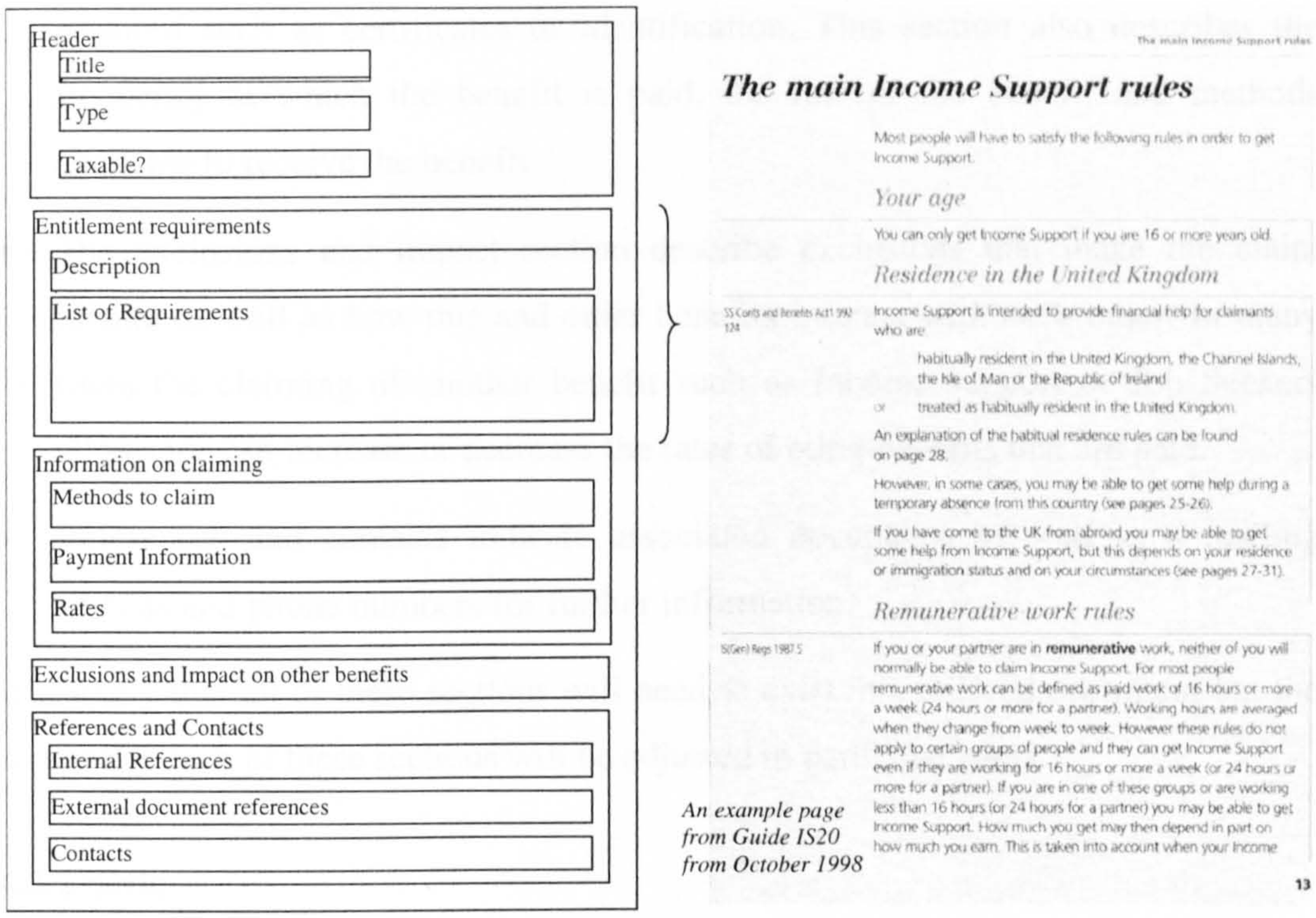
- The information is graphically enhanced and more pleasing to the eye, although this is unnecessary in getting the information across. Due to the large distribution of information it is not cost effective to produce paper based documentation that contains such graphical elements.

- Many pages contain cross-references to other benefit advice documents and by selecting the hyperlink the other benefit information is displayed. With paper documentation it is necessary to collect up-to 13 additional documents in order to have full coverage of the topic.

4.2 Advice Documents

4.2.1 Structure

The paper based and online advice documents are broken down into separate sections as shown in Figure 4.3. On the left is the general format for the entire document. On the right is a sample page that relates to a sub section of the document.



An example page from Guide IS20 from October 1998

(Reproduced in appendix)

Figure 4.3: *Advice Document Structure and example page*

The document structure in Figure 4.3 also corresponds with the format of the online pages (the advice document can be seen in the bottom 4 images in Figure 4.2 which all exist as a single online document). Some pages differ in structure but the differences are slight: for example some related documents will exclude the header information in subsequent pages. The components of the advice pages are as follows:

- The header contains the title of the benefit. This is the name of the benefit and used in the subject index. The type of benefit indicates if the benefit is contributory or income related. Some benefits are taxable, and this is also indicated in the header.
- The benefit entitlement section provides a brief description of the benefit and a listing of the requirements, many of which are cumulative. In many cases this list can be mapped directly into rules within a knowledge base.
- A section about claiming describes the methods that can be used to claim, indicating forms that need to be completed and other resources that will be

required such as certificates or identification. This section also describes the frequency at which the benefit is paid, the rate of the benefit and methods available to receive the benefit.

- The exclusions and impact section describe exclusions that make the claim invalid as well as how this and other benefits interact with each other. In many cases the claiming of another benefit such as Income support or Job Seekers allowance can increase or decrease the rates of other benefits that are paid.
- References and contacts indicate associated documents as well as a mailing address and phone numbers for further information.

It is likely that all of these sections will need to exist in a tailored document but the content of each of these sections will be adjusted to particular users.

4.2.2 Style

Each of the Benefit Advice documents focuses on a particular benefit that can be claimed. The documents are broken down into logical subjects. Each section in the document is broken down into sub headings such as 'Type', 'Who is entitled?', 'How to claim' and 'How is it paid', questions that may be commonly asked about a benefit. Under each of the headings, the subject has an introductory paragraph. The introductory paragraph usually leads into separate sentences that define the different cases for claiming or not claiming a benefit. This format is clearly shown in Figure 4.3.

The main style components used within the documents are limited to 4 main elements.

1. The main title is used to indicate the benefit that the following advice applies to. Titles are usually in large emphasised and bold Arial font.

2. Sub titles are used to highlight key areas of interest or question driven answers. It is common to see headings 'Who can claim?' which are then followed by paragraphs that answer the question of the title. The sub title is followed by a horizontal rule. In larger and more detailed documents the proceeding paragraph is indented from the left to leave room for citations or references to associated resources. The sub titles are in the same format as the main title but the font size is smaller.

3. Main body text contains little internal formatting or highlighting. In some rare cases key words are highlighted in a bolder font. The font used for the main text and citations is Arial.

4. There are a number of indented sections that are used to list a number of associated cases. These indented sections usually define rules for claiming a benefit.

The style of the paper based medium is also mirrored in the online medium, see Figure 4.2, which has a similar format for headings, sub heading and indented paragraphs.

4.2.3 Underlying sources

All benefits available in the UK have underlying legislation in place to regulate them. In the case of a legal decision being made, the legislation is authoritative over the advice documentation. It is important to remember that the advice documents are second hand representations of the rules and regulations controlling benefit.

The layout and structure of the legislative sources are very similar to the benefit advice documents. They follow the same sequence and in some situations use similar wording.

During the development of a knowledge representation opportunities should be taken to use the primary source of information such as legislation over secondary sources such as advice documents and handbooks. In the translation from legislation to

knowledge representation there are a number of steps that could introduce errors or misrepresentations and so the fewer such steps the better.

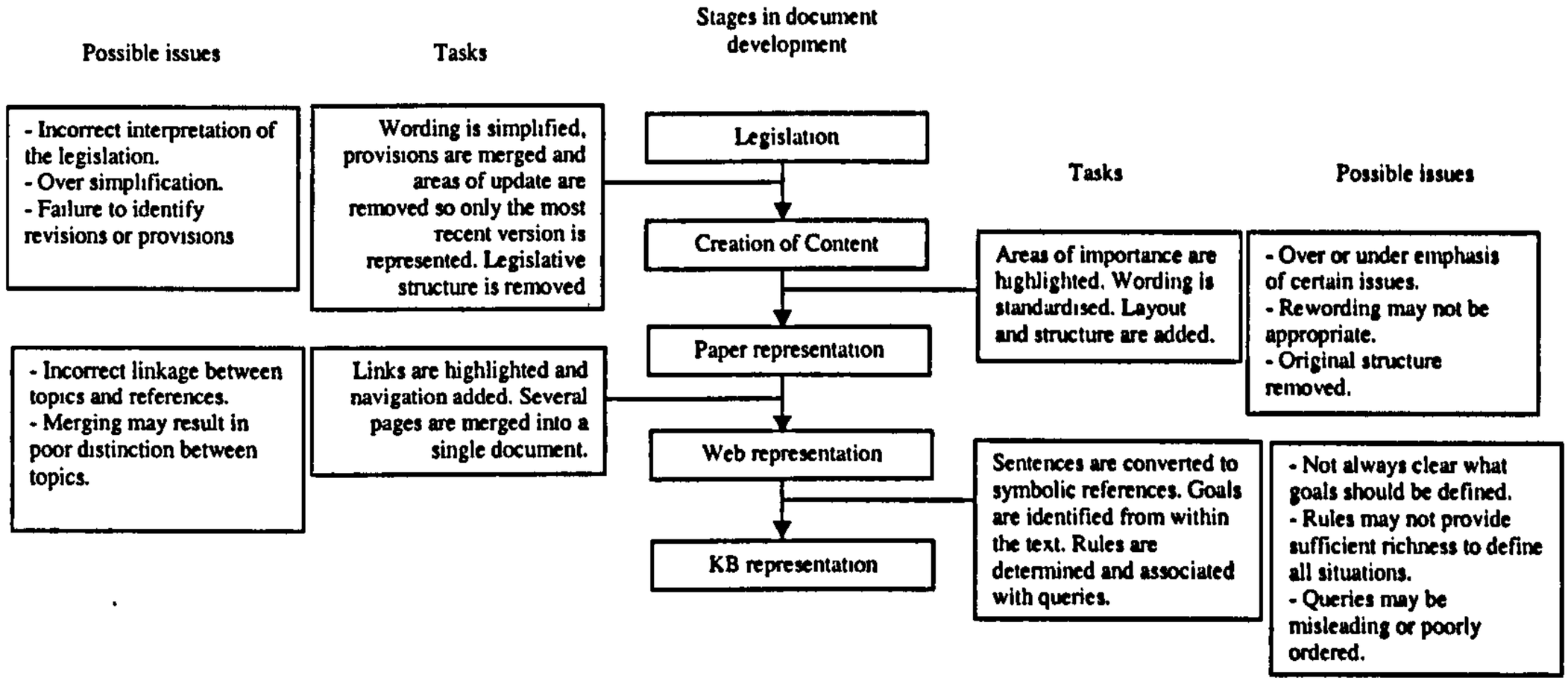


Figure 4.4: Layers of interpretation

At each step a new interpretation of the information trying to be presented provides opportunities for errors or poor interpretation to occur. By generating the KB representation from the source legislation, the chance of errors being introduced in the many conversions is far less likely. This is a long established principle in building legal KBS. See, for example, Johnson and Mead (1991) and Bench-Capon and Coenen (1992).

4.2.3.1 Example legislative sources

This section provides an example of benefit advice and the main legislative sources that it is based on. The example highlights some of the difficulties in interpretation of legislation and the advice based on it.

First consider this fragment of legislation relating to Child Benefit:

> *A person who is responsible for one or more children in any week shall be entitled, subject to the provisions of this Part of this Act, to a benefit (to be known as "Child Benefit") for that week in respect of the child or each of the children for whom he is responsible.* (Section 141, *Social Security Contributions and Benefits Act 1992)*

There are areas of uncertainty in the translation from legislative source to advice document. It is not necessarily the case that the advice is wrong, but the advice in some cases is no clearer if not more confusing than the legislation on which it is based.

The advice document (CH1) has the section

> *You may not be able to get child benefit if: ....*
>
> *or you (or your husband or wife) or your child have not lived in Great Britain for more than 26 weeks in the past 52 weeks ...*

This statement implies that if any party (husband and wife being assumed as parents) is outside the United Kingdom then you (the claimant reading the advice) may not be able to claim.

In contrast to this statement the legislation (Social Security and Benefit Act 1992) refers to this as

> *146 - (1) Regulations may modify the provisions of this Part of this Act in their application to persons who are or have been outside Great Britain at any prescribed time or in any prescribed circumstances.*
>
> *(2) Subject to any regulations under subsection (1) above, no child benefit shall be payable in respect of a child for any week unless –*
>
> *he is in Great Britain in that week; and*
>
> *either he or at least one of his parents has been in Great Britain for more than 128 days in 52 weeks proceeding that week.*
>
> *(3) Subject to any regulations under subsection (1) above, no person shall be entitled to child benefit for any week unless –*
>
> *he is in Great Britain in that week; and –*
>
> *he has been in Great Britain for more that 128 days in the 52 weeks proceeding that week.*

The statement in subsection 1 implies that regulations will apply if the person has been outside Great Britain (assuming the father has not been outside the UK therefore provisions will not apply to him)

The statement in subsection 2 implies that ... no child benefit shall be payable ... unless at least one of his parents has been in Great Britain for more than 128 days in 52 weeks. (lets assume the father has been in Great Britain!)

The statement in subsection 3 implies that ... no person shall be entitled to child benefit ... unless he has been in Great Britain for more than 128 days in 52 weeks. (the father has been in the UK for more than the required time!)

From the above statements and assuming that the father and child have been in Great Britain for more than 128 days in the 52 weeks proceeding that week then Benefit can be claimed. There is no requirement on the wife or mother of the child to be resident in the UK for any period before or during claiming of the benefit.

In contrast to the above statements, the advice document gives a different indication that if either spouse or child "you (or your husband or wife) or your child" has been out of Great Britain for more than the prescribed time then Benefit may not be claimed.

This is a single example of how it can be difficult to get the correct intention of meaning from Benefit Advice and an example of how second hand information could be misleading. Although the advice should be based on legislation it should also be easier to interpret otherwise it would be more beneficial and accurate for the reader to resort to interpreting the legislation directly. The discrepancies identified within the benefit advice show the importance of using underlying legislation to construct the knowledge representation used in QERIS.

Using QERIS to develop a knowledge representation of the advice while it is being authored provides a number of additional benefits. QERIS requires a mechanism for asking questions of the user and by generating a rule structure the questions and the goals for those questions need to be considered in more detail. Once goals and queries have been identified for the knowledge representation then the important elements can be fed back into the authoring process to improve the advice created. Errors in interpretation or translation from legislation to advice are easier to identify

because at each step rules can be written that break down the problem into smaller components. When rebuilt into a document these rules also provide a valuable mechanism for testing the advice under different conditions and in a controlled manner prior to publication.

4.2.4 Target audience

Benefit advice documentation has a very broad and diverse audience. The audience can be partitioned into a number of main groups. The main groups will change depending on the advice document being accessed as some advice is highly specialised. Each reader of the advice will have a specific goal for finding, reading and then interpreting the information. There are four main groups of people that have slightly differing goals when accessing benefit advice these can be defined as:

Advice – This group or readers does not know what benefits they can claim if any, and is using the documentation as an initial source of advice, possibly before getting more advice about a benefit or to get some background understanding. Even if the documentation answers the reader's questions then it is likely that they will seek further advice from an advisor. This group must be aware of which documents are associated with their particular situation. They also require awareness of which benefits they may be entitled to. The online benefit advice guides provide some support with identifying these by providing event driven navigation to benefits, e.g. advice aimed at those who have recently become unemployed.

Information – This group of people already know which benefits they can claim, and are accessing the documents for additional information such as the rate of a benefit, to determine how it will be paid or to familiarise themselves with other elements of the benefit. It is likely that once benefit information is explained to a claimant by an advisor then they would be given the related benefit advice documents to read in their own time.

Confirmation – This group of readers has been advised elsewhere but wishes to confirm what was said to them. They want to read particular areas of interest or concern to confirm from a second source that the information that was provided to them was correct. It is likely that they will make their own interpretation of the

advice provided in the documentation and may seek further advice from an advisor as and when questions are raised. There are many areas of the advice that are not straightforward to interpret and sometimes the groups that these difficult and specialised areas of advice are focused on will also find it harder to apply their own interpretation. A good example of this is non-UK residents that may require an international language version of the advice and have a more complex personal situation.

Interest – There are also groups of people that access benefit advice for interest such as research, analysis, news and formal advice. Benefit agency advisors need to be completely familiar with the benefit documentation so they must at sometime read and understand the information available. As new Benefits are made available then information about them will be published, and media sources will access the information to further publicise the changes. These groups are not after advice, indeed they are not after the informational content for themselves at all. They access the benefit advice documentation to pass on the information in another form, to those who it may concern.

4.2.5 Audience Characteristics

As well as having different goals for reading benefit advice documentation, there are also distinct groups of people that have very different characteristics. Key characteristics for a large volume of the benefit advice documentation includes:

Age – Age is a very important component in nearly all benefit claims. Benefits related to pensions only apply to claimants over pensionable age. Benefits claimed on behalf of a dependant, such as child support are limited by the dependant's age. There are very important age brackets for benefits such as income support and Jobseeker's allowance, where school-leaving age becomes and important element in how the benefit is calculated.

Dependants – Many benefits are specifically designed to help the claimant if the have dependants. This is true if the claimant is a carer or a parent. Benefits such as child benefit vary depending on the number of dependants of specific ages that the claimant is responsible for. Other benefits have increases for dependants.

Employment – Several benefits are based on the claimants income and others are designed specifically for people out of work. Employment plays an important role in claiming benefits and many of the means tested benefits are dependent on the amount of income and assets that the claimant has.

Marital Status – A large number of benefits vary in relation to marital status. This characteristic is closely associated with dependants where a single parent is usually given a higher level of benefit.

Residency – Nearly all benefits are specifically for UK residents only, although there are a number of exceptions to this rule such as residents of European Union countries, or to cover temporary absences.

These characteristics change at different rates. It is important that when developing an advice system to keep in mind how characteristics change. When building a user profile it should be able to adapt to the different rate of change and apply a degree of prediction to appropriate characteristics. There are three main groups of characteristics that change at different rates.

Constantly – Age changes constantly. Although the main stages of age such as 16, 18 or 60 occur at very predictable times it is important to be ready for the transitions and apply for the benefits accordingly.

Intermittently – Characteristics such as Employment, Marital Status and dependants can change unproductively and very rapidly. Characteristics such as Employment can also have a dramatic impact on somebody's life and it is important that these people are supported as quickly as possible.

Infrequently – It is unlikely that somebody's residency situation will change over a short period of time: there are usually lengthy procedures that must be applied before these characteristics change.

By combining all of these elements the key groups of people can be identified. It is clear that even at this high level there are a large number of different user profiles that could apply. There are also less significant characteristics that apply for particular benefits such as assets, health or housing. When combining all of these

different characteristics a good profile of the user can be built that can then be applied across all benefits.

4.3 Tailoring

In order to provide the benefit advice tailored to the user, information will be added, removed, highlighted or suppressed. The goal is to take an existing document and present it as a document but with added information in the textual styling and the content. Because QERIS provides tailoring mechanisms by manipulating the underlying mark-up content it is possible to create many different presentation tailoring mechanisms. It is even possible to integrate dynamic elements such as JavaScript, DHTML or XML into the content to provide even more advanced tailoring and presentation strategies.

There are three key mechanisms that will be employed in the following examples to improve page representation:

- Graphical manipulation will provide a means of highlighting and suppressing information.
- Content manipulation will provide a means of helping the reader interpret the information they are accessing.
- Navigation manipulation will provide a mechanism for associating documents and resources dynamically.

The following sections will focus on a simple example page and demonstrate a number of different approaches that can be used to improve the presentation and interpretation of the information.

The child should be:

under age 16;

or under age 19 and still in full-time education of at least 12 hours a week at a recognised educational establishment, studying for a qualification up to and including A level/(G) NVQ level 3 or an equivalent.

4.3.1 Graphical Manipulation

A page is represented as a picture by processing the underlying HTML content and rendering the textual and graphical elements within a window of the browser. Modifying the underlying HTML will have an impact on the graphical presentation of the page. There are a number of graphical effects that can be used to provide emphasis, focus, suppression and removal of content.

4.3.1.1 Content removal

Textual and graphical components can be completely removed from the document by wrapping them in the HTML comment tag:

```
<!-- ... -->
```

So if the system determines the age of the child as 14, then the example section could be rendered as:

The child should be:

under age 16.

but if the childs age is 18 then same text can be rendered as:

The child should be:

under age 19 and still in full-time education of at least 12 hours a week at a recognised educational establishment, studying for a qualification up to and including A level/(G) NVQ level 3 or an equivalent.

In a similar manner, if the child were aged 19 or over then the entire content would be suppressed.

4.3.1.2 Content Suppression/Emphasis

An alternative to content removal is content suppression. The intention is not to remove but to make the information less dominant on the page. There are two approaches that have been considered.

1. Change the font properties so that the text does not attract as much attention as other areas of the page.
2. Replace the content with a link that still provides access to the underlying content. This can be done dynamically using JavaScript and DHTML so that the content can be 'folded' or by inserting a URL to another document.

Based on the last example of the previous section, instead of removing the non relevant text, it is wrapped using the HTML style tag:

<FONT COLOR=gray> ... </FONT>

resulting in the following representation of the page:

The child should be:

under age 16;

or under age 19 and still in full-time education of at least 12 hours a week at a recognised educational establishment, studying for a qualification up to and including A level/(G) NVQ level 3 or an equivalent.

The advantage of suppression is the original source information is still avaialble for the reader to access, but it should be appreciated that this is background information. The original format of the content is unchanged, only the visual rendering of the document is modified. It is possible to apply this modification in several ways such as font size, colour or face.

The opposite of suppression is emphasis: it is possible to highlight both sections or even emphasis sections within non-suppressed sections to further highlight key elements. A child of age 18 who is studying A-levels would result in a fully suppressed and emphasised page may be represented as:

The child should be:

under age 16:

or **under age 19** and still in full-time education of at least 12 hours a week at a recognised educational establishment, studying for a qualification up to and including **A level**/(G) NVQ level 3 or an equivalent.

By combining removal, suppression and emphasis a different presentation of the page can be presented where information relevant to the readers circumstance is displayed while less relevant information is moved into the background, or removed altogether.

4.3.2 Content Manipulation

In additional to changing the graphical rendering of the HTML content it is also possible to modify the actual content that underlies the document. As well as using the HTML comment tags, it is possible to physically remove the content. In addition there are a number of other content changes that can be applied:

- Insertion of additional HTML content.
- Insertion of values entered by the user.

These mechanisms can be used to further improve the presentation and guide the interpretation of the reader.

4.3.2.1 Inserting HTML content

Consider the last comment from section 4.3.1.1, where the child is aged 19 or over. This would result in the complete removal of the section. This is not a good thing to do without providing some additional feedback to the reader. Insertion of content is applied directly by QERIS effect tags. For example the section:

The child should be:

under age 16;

or under age 19 and still in full-time education of at least 12 hours a week at a recognised educational establishment, studying for a qualification up to and including A level/(G) NVQ level 3 or an equivalent.

Could be represented as:

The child should be:

If your child is aged 19 or over then they are too old for Child Benefit to apply

This representation is a combination of content removal and insertion. By removing all of the information that is not relevant to this claim and replacing it with a sentence that explains that the customer cannot claim a tailored presentation is provided for that customer.

An alternative approach to this presentation is to use suppression and other inserted indications, where the child is aged 14 the following presentation may be used:

The child should be:

✓ under age 16;

× or under age 19 and still in full-time education of at least 12 hours a week at a recognised educational establishment, studying for a qualification up to and including A level/(G) NVQ level 3 or an equivalent.

You have indicated that your child is 14 years old, therefore the first case applies. You may proceed with this claim.

This provides indicators to show which of the cases are not met and a final statement that confirms which case has been met. In additional to providing these added comments the core content of the document is still available so that if the reader wishes to confirm the advice provided by the backend system, they have all the material to hand to still make their own interpretation. The answer provided by the user is also used in the confirming paragraph to further clarify the situation. Navigation Manipulation

In the last example, a URL is added as part of the conclusion. By inserting HTML content it is also possible to insert links to additional resources. The × symbol is also wrapped with a URL that links to the actual inference engine and requests additional information and confirmation about the rules used to reach this interpretation.

Depending on the content inserted, links to the next page or links to help pages can be added dynamically depending on the interpretations applied at every step.

4.3.3 Knowledge Base Representation

The knowledge base (KB) consists of the rules and mechanisms that are applied to change the presentation and content of the document. The KB is very closely related to the document content and in many cases is very simple, reflecting paragraph or sentences directly. The KB size is related to the complexity of the logic within a document as well as the degree and resolution of tailoring. There were several aims when developing and using the KB representation.

It should be expressive enough to be useful in a wide range of applications.

It should be easy enough to learn, create and understand.

The knowledge representation should not grow too large.

The knowledge representation should integrate with the underlying HTML mark-up.

The final representation used for creating a prototypical example is based on simple production rules. In the case of Benefit Advice documents many of the individual sentences can be identified as atomic elements. The structure and style of writing closely resembles a rule-based structure. Reading through the document can identify atomic elements that can be used to construct a rule. Each atomic element or rule is allocated a label. For example the sentence

"A customer who is responsible for a child living with them"

can be given a symbolic reference:

"R1"

Or a more expressive label:

"responsible_for_child_living_with_them"

It is important that if longer labels are used they are correctly spelt and the capitalisation of the string matches in every case the rule is referred to. It is also

possible to break sentences down into small components to provide more refined feedback. The above sentence could be broken into the following two sub elements.

"Responsible for a child"

"Child living with them"

The sentence may be broken in this way if the knowledge base engineer would like to query the user on these subjects independently or if in the final representation of the document these two sections of the sentence are treated separately.

4.3.3.1 Rules and Facts

Once individual components have been identified they can then be combined using rules. If the sentence is associated with a conjunctive term then they can be combined directly. Each atomic element is not described in the knowledge representation but within a rule. The following rule depends on the state of the elements E1 and E2.

```
R1 = E1 and E2
```

States of elements can be assigned directly, a fact can be represented by the representation

Each fact has:

- An identifier that is used within the representation to refer to the fact.
- A condition that contains the identifier of the element that should be assigned a state.
- An assignment, which contains the identification of an element and the assignment that it should be given.

```
<RULE ID="E1"

CONDITION="E1"

CONCLUSION="E1 is true">
```

Rules can be constructed from elements and they have:

- An identifier that is used within the representation to refer to the rule.

- A condition that contains a sequence of elements and connectives that associate them.
- An assignment, which contains the identification of an element and the assignment that it should be given

```
<RULE ID="R1"
CONDITION="E1 AND E2"
CONCLUSION="R1 is true">
```

Condition strings can include element identifiers, facts or rules; and connectives including AND, OR, EQUAL, IMPLIES.

Conclusions must include an element identifier, the string "is" and then a value. The value can be any string such as "blue", "yes", "true", "John". These values can then be checked within rule conditions

e.g.

```
<RULE ID="R1"
CONDITION="E1 AND E2"
CONCLUSION="R1 is value_of_R1">


<RULE ID="R2"
CONDITION="R1 is value_of_R1"
CONCLUSION="R2 is true">
```

If no extra terms are used for an element within a condition then the default values "true" or "false" are checked for.

4.3.3.2 Queries

If an element is reached that has not been defined then the state of this element must be determined. If no state can be determined then "false" is assumed. States for elements can be collected from the user by queries. A query is designed to interact with the user and collect information from them. A query is constructed from:

- An identifier that is used within the representation to refer to the query, which should match with another element used with a rule.
- A question that is presented to the user as a text string within a query dialog.
- An answer or set of answers to the question, which is the value assigned to the element indicated in the identifier.

```
<QUERY ID="E2"
QUESTION="What should the value of E1 be set to"
ANSWER="true,false,blue,yesterday">
```

For the above query, the following page will be presented to the user:

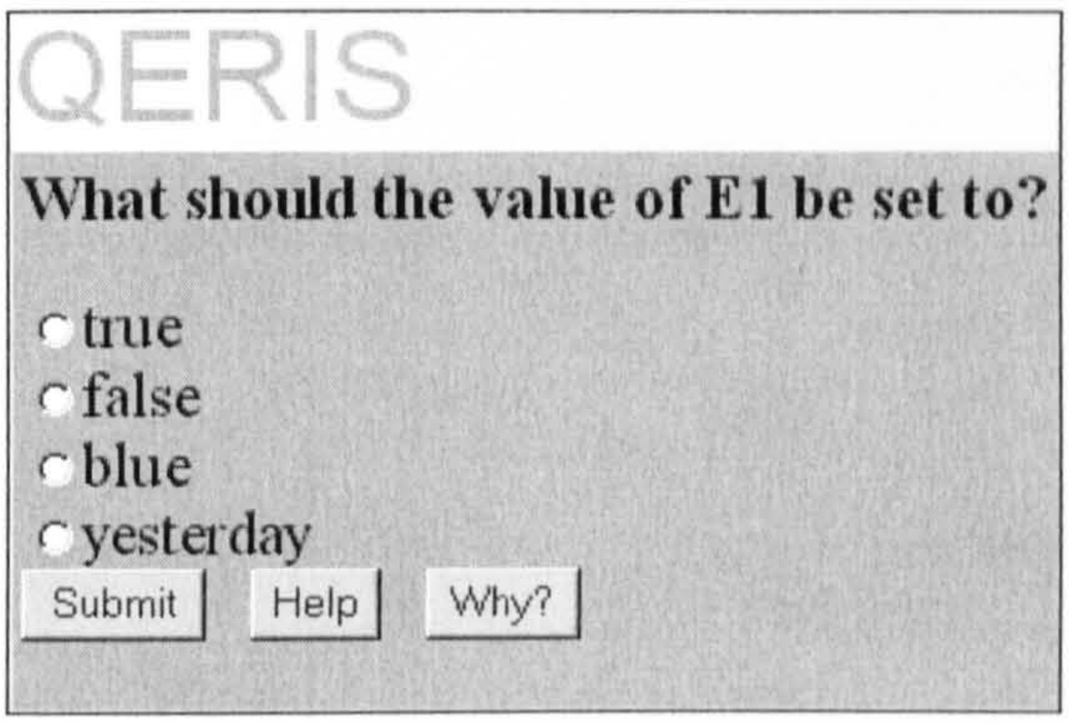


Figure 4.5 *Multi choice question dialog*

If no answers are provided in the NASWER attribute then an open text box is provided so an element can be given any value.

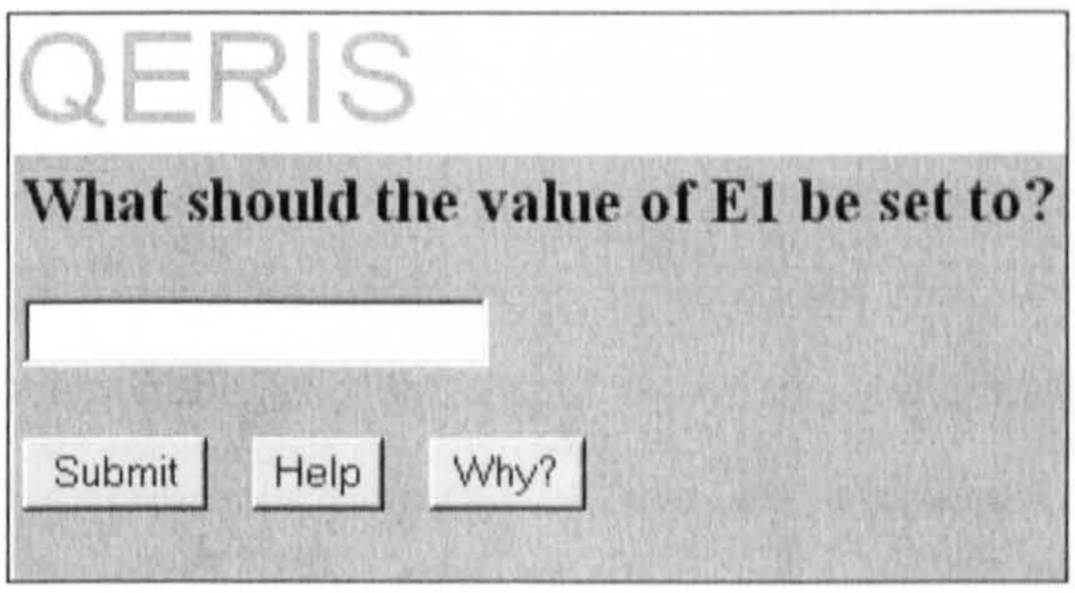


Figure 4.6 *Open question dialog*

Answers provided by the user are stored in the knowledge base as facts where the query element is assigned the value selected or entered by the user.

These simple representations within the knowledge base, the Fact, Rule and Query, provide a mechanism for creating complex interactions and a backward chaining rule representation of the document.

4.3.3.3 Effects

In order to apply alterations to the document a third mechanism is used, an effect. Unlike rules and queries an effect has a position within the document and instead of assigning a value to an element, it triggers a plug-in component that alters the underlying document mark–up.

An effect consists of:

- An identifier that is used within the representation to refer to the effect.
- A condition that contains a sequence of elements and connectives that associate them. This condition is the same as those used in rules.
- A plug-in name that indicates the function to apply to the document. There are a number of default plug-ins, or plug-ins can be document specific and downloaded and integrated with the processing engine on demand.
- Arguments to the plug-in. There are two arguments, a True argument that is applied if the condition evaluates to true and a False argument that is applied if the condition evaluates to false.

```
<EFFECT ID="E10"

CONDITION = "R1 is value_of_R1"

PLUGIN="insertString"

TRUE="<B>This will be inserted if R1 is <A
HREF='value_of_R1'>value_of_R1 </A></B>"

FALSE="<B>This will be inserted if R1 is <A
HREF='something else'>something else </A></B>" >
```

Any HTML mark-up or content can be included in the argument string and will be inserted into the document at the location of the original effect tag.

Currently the plug-ins available include:

insertFile – inserts a complete file containing either plain text or HTML mark-up. The attributes TRUE and FALSE should be set to the URL of the file to insert.

insertString – insert a text string at the current location within the document. The attributes TRUE and FALSE should be set to the strings to insert when these cases occur.

insertResult – inserts the value of an element into the document. Inserts the value of the condition for this effect. So if the condition is "age is 14" then the string "14" will be inserted.

removeLines – removes the indicated number of lines from the document.

return – terminates the consultation for this document. This can be used to immediately return the document and is useful for terminating consultations quickly.

In the following example the document will be broken down into components, which will be represented as elements and rules. The goal of the example is to query the user and highlight within the document text strings that support their claim and text strings that do not support their claim. General conclusions shall also be provided. All this is based on the information given to the system.

4.4 Example implementation

This section will describe a number of steps in developing a knowledge base for a benefit advice document. The main Benefit document that will be discussed is the Child Benefit advice document. Several example statements will be explained and the rules, queries and effects used to implement a new presentation will be explained.

4.4.1 An active session

The image in Figure 4.7 shows an active session where a user is accessing information about Housing Benefit. Although not visible on a black and white print, the page displayed shows some sections of text shaded red and other sections shaded

green. These indicate areas of the requirements that have and have not bee met by the user. In the DOS window it can be seen that several effects have been activated while others have been ignored. If an effect is activated then its TRUE action is applied to the document. If an effect is ignored then it FALSE action is applied. The example shows a local execution of the QERIS engine: it is possible to configure the engine to run locally or remotely, and it can also handle more than one user request at the same time.

Figure 4.7 *Screen shot of QERIS components*

When using the QERIS program, there are 4 important components that can be used during the session. QERIS is started using a standard windows DOS batch file to invoke the Java VM and start the main proxy server process. The DOS window that

is displayed shows in real time the activities of the proxy server and also any diagnostic functions that have been activated. The QERIS main panel provides quick access functions for controlling the proxy server from clearing the fact store to resetting all the active threads. These functions can be accessed off the pull down menus. There is also a small icon on the display that shows if any sessions are currently active. The hierarchical rule display is a graphical representation of the rules that can been processed by the engine. It is possible to see how the rules are associated, which rules have been proven and which queries have been asked. It is possible to select a rule and remove it from the display, which will also remove it from the knowledge base. This is useful for trying out different situations or for forcing the system to re read the rule from the next document that contains it. Although not implemented in the prototype it is proposed that this display could also be used to add or modify rules prior or during a consultation. Any internet browser can be used that supports HTTP proxy servers, and it is also possible to display questions in the internet browser as HTML pages although this requires multi-part MIME support so the session can be kept active and the questions can be pushed to the browser.

4.5 Child benefit (CH1)

The first case to be examined is the Child Benefit leaflet CH1 that provides guidance on claiming child benefit. This leaflet references 13 additional leaflets and is designed as a general guide only. The Child Benefit guide is relevant to everybody with children in the UK because it can be claimed irrespective of income level.

4.5.1 The document

The Benefit document has some important features:

- It is relevant to a broad section of society, because the benefit can be claimed by people responsible for children and is not based on an income.
- It includes characteristics of 2 people, the claimant and the child. This provides a high level partitioning of the document.

- It is question driven, since the document itself contains the questions that need to be asked of the user. It is therefore a case of identifying the questions and formulating them in a way that is easy to understand and follows a logical flow.
- It has many compound structures, A **OR** B, which can be used as logical blocks that can be defined in the knowledge base. The document can be broken down into manageable pieces that can be combined as a consultation progresses.
- Although a simple subject, the advice document is not easy to interpret in several places.
- It includes several pro and con statements for claiming child benefit, which can be used to build a supportive case or a counter case for claiming the benefit.
- Child Benefit is an important benefit and has several provisions associated with it, see section 4.2.3.

4.5.2 Example statements

Within the benefit document are a number of key statements that indicate if a claim for support is valid. There are some issues with the interpretation of some of the statements that highlight the issues with understanding, interpretation and relevance to the reader. There are also a number of questions that can be asked and issues raised with the way the advice is presented. For example:

Sponsorship

> *Can claim* if their child age 16 or 17, has recently left school and is registered for work or training at a careers office.
>
> *But cannot claim* if the child is on a training course sponsored by an employer or the Training Enterprise and Education Directorate.
>
> There is no mention of how much the child is being sponsored for and there appears no incentive to seek sponsored training. Also the term 'registered for work or training course at a careers office' is ambiguous. Does this mean that the child is doing work/training and has registered the fact or does it mean that the child has registered in order to seek work/training?

Training from home

Can claim if aged 16 or 17, has recently left school and is registered for work or training at a careers office.

But cannot claim if now studying at home, but was at school or college until the age of 16.

Does this mean they cannot undertake a training course where the training is done at home?

Resident in the UK

Can claim if from the UK but not if your spouse has not lived in Great Britain for more than 26 weeks in the last 52 weeks.

A claimant receiving single parent child benefit, marries a non UK national who has just moved from abroad, the claimant is no long able to receive child benefit for up to 52 weeks. There is no mention of the spouse's financial status or ability to support the child during this period.

4.5.3 Example Rule Structures

4.5.3.1 Basic Rules

By examining the Benefit document for Child Benefit the following atomic rules can be determined. These rules represent the smallest breakdown of each sentence and can be combined in several ways to represent the original text.

Rules that support claiming child benefit for the child (or)

Can claim if the child is

P1 – under 16

P2.1 – aged 16, 17 or 18

P2.2 – studying full-time up to and including A-level equivalent standard

P3.1 – aged 16 or 17

P3.2 – recently left school

P3.3 – is registered for work or training at a careers office.

Take the atomic rules and associate them based on their linkage within paragraphs.

P4 – If age < 16

P5 – If age < 19 and age > 15 and P2.2

P6 – If age < 18 and age > 15 and P3.2 and P3.3

The rules that represent a paragraph can be combined into a single rule, P7, which can be used to help interpretation of the text as it indicates all the supportive cases in claiming. If P7 cannot be proven then it is likely that the person may not be able to claim.

P7 – If P4 or P5 or P6

Rules that do not support the claiming of child benefit for the child

Cannot claim if the child is

C1 – on an advanced course of education higher than A-level.

C2.1 – now studying at home.

C2.2 – was at school or college until the age of 16.

C3.1 – aged 16 or over

C3.2 – has left school

C3.3 – has started work of 24 hours or more a week for which payment is received or expected.

C4 – on a training course sponsored by an employer or the training Enterprise and Education Directorate.

C5 – getting severe disablement allowance

C6.1 – getting Income Support

C6.2 – getting income based Jobseeker's Allowance

Take the atomic rules and associate them based on their linkage within paragraphs.

C7 - C2.1 and C2.2

C8 – C3.1 and C3.2 and C3.3

The paragraph rules can be combined into a single rule, C9, that indicates all states which do not support the claiming of child benefit. If C9 can be proven then the person may not be able to claim.

C9 – if C1 or C7 or C8 or C4 or C5

The two high-level rules that include all supportive and non-supportive cases can be combined into a rule to determine if a person can claim Child Benefit or not.

E1 - If P7 and not C9

These rules can be used to determine statement by statement, which apply to the child and which provide problems in claiming the benefit. The rules follow the flow of the document and an expert consultation would follow the sequence of steps laid out in the document. If each combined rule is associated with an effect wrapped around the referenced sentence then the sentences keyed to each rule combination can be modified to indicate a supportive or non-supportive case. For example, the following tags:

```
<EFFECT ID="E1"

CONDITION = "P7 is true and C9 is false"

PLUGIN="insertFile".

TRUE="can_claim.txt"

FALSE="cannot_claim.txt">

<EFFECT ID="E2"

CONDITION = "P7 is false"

PLUGIN="insertFile"

TRUE="no_support.txt">

<EFFECT ID="E3"

CONDITION = "C9 is true"

PLUGIN="insertFile"

TRUE="non-supportive_cases.txt">
```

The first effect identifies if a claim is possible, while the other two effects provide a break down to indicate which answers did not provide a supportive case. This can be expanded to include all combinations of answers supplied and can be used to provide an answer to a "why" query.

The above would generate the following HTML segment variations.

Supportive example

<P>
Based on the answers you have provided, there is evidence that you can claim for child support.
</P>
<P>
<A HREF=CBclaimForm.html>Steps to proceed with your claim</A>
</P>

Non Supportive example

<P>
Unfortunately based on the answers you have provided, you may not be able to claim child support
</P>
<P>
There were <A HREF=nonSupportAnswers.html>answers</A> that did not support the claim.
</P>

4.5.3.2 Category based rules

More complex structures can be built based on combining rules for pro and con into single, supportive and contradicting structures. By combining rules in this manner, users can interact with the system based on their desire to know what characteristics they have and what characteristics are lacking in order to claim.

By having a dummy effect that applies no changes to the underlying document the sequence in which questions are asked can be more finely controlled.

There are many rules in the above listing that can be grouped into categories or topics. By grouping rules such as those related to age it is possible to generate a simpler set of queries that focus on this topic before moving onto asking additional information.

Age based

P1, P2.1, P3.1, C3.1

Education Based

P2.2, P3.2, C2.1, C2.2, C3.2

Training Based

P3.3, C4

Work Based

P3.3, C3.3, C6.1, C6.2

Miscellaneous

C5

These rules can be combined in the following higher level rules.

(P1 – under 16 **or**

(P2.1 – aged 16, 17 or 18 **and**

P2.2 – studying full-time up to and including A-level equivalent standard) **or**

(P3.1 – aged 16 or 17 **and**

P3.2 – recently left school **and**

P3.3 – is registered for work or training at a careers office.)) **and not**

(C3.1 – aged 16 or over **and**

C3.2 – has left school **and**

C3.3 – has started work of 24 hours or more a week for which payment is received or expected.)

So for the situation above, a dummy effect can be created in order to fully process the age rules so controlling the sequence that questions are delivered in. It should be noted that effects are processed in the order that they occur in the document. The tags could be written as:

```
<EFFECT ID="age"
CONDITION = "age is under 16 years or
                    age is 16-17 years or
                    age is 18 years or
                    age is over 18 years"
PLUGIN="dummy">
 <RULE    ID="P1"    CONDITION="age    is    under    16"
 CONCLUSION="true">
 <RULE ID="P2.1" CONDITION="age is 16-17 years or age is
 18 years"
CONCLUSION="true">
 <RULE ID="P3.1" CONDITION="age is 16-17 years"
CONCLUSION="true">
```

```
<RULE ID="C3.1" CONDITION="age is 16-17 years or age is
18 years or age is over 18 years"
CONCLUSION="true">
<QUERY ID="age"
QUESTION="Is your child aged"
ANSWER="under 16 years,16-17 years,18 years,over 18 years">
```

The above tags will force all age related rules to be processed and the corresponding query asked of the user. Once the facts are set processing can continue into the next subject area.

The statements P 3.X and C3.X and then the requirements "aged 16" and "left school" appear in both supportive and contradicting arguments. Also the biggest element in this section is the distinction between a child of 16 or over registered for work and a child being paid for work. By writing questions that address as many of these cases as possible in a single set of initial queries there will be fewer prompts to the user later in the consultation.

4.5.4 Document to Reader interaction

In order to generate the tailoring of the document, the characteristics of the reader need to be captured. There are three main ways of identifying information about a reader:

1. The most common method of capturing new information from the reader is by asking a set of questions.
2. In some cases old information may be available and can be reused. As the user interacts with a system, it can learn about them and build up a profile. Information reuse can reduce the number of potential questions but can become unreliable over time as the reader's characteristics change.
3. Information can be retrieved from external sources. For example, if the user enters their postcode it is possible to identify their full address and from this, their home telephone number.

Questions can be presented in a number of ways. A question can be asked in a number of ways depending on the style and capabilities of the underlying inference engine.

Common single question styles include:

Open questions

What is the age of your child?

Closed questions

Single Selection, single choice

Is your child 16 years old?

Single selection, multiple choice

Is your child younger then 16, Aged 16, older then 16?

As well as single questions, a set of single but associated questions can ask in the form of a questionnaire.

Asking the reader a set of 4 questions can create the tailored example in figure 2.

For example:

1.	Is you child aged	< 16 years?
		16 –17 years?
		18 years?
		> 18 years?
2.	Is your child still at	School?
		College?
		Learning at home?
3.	Is your child	In full time education?
		Register for training with the careers office?
		Registered for work with the careers office?

4. Is your child studying upto A-level, standard?
   CSE/GCSE?
   AS-level/A-Level?

These questions can act as prompts for the user to focus their attention on a single element at a time. It is also possible at every query to provide more detailed help. There are a number of ways of achieving this:

1. Each question can have a help page associated with it using the HELP attribute of the QUERY tag. If a URL is provided as the value for this tag then the page is displayed when the help button on the query is pressed.
2. It is possible to construct rules and queries in such a way that in the case of a failure further questions are asked to identify if the user really understood what was being asked. By having an option "other" or "unknown", this can be identified and more information could be provided or a more detailed or rephrased question asked.
3. Queries can be completely modified and a HTML template created to represent the query page displayed. It is possible for complex questions to have detailed explanations about each choice the user can make.

4.5.5 Example consultation

The child benefit documentation is accessed and a number of questions are asked of the user in order to tailor the presentation specifically to them. The following consultation will focus on the rule described in the previous section demonstrating how the interaction will proceed and the possible outcome of answers to the consultation. From the user's point of view they will see a sequence of queries like those shown in Figure 4.8. As questions are answered, new questions will be presented until a full picture is available from which to develop a new document presentation. In some cases questions will be asked to provide presentation based tailoring, in other situations questions will be asked in order to help provide additional interpretation. The question diagram is a small section of a larger consultation that will be undertaken across several separate yet associated

documents. There is a document for single parents and one for people from abroad. Each of these documents has their own knowledge representation, which will reuse some of the questions provided in this example. All answers to questions associated with this document section can be reused in later consultations. For example, selecting to proceed with the claim may lead to a form for submitting a claim, and many of the answers that were provided during the consultation can be used to help fill in the form.

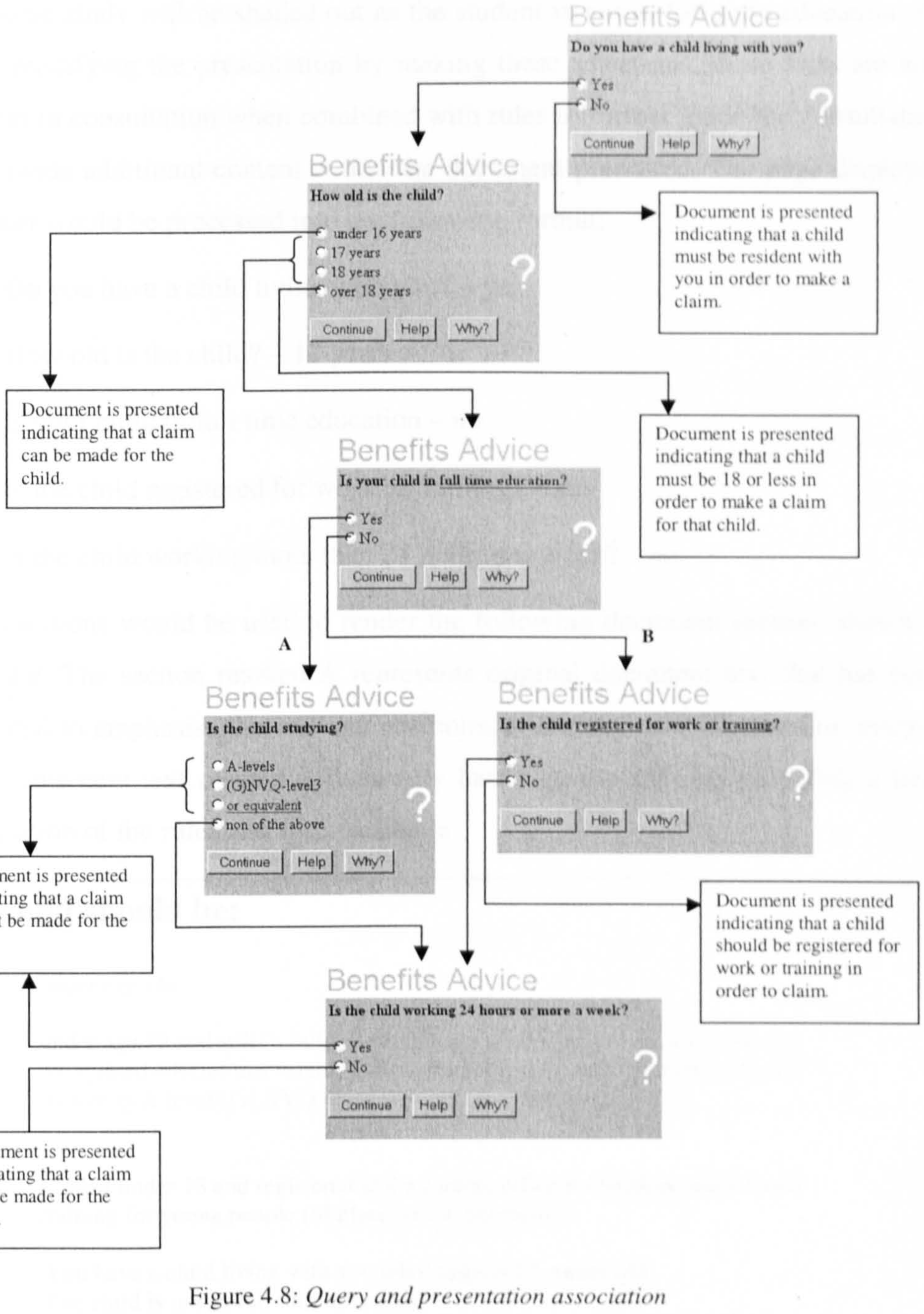


Figure 4.8: *Query and presentation association*

In particular situations such as "Is the child in full time education" the answer to this query will result in different degrees of formatting taking place. For example in Figure 4.8, selecting option A would result in the sentence associated with registering for work or training being shaded out as it will not apply if the child is in full time education. In contrast if option B is selected then the sentence associated

with course study will be shaded out as the student is not in full time education. As well as modifying the presentation by making these selections, these facts are also used later in consultation when combined with rules to further guide the consultation or to provide additional content within the document presented. The page displayed to the user would be processed into the following format:

Do you have a child living with you? – yes

How old is the child? – 17 years

Is your child in full time education – no

Is the child registered for work or training? – yes

Is the child working more than 24 hours per week? – no

These questions would be used to render the following document section, shown in Figure 4.9. The section marked **A** represents original document text that has been highlighted to emphasise the relevant attributes of the claimant. The section marked **B** represents new text created dynamically based on the answers providing a final interpretation of the rules and facts available.

The child should be:

under age 16;

or under age 19 and still in full-time education of at least 12 hours a week at a recognised educational establishment, studying for a qualification up to and including A level/(G) NVQ level 3 or an equivalent. } A

or Is aged under 18 and registered at the careers office for work or work-based training for young people (Skillseekers in Scotland)?

- You have a child living with you who's age is **17 years** old.
- The child **is not** in full time education
- The child **is** registered for work or training
- The child **is not** working 24 hours a week or more. } B

Based on these answers you should claim child benefit.

Figure 4.9: *Tailored document representation*

During the writing of this thesis the presentation and wording of the Benefit documents has changed. The examples given here are based on benefit advice documents available between September 1998 and September 2000. The advice provided over the internet is also simplified and does not contain the same details as information in paper based benefit advice guides.

4.6 Evaluation and Feedback

The Benefit Advice example was used to provide a demonstration of the QERIS system to different users in order to receive feedback and comments. Due to issues with the implementation of QERIS it was not appropriate to undertake a formal analysis of the approach. The issues included:

- Distribution and configuration required to provide interaction with the browser. The University network employed a firewall, which caused several problems to occur when accessing external sites. QERIS was modified to overcome redirection through a proxy after the demonstrations had been carried out.
- The Interface was created as a test and debug interface rather than being as intuitive. The interface would require further work to improve confidence in its user aspects.
- The system generates a lot of debug and trace information, prior to a full demonstration this should be suppressed so that performance and ease of interaction are improved.

Feedback was provided by users who had the opportunity to interact with the system while under the guidance of the author. Requirements for a more detailed study of the system will be described later in this section and represent further research that could be undertaken.

The aim of the evaluation was to indicate if the information presented was an improvement on the original representation. The Benefit example proved a good subject due to the poor quality of the original online format.

The following environment (Figure 4.10) was configured prior to evaluation.

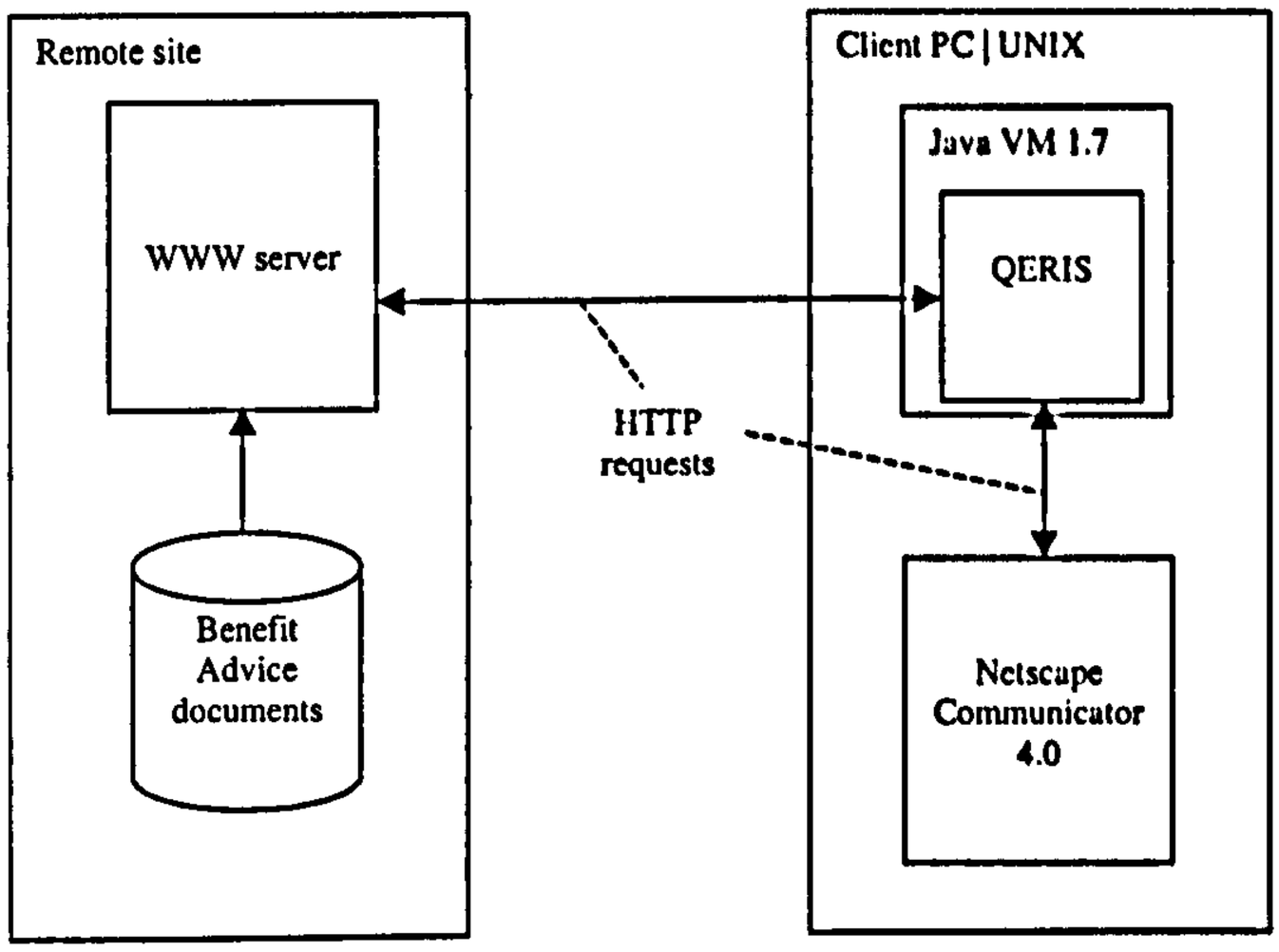


Figure 4.10 System Architecture

4.6.1 Professional feedback

Mr T. Craven, Advice Officer. *Benefits and Advice Service, Central Policy Unit*

The system was described and demonstrated to Mr Craven to provide a good insight into how it functioned and the results that could be achieved. Many aspects of the system were discussed, and positive feedback was provided. The demonstration was conducted in October 1998.

This section will list the main discussion points and then provide a brief explanation of what steps were taken to respond to the comments. The main comments include:

> *"Who is the system targeted at?" Is the system aimed at the client, employees providing advice, employees reviewing advice or those evaluating claims?*

At the time the demonstration was given, it was unclear if the research was to be limited to a particular group of users. The system was developed to be as general as possible and accessible by a very wide audience.

> *It is important to keep the approach simple. In many cases the people accessing information sources for Benefit Advice have difficulties using the computer even before they start using the software. "A solution should prompt and guide without becoming a choir".*

A key goal of QERIS was to provide a simple interface and knowledge representation to make the system approachable by as many users and developers as

possible. The interface in many cases is transparent to the user. The user only sees a sequence of questions followed by the tailored document they requested. The knowledge representation is very basic so that learning and combining the knowledge representation with a source document is a straightforward as possible. The difficult aspect is the design of the rules and their associations.

> *Reach a balance between providing useful guidance while limiting the amount of information the user has to provide. It is important to maintain the user's privacy and protect their information.*

The questions asked are the minimum set required to reach the goals listed in the document. If the goals are broken down into small well-defined blocks it is possible to ask a few questions per document and re-use facts and rules that have been associated with the user from earlier consultations.

- *There are already plans for providing computer terminals in Benefit Offices and such a system would be appropriate for such a location. It is important thought to consider where the user's information will be stored.*

An aspect of QERIS is the association of facts with a user in order to build up a persistent representation. The QERIS model does not work well in a public environment where many users will access a single information point. There are solutions to this situation such as:

- Losing the advantage of long-term persistence and requiring the user to enter their details for each session.

- Allowing the user to store their information on a central data-store that they can access by username/password.

- Storing the user information on a smart card or floppy disk that they can then take with them (and even access the same resources from a different location but with the same user model).

> *Another area that would be appropriate for automated support is Debt advice (a document was provided to demonstrate the type of advice currently provided. Dealing with your Debts, Money Advice Trust, 1998). There are a large number of users that make inquires in this area and at the moment the amount of advice is limited. The volume of advice available is high because of the many different types of debt but given very limited information it is usually possible to limit the amount of advice that is relevant.*

No demonstrations were developed based on Debt advice but the flexibility of QERIS certainly makes the development of an interactive advice system possible. The changes in a persons financial status could also be taken into account over a longer period and the advice would change based on the remedies and actions taken.

> *An area which proves difficult is providing a multi lingual system. It is important to keep in mind the many different language's that the advice is published in.*

This is not a key focus of the QERIS system but the most direct approach would be to use the content modification techniques to retrieve and combine document sections from database sources. By asking an initial question about the language to use, this fact could then be stored in the users profile and any further processing carried out in the selected language. This is certainly a section of the research that would benefit from further work.

4.6.2 User feedback

Connect Summer Courses (Retraining program for unemployed Professionals)

During the three month period of the course, the QERIS solution was informally demonstrated to a number of the students and was accessible via the University of Liverpool UNIX network. The examples demonstrated included both Child Benefit advice and housing Benefit advice documents reformatted and containing QERIS markup content.

In many cases the people using the system were claiming or eligible to claim the benefits that were being demonstrated.

The focus of the demonstrations was how the system was used by the readers accessing the pages to determine if they could understand the queries being asked and provide feedback about the questioning mechanism. Both mechanisms for

interacting with the reader were presented. The local mechanism where the pages are displayed by the Java application in separate dialog boxes, and the remote mechanism where the queries are presented in the Internet browser.

When the queries were presented as dialogs from the Java application, the readers were in some case surprised and disoriented that another application other than the browser was interacting with them. Their initial response was that they had done something wrong. The dialog box was usually moved out of the way to see if a result had been displayed in the browser. The remote mechanism where dialogs were presented in the browser were preferred by the readers and they interacted with them much more quickly.

A surprising result of these demonstrations was the response that several of the people using the system were not aware of some of the benefits they could claim. Two of the readers successfully identified a benefit they could claim after using the brief and simple demonstrations. This was not specifically because of the interaction and guidance provided by QERIS but because they were not aware prior to the demonstration that they could claim the benefits and had not read the relevant benefit advice. This emphasis's one of the problems with accurately determining if the approach taken by QERIS can be evaluated successfully. There are many aspects that have no direct bearing on the actual functions of QERIS but which could dramatically effect a reader's interaction and opinion of the system. Aspects include:

- The layout and aesthetics of the HTML content and WWW pages, which can vary between implementations and even browsers, used to access the resource. The actual layout and design of the HTML must be good in order to benefit from the QERIS approach.

- The collection of facts and user attributes requires a lot of initial questions, which then lessen as a more complete set of facts are built up. This only really becomes evident after a longer period of using the system. Readers using the system for a short period may not realize the longer-term benefits.

- It is not a simple task to determine if the users goals have been met and if those goals have been met in a more appropriate manner than simply providing the textual content.

4.6.3 Approaches for further evaluation

Although the testing and evaluation of the system is limited, there are a number of areas and approaches that can be taken to evaluate the system further.

4.6.3.1 Method 1 - System test

Check the correctness of answers supplied by the system to a set of generated questions. This evaluation is designed to check that a set of sample questions provides, correct, sufficient and useful information as output. This method of testing would be applied as an extension of the verification of the knowledge base and inference engine. The following steps would need carrying out.

- The development of the test knowledge base and question set for use in the analysis.
- A set of answers to the questions that would represent a broad sample of the potential users of the application.
- A set of correct answers to the questions to be posed by the system created by an advisor. These answers would act as a control and ensure that the system met the required output.

A comparison between the results returned from the system during application of the set of answers, against the correct human generated samples would indicate a variance in the question/answer process and could be investigated further.

4.6.3.2 Method 2 - Blind evaluation

Check that the information provided by the system is presented in a better style than information in its original format.

Carry out a pen and paper exercise by querying a group of readers based on a set of questions poised by the existing benefit advice documentation. Take the answers provided and feed them into the system noting the output produced.

The corresponding original advice document and dynamically generated document could then be shown to the user to determine which document represented their required information in the most appropriate form. The aim is to determine:

- Which document was easier to read?
- Was there sufficient information in each document or any conflicts in any conclusions reached?

Without taking into account the user interaction or modelling aspects of QERIS.

4.6.3.3 Method 3 - Interactive analysis

This would place users in a position where they interacted with the system. There are a number of scenarios that could be enacted.

1. A user is given the task of identifying the answer to a posed question, first using the original full text version, and then using the interactive system. Success can be marked with respect to the time required to search, and satisfaction with the information provided.
2. The user is allowed to interact with the system freely, there reactions to the system would be noted and used to evaluate the user interface elements and response to the information returned.

The evaluation would help answer some of the following questions.

1. Does the system provide an enhanced mechanism for information presentation specific to the presentation of legal information.
2. Is the system appropriate for the environment and source it is targeted at, for example, presenting information to the general public.
3. Can knowledge bases be developed for the system that are easy to build, maintain and verify.
4. General feedback about the interface, interaction and feedback elements of the system.
5. Test the speed of the user in identifying and interacting with the system. For example, can they find the information quicker than using standard navigation or reading strategies currently expected of the reader.

4.7 Summary

This chapter has provided further details of the concept of tailoring online presentations and has provided an implemented example of tailoring online Benefit Advice so that the presentation of the advice is more interactive, provides better guidance and helps identify and interpret key information.

Section 4.1 provided an outline of benefit advice and the different sources that are currently available. The two main forms of documented advice, traditional paper based advice booklets and the more recent (since 1998) online benefit advice pages are described and examples given.

Section 4.2 focused further on the format and structure of advice documents. The visual representation and underlying format of the documents is discussed followed by some example legislative sources. Some issues with interpretation and translation between legislation and advice is provided. An outline of the target audience is also provided, highlighting the differences in goals and characteristics.

Section 4.3 provided details of how the documentation is tailored in different way to provide a more detailed representation of the information. Graphic and content manipulation techniques are described and a number of examples are provided. The section also provides an overview of the different tags used to build the knowledge base.

Section 4.4 and 4.5 conclude the chapter by providing an example of the running system. This section builds on the previous section by providing an outline of possible rules that can be defined and different strategies for combining the rules into a consultation. Example queries and consultation sequence are shown followed by an example of how the document may be represented based on the queries asked.

The other leaflets referenced in CH1 that should be available in order to create a complete picture of the general advice being provided are listed below. In a complete system, these would be connected to the child benefit pages through hyperlinks.

GL23 Social Security benefits rates
CH7 Child Benefit for children aged 16 and over

CH4A Social Security and children being looked after by a local authority

CH5 Child Benefit for people entering Great Britain

CH11 Child Benefit for lone parents

CH6 Child Benefit for people leaving Britain

BC1 Babies and children

FC1 Family Credit

JSA8 Jobseeker's Allowance and Income Support

JSAL5 Jobseekers Allowance – helping you back to work

NI14 – Guardians Allowance

HC11 – Are you entitled to help with health costs.

SD3 – Long-term sick or disabled.

CHAPTER 5

CONVENTION FOR THE INTERNATIONAL SALE OF GOODS

5 CISG Development

5.1 Overview

The most recent act of the United Nations Convention for the International Sale of goods was passed in Vienna in 1980 and is still used today. The convention is available in several languages and, more importantly for our purpose, in plain text and hypertext versions over the Internet.

5.1.1 Internet Availability

The convention is available in two main formats, as one large document with a set of hyperlinks located at the top allowing the user to move through the document easily and in a second format where each independent section is stored and accessed as a separate page. Internal references are also usually hyper-linked allowing quick access to the related article or section. The second version is where the convention has been broken down into different sections providing a navigation of the contents that links to these separate sections. There is currently no evidence of an attempt to provide a better form of presentation or to extend the functionality of the medium to improve the presentation and interpretation of the information. One of the biggest online resources is the CISG online site (http://www.jura.uni-freiburg.de/ipr1/cisg/) that hosts a very large database of cases and acts as a comprehensive archive and forum for all CISG related information.

5.1.2 Research Interests

Off line there is interest in the logical and structural elements of the CISG with numerous legal texts analysing the document in fine detail. There are also groups interested in the relationships between the CISG, computing and artificial intelligence. This chapter brings some of the ideas in the field of AI and the CISG into the field of Internet presentation. There are a number of active AI research projects that concentrate specifically on the CISG as a source material. The interest is focused by a number of international workshops specific to Expert Systems and the CISG. The leading light of these workshops is Hajime Yoshino whose research has been directed towards providing AI solutions to the representation and interpretation of the CISG. The following chapter and work described builds on Yoshino's series of works articulating the logical structure of Contract Law, Daskalopulu (1999) and the CISG, Yoshino(1997). By breaking the CISG down into sentences, each sentence representing certain events within a contract life span, Yoshino has provided a good foundation from which to develop a concept of tailoring the large CISG convention and associated resources into reader specific variants. As well as considering the entire CISG as a document, there are a number of works that focus on specific areas of the convention such as Article 14, Allen L. E. (1997) that have provided an insight into how complex the smallest elements of the CISG can become.

5.1.3 Goals of this Example

The CISG example will demonstrate:

- Producing documents specific to different readers that have conflicting interests. Due to the nature of the CISG readers will want to read groups of Articles that are different depending on their aims. They may wish to identify their obligations in respect to the convention; they may wish to determine remedies for breach of contract; it may also be necessary to gain an overview of obligations for all parties if a certain state exists. There are three main reader groups, the buyer, the seller and an external party such as an advisor or solicitor. Each group will have different goals when accessing the CISG and these different goals are considered.

- The changes in the Articles of the CISG relevant to users over time as contractual obligations vary. Because contractual obligations change depending on different occurrences such as delivery, dispatch or receipt of goods, it is necessary to reflect this in the Articles that are considered relevant and displayed. Although exact times and dates are not examined in this example, the concept of sequence and order is described.

- The integration of additional database and CGI resources with QERIS. The distributed nature of the Internet allows many different servers and sites to collaborate on a single project. A means of doing this is providing simple interfaces to database resources that allow one solution to use the source information and data from another. The concept is similar to program APIs used to provide an interface to the internal workings of applications. CGI solutions should also strive to provide interfaces designed so the responses can be processed by other systems without having the extra HTML mark-up for style and presentation that is most evident in online CGI database front ends. The presentation of the information should be clearly separated from the retrieval and delivery of the source information.

- The generation of new documents rather than the tailoring of existing documents. In the CISG example, templates are used to provide an interface to the reader and encapsulate the knowledge bases used. In the CISG application most of the required components are loaded from a supporting database and integrated into templates.

5.2 Convention structure

The CISG is broken down into a multi-level structure consisting of key elements. Elements such as articles and sections are usually grouped into related subjects and many cross-references exist between articles. The convention brings together a number of concepts that include:

- Determining the application of the convention to the contract being proposed. The convention does not apply to all contracts and it is important to establish

'ground rules' before progressing with the creation, execution and termination of such a contract for the international sale of goods.

- The nature of a party's involvement in the contract – is the party selling or buying goods? By identifying the party and their obligations it is then possible to determine the goal of that party. Do they wish to know their own obligations or the obligations of the other party? Consider **S** is a seller but would like to determine the obligations of **B**, a potential buyer. **S** needs to be aware of their obligations to **B** as well as having access to information associated with **B**'s obligations to **S**. Any inference to determine relevant articles would require not only a model of the seller's characteristics but also those of the buyer as the two parties may be in different stages of the contract.

- Identification of key stages common to both parties – The convention also has sections relevant to both parties but specific to a particular instance of a contracts life cycle. The life cycle of a contract will be discussed in more detail in section 5.3.1. But articles relating to actions such as the creation of a contract will only be relevant if a contract has yet to be created or a party is interested in determining that a contract was established in the correct manner.

- General provisions and global information – There are a large number of articles that are relevant throughout the contract. Many of these articles are used to determine the application of the contract, definition of key terms or the definition of the contract and although relevant at all times can provide too much information to specific requests.

The actual convention is broken down structurally into the following key elements:

Preamble – this is a short paragraph that provides an introduction to the convention as a means of establishing uniform rules in the international trade of goods.

Parts – There are 4 main Parts to the convention and this provides the highest level breakdown of the document. The Parts consist of, sphere of application, formation of the contract, sale of goods and final provisions.

Chapters – some of the Parts are further broken down into chapters that provide a second level heading and focus on key subjects such as distinctions between the parties.

Sections – Some Chapters are broken down further into sections that again focus on particular subjects within a chapter such as the obligations of the seller on taking delivery of the goods.

Articles – Articles provide key concepts within the convention. There are 101 articles and each one covers a very specific area. It is intended that articles will act as the atomic elements of a tailored representation and will be selected based on the characteristics of the parties and the stage of the contract.

Text – Each article is broken down into paragraphs, bullet points and sentences. Although it is possible to highlight important and relevant components within each article this, is handled at a different level to the article selection.

It is important to determine the integration and association between these as it will shape the presentation of the final tailored presentation and also determine how external resources such as the database discussed in section 5.5.2 is structured. The actual key subjects defined in the above structure is shown in Figure 5.1

Part	Chap	Sect
1. The sphere of application		
	1.1 Sphere of application	
	1.2 General provisions	
2. The formation of the contract		
3. The Sale of goods		
	3.1 General Provisions	
	3.2 Obligations of the Seller	
		3.2.1 Delivery of the goods and handing over of documents
		3.2.2 Conformity of the goods and third party claims
		3.2.3 Remedies for breach of contract by the seller
	3.3 Obligations of the Buyer	
		3.3.1 Payment of the price
		3.3.2 Taking delivery
		3.3.3 Remedies for the breach of contract by the buyer
	3.4 Passing the risk	
	3.5 Provisions common to the obligations of the seller and of the buyer	
		3.5.1 Anticipatory breach and instalment contracts
		3.5.2 Damages
		3.5.3 Interest
		3.5.4 Exemptions
		3.5.5 Effects of avoidance
		3.5.6 preservation of the goods
4. Final provisions		

Figure 5.1: *Top level outline of the CISG*

The sections described above can be used to establish that the CISG applies to the contract and that the provisions of the convention apply, to determine the obligations for the creation of the contract and to define the scope and application of such a contract, and to determine the actual actions that must be carried out in order to deliver, receive and pay for the goods. Section 3 also provides remedies for situations where the contractual agreement may be breached, and defines outlines of which parties are responsible for loss or damage of the goods during the period of the contract, And provides general provisions that focus on breach and remedy that are appropriate to both parties under contract. Elements regarding the protection and transport of the goods are also discussed. Finally additional provisions that apply to the contract and the states within which the contract is made rather than the parties involved in the contract are described.

5.3 A CISG contract

Establishing a contract requires the understanding of key concepts. To establish a contract involves a set of stages where obligations of each of the parties concerned will change. The contract will then be formed and the offer becomes valid. It is then the responsibility of the parties to meet the offer and fulfil the transaction of the goods arises. By determining the characteristics of the contract from the point of view of each party a better image and a more appropriate representation of the convention can be provided. The changing characteristics of the contract can be considered its lifecycle.

5.3.1 The contract lifecycle

Each contract created for the international sale of goods has a lifecycle, from the first offer made through to the conclusion of legal obligations between the parties, perhaps with delivery and payment. This section focuses on the close association between the relevance of articles and the stage of a contract in its lifecycle. By identifying which events have occurred and at which stage in the sequence a contract is, it is possible to direct an interested party to more relevant sections of the convention. The diagram in Figure 5.2 shows a small section of the possible events that may occur during the execution by two separate parties of their obligations under a contract. Each event provides a milestone within the overall life cycle of the contract that has associated implications with the events that will follow.

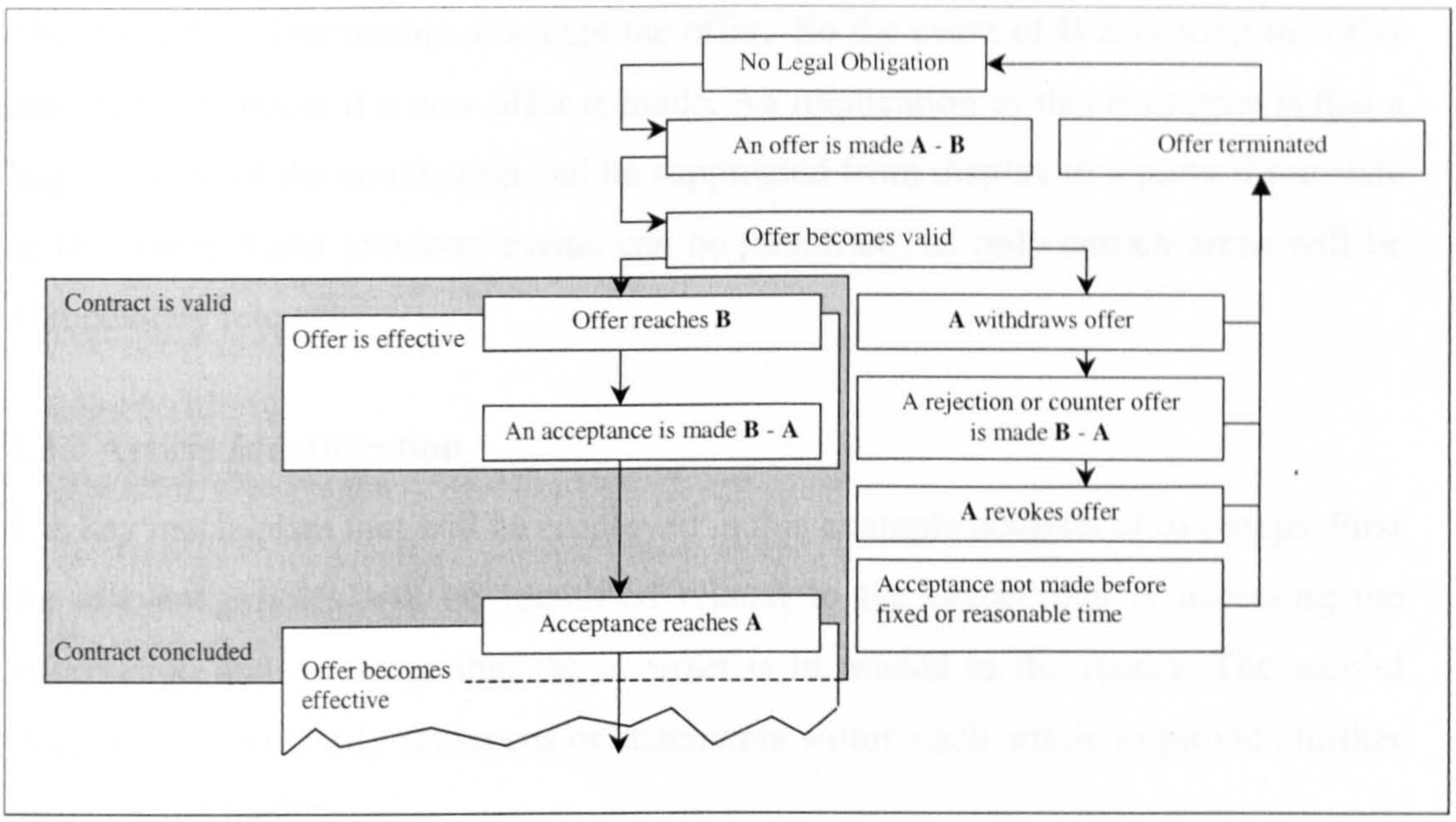


Figure 5.2: *Initial events in a Contract lifecycle*

Each of the events indicated in Figure 5.2 has a number of articles, or paragraphs and sentences within articles, that will be relevant to it. An example of this association is the event "**A** withdraws offer which has a direct association with the Article 22 of the CISG.

> *Article 22 "An acceptance may be withdrawn if the withdrawal reaches the offeror before or at the same time as the acceptance would have become effective."*

In terms of its relevance to **A**, **A** must be aware that a withdrawal must reach **B** either before or at the same time the acceptance becomes effective. An acceptance becomes effective if the acceptance reaches A, unless some later date is stated**.** The relevance to **B** is to inform **B** that once their acceptance reaches **A** and no withdrawal is received then the offer becomes effective. Each of the two parties has a different view of each event and in many cases each event has a different set of relevant articles for each party. A needs to know whether he is permitted to change his mind, and B needs to know what he can be confident that the contract is definite.

Another important aspect of the lifecycle is the different paths that can be followed based on decisions made or events that have occurred that exclude certain situations from arising. For example, if **A** withdraws an offer made to **B** before **B** accepts the

offer then **B** will be unable to accept the offer. So the event of **B** accepting the offer can then only occur if a new offer is made. An implication of this exclusion is that a large volume of the convention can be suppressed from display to a party if the state or the contract and previous events can be identified, as only certain areas will be immediately relevant.

5.3.2 Article Identification

The key mechanism that will be employed in this example consists of two steps. First the relevant articles will be identified related to the reader that is accessing the information ánd the stage that the contract is in related to the reader. The second stage is then to identify sentences or statements within each article to provide further focus for the reader.

Articles within the CISG are related to each other in a number of ways:

- They may be clustered together under general topic subjects such as the group 11-13 that focus on the means of communication between the parties and clarifies the term "writing" which is used throughout the convention.
- Many articles use related terms such as "Place of business" which occurs in 10 separate and widely distributed locations throughout the convention. So to determine if a place of business will have an impact on the convention the articles must be searched and all related entries read and understood in order to gain a complete representation.
- In many situations articles constitute supporting and contradicting groups. Articles 1 and 2 are good examples of how articles can be exclusive. Articles 1 describes which contracts the CISG can be applied to while Article 2 describes contracts, which the CISG cannot be applied to.

A reader may be interested in the actual event articles, the alternative event articles, referenced articles, articles within close proximity to a returned article and also the section from which the article was returned. Retrieval of articles that fall into these groups is discussed and an implementation described in section 5.5.2. In an online context it is possible to present the reader with a page or query that lists possible events such as Offer received or Acceptance received. These events would then lay

the foundation for further consultations that may be required to identify additional articles within the convention. The solutions to these queries would then activate or modify the display of the articles to identify a degree of relevance to the current event based on additional reader characteristics. An example consultation is described in section 5.5.4.

5.4 Analysis

Consider the convention from the point of view of a reader hoping to apply the convention to a contract they wish to undertake. A contract is usually a document drawn up between two parties, the buyer and the seller who wish to carry out a transaction of goods. As discussed earlier, the convention is broken up into sentences and paragraphs but, more importantly, articles that contain collections of related paragraphs. By providing a contracting party with the articles most relevant to their characteristics and the temporal stage of the contract the supported identification and access to relevant articles can be accomplished.

There are a number of stages that define the relevance of the convention, for example:

- Is the convention relevant at all? The convention only applies to contracts where the sale is international and not domestic.
- Which major subsets of the convention apply to which reader? Is the reader a buyer or seller?
- Which Chapters, Sections and Articles are relevant to the reader?
- At what stage in development is the contract being undertaken?
- What are the next steps in the contract life cycle?

If no contract exists then the first and second sections are of most relevance. These sections indicate that the convention can be applied to the contract under development and also the means of creating a contract of the required type. Once a contract has been formed then the actual mechanics of the contract are of more interest, the actual sale of the goods in conformance with the contract. The sale of

goods section is broken into three main sections, sections relevant to the buyer, seller and to both. This does not mean that if the reader is a seller then the section dealing with buyers is of no interest, but it does provide a greater separation of distinct topics.

The convention can be examined at several different levels to provide presentation and interpretation support for the user.

1. The first stage is identifying if the convention applies to the contract under evaluation. This will determine if the rest of the articles are relevant. It may be, however, that a contract does not exist and the reader is determining the requirements prior to its creation or that the contract already exists and the reader is determining whether it is controlled by the CISG.

2. The second stage is to reduce the number of possible articles by identifying large blocks that are related or not related to the reader. There are two methods of determining their relevance: the users role in the contract, and the contracts current state, for example that an offer has been made but no response to the offer has been received.

3. The third stage is to determine unique characteristics about the reader article by article guiding the reader through an intuitive consultation to determine the articles most relevant. Aspects of the contract life cycle as well as the reader's characteristics may support this consultation.

Let us examine the first level and start formulating questions to determine if the convention is relevant to the contract under consideration by the reader. The convention will apply only if two basic requirements are met:

The sale must be international – the seller and buyer must have their "places of business in **different** states".

The sale must have a prescribed relationship with one or more States that have adhered to the Convention.

So by asking only two questions, a rudimentary decision can be made as to the relevancy of the convention.

Question 1 - Are the places of business of the seller and buyer in different states?

Yes – Then the convention may apply to a contract between the parties – proceed to q2

No – Then the convention does not apply to a contract between the parties.

Unknown – Sections 1, 2, and 10 are relevant – it may be that the parties have places of business in multiple states or the location of the business may not be provided by a party until later in the contracts life cycle.

Question 2 - Is there a relationship between the sale and a State that adheres to the Convention?

Yes – Then the convention applies in this case

No – The convention does not apply to a contract between the parties.

Unknown – There may be a lack of understanding of the question or lack of knowledge about the contracting States. Further explanation may be provided on request to determine the application of the convention.

There are now two options in further breaking the convention down. The first approach is to work from the top, article 1 down through to article 101. A far more interesting approach is to break the convention down in a hierarchical fashion bounded by the main section and chapter separations refining the hierarchy based on the reader characteristics. This approach will highlight/diminish much larger sections of the convention at each iteration while still providing a means of highlighting individual articles that occur outside generally relevant sections. The hierarchical approach also fits well with the navigational and structural design of the convention. so providing a convenient and more natural navigation.

It must be remembered that while the benefit prototypes were aimed at the general public, it is less likely that a similar audience would want to read this convention. The target audience for the use of the CISG would be companies or groups who regularly undertake international sales. The system can be viewed as a support tool

alongside case databases, search engines and hypertext linking. The goal of the tool is to aid the development and execution of a contract of international sales.

A contract evolves by going through a set of stages which may be repeated a number of times during the duration the contract is valid. A contract occurs between two parties and each of the stages is slightly different for each party and different obligations apply. A contract follows a set of main stages or milestones:

Offer – A party makes an offer for the sale of goods to another party. At this time the contract becomes valid and the provisions of the convention apply. Once an offer is made, if it is not withdrawn beforehand, then the next event, acceptance, may occur.

Acceptance – It is the role of the party receiving the offer to accept or reject it. The receiving party can also make a counter offer, which then would be returned to the original party for consideration and acceptance. Prior to returning an acceptance to the offering party, there is a period of consideration.

Consideration – The period of consideration allows a party to evaluate an offer made to them. There is usually a limit on the period of consideration. Once an acceptance is made then the contract become effective and any breach of contract would result in appropriate remedy. Once and acceptance has been received by the offering party the offer becomes effective and the contract is concluded.

Intention – this is a state where intention is indicated of the offering party to be bound by the acceptance of the offer. It is now a situation where the offering party is bound to deliver the goods based on the initial contract and offer made.

Capacity – The next stage is the actual transaction and exchange of the goods and all associated resources in compliance with the provisions of the convention.

At each of these stages different sections of the CISG apply to the different members of a contract. The CISG as a single document does not aid the identification of

relevant sections at particular times in the contracts life cycle. A contract can be described as having state: the dates and times that define the limits of these states are very important in the development of the contract. At each state it is possible to identify sections of the CISG that are directly relevant to both parties in the contract. If this information is presented online then the QERIS engine is well suited to identifying the relationship between a contracts state and the sections of the CISG that relate to that state and each client.

5.4.1 Contract Events

The events in the lifecycle of a contract are key in determining the obligations of a party to the other contracting party. As each event occurs then different sections of the CISG change in their importance and relevancy to the reader. In his work Yoshino focuses on a single case and a single thread through the life cycle of a contract. . The events represent stages of the contract life cycle and at each event new articles of the CISG become relevant, others may become less or not relevant. The events become milestones and bookmarks within the CISG allowing the retrieval and presentation of the corresponding articles. The table in Figure 5.3 shows the key events identified in Yoshino (1997) in chronological order.

	Date	Event
0	4/5	No relation
1	4/8	Offer reaches B
2	4/9	Acceptance reaches A
3	5/1	A hands over goods to the first carrier Japanese Container Ship
4	5/10	B pays the price
5	5/31	The goods are delivered to B
6	6/5	B examines the goods
7	8/10	The machine is operating out of order, B noticed to A specifying the nature of lack of conformity
8	9/1	B asked A to repair the machine within one month
9	10/1	The additional period expired
10	10/10	B declared the contract avoided
11	12/10	A recovered the damage, B restituted the machine
12	12/20	A restituted the price paid by B

Figure 5.3: *Events in a CISG contract*

The example events in Figure 5.3 reflect a single path through a single example case. By examining the articles in more detail by their heading, topic and content they can

be grouped into ranges that are appropriate for a particular contractual event. By identifying the key events and also possible alternative events in the contract development a more general listing of the events and associated articles can be represented as in Figure 5.4.

	Generic Events	Alternate Events	Related Articles
1	Offer received	Offer lost	11-16/24
2	Acceptance received	Offer declined or Acceptance lost or Acceptance late	11-13/17-24
3	Goods passed to carrier	Goods delayed	31,32,67
4	Price paid	Price not paid	50,64,78,81
5	Goods received	Goods not received or damaged	82,84
6	Goods examined	Goods not examined or damaged	38,58
7	Lack of conformity	Goods conform	37,39,50
8	Request for repair (remedies for lack of conformance)		13/45/46,86
9	Lack of conformity not remedied	Lack of conformity remedied	34,45,46-52/74-77,78,79
10	Contract avoided	Contract not avoided	49,81-84
11	Damages recovered or goods restituted	Damages not recovered or Goods not restituted	34,37,44
12	Price restituted	Price not restituted	75,76,78,84

Figure 5.4: *General events in a CISG contract*

Alongside each event is the contrary event that can occur and may apply to the other party. For example either an acceptance is received OR the offer is declined or lost or arrives late. All of these cases fall into a range of related articles 11-13 and 17-24. This shows that as these events occur, a subset of articles becomes relevant to each party and another subset becomes relevant to both parties, while others cease to be relevant.

This textual material differs from the Benefit Advice documents in that the text does not contain direct queries and solutions to those queries. The interaction is thus unlike the Benefit Advice leaflets where a reader would navigate to a document and then based on a consultation would receive advice and a tailored result. The CISG is a large single document and the goal of tailoring is to remove non-relevant articles while providing focus to articles that are relevant to a party at a particular moment in the contract life cycle. In the Benefit Advice example, characteristics of one or more people were determined. In some cases information about a dependant or a spouse is also be required in order to help in interpretation and presentation. In the CISG

example there are two key parties, a number of possible external observers and the contract itself.

5.5 Example implementation

The CISG is stored and processed differently to the Benefit advice example in the last chapter (see 5.5.2 for more details). Due to the many different document sections that could be brought together into the final document presented to the user it was necessary to store each article of the contract so it could be selected in a number of different ways for processing by QERIS.

5.5.1 Architecture

The CISG is a large document that consists of well-defined parts, sections and articles. Unlike the benefit advice documents where the information existed as a single document that could be tailored inline, the size and format of the CISG document requires a different approach. The QERIS engine acts as a hub into which many different resources can be combined into a single HTML document. The document will be based on the unique attributes of the user but can also be controlled by which links and navigation options they choose.

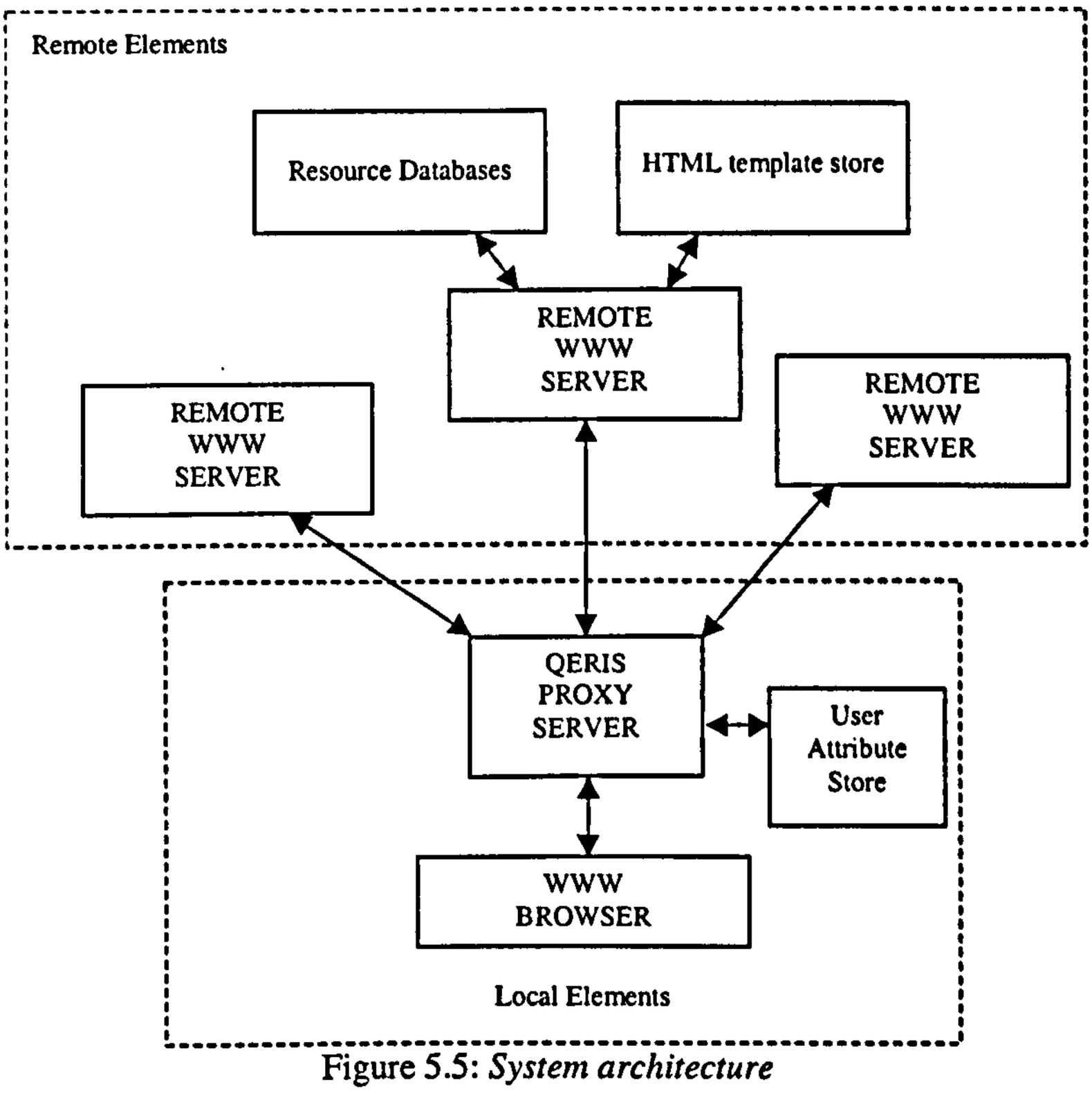


Figure 5.5: *System architecture*

The architecture that will be used for the CISG is shown in Figure 5.5 There is a reliance on external resources from both a document file store on a web server and also a CGI driven database interface. The Database interface is discussed in section 5.5.2 and is used to selectively incorporate articles and CISG sections into templates stored on the webserver. The proxy server acts as a local application that has access to the two separate sources of information and also controls the processing and integration of these components before the final page is sent to and the rendered on the browser.

5.5.2 The Database

The CISG is broken down into the separate articles and these articles are stored in an Access database. By storing the Articles in this way they can be accessed in a number of different ways. An important function of a database representation is it is possible to search the text of the articles for key words. To integrate the Database representation with QERIS, a CGI style of interaction was developed using Active Server Pages (ASP). A single ASP can be used to carry out a number of different

queries on the database by providing different CGI query strings. A query string looks similar to:

```
http://royles.phd.csc.liv.ac.uk/database/cisg/c
isg.asp?rangeStart=30&rangeEnd=52&index=1
```

This query retrieves all Articles in the range 30 to 52. The index clause also tells the CGI request to generate an active index that allows navigation through the retrieved Articles. This query can be used to retrieve all articles relevant to a Seller.

The CGI functions available include:

Range Limits – Retrieve a range of articles, which can be used to display a section, part of or a whole chapter of the convention For example, to display all Articles that discuss the Sellers obligations:

```
rangeStart=30&rangeEnd=52
```

Search – Search the articles and display only those containing the key word. To display all Articles which contain the word observer:

```
search=observer
```

Headers – Generate title and navigation information for the articles retrieved. To retrieve all articles containing the word observer and generate an index and headers for the articles:

```
search=observer&headers=1&index=1
```

Index – Generate an index of all retrieved Articles. To retrieve all articles containing the word observer and generate an index of the articles:

```
search=observer&index=1
```

Get Articles – Retrieve only the articles indicated in the arguments. The number of articles is given by *articleCount*, and the Article numbers are then selected by *articleNum*. To retrieve articles 1 and 100:

```
articleCount=2&articleNum=1&articleNum=100
```

These five queries are powerful enough to generate a wide range of different representations of the convention and could be extended further to provide additional functions. Additional functions could include retrieval of cases related to a particular article, a history of any amendments or notes relevant to an article or different language versions of articles. The use of a CGI interface to a database also shows the flexibility of the QERIS engine in accessing additional resources. It is also recommended that CGI resources such as online case databases should have a simple text interface that can be easily integrated with other engines such as QERIS. The CGI process should not always be responsible for rendering the structure and style of a page but used to serve raw data for other systems to tailor and process. By providing data in a plain format the distribution and reuse of online databases and resources would prove far easier and more flexible.

The database contains a number of tables that contain the text of each article and a number of keys that can be used to retrieve the titles associated with an article at different levels, article, part and chapter. Depending on the request made the article is displayed in a number of different formats. If a single article is requested then this article will be provided with all relevant titles.

e.g. The query articleCount=1&articleNum=1&headers=1&index=1

Will generate the HTML shown in Figure 5.6.

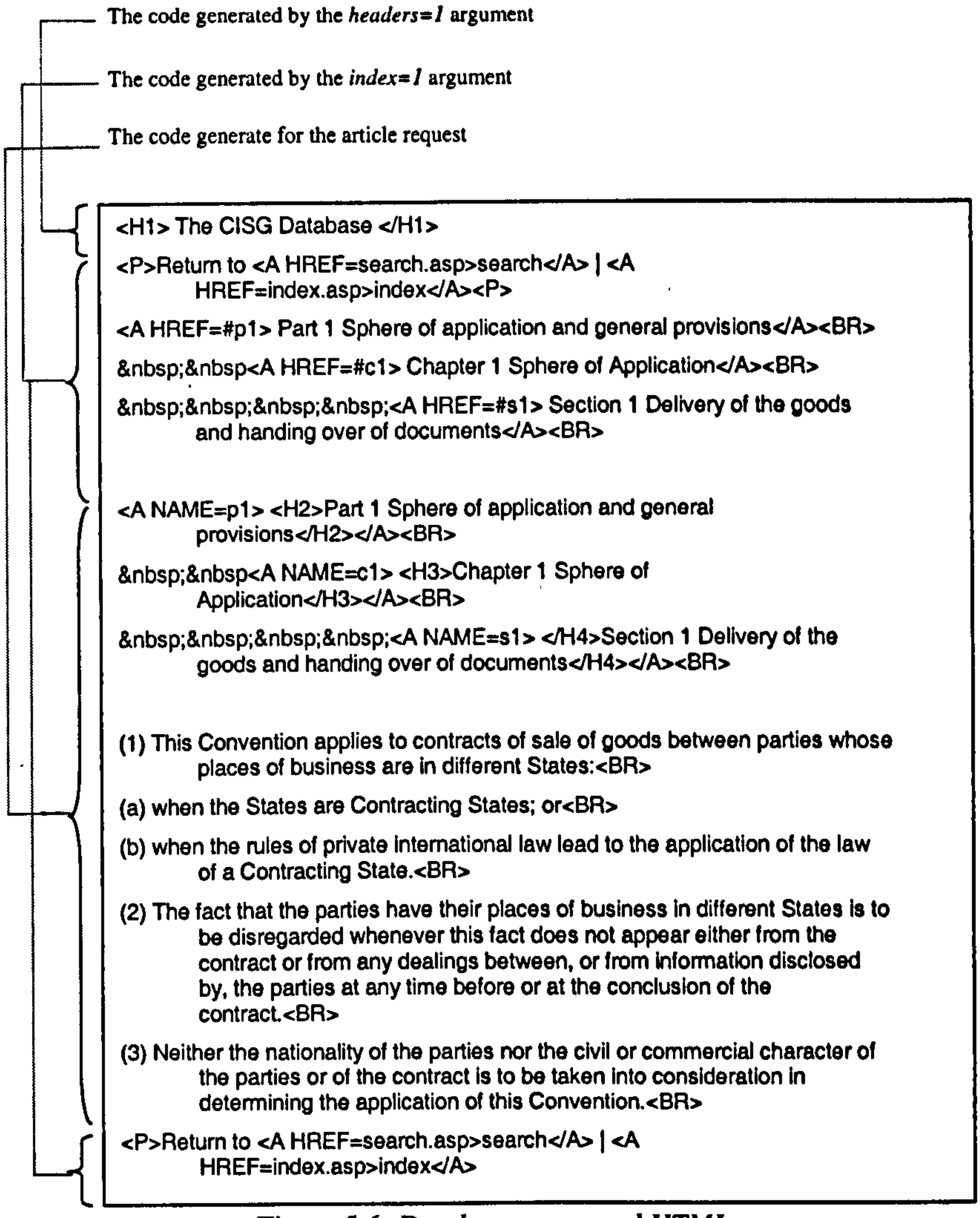


Figure 5.6: *Database generated HTML*

To provide navigation and references within the Article text, all sub references are designed so that when they are selected they make additional queries to the CGI script. E.g. A reference to article 29 is represented in the database as a anchor

```
<A HREF=cisg.asp?articleNum=29>article 29</A>
```

References to entire sections, parts or chapters will also retrieve the entire sub set of articles relevant to the request. Figure 5.7 shows the form content used to exercise the database interface.

```
<FORM name=cisg.asp action=cisg.asp method=post>
        Include dynamic index in returned results?
                <INPUT TYPE=checkbox NAME=index><P>
        Get a single Article (Enter the article number)<BR>
                <INPUT TYPE=text NAME=articleNum SIZE=4><P>
        Return articles from ? to ? (Enter a lower and higher bound)<BR>
                <INPUT TYPE=text NAME=rangeStart SIZE=4>
                <INPUT TYPE=text NAME=rangeEnd SIZE=4><P>
        Search on word (Enter a word or sequence of words)<BR>
                <INPUT TYPE=text NAME=search SIZE=20><P>
                <INPUT TYPE=submit>
</FORM>
An example of an internal link that accesses the same database using
the above techniques.<BR>
<A HREF=cisg.asp?articleNum=12>
Article 12</A>
```

Figure 5.7: *Article and section form syntax*

5.5.3 The Templates

The CISG has not been constructed around a document model as with the benefit advice documents in the previous Chapter. Instead the CISG is presented to the reader as a set of templates that when processed by QERIS and combined with entries from the CISG database will provide the final page representation. The templates are used to add intelligence to the retrieval and integration of database components. There are many examples of databases being used with templates to generate fixed page representations but, by adding QERIS tags to the templates, finer control over the integration can be achieved. The templates use effects that integrate the database elements by using the following syntax:

```
<EFFECT ID="<effect identification>"
CONDITION ="<condition>"
```

```
PLUGIN="insertFile"

TRUE="http://localhost/database/cisg/cisg.
asp?<database query arguments> "

FALSE="">
```

The integration of the database elements can be fully controlled as part of the database query string so if this effect is activated then the articles defined by the query string are integrated into the template. There are a number of cases where information is loaded into the knowledge base of QERIS using facts within pages so that as a reader navigates from page to page the actual visiting of a page is registered using a fact. It is possible to 'remember' that a reader has been to a page by using an effect and rule combination that loads this information into the knowledge base for use within other rules and effects later in navigation or inference. A fact can be registered using the following syntax.

```
<RULE ID ="page1"

CONDITION="page1 is visited"

CONCLUSION="page1 is visited"


<EFFECT ID="page1"

CONDITION ="page1 is visited"

PLUGIN="insertString"

TRUE=""

FALSE="">
```

The effect is used to trigger processing of the rule. As there are no arguments for the effect the document will not be modified. The rule will be called and the conclusion for the rule set to "page1 is visited". It is then possible to use the information later in navigation by placing an entry "page1 is visited" into the condition of a rule or effect. In the same way as the Benefit Advice example in Chapter 4 queries, rules

and effects can also be added to provide more complex, direct and interactive consultations.

5.5.4 Example Consultation

5.5.4.1 Initial access

The actual access and navigation through the site is broken down into seven stages where at each stage a reader will provide a different set of facts to the expert system. These facts can be set through questions but in the case of the first few pages some initial information will be set as facts directly into the rule base by accessing certain pages. As each page is traversed a clearer picture of the reader and their goal should be determined. As the reader navigates the initial access pages their characteristics will be altered and updated in preparation for asking more detailed information about the contract. Figure 5.8 shows how pages will be linked together to provide guidance to the article presentation and to show how the consultation will be broken down across several pages.

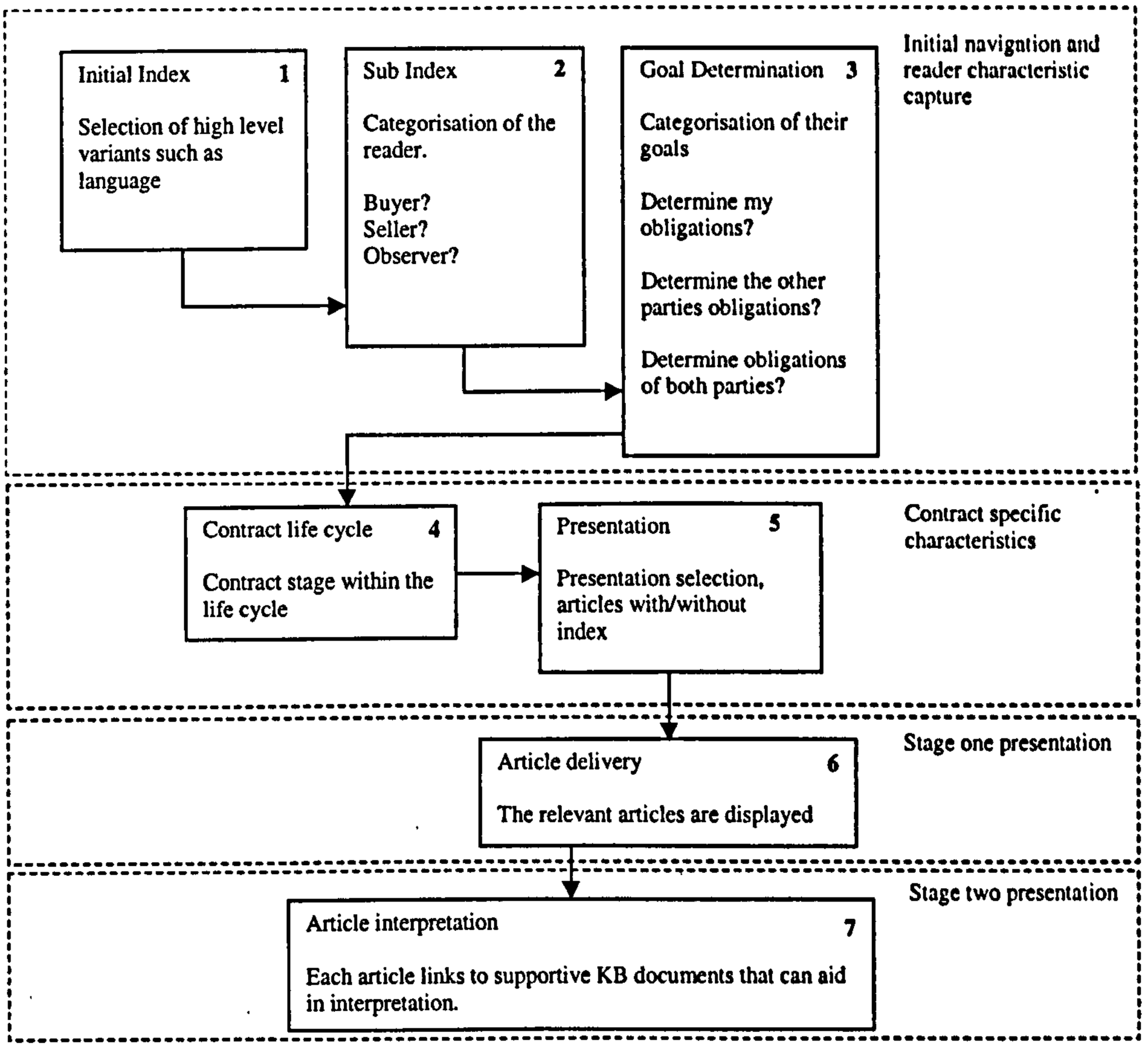


Figure 5.8: *Outline of navigation strategy*

Pages indicated in 1,2 and 3 are used to set facts about the reader into the knowledge base. As more information is gathered a better understanding of the reader can be created. The reader simply selects URLs and does not have to answer any questions at this stage. Information such as their intention, goals and default settings are added to the knowledge base as the reader moves from page to page. When entering page 4 some additional processing is required to determine at which stage the reader and also the other party is at. The system has already determined that the reader is a buyer, and therefore they may have received an offer to purchase goods. If this is the case then initial assumptions can be made that the reader has received the offer but has not yet returned an acceptance, the reader is in a period of consideration.

5.5.4.2 Initial consultations

The system can identify the reader as a buyer based on the facts set during earlier navigation. Based on this information the consultation that follows would be limited to a single template associated to establishing the state of the contract and determining if an offer had become valid.

Have you received an offer?

This would indicate that no obligations currently exist and provide information regarding the construction of a contract under the CISG

Have you returned an acceptance to the offer?

If an acceptance has been returned then the buyer should be made aware that upon receipt of the acceptance the offer becomes effective. They should also be informed that they could withdraw the acceptance as long as the withdrawal reaches the offering party before or at the same time as the original acceptance.

Did you receive a revocation of the offer before submitting the acceptance?

In addition they need to be aware that a revocation of the offer received before the acceptance was dispatched would conclude the contract and result in no further obligations.

Has the offer been withdrawn?

If no acceptance has been made then it is still important to be aware that the offer can be withdrawn or revoked until the receipt of an acceptance by the offering party.

Do they intend to issue a counter offer?

If no acceptance has been made then the reader may wish to issue a counter offer, relevant information about such a process should be provided and again the information about the offer being withdrawn or revoked will apply.

From the point of view of a seller, the consultation would be different because their obligations are different and some of the events will be in contrast to the buyer. Again determining if the reader is a seller would be carried out by facts set during earlier navigation or consultation For example the following consultation would be more appropriate.

Have you issued an offer?

This would indicate that no obligations currently exist and provide information regarding the construction of a contract under the CISG.

Have you received an acceptance?

If an acceptance has been received then it is important to inform the seller of their obligations as the offer is effective unless: 1. They had withdrawn the offer and the withdrawal had reached the buyer before or at the same time as the offer. 2. The offer has been revoked and the revocation is received prior to the dispatch of an acceptance.

Do you intend to withdraw the offer?

If the seller intends to withdraw the offer then they need to be informed that their withdrawal must reach the other party before or at the same time as the offer. If a period of time has elapsed since sending the offer then the seller should be informed that they can revoke the offer but only if it is received prior to the dispatch of an acceptance. It is obvious that time and events play a significant role in the obligations of the seller at this point in the transaction.

Have you received a counter offer?

If a counter offer is received then to some degree the roles are reversed and events associated with the buyer now also apply to the seller. This counter offer and role reversal may occur more than once as an offer is refined.

Have you received a revocation of acceptance?

If the seller receives a revocation then they should be informed that a revocation of acceptance should reach them before or at the same time as the acceptance.

The nesting of the queries indicates a dependency on those below, for example an offer must have been made for the receipt of acceptance to apply. By extending this questioning to include the period of an offer and the transaction, transport and receipt of goods and also include the remedy of breach then a complete picture not only of the reader but also the state of the contract can be constructed.

There are a number of articles in Part2, Chapter 2, Obligations of the Seller that can be translated directly into appropriate queries and linkage to associated articles. For example possible queries that may be appropriate during the delivery phase will include:

Have the goods been dispatched?

The goods should have been hander over to a carrier or made available for the buyer to collect from the sellers place of business.

Were the goods clearly identified?

Notice of the consignment needs to be sent to the buyer

Were the necessary contracts of carriage appropriate?

It is the seller's obligation to seek appropriate contracts of carried.

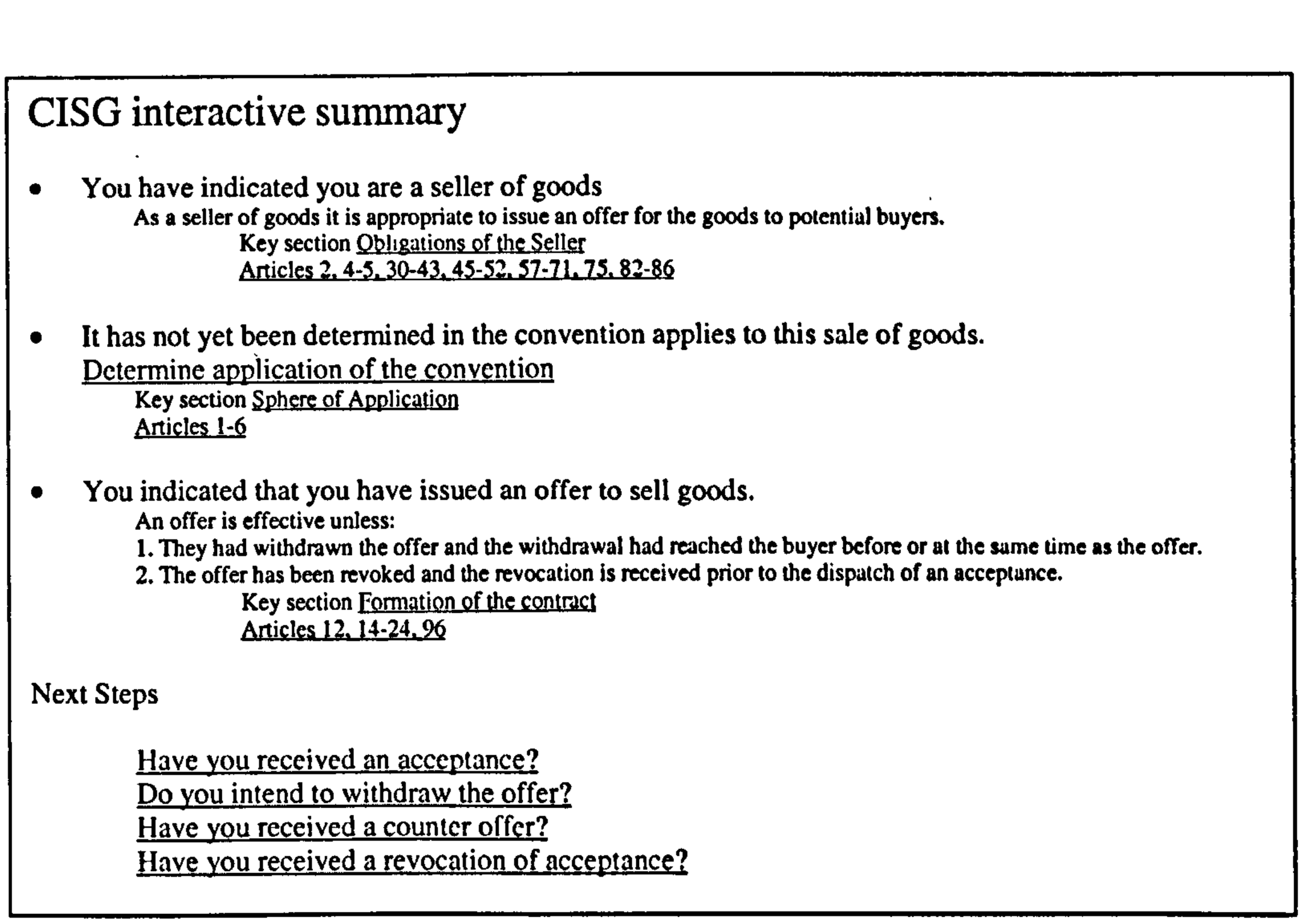
Are the goods insured

If the seller is not bound by the contract to provide insurance then they should make available sufficient details so that the buyer can gain insurance. Reference to article 32 is appropriate because this is the only instance where insurance of goods is discussed.

In many cases it is not appropriate to ask these questions individually but to embed associated queries such as those displayed above into a single document associated with the obligations of the seller. The queries can be broken down into appropriate groups and located in a single template. When the time comes to proceed to the

delivery phase, the reader would be presented with a new link to the next step, this would point to the template containing the delivery obligations and in turn load the relevant articles.

Because the CISG can be broken down into logical topics it is also possible to break the template architecture down into logical groups as well. The above consultations show two possible templates that could be constructed, one that focuses on establishing the state of a contract and offer and another that focuses on the obligations of a seller once and offer is effective and the goods are to be delivered. As each step is accessed and the consultation followed, links to later templates can be incorporated into the returned document. These links would provide the 'next step' in the consultation using facts and rules carried over from previous templates. An example consultation such as the first one involving the seller may result in a HTML document that would be presented as shown in Figure 5.9.



CISG interactive summary

- You have indicated you are a seller of goods
  As a seller of goods it is appropriate to issue an offer for the goods to potential buyers.
  Key section Obligations of the Seller
  Articles 2, 4-5, 30-43, 45-52, 57-71, 75, 82-86

- It has not yet been determined in the convention applies to this sale of goods.
  Determine application of the convention
  Key section Sphere of Application
  Articles 1-6

- You indicated that you have issued an offer to sell goods.
  An offer is effective unless:
  1. They had withdrawn the offer and the withdrawal had reached the buyer before or at the same time as the offer.
  2. The offer has been revoked and the revocation is received prior to the dispatch of an acceptance.
  Key section Formation of the contract
  Articles 12, 14-24, 96

Next Steps

Have you received an acceptance?
Do you intend to withdraw the offer?
Have you received a counter offer?
Have you received a revocation of acceptance?

Figure 5.9: *CISG HTML document, issuing an offer*

The template indicated in Figure 5.9 is focused on determining that the sale will be across international boundaries and that the CISG applies: it can determine that the

reader is a seller from previous navigation and facts set as during navigation and as introductions have been presented. The reader has also indicated that they have issued an offer. Based on these facts within the KB the page above is generated by activation of appropriate effects within the template. A brief sentence is given that describes the known information so the reader can confirm that the page applies to him. Along with each sentence is a brief paragraph giving key topics. A number of links are also associated with each sentence which indicate the core sections and individual article ranges relevant within the CISG and these link to CGI and database generated content and indexes. At the bottom of the page are a number of options for the reader to select. Each link points to a template that will further the consultation and at each stage focus on a particular subject. Links are only generated if appropriate events have occurred within an earlier template. In this way the information required to proceed will be known and stored in the knowledge base. If information is missing then links can be added to indicate this and the reader routed back to the appropriate stage (this is indicated in Figure 5.9 by the application of the convention sentence) in the template. Later consultations already accessed may already be concluded and the inference can be repeated without additional reader intervention.

Building on the linkage of HTML documents, documents containing a segment of a knowledge base, and carrying facts and characteristics from one page to the next, provides not only an online legal presentation solution but also a more general solution. Such an approach can be used for the presentation of information to many emerging Internet associated technologies such as interactive televisions, personal data Assistants and next generation WAP mobile phones. The open and simple concepts described in the previous sections can be used for many applications and for many varying consultation, navigation based or interactive tailoring of single documents or entire web sites. Although QERIS is in an early development phase and there are a number of issues still to address, there is potential in the expert system/Internet integrated model described and many possible applications that have still to be considered. QERIS is currently a tool for tailoring documents but also shows capability of providing a more interactive interface between web pages and the end user.

5.6 Summary

The CISG is a good example of a legal subject that can benefit from improved presentation when distributed via the Internet. This chapter provides an explanation of the CISG and indicates techniques for employing a QERIS knowledge based within the convention to improve its presentation to readers. This chapter has provided an overview of related topics including the CISG structure, contract lifecycle and a proposed QERIS based architecture. The CISG differs from Benefit Advice documents in that it consists of a single document containing separate sections. Like benefit Advice, certain sections apply based on user characteristics but the convention also applies throughout the evolution of a contract where obligations from both parties can change very rapidly.

Section 5.1 provides an overview of the CISG as a resource on the Internet providing an overview of research interests in the field and the goals that are used to define the CISG example.

Section 5.2 details the structure of the convention and how the document is formatted into topics and sections that are relevant to readers of different groups and at different times during the contract life cycle.

Section 5.3 Discusses in more detail the concepts of the contract life cycle by providing a number of example events and relationships between those events and related section of the convention.

Section 5.4 Provides a more detailed analysis of the convention, identifying the separate elements that can be extracted and used as part of a QERIS solution.

Section 5.5 details an example implementation where particular consultations with different parties are extracted, access to external database resources is explained and a final representation from a readers point of view is discussed.

CHAPTER 6

CONCLUSIONS

6 Conclusions

This work has brought together three separate areas of research, providing a means of improving the presentation of legal information delivered over the Internet. The three key areas and a summary of the issues include:

- The Internet, a delivery medium of information to a wide a varied audience - The Internet provides a method of storage, representation and distribution. It is limited by the technologies and standards used to define the presentation of documents and the methods of transmitting and rendering content. Although the Internet contains answers to many questions it is limited not by its content but by the interfaces used to access the information. The use of computers to present the information provides scope for using the processing capability not only for converting one representation (HTML) into another representation (screen image) but also intelligent adaptation of the content to an individual user.

- Artificial Intelligence as a means of adaptation to an audience and their varied characteristics. - Although several examples of artificial intelligent technologies are available over the Internet there are few examples that have succeeded and entered the mainstream. This work provides a review of relevant online examples in Chapter 2. It was surprising to find the lack examples available that demonstrated a focus on the presentation and support of interpreting documents that are already available online.

- Legal information and document presentation as a complex and relevant domain. - The presentation of legal information is an area of active research. It is appropriate to consider the presentation of legal information due to its relevancy to a broad audience. It is important that certain legal resources are available to the

general public but few attempts have been made to represent legal information to readers that may have difficulty interpreting and using the information?

This thesis has provided a detailed explanation of the techniques currently used in intelligently presenting information. A lack and maturity of examples employed over the Internet has reinforced the initial statement that there is an inherent difficulty in bringing together the two domains of AI and Internet presentation. This thesis has proposed a solution for Internet presentation (QERIS) that provides a mechanism for combining document content with an author's knowledge. In the undertaking of this thesis and the development of a proposed solution, several questions have been answered. A set of questions were put forward in Chapter 1, these have been addressed in the following way.

> *"To what extent are Intelligent-processing techniques currently used in conjunction with Internet information delivery?"*

There are very few examples of the conjunction between AI techniques and the Internet. Although a couple of examples described in Section 2.7 do exist, they are cumbersome in their integration between the Internet and the related AI solution. The trend is to use the CGI to provide client-server architectures that have several problems associated with them as described in section 3.2. The key issues include:

- The speed of interaction where each transaction (question/answer) is highly dependant on the network latency.

- The reliance on sending all user characteristics and data over the Internet for storage and processing by remote systems (see Section 2.7.1.2).

- Being able to successfully maintain short-term state between subsequent requests and also provide longer-term persistence of end user characteristics (see Section 2.7.1.1).

"How can Internet distributed information be better presented to the end user? Can information available over the Internet be presented in a form specific to the reader and the characteristics of the reader, their environment and context?"

Information presented over the Internet is largely marked up using HTML (and more recently XML). By intelligently modifying the underlying markup it is possible to provide many different variations in the final presentation of a document. The development of QERIS (see Chapter 3) and the demonstration of its use in two legal domains provides a good indication that:

- Information being presented can be tailored specifically to the characteristics of the end reader based on knowledge represented and embedded in the document they are accessing.

- An architecture where the intelligence and user model is deployed to the end users machine provides far more control over the maintenance, security and persistence of the user model. The characteristics and facts that are determined for one document could be employed in processing further documents.

"Is it possible to provide interactive online advice developing the concept that, an author can integrate their knowledge and experience of a subject domain with the document that describes that domain?"

The combination of the source document and a knowledge representation provides a powerful model for embedding an author's knowledge of a domain into the actual document describing the domain. There are suitable subject domains within law that provide a strong example of the integration between a documents content and an authors (or developed from legislative sources) knowledge representation. These subject domains include.

- Providing interactive consultative advice for Benefit claimants (see Chapter 4).

- Demonstrating that it is possible to maintain long term persistence that can be used to develop different document presentations based on the characteristics of the end user along with contextual states in the life cycle of a contract or entity that changes over time (see Chapter 5).

- A knowledge representation can be successfully integrated with an online document for later processing to determine the presentation of the document (see Section 3.5.8).

"To what extent can the presentation of legal content be improved, by employing AI techniques in its presentation over the Internet?"

The demonstrations in Chapters 4 and 5 indicate the initial capabilities of the QERIS system. Although very limited in its functionality and constrained by its lack of integration into Internet technologies such as web browsers, this very initial demonstration shows there is scope for moving these techniques forward. With the wide uptake of XML and the presentation of information to different devices, tailoring the presentation of content of a single document instance into many different representations will become an important requirement in the next generation of Internet applications.

In terms of improving the presentation of legal information, the two examples described in Chapter 4 and 5 demonstrate that the legal domain can benefit from improved information presentation. There is a need to reduce the shear volume of information by presenting the content that is more relevant to the user and their goals. This is a powerful concept for both practitioners in the legal domain and more importantly the general population. Taking a very complex and large document and distilling the content based on an the content being processed as well as the characteristics of the reader is an

interesting step in making information more accessible to a much wider audience.

6.1 Contributions

This research has taken a fresh perspective on the methods used to interact with expert systems currently employed over the Internet. There are a number of contrasts that can be drawn between existing techniques (1998 - 2000) and the techniques employed in the conceptualisation and realisation of QERIS.

- Rather than following the current methodology of a remote server that has full control of the inference and content, the 'intelligent' elements of the system have been relocated on the users own machine. In addition, rather than integrating the intelligence into an applet within a document where the applet must be downloaded, the emphasis has been on integrating it instead into the browser itself.

- Rather than carrying out consultations using traditional CGI approaches that suffer from network, security and trust issues the consultation occurs between the user and their own machine/software. Due to this close proximity of the user, the inference engine and knowledge base a far quicker and more immediate interaction can take place.

- The user should be given access to and ownership of the facts and characteristics that have been collected and assumed about them. The approach described in this thesis gives the user that control and also ensures that no information (except the history of page access and database queries) is sent over the Internet unless absolutely necessary. This concept is very important with legal information served to the general public. There may be concerns from the user about answering questions related to Social Security issues and it is possible for a user to mislead themselves by not wanting to be completely open or honest with a remotely administrated and faceless system.

- The Knowledge representation is minimal and simple. This concept follows on from the initial intention of the development and ultimate success of HTML. The goal when creating HTML was to create a mark-up that would

allow the fast and easy dissemination of information. The intention of QERIS knowledge tags is to extend HTML while maintaining as simple yet powerful methodology.

- There are few limitations in extending QERIS to provide additional document tailoring mechanisms that 'plug in' to the existing architecture.

- QERIS goes beyond just providing a consultative expert system, instead it provides a mechanism for developing document drafting, boilerplate document construction, expert system consultation, document tailoring and many other applications types common to the AI and Law field.

The last stages of this work focused on two main sources of legal information currently available in accessible forms on the Internet, proposing new approaches the way these materials are presented.

The concepts described above were used to implement a prototype system, QERIS that provides a HTTP proxy gateway that can process the HTML documents requested, push queries to the browser, and then deliver the final HTML content. The QERIS prototype was used to help visualise two example solutions for tailoring and improving the presentation of Social Security benefit Advice and also the CISG.

The Benefit Advice documents provide information to a very wide and varied audience about legal information. Benefit documents are written to help people to understand the legislative sources that underlie them and help people understand their rights and levels of benefit. Attempts have been made to simplify the online information and to guide the user by event oriented navigation yet the benefits and processing power available to manipulate and aid the reader in their understanding and interpretation of such information is not exploited. QERIS provides a mechanism for highlighting relevant clauses and also combining logic statements encoded in the wording and sentences of the advice to provide aid in the interpretation of the text. Each document contains a knowledge representation of the textual semantics and even the underlying legislative interpretations in order to provide a more interactive and improved presentation of each document. Switching the processing mechanism off still provides the original document. Activating the processing allows the system to query the user and build a model of their characteristics and requirements. Once

the characteristics are known they can be used to remove, add, and highlight textual segments or to add additional hypertext components such as dynamic links to the document. The result is not a benefit advice document written for the general population containing general rules and irrelevant content, but a very focused representation of the readers requirements that is concise and relevant.

The discussion and inclusion of the CISG example is to illustrate other attributes of a document such as the period for which different scenarios will exist and show how those attributes can change with time. The CISG also provides a good example for demonstrating the combination of other Internet resources such as content databases and how such resources can be integrated with QERIS to generate tailored documents.

This thesis has taken the key concepts and the prototype QERIS implementation and shown a number of benefits of integrating AI with Internet technologies.

- There are few examples of expert systems delivered over the Internet, examples that are available are based on traditional single user, single session consultations using the Internet as a new means of remotely accessing these systems rather than leveraging the capabilities of the internet to extend the expert systems capabilities.
- There are few examples of systems that can intelligently tailor documents received over the Internet to more closely resemble a document relevant to the reader. This approach although not original in other application areas is currently not evident on the Internet. The introduction of technologies such as XML are helping to increase the momentum of browser side intelligence but this is taking time to emerge into the main stream.
- Introduced the concept of intelligent presentation of Internet content., moving away from traditional expert system consultation as a means to reach a conclusion, instead using it as a means to reformat a document.
- Overcame several architectural issues such as integration with current web technologies, session management, user information management, content parsing and document tailoring.

- Demonstrated with the use of two examples in the legal domain the implementation of, and advantages related to tailoring the presentation of the content using the QERIS prototype.
- The tailoring of documents such as online benefit advice does provide a more focused representation of the information to the user. It is possible to highlight to the user areas that benefit and don't benefit their claim. It is possible in many situations to take the textual content and rules defined in the text and provide inference and conclusions that can be incorporated into the document. It is possible to generate a personalised summary of a larger resource in an interactive, quick and simple manner.

The thesis contains:

Chapter 2, a review of related research and technologies in AI, Internet and Law and an analysis of current techniques for providing online intelligent interaction. Many issues in all the relevant areas are discussed.

Chapter 3, Proposes a number of concepts that can be used to define a system for intelligent tailoring of online content providing a discussion of the specification and implementation of a prototype system for intelligently tailoring legal online content. The key concepts and methods used to realise them are described. A number of questions raised in Chapter 2 are answered by the QERIS solution.

Chapter 4, An example of an online legal resource, Social Security Benefit Advice is described. The issues with current paper and online versions are discussed. Methods of analysing the documentation and providing the integration with QERIS are provided. The chapter concludes with an example implementation of a Child Support advice document, how it is interpreted and then represented using QERIS mark-up and the possible interface that a potential user would interact with.

Chapter 5, An example of a different type of online legal resource, Convention for the International Sale of Goods is described. This example varies from that described in chapter 4 by including contextual elements such as the role of the contract on the inference and the use of external resources to augment QERIS. Rather than acting as

just a tailoring mechanism, QERIS also provides a degree of intelligent hypertext document construction.

6.2 Critique

This work provides a good snapshot of technologies in the associated research areas and describes a number of important examples. The work also introduces and demonstrates from concept through development to example implementation a proposed solution to the questions poised.

There are a number of areas where the work can be criticised or seen to be lacking. The most obvious observation is the high level and practical approach taken. This is a reflection on an experimental approach where the interactions between certain systems were uncertain and the most reliable way to be confident that proposed solutions would work is to develop a prototype and try it.

Within the AI and Internet technologies, many of the most prominent applications and solutions have been developed outside an academic environment. It was at times challenging to find reliable academic sources for many of the applications discussed. The lack of interest in AI and Internet integration especially amongst Internet research circles and conferences was also surprising.

Although two examples of am implementation using QERIS are provided there are no user or interaction studies to validate claims that a tailored document is better or easier to understand than the original source. There are a number of reasons such studies were not made but the most prominent is the lack of a fixed example that would test the QERIS concept rather then the HTML layout and page designs. Many of the functions of QERIS are transparent to the user and difficult to quantify. The vast variations in document and knowledge base designs also would provide other problems as it would be unclear if tests were swayed by poor layout of the original document or poor design of the content rather than the tailoring mechanisms employed. Although these reasons exist the work could be strengthened by taking one of the examples further and providing a real online test bed of the solution. In order to do this the QERIS prototype would need further development to improve its

robustness and the knowledge bases incorporated into the documents would need extending.

The implementation is limited by a simple knowledge representation and inference engine, in some applications the knowledge representation will be lacking. For example the inability to handle conditions such as 'A < 10' makes it difficult to check numerical attributes. During the development of this work it was clear that the goal was not to create another expert system shell but rather to demonstrate that a general expert system shell incorporated into a web browser would be able to provide very powerful processing and presentation capabilities.

6.3 Further work

The area of AI, Law and Internet integration is very large and there are many areas that this research has not touched on in order to maintain focus on the original concepts rather than established and well understood issues. An example of an area that has seen little discussion in this work is the creation and maintenance of the knowledge representation within the content. The distributed nature of the knowledge representations prompts a number of important questions. The concepts of reusing the knowledge gained about a reader during a consultation and how that knowledge can be shared across resources and Internet sites would benefit from consideration. There are other interesting application areas where QERIS could prove very useful, areas that could be investigated further include:

- User support in filling out online forms.
- An online consultative expert system implementation that provides advice and conclusions rather than focusing on document tailoring.
- Integration of QERIS with offline or non-legal based administration systems.
- Contract creation and boilerplate construction of legal texts.
- It is clear that this work touches on the tip of a larger research area in online expert systems, intelligent Internet interfaces and the integration of interactive assistants into information portals.

There are a number of practical development areas that could be developed further and these include:

QERIS uses a modular construction and it would be interesting to integrate an existing knowledge representation and expert system shell into QERIS rather than the current limited implementation. A good example of reusing existing expert system shells is the Jess implementation that uses a CLIPS expert system shell at its core. By extending the capabilities of the inference engine and knowledge representations in this way it would be possible to develop more complex and comprehensive knowledge bases. The disadvantage of such an integration is that the simplicity in learning, creating and using the current knowledge representation would be lost.

The development of QERIS as a proxy server was a prototypical implementation, the longer term plan and ultimate goal of developing QERIS was to integrate it directly into the web browser. Initial experiments early in the development stage considered using the plug-in or API capabilities of modern web browsers. The introduction of XML in the 2nd year of this thesis (1998) has provided impetus to include scripting capabilities such as XSL into web browsers. It is also clearer that the disparate mark-up variants emerging will require a degree of intelligence within a web browser to handle the processing and presentation of the content.

The concepts associated with maintaining and developing the user characteristics and profiles needs a lot of consideration and further work. The user requires a good method of viewing and modifying the information stored about them so that they can maintain and update their characteristics as they change. The representation of these profiles and characteristics externally to QERIS so they may be reused by other applications is also an area of interest. It is considered that XML could be used not only to represent the knowledge tags used to build the knowledge base, but also as a means of representing and storing the users profile and characteristics.

The creation and maintenance of the knowledge representation has not been considered in detail. It is believed that a degree of automation could be implemented to provide automatic generation and integration of knowledge representation with the source documentation. If automation is not possible at this stage then there is

certainly scope for developing tools to aid the integration of an author's knowledge into the underlying mark-up of a document.

The introduction of XML provides a very interesting development over HTML and validates the intention of extending HTML to include knowledge representation tags. XML provides a means of defining a broad range of Internet mark-up languages and an example of an XML DTD for the knowledge representation is provided in the Appendix.

The concepts, ideas and proposals outlined in this work provide a new perspective on intelligent interaction between information and readers. We believe that this work outlines a novel approach to integrating AI technologies into the Internet and opens up many avenues for continued research and the development of these concepts. It is clear that as the Internet evolves and the impetus grows aided by the vast number of connected users, the many different applications and the growing number of access methods. Developing ways of helping the user understand the information they are accessing, helping them to find the relevant source amongst growing volumes of irrelevant sources will be a very important issue in the future of the Internet. The Internet is becoming the primary source of information, in many cases, the first port of call to determine an answer to a question, yet the methods of accessing such a key resource are still young and uninspiring. Over the next few years exciting things will have to happen in order to move the Internet forward and fully realise its potential.

7 References

Aleven V. (1997). Teaching Case-Based Argumentation Through a Model and Examples. Ph.D. Dissertation.

Allen L. E.. (1997). Anatomy of a MINT system for Article 14 of the United Nations Convention on Contracts for International Sale of Goods. The Forth International Workshop on Legal Expert System for the CISG, Legal Expert System Association (LESA), Japan, pp 14-27.

Ashley K. D. (1989). Toward a Computational Theory of Arguing with Precedents: Accomodating Multiple Interpretations of Cases. in Proceedings of the Second International Conference on AI and Law, ACM Press.

Ashley K. D & Aleven V. (1997). Evaluating a Learning Environment for Case-Based Argument Skills. in Proceedings of the Sixth International Conference on AI and Law, ACM Press, pp 170-179.

Ashley K.D. & Aleven V. (1991). Toward an Intelligent Tutoring System for Teaching Law Students to Argue with Cases. in The Third International Conference on Artificial Intelligence and Law, ACM Press, pp 42.

Belew R.K. (1987). A Connectionist Approach to Conceptual Information Retrieval. in Proceedings of the First International Conference on AI and Law, ACM Press, pp 116-126.

Bench-Capon T. J. M & Staniford G. (1995). PLAID Proactive Legal Assistance. , ACM Press, pp 81-87.

Bench-Capon T. J. M, Coenen F. P. (1992). Isomorphism and Legal Knowledge Based Systems. AI and Law, pp 65-86.

Bench-Capon T. J. M, Lowes D, McEnery A. M.. (1991). Argument-based Explaination of Logic Programs. Knowledge Based Systems, Butterworth-Heinemann Ltd, pp 177-183.

Bench-Capon T.J.M (ed.). (1991). Knowledge Based Systems and Legal Applications. Academic Press.

Bing J. (1987). Designing Text retrieval Systems for 'Conceptual Searching'. in Proceedings of the First International Conference on AI and Law, ACM Press, pp 43-51.

Born R (ed.). (1987). Artificial Intelligence, The Case Against. Croom Helm.

Branting K. L, Callaway C. B, Mott B. W, Lester J. C. (1999). Integrating Discourse and Domain Knowledge for Document Drafting. to appear in Proceedings of the International Conference on AI and Law, ACM Press.

Branting L. K, Lester C. J, Callaway B. C. (1997). Automated Drafting of Self-Explaining Documents. in Proceedings of the Sixth International Conference on AI and Law, ACM Press, pp 72-81.

Buchanan B. G, Shortliffe E.H. (1984). Rule Based Expert Systems. Addison-Wesley.

Buchanan, B. G., and E. A. Feigenbaum. (1978). DENDRAL and META-DENDRAL: their applications dimensions. Artificial Intelligence, pp 5-24.

C Grice & S Ard. (2000). Hoax briefly shaves $2.5 billion off Emulex's market cap. CNET news.com (news.cnet.com).

C. S. Horstmann & G. Cornell. (1998). Core Java 2 Volume1: Fundamentals, 4th Edition. Prentice Hall.

Capper P. (1987). Latent Damage Act 1986, A Legal Studies and Services Special Report. London, Legal Studies and Services.

Culley A. (1995). Intelligent Template Document Automation? What is HotDocs!. Law Technology Journal.

De Mulder R.V, Van den Hoven M. J, Wildemast C. (1993). The concept of Concept in 'Conceptual Legal Information Retrieval'. Law Technology Journal.

Daskalopulu A. (1999). Logic-Based Tools doe the Analysis and Representation of Legal Contracts. PhD Thesis, Imperial College London.

Daskalopulu A, Sergot M. (1995). A Constraint-Driven System for Contract Assembly. ACM Press, pp 62-70.

DHSS. (2000). CH1 Child Benefit Leaflet.

Dick J. P. (1991). Representation of Legal Text for Conceptual Retrieval. in Proceedings of the Third International Conference on AI and Law, ACM Press, pp 244-253.

Dick J. P. (1987). Conceptual Retrieval and Case Law. in Proceedings of the First International Conference of AI and Law, ACM Press, pp 106-114.

Duda, R., Gaschnig, J. and Hart, P.,edited by Michie, D.. (1979). Model design in the PROSPECTOR consultant system for mineral exploration, in: Expert systems in the micro electronic age. Edinburgh University Press, Edinburgh, pp 153-167.

Fisher R. (1997). Determination of Residence Status for Taxation Law: Development of a Rule-based Expert System. in Proceedings of the Sixth International Conference on AI and Law, ACM Press, pp 161-169.

Gelbart D, Smith J. C. (1991). FLEXICON, 'A Legal Text-Based Intelligent System. In Proceedings of the Third International Conference on Artificial Intelligence and Law, ACM Press, pp 225-233.

Gordon C & Veloso M. (1999). C Gordon, M Veloso et al, Guidelines in Healthcare: the experience of the Prestige project. , MIE 99.

Gordon C & Veloso M. (1996). The PRESTIGE Project: Implementing Guidelines in Healthcare. in, MIE 96.

Gray, N. A. B., D. H. Smith, T. H. Varkony, R. E. Carhart, and B. G. Buchanan.. (1980). Use of a computer to identify unknown compounds: the automation of scientific inference. Chapter 7 in G. R. Waller and O. C. Dermer, eds., Biomedical Application of Mass Spectrometry, Academic Press, New York.

Greenleaf G, Mowbray A. (1993). Controlling and Augmenting Legal Inferencing: 'ysh, a case study'. in Proceedings of the Fourth International Conference on AI and Law, ACM Press, pp 162-166.

Greenleaf G, Mowbray A & Tyre. (1991). The DataLex Workstation - Integrating Tools for Lawyers. in Proceedings of the Third International Conference on AI and Law, ACM Press, pp 215-224.

Greenleaf G, Mowbray A, King G, Cant S, Chung P. (1997). More than wyshful thinking: AustLII's legal inferenceing via the World Wide Web. in Proceedings of the Sixth International Conference on AI and Law, ACM Press, pp 47-55.

Hafner C.D. (1987). Conceptual Organization of Case Law Knowledge Bases. in Proceedings of the First International Conference on AI and Law, ACM Press, pp 35-42.

ISO 8879. (1986). Information processing -- Text and Office Systems -- Standard Generalized Markup Language (SGML). First edition -- 1986-10-15. [Geneva]: International Organization for Standardization, ISO (International Organization for Standardization).

Jackson, P. (1990). Introduction to expert systems (2nd ed). Addison-Wesley.

Japp de Wildt, Schmidt A, Quast J, Japp van den Herik, Robert van Kralingen, Visser P, Wim de Vries. (1991). The Evolution of Research Aims. Legal knowledge Based Systems, Vermande, pp 106-117.

Johnson P & Mean D. (1991). Legislative Knowledge Based Systems for Public Administration - Some Practical Issues. in Proceedings of theThird International Conference of AI and Law, ACM Press, Oxford, pp 108-117.

Knuth D. (1984). TeX is The TeXbook. , Addison Wesley.

Krause P & Clark D. (1993). Representing Uncertain Knowledge, An Artificial Intelligence Approach. , Kluwer Academic Publishers.

LEXIS-NEXIS. (1998). LEXIS-NEXIS, Learning LEXIS, A Handbook for Modern Legal Research. , LEXIS-NEXIS, 9443 Springboro Pike, P.O. Box 933, Dayton, OH45401.

Lindsay, R. K., B. G. Buchanan, E. A. Feigenbaum, and J. Lederberg. (1980). Application of Artificial Intelligence for Chemistry: The DENDRAL Project. McGraw-Hill, New York.

Luger G. F & Stubblefield W. A. (1997). Artificial Intelligence: Structures and Strategies for complex problem solving (3rd ed). Addison Wesley.

McDermott, J. P. (1993). R1 ("XCON") at Age 12: Lessons from an Elementary School Achiever. Artificial Intelligence, vol 59, pp 241-247.

McDermott, J. P. (1981). R1: The Formative Years. AI Magazine, vol 2:2, pp 21-29.

McNeely, M.D., & Smith, B.J. (1995). An interactive expert system for the ordering and interpretation of laboratory tests to enhance diagnosis and control utilization. in Canadian Medical Informatics, vol 2:3, pp 16-19.

Morris A (ed.). (1992). The Applications of Expert Systems in Libraries and Information Centres. Bowker-Saur.

Morris A (ed.). (1992). The Application of Expert Systems in Libraries and Information Centres. , Bowker-Saur.

Paterson A & Susskind R. (1993). Technology, Lawyers and the Atlantic Divide. Law Technology Journal, vol 3:2.

Poulin D, Huard G, Lavoie A. (1997). The other formalization of Law: SGML modelling and tagging. in Proceedings of the Sixth International Conference on AI and Law, ACM Press, pp 82-88.

Punin J. R & Krishnamoorthy M. S. (1994). A Simple HTML Editor. in Proceedings of the Second WWW Conference '94.

RFC2109. (1999). HTTP State Management Mechanism, http://www.faqs.org/rfcs/rfc2109.html. , Bell Laboratories.

Rissland E. L & Ashley K. D. (1987). A Case-Based System for Trade Secrets Law. in Proceedings of the First International Conference on AI and Law, ACM Press.

Royles C, Bench-Capon T. J. M. (1999). Customising the Presentation of Legal Docuemnts over the World Wide Web. in Infomration & Communications Technology Law, vol 8:2.

Rt Hon Lord Justice Brooke. (1998). IT and the English Courts: the next ten years (http://www.bileta.ac.uk/98papers/brooke.html). 13th Annual BILETA Conference "The Changing Jurisdiction", Trinity College, Dublin.

Sehmi A & Kroening M. (1996). WebLS: A Custom Prolog Rule Engine for Providing Web-Based Tech Support. in 1st Workshop on Logic Programming Tools for INTERNET Applications, JICSLP '96.

Soper P & Bench-Capon T. J. M. (1994). Coupling Hypertext and Knowledge Base Systems: Two Applciations in the Legal Domain. in Artificial Intelligence and Law, vol 2, pp 293-314.

Soudakoff A & Lauritsen M. (1997). Shopper's Guide to Legal Document Assembly (http://www.docauto.com/locart.htm). in Law Office Computing (October/November 1997), pp 69.

Spirgel-Sinclair S. (1988). The DHSS Retirement Pension Forecast & Advice System: an update. in Proceedings of the Second European Conference on KBS in Government, Blenheim online Ltd, pp 89-106.

Sprowl J. A. (1980). Automating assembly of legal documents. Computer Science and Law (Niblett B, ed), Cambridge University Press, New York, pp 195-205.

Stefik M. (1995). Introduction to Knowledge Systems. Morgan Kaufmann.

Susskind R. (1989). Expert Systems in Law, A Jurisprudential Inquiry. Clarendon Paperbacks.

W3C, Extensible Markup Language (XML) 1.0 Recommendation. (1998). W3C, REC-xml-19980210, http://www.w3.org/TR/1998/REC-xml-19980210. W3C.

W3C, Extensible Stylesheet Language (XSL) 1.0 Working Draft. (2000). W3C, WD-xsl-20000327, http://www.w3.org/TR/2000/WD-xsl-20000327. W3C.

W3C, HyperText Markup Language (HTML) 4.01 Specification. (1999). W3C, REC-html401-19991224, http://www.w3.org/TR/1999/REC-html401-19991224. W3C.

Widdison R. (1993). Databases in Law Firms - An Overview. Law Technology Journal, vol 3:1

Widdison R, Pritchard F, Robinson W. (1992). Expert System Meets Hypertext: The European Conflicts Guide. Information Technology and Legal Education.

Yoshino, H. (1997). On the Logical Structure of Contract Law. The Forth International Workshop on Legal Expert System for the CISG, Legal Expert System Association, pp 43-54

Appendix A

Application and resource dependencies

	Application location			
Attribute	*Client Machine*	*Intermediate Machine (proxy)*	*Remote Machine (server)*	Language (e.g. JAVA applet)
Simultaneous users	1	≥1	≥1	1
Download requirements	Page contents + KB	Page Contents + Queries + Responses	Page Contents + Queries + Responses	KB – limited page control
Disk access	Local HD + removable	Intermediate HD	Server HD	Limited
User feedback	GUI and WWW Browser	WWW Browser	WWW Browser	WWW Browser
Access to client applications	Yes	No	No	No
Access to server applications	Yes via CGI	Yes on intermediately	Yes	No
Information privacy	Yes – not broadcast	No	No	Yes
Query delay	1 Client dependant	≥1 Clients + network	≥1 Clients + network	1 Client dependant
Client install	Yes	No	No	No
Update strategy	Patches and updates	Remotely upgradable *	Remotely upgradable *	Remotely upgradable

* If implemented

KB – knowledge Base

HD – Hard disk

Appendix B.1

XML DTD representation of QERIS effect, rule and query tags

```
<?XML version="1.0" ?>
<!DOCTYPE knowledgebase [

<!ELEMENT main(rule|query|effect)*>
<!ELEMENT rule (#PCDATA)>
<!ATTLIST rule
  id ID #REQUIRED
  condition PCDATA #REQUIRED
  conclusion PCDATA #REQUIRED>
<!ELEMENT query (#PCDATA)>
<!ATTLIST query
  id ID #REQUIRED
  question PCDATA #REQUIRED
  answer PCDATA #IMPLIED
  help PCDATA #IMPLIED>
<!ELEMENT effect (#PCDATA)>
<!ATTLIST effect
  id ID #REQUIRED
  condition PCDATA #REQUIRED
  plugin
(insertString|insertFile|debug)"insertString">
  true CDATA #IMPLIED
  false CDATA #IMPLIED>

]>

<main>
  <rule id="1"
    condition="dwelling_occupied_home is yes"
    conclusion="condition1"/>
  <query id="dwelling_occupied_home"
```

```
    question="Do you occupy the dwelling as
your home"
    answer="yes,no"

help="http://www.somewherelse.com/helpfile.html
"/>
  <effect id="condition1_text"
    condition="condition1"
    plugin="insertString"
    true="<FONT COLOR=green>The person meets
the first requirment.</FONT>"
    false="<FONT COLOR=red>The person does
<B>not</B> meet the 1st
      requirment.</FONT>"/>
</main>
```

Appendix B.2

Body content of an example XML document that employs the DTD defined above.

```
<main>
<rule id="rule1">
  <condition>
    <variable     name="dwelling_occupied_home"
value = "yes"/>
  </condition>
  <conclusion>
    <variable name="condition1"
  </conclusion>
</rule>
<query id="query1">
  <question>
    Do you occupy the dwelling as your home?
  </question>
  <response type="multi">
    <answer>yes</answer>
    <answer>no</answer>
    <answer>unknown</answer>
  </response>
  <help>
<helpref="http://www.somewhere.com/help1.html">
dwelling info</helpref>

<helpref="http://www.somewhere.com/help2.html">
occupation info</helpref>
  </help>
</query>
<effect id="effect1" type="inline">
```

```
<condition>
  <variable name="condition1">
</condition>
<plugin>
  <type value="insertString">
  <true value="Yes you are resident">
  <false value="No you are not a resident">
  </plugin>
</effect>
<effect id="effect1" type="block">
  <condition>
    <variable name="condition1">
  </condition>
  <plugin>
    <type value="insertTag">
    <true   open="<relevant   type='enhance'>"
close="</relevant>">
    <false open="<hidden>" close="</hidden>">
  </plugin>
  <content>
    This section of markup can be processed by
the surrounding effect tag. If
    the condition1 succeeds then this text will
be marked up by the "relevant"
    tag, if   condition1 fails then the text
will be marked up by the "hidden"
    tag.
  </content>
</effect>
</main>
```

Appendix C

Online Child Benefit document CH1 (1998). This document has now been replaced (Sept 2000) with an updated and simplified version. It is reproduced here as reference for sections that were developed early in this research when this page was still current.

Child Benefit

Type: Non-contributory. Not income related.

Taxable: No.

Who is entitled ? A customer who is responsible for a child living with them; or

is contributing to the child's maintenance at a rate of at least equal to the Child Benefit payable for the child; and

the customer's right to enter or reside in Great Britain is not subject to a limit or condition, and the customer or the child have been resident in Great Britain for a total of more than 26 weeks in the preceding 52 weeks. A person coming to Great Britain from abroad can claim Child Benefit if he or she:

has been granted refugee status; or

has been given exceptional leave to remain in the United Kingdom; or

is a national, or member of the family of a national of an EEA country; or

is lawfully working in Great Britain and who is a national of Algeria, Morocco, Slovenia or Tunisia, or a member of the family of such a person living with them; or

is covered by one of Great Britain's agreements on Child Benefit with other countries; or

has leave to enter or remain in the United Kingdom which is not subject to any limitation, for example, has been given indefinite leave to enter or remain or the right of abode in the United Kingdom.

Asylum seekers are not entitled to Child Benefit.

An employed or self employed person, who comes under the United Kingdom National Insurance arrangements, can be paid Child Benefit during the first 26 weeks of residence in Great Britain if they intend to stay here for at least six months.

The child should be:

under age 16; or

under age 19 and still in full-time education of at least 12 hours a week at a recognised educational establishment, studying for a qualification up to and including A level/(G) NVQ level 3 or an equivalent.

How to claim: Form CH2 (for an initial claim) or form CH497 (for a subsequent child/children) should be completed and returned, with the child's birth certificate to the Child Benefit Centre.

A claim can be made at any time after the child's birth but payment can only be backdated for a limited period. The customer should not delay in claiming if the birth certificate is not available.

How is it paid ? Every four weeks in cash at the post office, or by direct payment into a bank or building society account.

If the customer receives

Income Support; **or**

Jobseeker's Allowance (Income based); **or**

Family Credit; **or**

Disability Working Allowance

they can ask to have the benefit paid weekly at the Post Office.

In some areas Child Benefit will be paid by the card method of payment replacing the use of order books.

Customers will be advised of their method of payment, or any change to their existing method of payment.

Rates of Child Benefit

Effective from 6 April 1998

£11.45	**Eldest or only child (standard rate)**
£17.10	**Eldest or only child (lone parent rate) (for those bringing up children alone)**
£ 9.30	**Each subsequent child**

Other information

The customer will be excluded from the right to Child Benefit if:-

the child has been looked after by the Local Authority or been in prison/legal custody for more than eight weeks;

the child is boarded-out with foster parents and an allowance for the child's maintenance is being paid by the Local Authority;

the child is married to or living with a partner, (unless the spouse or partner is receiving full-time education). If the child is married, but not living at the same address as their spouse, someone else may claim the benefit;

the child is receiving Income Support or Severe Disablement Allowance;

their earnings (or their spouse/partner's earnings) are exempt from UK tax;

the child is aged 16 or over, has left school and is engaged in remunerative work for 24 hours or more a week or is receiving a Youth Training Allowance; or

the child stays in hospital for more than 12 weeks and they are not providing for the maintenance of the child.

To obtain further information ring:-

Child Benefit Centre
0541 555540

or write to:-

Child Benefit Centre
PO Box 1
Washington
Newcastle-upon-Tyne
NE88 1AA

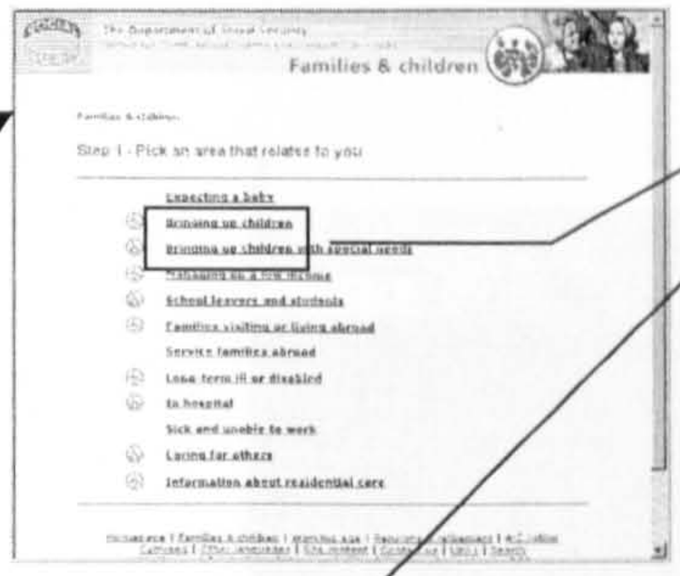
Main leaflets: CH1, CH4, CH4A, CH5, CH6, CH7, CH8 (other languages), FB8, FB27, NI17A.

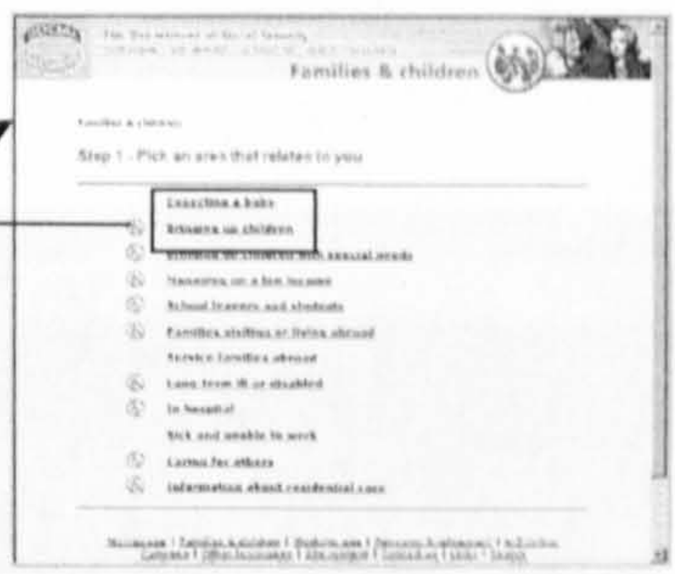
See also:

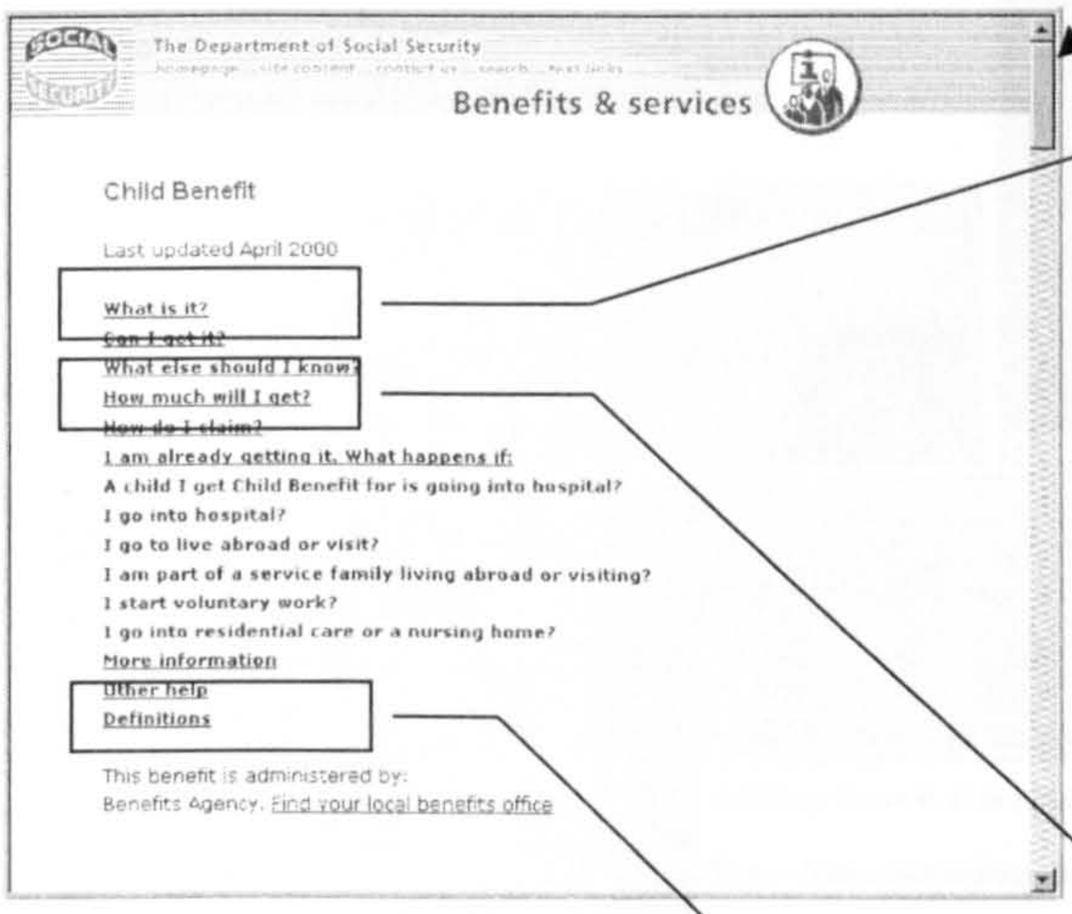
Child Benefit (Lone Parent)

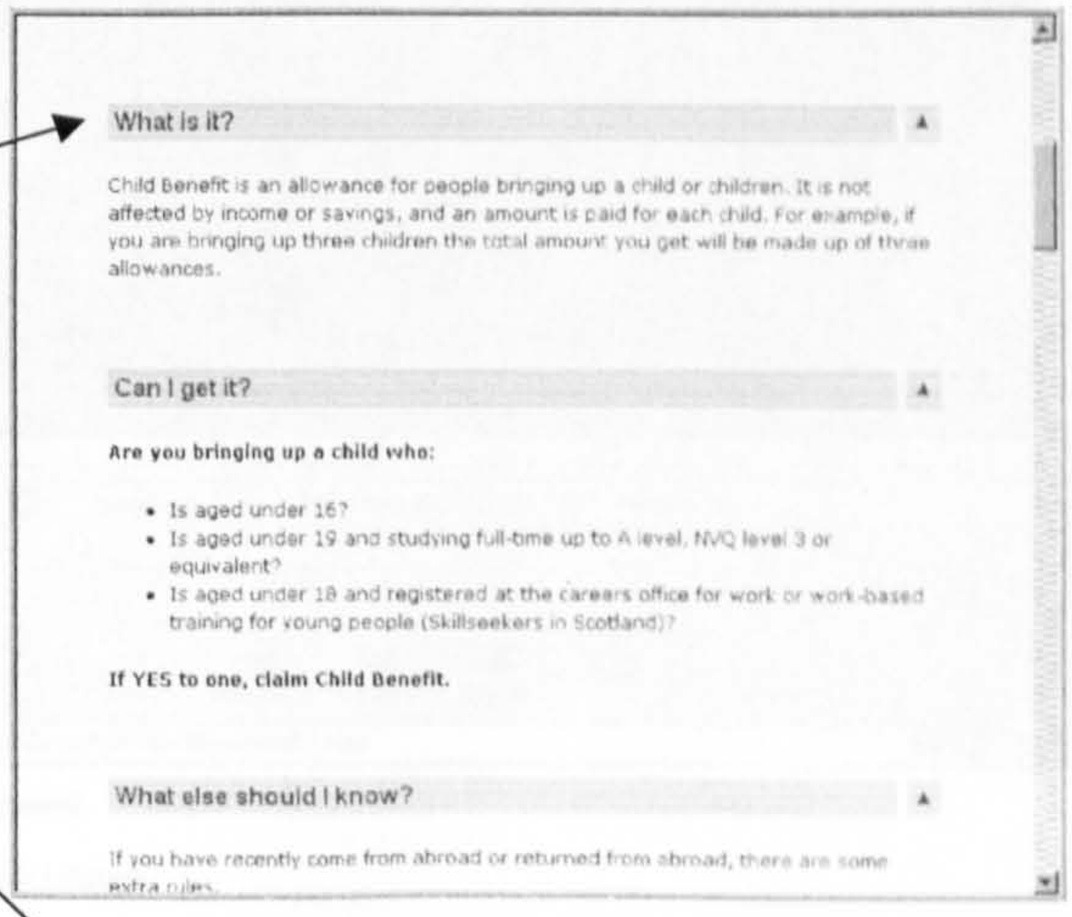
Appendix D

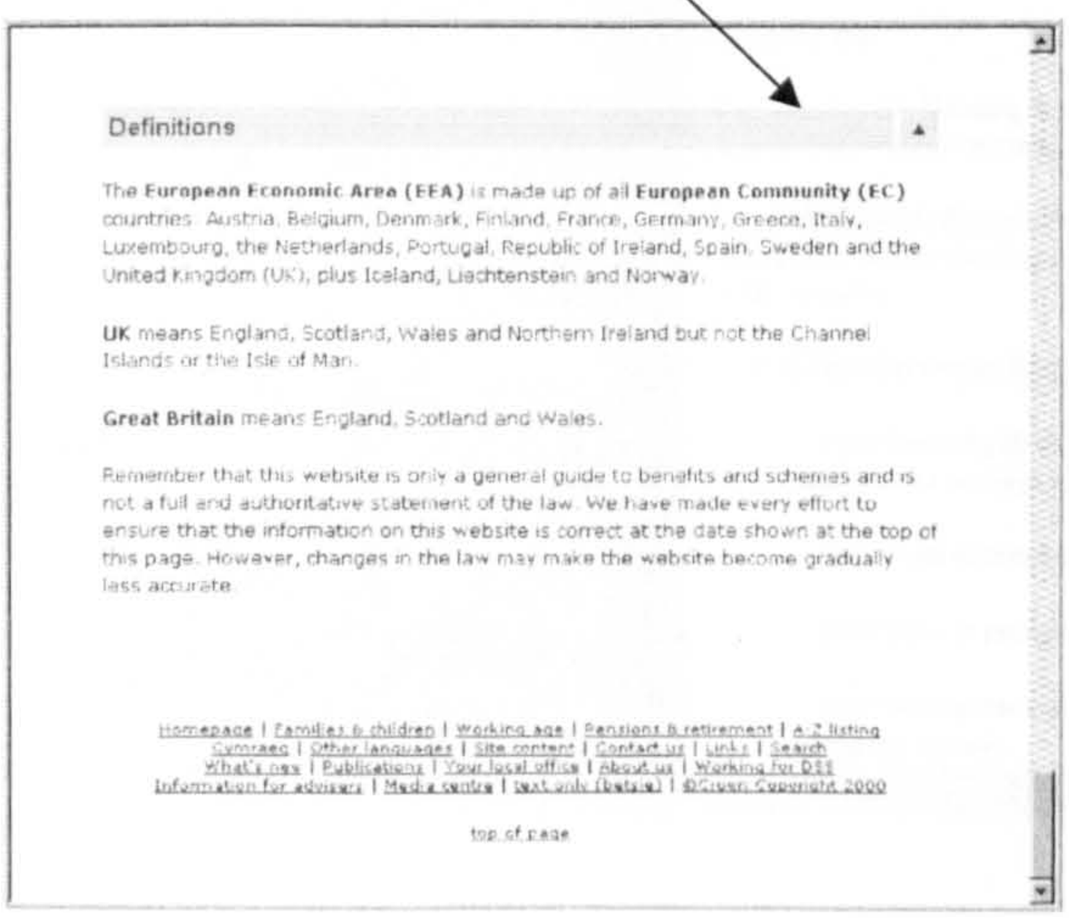
Large format picture of Benefit Advice Internet presentation

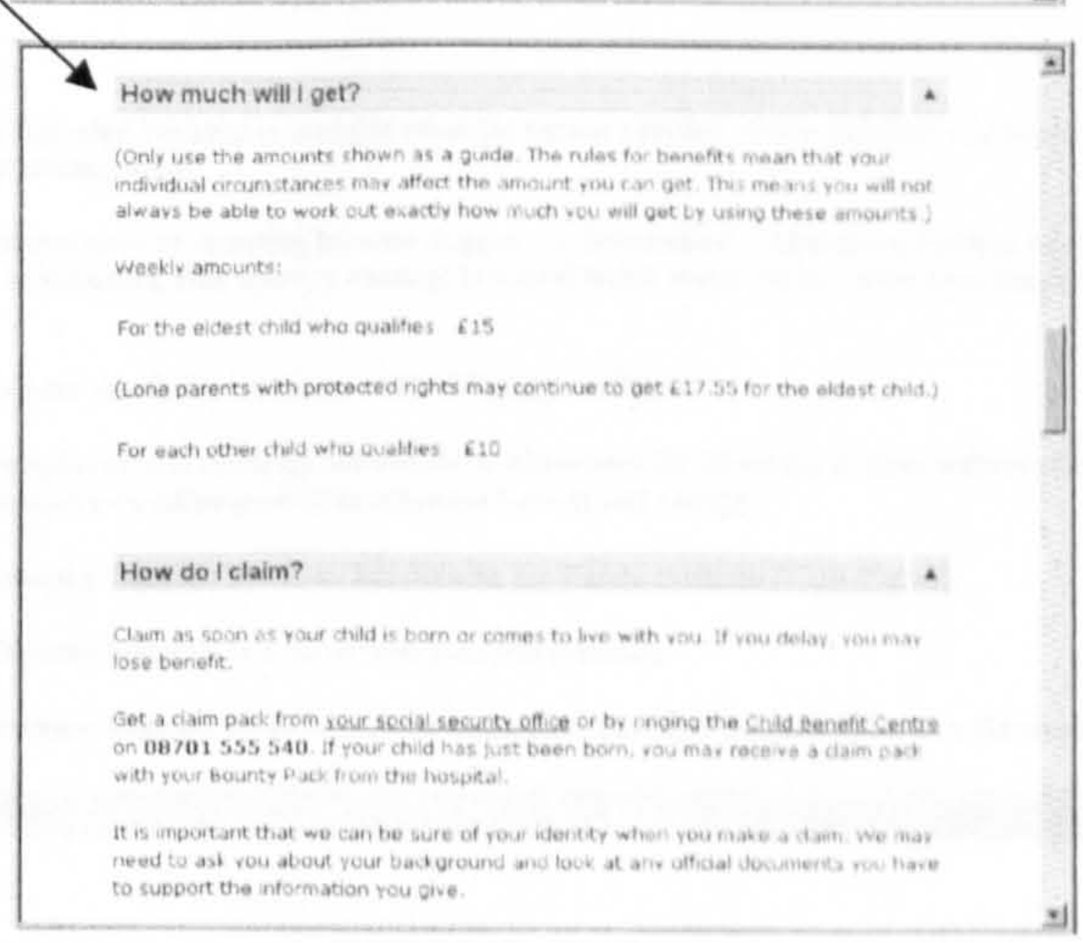












Appendix E

Large format picture of QERIS components.

1. QERIS control panel
2. Java DOS window
3. Rule hierarchy display
4. Standard Internet browser