JointZone
An Adaptive Web-based Learning Application

Muan Hong Ng [1], Ray Armstrong [2], Wendy Hall [3] and Pat Maier [4]

[1,3] Intelligence, Agents and Multimedia Research
University of Southampton
SO17 1BJ Southampton, United Kingdom
Email:{mhn99r [1], wh[3]}@ecs.soton.ac.uk

[2] Rheumatology Unit, Mail Point 78
Southampton General Hospital
SO16 6YD Southampton
United Kingdom
Email:rayarmstrong@btinternet.com

[4] Center of Learning and Teaching
University of Southampton
SO17 1BJ Southampton, United Kingdom
Email:pjm@soton.ac.uk

Technical Report Number: ECSTR-IAM02-002
ISBN:0854327711

Abstract

This report presents the work involved in developing an adaptive web site, JointZone that personalises student learning on the web. The work combined user modeling, domain modeling and adaptive hypermedia techniques to deliver a personalised web-based learning environment. The idea of keyword indexing and the site layout structure was used to model the domain giving a conceptual and structural representation of the content. The student model involves the novel idea of using effective reading speed to better gauge if a student has read a page. The project applied the combination of adaptive link hiding and link annotation on a fully functional web site to present an adaptive web-based learning environment.

Contents

List of Figures

1 Introduction

Resources in an information-rich hypermedia environment allow users to browse the material in an informal way until something of interest commands greater attention. Alternatively, they may search the material in a goal-directed manner to find the information they need. The former is analogous with an informal learning environment (Duchastel, 1989) where one's attention skims a wide range of information. A corpus of information in a hypermedia environment is not necessarily an ideal learning environment as it is essentially 'non-pedagogical', and generally provides minimal structural support (Duchastel, 1992) . However, this mode of learning can be effective if the reader already has a fairly well developed schema (a personal construct of related information) for the material being read. A schema acts to filter incoming information and provides a framework for processing information. For a well-developed schema, paying scant attention to material as often happens when browsing, is very often enough to modify a schema, and hence learn. Practitioners, or those knowledgeable in a field therefore, may use this technique in a casual manner and informally upgrade their knowledge of a subject. Novices with less well-developed schemas are more easily overwhelmed and quickly suffer from cognitive overload. General browsing therefore is not necessarily an effective learning method for them (Mayes, Kibby, & Anderson, 1990) . Goal-directed learning on the other hand is more focused, affording a greater opportunity to attend to the material. It is important therefore to provide guided support, or scaffolding for novices while allowing the more advanced reader to move more quickly through the material. The nature of hypermedia as a searchable network of information can enhance goal-directed learning through effective search facilities while also allowing easy access to a wide variety of material for browsing. However, both these functions of hypermedia can still lead to a cognitive overload (Thuring, Hannemann, & Haake, 1995) as even with goal-directed learning a vast array of linked resources can be presented to the learner. Adaptive hypermedia is one approach that takes into account the reader's differences (i.e. browsing history, knowledge level, preferences) and can present an appropriate selection or an adapted set of content nodes - adaptive presentation. It can also provide adaptive navigational support offering a selection of links based on reader knowledge and, or readers' browsing history (Brusilovsky, 2001). In both these cases, adaptive mechanisms support the reader by offering a guided set of documents or links with the aim of reducing the cognitive overload, thus enhancing the reader's ability to attend to pertinent documents and learn.

1.1 Project Description

This work stems from a project funded by the Arthritis Research Campaign (ARC) to develop a web site, JointZone, for the study of Rheumatology primarily for undergraduate medical students. A table of contents sets out various sub-categories which comprise some basic science, a section describing rheumatic disorders, a description of history taking and examination of the locomotor system, the investigation of rheumatic disease and disease management. In an additional section, there is a total of 30 case histories. These ”virtual patients” present rheumatic disorders represented in a variety of clinical scenarios, most of which arise frequently in the typical District General Hospital rheumatology clinic. The cases are subdivided into three groups designated ”Beginner”, ”Intermediate” and ”Advanced”. The layout of these cases differs according to category and the nature of the interactivity reflects the assumed degree of expertise of the user. The remainder of the content exists in the form of an ”electronic online textbook” which is liberally illustrated. More than 80 documents and over 700 illustrations have been included. There are also 43 video clips, all but one of which have been especially produced for this application. The user is given the choice of either streaming these across the internet from the website server or playing them from a CD-ROM. The latter provides high resolution video playback. The majority of the video clips illustrate technique for examining the musculoskeletal system and some illustrate specific findings in patients with various rheumatic disorders. In addition to this, there is a detailed illustrated glossary. Hyperlinks to this glossary exist throughout the website in the form of highlighted words in the text. The site has a sophisticated search engine and there are context-sensitive links to educational material held on the external ARC website.

1.2 The Pedagogical Rationale

JointZone provides a rich source of material to enrich users' domain (declarative) knowledge in: basic science, rheumatic disorders, approach to patient, investigation and disease management. This network of information can be used by all readers in a browsing or goal-directed learning mode. The materials are suitable for medical students and practising doctors. However, domain competence is more than domain knowledge, it also comprises a whole range of skills and for medicine, skills based competence is vital. These skills are:

- *physical*, e.g. equipment and procedures and
- *cognitive*, e.g. analysis, interpretation and decision-making.

It is the cognitive skills that demand a more sophisticated learning process if they are to develop (Kinshuk, 2001) and with adaptive hypermedia mechanisms we are able to match the content level with the student's knowledge through interactive case studies. A case study approach was adopted for JointZone as it can holistically represent the complexity of clinical reasoning. Students are then able to develop their procedural knowledge, which elaborates how doctors approach problems, interpret clinical information and make decisions (DesCoteaux & Harasym, 1998).

The case study in JointZone are graded into three levels: advanced, intermediate and beginner to develop students' clinical reasoning skills within the domain of Rheumatology. The *advanced* case studies were developed to simulate a clinical decision making process by an expert practitioner.

Our analysis reflects the cognitive skills identified by (Nkanginieme, 1997), namely that physicians:

- obtain and recognise symptoms in the patient,
- identify the appropriate system involved,
- speculate on the pathological processes,
- differentiate pathological processes,
- identify the possible causes of the pathology,
- Evaluate all pieces of information and make a clinical diagnosis.

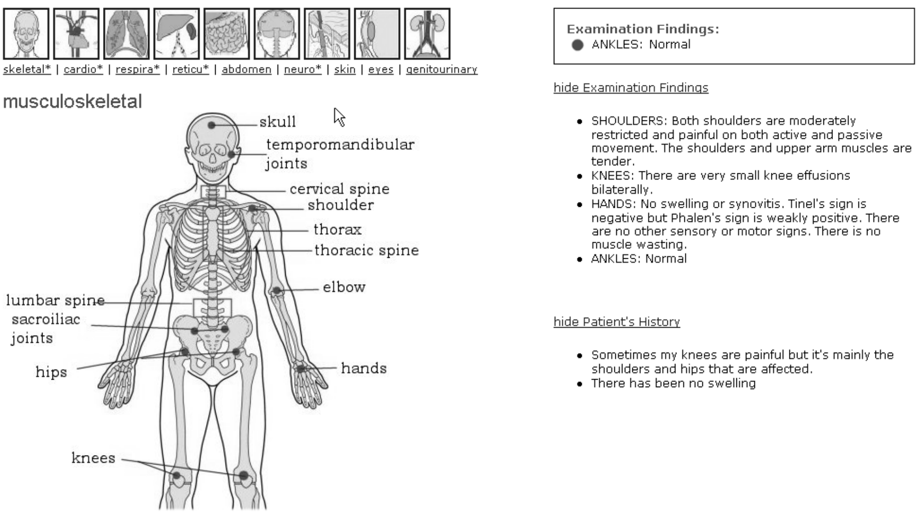

Figure 1: The examination stage of a case study for advanced level

On working through these stages, students are confronted with a wide range of options which give students information pertinent to the case under study. The clinical observations they gather by working through the case study are recorded so the students have ongoing access to this information as they proceed (figure 1). At the diagnostic stage, students are given a complete record of the examination findings and investigations. If the system feels insufficient or inappropriate information has been gathered this will be pointed out to the student. From a list of possible diagnoses, students select their preference based on their interpretation of clinical information gathered. They then obtain subsequent feedback on why that diagnosis is likely to be appropriate. This feedback, and the feedback across all case studies, makes explicit the expert's heuristics of clinical reasoning. This is an attempt to share that clinical expertise through a modelling or a cognitive apprenticeship (Collins, Brown, & Newman, 1998) approach, and it displays the underlying principles and rationale of the clinical diagnosis and completes the process. This enables students to develop their cognitive skills, providing a framework for problem solving and hence clinical reasoning. Case studies for beginners and in-

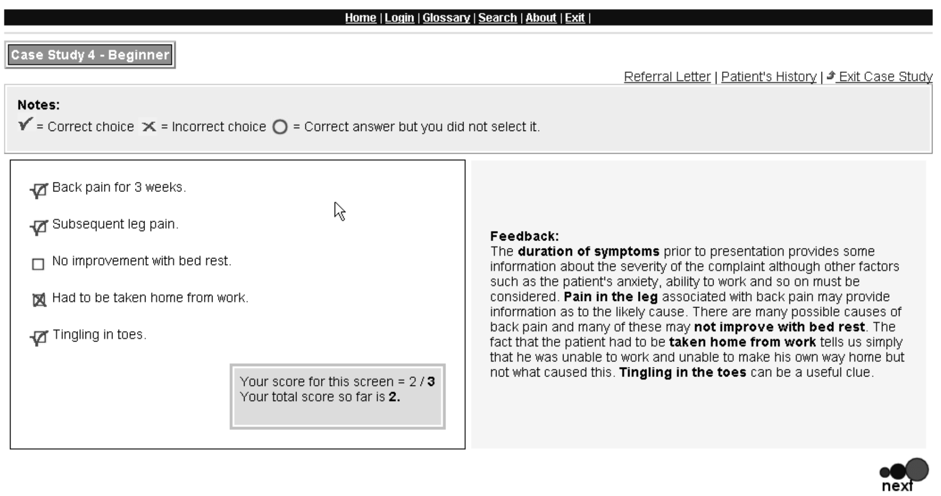

Figure 2: The feedback mechanism in beginner level of case study serves to model expert thinking in problem solving

termediate level students are much narrower in scope and the choices are more restricted. However, as with the advanced case studies, the feedback mechanism serves to model expert thinking in how to solve these problems (figure 2). The goal for *beginner* case studies is to recognise the value of particular pieces of clinical information, identify the pertinent investigations and make a diagnosis. For the *intermediate* students, the goal is to reassess a case in which the original diagnosis is under question. For both these levels, students are given scores on their performance in reaching a diagnosis. Through these case studies, students learn to recognise and interpret the value and relevance of disparate pieces of information that are integral to any clinical scenario. In this way students construct their own understanding of the relationship between the pieces of information rather than having it imposed upon them. This encourages a cognitive flexibility (Spiro, Feltovich, Jacobson, & Coulson, 1992) that is vital for the complex world of the physician.

2 Implementation

JointZone is implemented in a web-based environment, built within the Java Server Pages(JSP)[1] platform. The content of the website is structured using Extensible Markup Language(XML)[2] and displayed through the Extensible Style-sheet Language(XSL)[3]. The user database for adaptation and linkbases for link storage are implemented using a commercial database which offers fast retrieval and an efficient search mechanism.

JSP offers the facility to generate dynamic web content, a fundamental element in building adaptive hypermedia application. JSP allows Java programming code to be embedded into the web pages to describe how content is to be generated. JSP also provides the mechanism of *session tracking* to identify individual users across multiple HTTP requests. Session tracking is conveniently incorporated in the session management component of JSP, a layer transparent to the web developer. The session management maintains a state or identity of a user across multiple HTTP requests through the idea that all of the user's request for web pages during a given period of time are included in a *session*. This process establishes a unique session for each individual user during a client-server interaction, fulfilling the inadequacy of HTTP stateless protocol. On the client side, a session starts with the loading of a web browser and normally ends with the closing of the browser, unless intentionally disconnected by an external program.

2.1 The System Architecture

The system architecture as shown in figure 3 presents a view of *separation of concern* in the implementation of Jointzone. Content provided by the domain expert is structured using XML which is rendered by the web browser based on the XSL stylesheet . This process allows presentation code to be separated from information and encourages information re-use. For example, the same XSL style sheet can be used to display more than one XML page. Likewise, only one copy of XML content is maintained even if it is presented in different ways in the browser. The structural representation of content using XML will also facilitate content customization for the purpose of adaptation(see section 3.1).

The JSP acts as a shell to include the XML content, XSL presentation codes and Java codes before the web pages are presented to the users. The JSP pages

[1] JSP is a Java-based technology developed by the Sun Microsystems to simplify the process of developing dynamic web sites.

[2] XML, the universal format for structured documents and data on the Web, developed by World Wide Web Consortium (W3C). More on http://www.w3.org/XML/

[3] XSL, a language for expressing stylesheets. More on http://www.w3.org/Style/XSL/

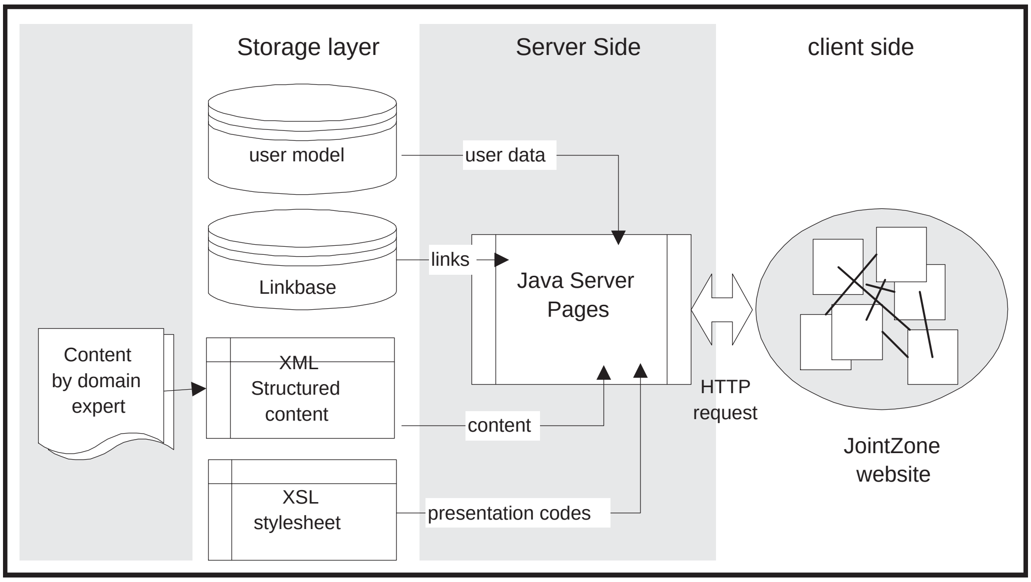

Figure 3: The system architecture of JointZone

also connect with the database through the JDBC driver[4] to pull in links or user information during an adaptive decision making.

2.2 The Adaptation Model

Figure 4 shows the adaptation model of JointZone. It comprises three essential building blocks for an adaptive system: a domain model, a user model and an adaptive layer. The *domain model* is derived from an offline process that draws its input from XML formatted content. The domain modeling is described in detail in section 3.1. The *user model* establishes dual-communication channels by providing user data to the adaptive layer during an adaptive decision making. In turn, all changes of a user's status during a session of user-application interaction are fedback to update the user model. The user modeling of this application is described in section 4. The *adaptive layer* is responsible for mapping user data to the domain model in the process of adapting information to users. It is achieved by JSP programming that involves if/else Java statements to make decisions on what links or content to be shown to the users. An example of this is shown below, the Java statements (<% ... %>) are embeded among the html codes in the Java server pages to provide different links and content based on user's status:

```
<% //decide on query to use based on usergroup %>
<% if (usergroup.equals("advanced")){ %>
```

[4]an API that provides cross-DBMS connectivity to a wide range of SQL databases

```
        <% query = queryExpert; %>
        <A href="?usergroup=beginner">Beginner</A>
        <A href="?usergroup=intermediate">Intermediate</A>
        <SPAN class = "big">Advanced</SPAN>
<% }else if(usergroup.equals("beginner")){ %>
        <% query = queryNovice; %>
        <SPAN class = "big">Beginner</SPAN>
        <A href="?usergroup=intermediate">Intermediate</A>
        <A href="?usergroup=advanced">Advanced</A>
<%}else if(usergroup.equals("intermediate")){ %>
        <% query = queryInt; %>
        <A href="?usergroup=beginner">Beginner</A>
        <SPAN class = "big">Intermediate</SPAN>
        <A href="?usergroup=advanced">Advanced</A>
<% } %>
```

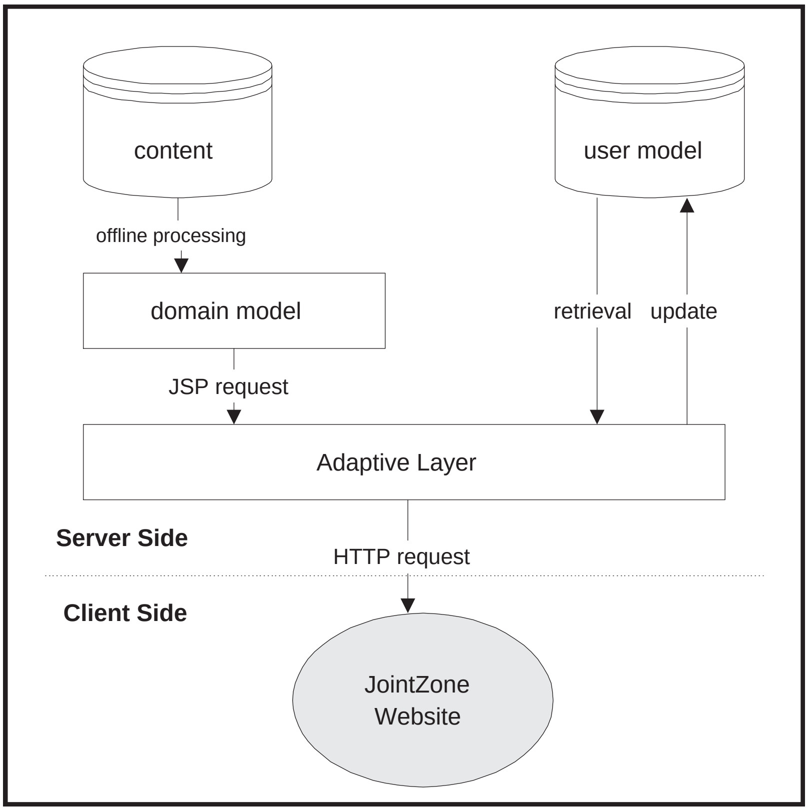

Figure 4: The adaptation model of JointZone

2.3 The Authoring of JointZone

Web technology offers JointZone a built-in browsing and navigation functionality. This basic navigation mechanism is built upon simple linking strategies where a simple link is generated from a single source to a single destination. In authoring JointZone, we have expanded on this basic notion of linking using linkbases. This idea is adopted from Microcosm (Fountain, Hall, Heath, & Davis, 1990) where links are maintained in databases(linkbases) separated from the documents to which these links apply. JointZone employs the use of different types of links in order to better manage the linking of information piece. These links are classified into structural links, referential links and associative links, adopted from the link taxonomy by (Lowe & Hall, 1996).

Structural links provide an illustration of the structural layout of the web site. They are built upon the simple one to one linking mechanism. These links include 'jump' links within a page to provide direct access to page fragments,'next' and 'previous' page links that ensure coherency between the documents, and the provision of the 'table of contents' to give a meaningful overview of the information site.

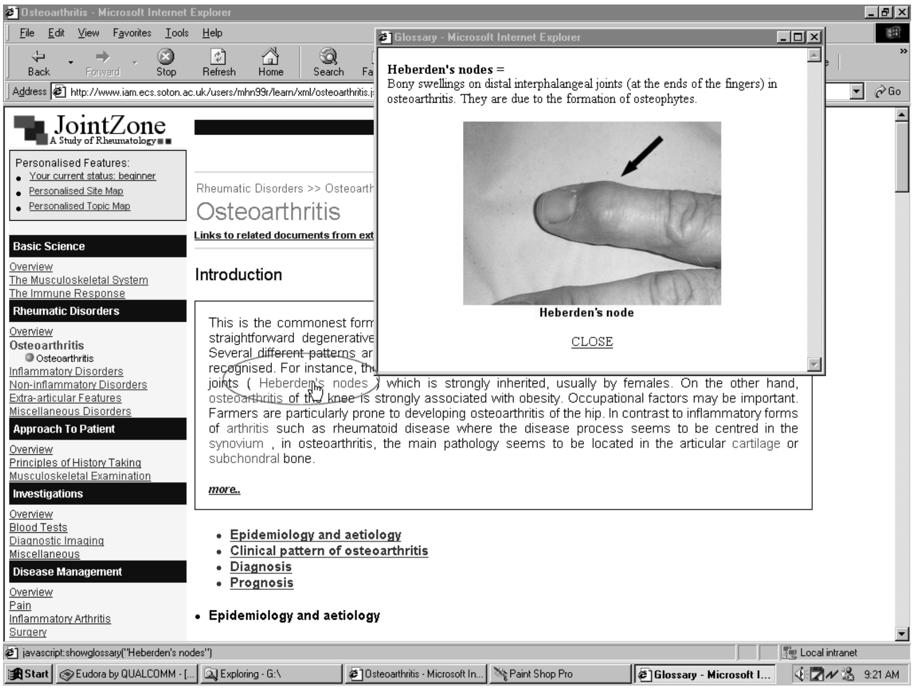

Figure 5: An example of a referential link

Referential links in JointZone are essentially glossary links that relate a word in context to its definition. They are generated using the idea of generic linking

(Hall, Davis, & Hutchings, 1996) where links are stored in linkbases and they can appear anywhere in any document where a glossary word exists. These links are not handcrafted but automatically generated using an offline processing mechanism. A document is first parsed to insert a glossary tag for any glossary word that appears on this document. The glossary links will appear highlighted in green on the web browser to differentiate them from normal links. They are resolved in real time from the linkbases once the links are followed by the users. The effect will be a pop up window explaining the word that a user has selected (see figure 5).

Associative links are stored in a linkbase where each of them is indexed or retrieved based on a keyword or concept. They are so called because they associate two or more documents based on a common concept. They are generated based on the usual phenomenon of users browsing in the current context wanting to find more information about a particular concept. For example, when users are reading a document or solving a problem that concerns concept 'X', links to other documents are suggested, which they could find out more about 'X'. In JointZone, associative links are used to provide context-sensitive links to documents from external sites. This mechanism expands a user's reading scope beyond JointZone so that they can gain more understanding on the current concept (see figure 6).

Associative linking also supports the adaptive mechanism of aggregating links based on the users' current knowledge. It is used to generate links based on concepts derived by the domain expert in accordance with the user's knowledge level. This is described further as part of the adaptive features in section 5.2 and 5.3.

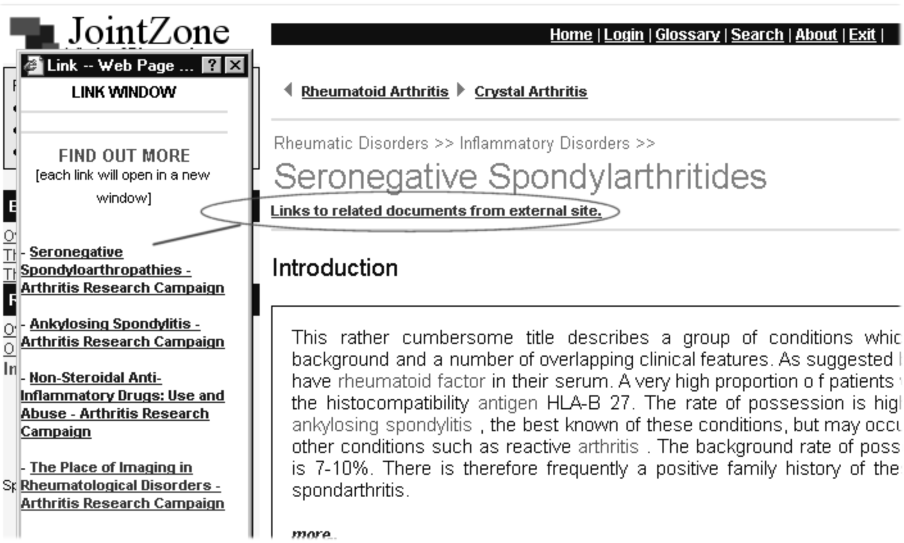

Figure 6: The associative link resolves to a list of suggested documents from an external site. These external documents share the same context with the internal document.

3 Domain Modeling

3.1 Structural and Keyword Representation of Content

This chapter explains how content in JointZone are organised to make adaptation feasible. In the domain modeling of JointZone, the idea of text processing is adopted from the area of information retrieval to identify the semantics of the contents. Text processing is applied to the domain in order to extract keywords that can be used as indexing terms for the documents. Indexing terms are (document) words whose semantics reflect the document's main themes, they are used to index or summarize a document (Baeze & Ribeiro, 1999). The keywords extraction methods used produce both single word and dependent bigrams[5] indexing.

In the case of single word indexing, a medical dictionary (Dark, 2000) is used to identify significant words in the documents. In order to reduce the set of representative keywords, the documents are first parsed to eliminate stopwords (such as articles and connectives). Stemming is used to reduce distinct words to their grammatical roots. Each word in the remaining text is then checked with the dictionary, if it exists in the dictionary, it is chosen as a keyword. In terms of extracting bigrams, a statistical method called fisher exact test(Pedersen, 1996) is used to extract dependent bigrams. A *dependent* bigram is one where two words occur together not purely by chance, for example this 'fine wine' or 'rheumatoid arthritis'. The fisher exact test decides if a bigram (XY) is dependant based on the frequency of appearance of each word that constitutes the bigram (X, Y) in all documents in the domain. The test computes the frequency distribution of XY, $X\overline{Y}$, $\overline{X}Y$ and $\overline{XY}$. For example for the bigram 'joint importantly', the variable X is used to denote the presence(X) or absence($\overline{X}$) of 'joint' in the first position of each bigram. Likewise, the variable Y is used to denote the presence(Y) or absence($\overline{Y}$) of 'importantly' in the second position of each bigram.

Table 1: A contigency table denoting the frequency of possible combination of the variable values for bigram 'joint importantly' and 'crystal arthritis'.

joint importantly	joint ?	? importantly	??
71	1384	2296	84179
crystal arthritis	crystal ?	? arthritis	??
91	91	355	84179

As shown in the table 1 below, the relationship of the frequency distribution will decide the degree of dependance of the bigram. In this case, the first bigram 'joint importantly' will be rejected as a dependent bigram because both 'joint'

[5] bigram, two consecutive words that occur together in a text

and ’importantly’ appear to be a common variable to form other bigrams rather than ’joint importantly’. On the other hand, ’crystal arthritis’ is accepted as a dependent bigram because the combination of crystal and arthritis as a bigram is significant since almost all appearances of the word ’crystal’ are followed by the word ’arthritis’.

The keywords with their respective document URLs are then stored as links in a database known as a linkbase(Carr, De Roure, Hall, & Hill, 1995).

```
<link>
    <id>179</id>
    <concept>bone scan</concept>
    <context>Miscellaneous Disorders in Rheumatic Disorders</context>
    <title>Algodystrophy Reflex Sympathetic Dystrophy</title>
    <url>http://softly:8080/xml/algodystrophy.jsp#intro</url>
</link>
<link>
    <id>2156</id>
    <concept>Osteoporosis</concept>
    <context>Non-inflammatory Disorders in Rheumatic Disorders</context>
    <title>Bone Disease</title>
    <url>http://softly:8080/xml/bonedisease.jsp#Osteoporosis</url>
</link>
```

As shown from the two examples of link entries from the linkbase above, each keyword extracted from a document contributes to a link entry.

- the *id* is an auto-generated unique number for each entry.

- the *concept* is a keyword extracted from a section on the document through text processing as described above.

- the *context* is formulated from the structural layout of the document where the keyword exists. For example, the entry with the title ’Bone Disease’ appears under the subheading ’Non-inflammatory Disorders’ which in turns comes under the heading of ’Rheumatic Disorders’. Hence, ’Non-inflammatory Disorders in Rheumatic Disorders’ is the context for this entry.

- the *title* refers to the title of a section in the document where the keyword appears.

- the *url* consists of an address pointing to not just the document but specifically the section of the page where the keyword exists.

The use of these link entries constructed from the conceptual representation (keywords-concepts) and structural-based labeling (title, context, url) of the content have made the content organisation feasible. This in turn creates opportunity for adaptation since different link entries can be shown to different users. This is achieved by adaptive link hiding or incremental linking to provide link-level adaptation. The provision of incremental linking puts restrictions on the number of links displayed based on the user's knowledge level. An empirical study by (Specht, 1988) has shown that there is an improvement in learning for students who use the incremental linking of hypertext. Advanced users are given a wider range of documents than those with less knowledge. For example for the learning topic of 'psoriatic arthritis', advanced students are provided with access to documents about the concepts: 'psoriatic, psoriasis, spondylitis, pitting and onycholysis' from the linkbase while intermediate students are directed to less concepts namely 'psoriasis arthritis, pitting and onycholysis', and beginners are simply shown 'psoriasis arthritis'. The domain expert identifies these three sets of concepts for each case study, which is an easier task for them compared to delivering a prerequisite network of documents. However, in order to be consistent with the philosophy of an unobtrusive learning environment, there is no restriction for any user group should they wish to investigate further, i.e. the sub-concept titles for advanced level are still accessible to a beginner.

3.2 Discussion

3.2.1 Previous Work on Domain Modeling

Keyword indexing provides an alternative to the involvement of a domain expert in organizing and labeling the content. The literature shows that in most educational AH systems the idea of curriculum sequencing is used to model the domain. Materials are grouped manually by the domain expert into prerequisites, remedial and outcome topics (Brusilovsky, Schwarz, & Weber, 1996) (Stern & Woolf, 2000) (Brusilovsky, Eklund, & Schwarz, 1998) (De Bra & Calvi, 1998) ; these aggregated learning materials are then presented adaptively to the students based on their current knowledge level. From the perspective of authoring, this method increases the burden on the domain expert to manually design a prerequisite network of the content. There is a solution in some cases where the application designers resolve to be the experts in the information resources themselves, a reason why many educational AH applications are built in the domain of computer science (Brusilovsky et al., 1996) (Stern & Woolf, 2000) (De Bra & Calvi, 1998). In JointZone, the content is in the medical domain and there is a lack of commitment or time from the domain experts in manually sequencing the content, hence the need for automatic text processing to support content organisation.

3.2.2 Scalability

The use of text processsing has long been establised in information retrieval research to provide indexing terms for online text. However this work reflects a pioneering use of the method in educational adaptive hypermedia for organising and labelling content. This method promised a degree of scalability because it can be used on any domain. In short, the idea is transferable to any adaptive hypermedia system regardless of the type of content. The type of dictionary used is however dependent on the type of domain. In our case, a medical dictionary is used because it contains terms related to acronyms, jargon, theory, conventions, standards, institutions, projects, eponyms, history, in fact anything to do with medicine or science. Hence, for a different domain such as computer science or history, a domain-specific or a general dictionary is perhaps more appropriate so that there is more mapping of words between the documents and the dictionary.

4 User modeling

A user model is a database storing information about users, accessed by the functions of initialization, update and retrieval. User modeling is a major component of any adaptive hypermedia system since data about users needs to be learned before any information can be adapted to the users. In JointZone, the user model captures two aspects of the students' differences: their individual browsing history and their knowledge level in the Rheumatology domain. Our model presents the novel idea of using individual effective reading speed to identify pages which the users have or have not read. The user's knowledge level is initialized based on his/her first entry registration details and this knowledge value can evolve through the user engagement with the application.

4.1 Modeling Knowledge level

Users' knowledge levels in the domain are captured in order to adapt them to the various difficulty levels of learning materials. The users are stereotyped into one of the following categories: expert, intermediate or beginner, upon their registration into the application. This categorization is based on two questions asked in the online registration:

- Are you a medical professional, medical student or non-medical professional?
- How much do you know about Rheumatology? (none/ a little/ average/ a lot)

A heuristic categorization is given based on the answers. Please see table 2:

Table 2: Heuristic Categorization of user group based on registration

Prior Knowledge / Status	Medical Professional	Medical Student	Non-medical Professional
none	intermediate	beginner	beginner
a little	intermediate	intermediate	beginner
average	expert	intermediate	beginner
a lot	expert	expert	beginner

However, this self-rating is not very accurate since users might not have a full understanding of their knowledge in Rheumatology. For this reason, they are prompted to an optional prior knowledge test at one stage in the application in order to make more precise assumption of the users' knowledge level. In this test, users are required to answer ten questions concerning the general topic of

Table 3: User categorization based on the prior knowledge test

Score	User Category
$<$30%	beginner
30%-70%	intermediate
$>$70%	expert

Rheumatology. The categorization is based on their scores (see table 3) and will overwrite the initial categorization in the registration.

The system assumption of users' knowledge can evolve as the users engage in the interaction with the application. This is based on their performance in the case study. Thresholds are set to migrate users from one group to another. A user's current knowledge is measured based on the scores of the attempted case studies in his/her user group. If more than a 50% score is achieved for every case study (10 cases in the current level), he/she will be promoted to the next level of user group in the hierarchy. In this case, the user will be informed of any change of status should any user group migration take place.

4.2 Modeling Browsing History

The modeling of browsing history in JointZone is achieved through an effort index (0-100%) estimated for each page in the domain. This index is the system's estimation of the 'degree of effort spent' by a user on each page. This is actually calculated by comparing the display time of a web page with an individual's optimal reading time. The optimal reading time for a page is identified for each user depending on the individual's reading rate, comprehension rate and prior knowledge of the domain. The user's characteristics are obtained when they use the system for the first time. Before their first entry to the application, a user will be asked to complete a knowledge test (10 multiple choice questions based on the general concept of the domain) to give a measurement of their prior knowledge (0-100%). Then each user has to take a 'reading speed test' where their reading rate is captured in the number of words read per minute. The reading time is recorded through a start and stop button on the web page. This is followed by a comprehension test (six true or false questions on general ideas in the text) to give an indicative score of their comprehension rate (0-100%). The combined measure of reading rate and comprehension gives the index of effective reading speed (Jackson & McClelland, 1979). This index is commonly used in commercial products for speed-reading courses (TurboRead.com, 2002) (ReadingSoft.com, 2000). We have however, modified the equation by adding the factor of prior knowledge (see Equation 1).

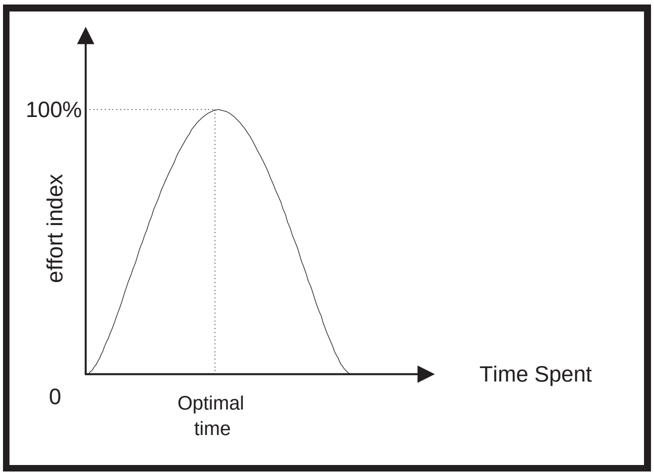

Figure 7: The Gaussian graph for the effort index estimation

$$\text{Effective reading speed} = \text{reading rate} * \frac{1}{2}(\text{comprehension rate} + \text{prior knowledge}) \quad (1)$$

$$\text{Optimal reading time for a page} = \frac{\text{total number of words in the page}}{\text{Effective reading speed}} \quad (2)$$

$$\text{Effort index estimation}, \mathbf{G}(x) = \exp\left(\frac{-(x - \text{optimalTime})^2}{2\sigma^2}\right) \quad (3)$$

During a browsing session the optimal reading time for any page in the domain is calculated in real time using Equation 2. The effort index on each page is then estimated using a Gaussian function (see Equation 3) by comparing the actual time spent on a page (x) with the optimal reading time. If 'x' approaches the optimal reading time, a high effort index is assumed (see figure 7). However, one problem still exists, as it is difficult to determine if a user has indeed read a page when it is displayed on the screen. To tackle this, a heuristic cutoff point is used to give a zero effort index to cases where display time falls below eight seconds or three times the optimal reading time. On the other hand, the use of individual optimal reading times enables us to skew the effort index based on a user's prior knowledge since a good student who skims a page will gain a higher effort index (with relatively lower optimal reading time) as compared to a novice student who spent the same amount of time on the page.

4.2.1 Motivation for Effort Index Generation

Literature in adaptive hypermedia has shown some unreliable solutions in identifying whether a page is read by the users. A general method is to use the viewing of a page as an indicator that the page has been read and understood (e.g. InterBook (Brusilovsky et al., 1998)). This method is obviously inefficient as the student might only be surfing through the pages without paying much attention to the content. In MANIC (Stern & Woolf, 2000), the system assumes a high studied rating for a page if a user spends an optimal amount of time on it. The optimal time that is generated, based on the length of the content, is static for every user irrespective of their individual reading speed. However, psychological research has shown that there are ”astonishing differences in the rate of reading speed” among individuals (Romanes, 1885). It has also been found that two individuals reading at the same speed can have different rates of comprehension (Huey, 1908). Our motivation for the index effort generation comes from users' unpredictable browsing behavior on the web, particularly the speed of reading web pages. Research has shown that users frequently skim read web pages instead of reading them in detail (Horton, Taylor, Ignacio, & Hoft, 1996). Studies have also shown that there is a general decline in the level of comprehension (especially in the recall of specific details) with faster reading as compared to reading at a normal speed (Dyson & Haselgrove, 2000) (McConkie, Rayner, & Wilson, 1973) (Poulton, 1958). Hence the reading speed, the time spent and the level of understanding on each page varies among individuals. The effort index captures this aspect of users' differences in reading speed and uses the differences to calculate how well they have read a web page based on the display time of the page. A validation test(Ng, Hall, Maier, & Armstrong, 2002) has been carried out for this work and the result shows a significant correlation between the effort index and the user's understanding test score of the page read.

5 The Adaptive Features

JointZone offers some adaptive features based on the effort index (see section 4.2) and knowledge level (see section 4.1) modeled from the users. Two adaptive techniques are used:

- *Adaptive link annotation*(Brusilovsky, 2001) is achieved by annotating extra information about the links. In JointZone, links are annotated with the users browsing history to form the history-based annotation

- *Adaptive link hiding* (Brusilovsky, 2001) is performed by hiding irrelevant links from users in an attempt to limit the users to only the set of links relevant to their interest or goal. In Jointzone, links are hidden or shown based on the users' knowledge level in the domain.

The effort index is used as a basis to form the history-based link annotation in JointZone. The users see this as a tri-value reading gauge , a visual representation of the amount of time they have spent reading individual pages (see figure 8). On the other hand, the user's knowledge level captured in the user modeling is used in link hiding. In general, users with different knowledge levels (beginner, intermediate or expert)see different sets of links to documents. This incremental style of linking mechanism attempts to show beginners less documents and knowledgeable users more documents.

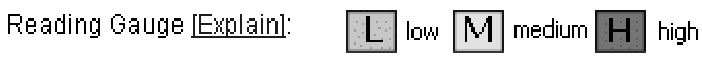

Figure 8: The tri-value reading gauge: indicated by low(effort index = 0-30 %), medium (effort index $\leq$ 30 %) and high (effort index $\geq$ 70 %)

5.1 Personalised Site Map

The site map in JointZone presents links to all documents that appear on the site. It is personalised by the set of links that are made adaptive to the individual's browsing history using history-based link annotation. As seen from figure 9, each document title is marked by a reading gauge icon to inform the user how much they have read the page before. Hence, this site map acts as a navigational support mechanism to give users a quick overview of the domain with relation to pages they have read or not read.

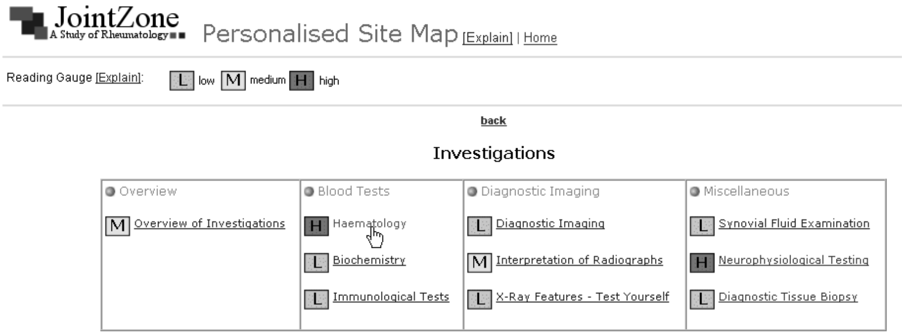

Figure 9: A screen shot of the personalised site map where history-based link annotation can be seen.

5.2 Personalised Topic Map

The adaptive mechanism of the personalised topic map is based on the user's browsing history , knowledge level and learning goal (a topic of learning, e.g. 'gout'). A combination of history-based link annotation and link hiding techniques is used in this case. As seen in figure 10, users need to manually select a

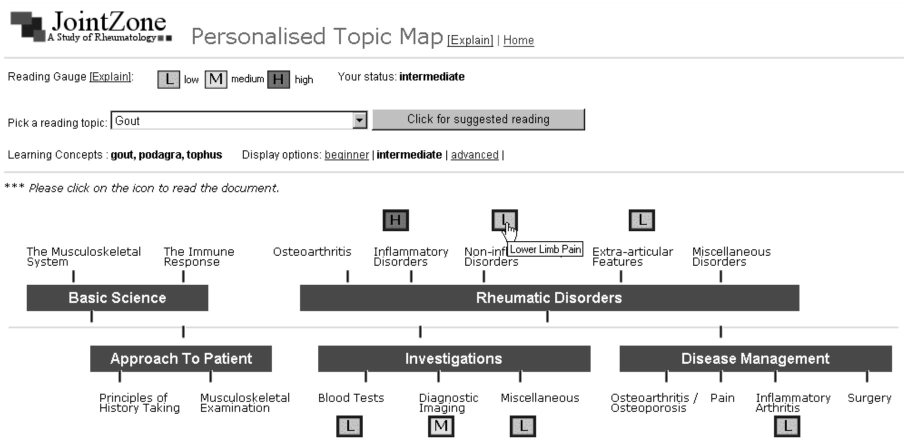

Figure 10: A screen shot of the personalised topic map with the combination of link hiding and history-based link annotation

learning topic (short term goal) from a pull-down menu, in this case, the topic selected is 'gout'. The topic map will then show links to documents about gout that are relevant to the user's status as an intermediate. At the same time, the history-based link annotation provides extra feedback to the user on which documents he or she has studied previously.

5.3 Adaptive Reading Room

The adaptive reading room appears at the end of every case study on the site (see figure 11). Its purpose is to provide an opportunity for further reading in order to find out more in depth of a topic or concept related to the case study.

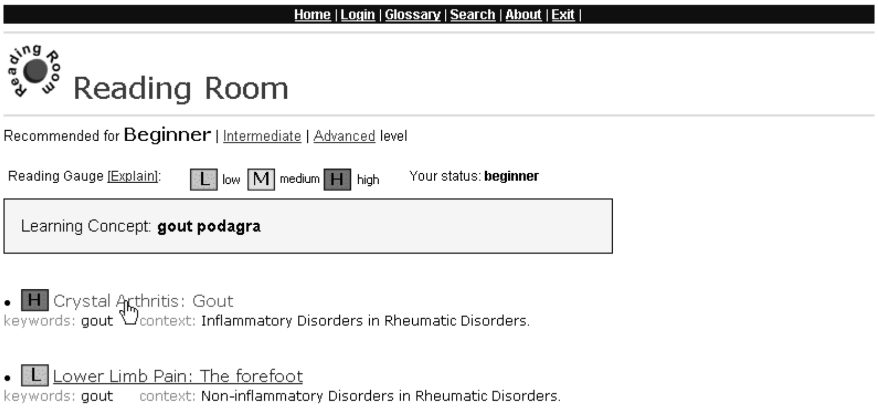

Figure 11: A screen shot of the adaptive reading room with the combination of link hiding and link annotation

Again, link hiding and history-based link annotation are used to suggest a personalised list of documents for further reading on the site. The number of documents shown is reflected by the users' status and each link is marked by the reading gauge icon.

5.4 Personalised Messages

Personalised messages appear mainly on the 'home' page (see figure 12) where the user first enter the site after the log in procedure. This includes a personal greeting to individual to provide an atmosphere of personalised learning environment. The messages inform users on their latest status (i.e a promotion to the next level in the user group hierarchy) and also feedback on their previous visit, which provides a pointer to the last page they have visited. The feedback also provides a detailed account of the case studies they have completed. From this, users are able to draw guidelines on whether they should retry the completed case study(if they have performed poorly), do further reading on a topic or proceed to other cases.

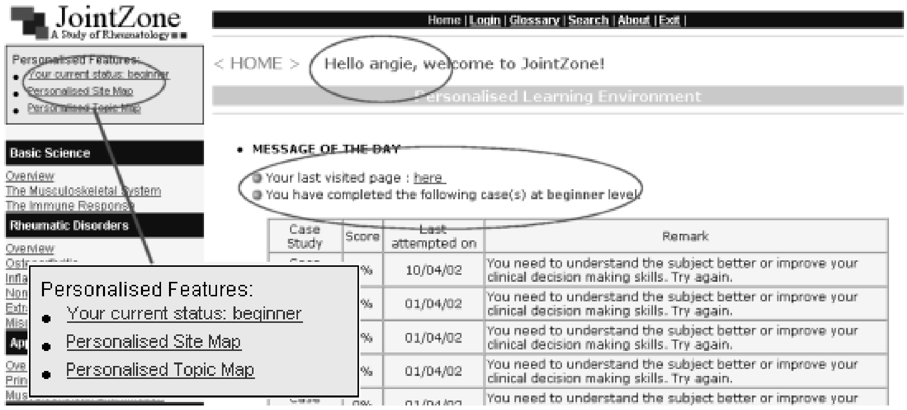

Figure 12: An example of some personalised messages on the 'home' page

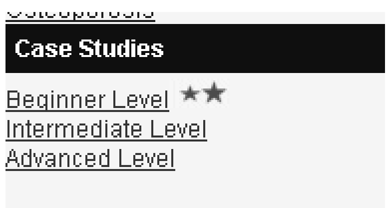

Figure 13: Adaptive assignment of case study: A screen shot of the table of content showing the annotated links for case study, in this case, the user is prompted to the beginner level

5.5 Adaptive Assignment of Case Studies

In JointZone, the system assigns case studies adaptively to the user's knowledge level. This is indicated by the adaptive link annotation on the table of contents (figure 13). Users are encouraged to try out the recommended level of case studies but nevertheless they are not prevented from trying non-recommended levels.

6 Discussion

The primary aim of making web-based learning adaptive is to reduce cognitive overload, prevent 'lost in hyperspace' and improve the process of learning on-line. The integration of adaptive mechanisms into JointZone is done in an unobtrusive manner. This can be seen from the options provided to the users: to register on the site or select free browsing, in which the web site can be used without its adaptive features. The controversial use of link hiding, as some users would want to see all links that other users see, is neutralised through the option offered to users to see links that are hidden from them.

On-going research in adaptive hypermedia has given contradictory results with regards to its benefits to users. Among the most widely used approach is to evaluating adaptive hypermedia systems is the 'with' and 'without' adaptivity evaluation design. This raises issues with regard to the functionality of the application without the adaptive features. Most adaptive systems are developed with the design of adaptivity in mind, hence if the adaptive features are removed, the systems basically loose their efficiency in their basic functionality. Hence, any evaluation approach using both adaptive and non-adaptive version of the system will not bring a fair comparison. To avoid this, the basic non-adaptive design of JointZone was optimised before any adaptive mechanism was integrated into the site. We ensure the basic functionality of the website is not overshadowed by the advantages given by the adaptive mechanisms. This can be seen from the evidence that options are given to users to use the non-adaptive version of the site.

A usability study of the website with adaptive features has been carried out with 29 students who were all taking a computer science module on human-computer interaction. The result of this study showed a positive degree of usability in JointZone and helped us to erradicate some minor usability problems. This ensures that the usability aspect of the website will not be an issue should there be any negative outcome from future evaluation of the adaptive mechanisms in JointZone.

An evaluation is currently being undertaken with the real users (medical students) to see the outcome of the adaptive mechanisms integrated into JointZone. The evaluation is adopted from the layered evaluation approach proposed by (Karagiannidis, Sampson, & Brusilovsky, 2001). This approach suggests two layers of assessment:

- Interaction Assessment Layer - Assess if the conclusion drawn by the system concerning characteristics of user-computer interaction is valid. e.g. the reading gauge validation test

- Adaptation Decision Layer - Assess if the adaptation decision is valid and meaningful, for selected system assumptions. e.g. evaluation of various adaptive features

The benefits of this layered approach can be seen when adaptation is unsuccessful since problem findings can be done in both layers. The layered approach gives two possibilities for adaptation failure. Firstly it can be the case that the adaptation decisions are reasonable, but they are based on incorrect system assumptions; or that the system assumptions are correct but the adaptation decision is not meaningful (Brusilovsky, Karagiannidis, & Sampson, 2001).

Our evaluation followed the guidelines above by concentrating on first finding out whether the system makes a correct assessment or assumption of the user's reading gauge, their current knowledge level and the system's matching of concepts to documents. The second phase of the evaluation will focus on the adaptation decision making, basically finding out whether the adaptive features are meaningful to users. The result of this evaluation will provide some answers to the following hypothesis:

- The adaptive features fascilitate the process of learning in Jointzone by reducing the time taken and avoid cognitive overload.
- The adaptive features improve learning in JointZone as compared to learning without the presence of any adaptive mechanism.

This evaluation will hopefully give us some insights into whether adaptive web-based learning really benefits the students or whether it is merely a technology-driven technique in a learning paradigm.

7 Conclusion

This report has described an adaptive web-based learning application known as JointZone for the study of Rheumatology, primarily for medical students. It has combined the existing adaptive hypermedia techniques of link hiding and link annotaion to make otherwise static web pages adaptive to users. The work includes a novel idea of using individual effective reading speed to model users' browsing history on the web. In domain modeling, the structural representation (based on document layout) and keyword representation (implemented by text processing technique) of the content are used to model the domain. This provides an alternative to the conventional domain modeling methods and successively produces a model that reduces the involvement of prohibitive effort from the domain expert.

8 Acknowledgements

This research was supported by Arthritis Research Campaign, UK, educational project grant A0549.

References

Baeze, Y. R., & Ribeiro, N. B. (1999). *Modern Information Retrieval*. Addison-Wesley.

Brusilovsky, P. (2001). Adaptive hypermedia. *User Modeling and User Adapted Interaction*, *11*, 87–101.

Brusilovsky, P., Eklund, J., & Schwarz, E. (1998). Web-based education for all: A tool for developing adaptive courseware. In *Proceedings of Seventh International World Wide Web Conference*, pp. 291–300.

Brusilovsky, P., Karagiannidis, C., & Sampson, D. (2001). The benefits of layered evaluation of adaptive applications and services. In *Workshop on Empirical Evaluations of Adaptive Systems, UM2001*.

Brusilovsky, P., Schwarz, E., & Weber, G. (1996). Elm-art: An intelligent tutoring system on world wide web. In Frasson, C., Gauthier, G., & Lesgold, A. (Eds.), *Intelligent Tutoring Systems, Lecture Notes in Computer Science*. Springer Verlag, Berlin.

Carr, L., De Roure, D., Hall, W., & Hill, G. (1995). The distributed link service: A tool for publishers, authors and readers. *World Wide Web Journal*, *1*(1), 647–656.

Collins, A., Brown, J. S., & Newman, S. (1998). Cognitive apprenticeship: Teaching the craft of reading, writing & mathematics.. In Resnick, L. B. (Ed.), *Knowing, Learning and Instruction: Essays in honor of Robert Glaser.* Hillsdale, New Jersey: Erlbaum.

Dark, G. (1997-2000). Online medical dictionary. http://cancerweb.ncl.ac.uk/omd/.

De Bra, P., & Calvi, L. (1998). Aha: a generic adaptive hypermedia system. In *Proceedings of the 2nd Workshop on Adaptive Hypertext and Hypermedia* Pittsburgh, USA.

DesCoteaux, J. G., & Harasym, P. (1998). Procedural knowledge! how important is it to converting novices to experts?. *The Bulletin*, *4*(6).

Duchastel, P. (1989). Formal and informal learning with hypermedia. In Jonassen, D. H., & Mandl, H. (Eds.), *Designing Hypermedia for Learning*. Springer-Verlag, Berlin, Heidelberg, New York.

Duchastel, P. (1992). Towards methodologies for building knowledge-based instructional systems.. *Instructional Science*, *20*(5-6), 349–358.

Dyson, M., & Haselgrove, M. (2000). The effects of reading speed and reading patterns on our understanding of text read from screen. *Journal of Research in Reading*, *23*, 210–223.

Fountain, M. A., Hall, W., Heath, I., & Davis, H. C. (1990). Micrososm: An open model for hypermedia with dynamic linking. In Rizk, A., & Andre, J. (Eds.), *Proceedings of Hypertext: Concept, Systems and Applications*.

Hall, W., Davis, H., & Hutchings, G. (1996). *Rethinking Hypermedia The Microcosm Approach*. Kluwer Academic Publishers, Dordrecht, Netherland.

Horton, W., Taylor, L., Ignacio, A., & Hoft, N. L. (1996). *The Web page Design Cookbook*. John Wiley, New York.

Huey, E. B. (1908). *The Psychology and Pedagogy of Reading*. The Macmillan Company.

Jackson, M. D., & McClelland, J. (1979). Processing determinants of reading speed. *Journal of Experimental Psychology: General*, *108*, 151–181.

Karagiannidis, C., Sampson, D., & Brusilovsky, P. (2001). Layered evaluation of adaptive and personalised educational applications and services. In Gilbert, J., Hubscher, R., & Puntambekar, S. (Eds.), *Workshop on Assesment Methods in Web-based Learning Environement and Adaptive Hypermedia at AI-ED 2001*.

Kinshuk (2001). Intelligent learning systems for cognitive skills development. In *Informing Science: Challenges to Informing Clients: A Transdiciplinary Approach* Krakow, Poland.

Lowe, D., & Hall, W. (1996). *Hypermedia and the Web - An engineering Approach*. John Wileys & Sons.

Mayes, T., Kibby, M., & Anderson, T. (1990). Learning about learning from hypertext. In Jonassen, D., & Mandl, H. (Eds.), *Designing Hypermedia for Learning*. Springer-Verlag, Berlin, Heidelberg, New York.

McConkie, G., Rayner, K., & Wilson, S. (1973). Experimental manipulation of reading strategies. *Journal of Educational Psychology*, *65*, 1–8.

Ng, M., Hall, W., Maier, P., & Armstrong, R. (2002). Using effective reading speed to integrate adaptivity into web-based learning. In *Proceedings of 2nd International Conference on Adaptive Hypermedia and Adaptive Web Based Systems* Malaga, Spain.

Nkanginieme, K. E. O. (1997). Clinical diagnosis as a dynamic cognitive process: Application of bloom's taxonomy for educational objectives in the cognitive domain.. *Med Educ Online*, *2*(1), 1–8.

Pedersen, T. (1996). Fishing for exactness. In *Proceedings of the South-Central SAS Users Group Conference* Austin, TX.

Poulton, E. C. (1958). Time for reading and memory. *The British Journal of Psychology*, *49*, 230–245.

ReadingSoft.com (2000). Speed reading test online. http://www.readingsoft.com.

Romanes, G. J. (1885). *Mental Evolution in Animals*. Appleton & Co, New York.

Specht, M. (1988). Empirical evaluation of adaptive annotation in hypermedia. In *Proceedings of EdMedia and Ed-Telecom*, pp. 1372–1378 Freiburg, Germany.

Spiro, R. J., Feltovich, P. J., Jacobson, M. J., & Coulson, R. L. (1992). Cognitive flexibility, constructivism and hypertext: Random access instruction for advanced knowledge acquisition in ill-structured domains.. In Duffy, T., & Jonassen, D. (Eds.), *Constructivism and the Technology of Instruction*. Hillsdale, NJ: Erlbaum.

Stern, M., & Woolf, B. P. (2000). Adaptive content in an online lecture system. In *Proceedings of the International Conference on Adaptive Hypermedia and Adaptive Web-Based Systems*, pp. 291–300 Trento, Italy.

Thuring, M., Hannemann, J., & Haake, J. M. (1995). Hypermedia and cognition: Designing for comprehension. *Communication of ACM*, *38*(8).

TurboRead.com (2002). Turboread speed reading. http://www.turboread.com.