The Impact of Empirical Studies on the Design of an Adaptive Hypertext Generation System

Kalina Bontcheva

University of Sheffield, Regent Court, 211 Portobello St., Sheffield S1 4DP, UK
kalina@dcs.shef.ac.uk

Abstract. This paper presents two empirical usability studies based on techniques from Human-Computer Interaction (HCI) and software engineering, which were used to elicitate requirements for the design of a hypertext generation system. Here we will discuss the findings of these studies, which were used to motivate the choice of adaptivity techniques. The results showed dependencies between different ways to adapt the explanation content and the document length and formatting. Therefore, the system's architecture had to be modified to cope with this requirement. In addition, the system had to be made adaptable, as well as adaptive, in order to satisfy the elicitated users' preferences.

1 Introduction

The aim of our research was to design and implement an adaptive hypertext generation system, HYLITE+, which generates factual explanations of domain terminology. The corpus analysis of online encyclopaedia and previous empirical studies (e.g., [13]) have shown the positive effect of additional information – e.g., definition of key vocabulary, less technical content, supply of background information and illustrations – on the subjects' reading comprehension and reading behaviour. On the other hand, hypertext usability studies [10] have shown that hypertext needs to be concise with formatting that facilitates skimming. Therefore, we performed empirical studies to test users' preferences and their perception of several adaptivity techniques. The results were used to establish a set of requirements for HYLITE+ and influenced the choice of adaptivity techniques adopted in the implementation.

For instance, the experiment showed that users prefer different additional clarifying information depending on the chosen formatting and desired explanation length. Another, somewhat unexpected, result was the strong desire of users to control the personalisation techniques applied by the system. Consequently, HYLITE+ was designed to be **adaptable** as well as *adaptive*.

The main difference between our approach and other existing adaptive hypertext generation systems (e.g., [7, 8]) is the use of results from hypertext usability studies, user trials with similar software products, mockups and walkthroughs during system design. The use of these techniques, together with corpus analysis, which is traditionally used in the design of language generation systems [14], enabled the choice of adaptivity techniques, tailored to and by the user.

2 The Empirical Studies

Hypertext readability studies [10] have shown that people read 25% slower on the screen and dislike scrolling. Therefore, unlike printed material, people prefer hypertext with concise, objective content and scannable layout, i.e., the length and formatting of the hypertext are very important. For our system these requirements translate as:

- *brevity* – do not exceed one or, if a more detailed explanation is needed, two pages;
- *structuring* – use formatting that makes it easy to pick out the important information while skimming the text, e.g., bullet lists.

2.1 The First User Experiment

Research in usability engineering [9] has shown that empirical user tests on existing similar products are a productive way to elicit user requirements and facilitate system design. Therefore we performed a limited user trial with an existing electronic encyclopaedia: The Encyclopaedia Britannica CD-ROM [4]. The goal of the experiment was to gain insight into the ways users browse encyclopaedic hypertext, the types of information they prefer, and the best ways to present it.

8 subjects (4 male and 4 female) were asked to find and browse articles related to dispersion (physics sense) and computer memory[1]. The subjects were asked to think aloud and were also interviewed at the end of the session. Their path through hyperspace was logged using software that intercepts the Web browser, and the sessions were also recorded on audio tapes.

The subjects were not given a strict time limit because the idea was to let them decide when they had got enough information since encyclopaedia browsing often does not have a well-defined goal and different people might have different strategies depending on their personalities and interests (e.g., skimming versus in-depth reading of all relevant articles)[2].

The interviews and transcript analysis showed a set of problems which was consistent among the users:

1. The multimedia software did not always show visited links in a different colour, so sometimes users could not recognise easily whether they have already followed a link.
2. Links need to be informative in order to help users decide whether they want to follow them or not (bad examples are Figure 1 or Ref. 2). Britannica also has links to index pages (e.g. table of contents) which users preferred to be distinguishable from links pointing at text pages.

[1] In this research we followed the discount usability engineering practices [9] which have shown that a small number of expert users is sufficient for this task.

[2] Previous studies of hypertext usability have already established that most users fall into two broad categories – skimmers (79%) and word-for-word readers [10, p.104].

3. Users do not like following long sequences of links away from the page they are reading since they feel distracted from the main topic.
4. Most users first skim a page to assess whether it is relevant and only afterwards read in detail the parts they are interested in. Consequently, they prefer formatting which facilitates skimming, not the usual mostly textual pages. Most users also decide which links to follow only after skimming the entire article first.
5. Most users found the articles too detailed and expressed a preference for having unimportant information on separate pages connected with links.
6. Users found that links incorporated in the text are more helpful than those provided in the index, since the text makes the connection with the current topic more apparent.

In addition, users with background or interest in the subject area (i.e., more familiar with the terminology) found it much easier to navigate through the hyperspace and looked at less pages since they ignored links to already known terms and also judged better whether a link is likely to lead to relevant material. Unlike them, novice users had problem navigating because most links contained unfamiliar specialised terms. They also showed a preference towards examples and figures which help them understand dry, abstract domains (physics, computers).

2.2 The System Mockup Experiment

The user study, the analysis of encyclopaedic texts[3] and previous research in dynamic hypertext (e.g., [7, 8]) suggested various ways for adapting the generated explanations:

1. provide the user with *definitions of* important *unknown terms* in brackets (used in encyclopaedias to facilitate users' text comprehension; our study suggests it might also improve users' navigation);
2. provide the user with a *familiar superconcept* in brackets to clarify unknown terms (same as above);
3. *omit already known information*, e.g., omit mentioning computer parts when describing a computer if the user knows them already;
4. contextualise the explanation by *referring to previously seen material*. For example, use phrases like 'As already mentioned' at the beginning of an already visited page or an already seen fact.
5. use syntactic structures that refer explicitly to *previously seen material* when it is part of a sentence (e.g., 'Besides dispersion, other characteristics...');
6. when a user is returning to an already visited page, modify its content to take into account what was seen in the mean time;

[3] We analysed a corpus that included texts from Encyclopaedia Britannica Online (`www.eb.com`), Microsoft Encarta (encarta.msn.com), and Harcourt Academic Press Dictionary of Science and Technology (`http://www.harcourt.com/dictionary`).

7. include *links to other related material* or if there is space, include this material on the page (e.g., information about the subtypes of the explained concept).

Before implementing these features in HYLITE+, we decided to test the user perception of their usefulness, that of adaptivity in particular, because most of them were derived from research on text/dialogue generation and might not fit well with the user expectations about encyclopaedic hypertext. Previous work on dynamic hypertext (e.g. [7,8]) has explored some of these alternatives but user acceptability has not been measured[4].

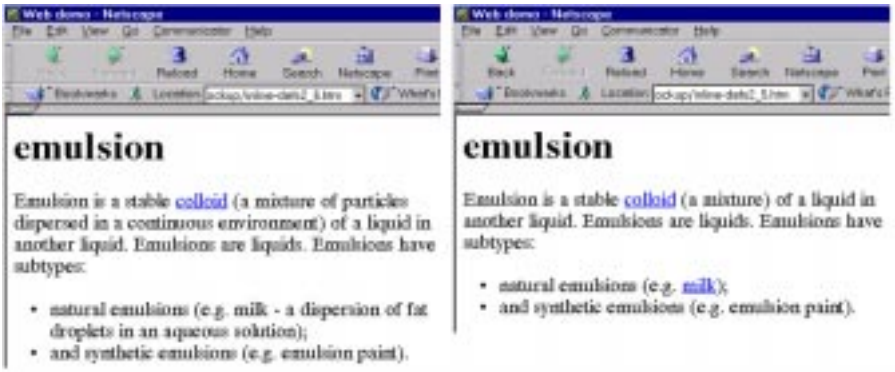

Fig. 1. The preferred version with the clarifying definition (left) and the one with a familiar supertype (right)

One fast, yet effective, way to test this is by using *predictive evaluation* techniques [12], which involve a small number of users that test a set of scenarios, realised as paper mockups and walkthroughs. We designed paper mockups for several user interaction scenarios which were used to test users' preference of different adaptivity techniques. Each scenario consisted of one or more set(s) of hypertext pages which presented approximately the same information in alternative ways (see Figure 1 for an example of two such alternatives (out of 6 in total for this scenario)). The scenarios which tested different ways to adapt the presentation depending on previously seen material consisted of a sequence of pages which were given to users one after another as they pretended to follow a given link. The scenarios were desinged so that they only focus on one ot two adaptivity techniques at a time.

[4] Where empirical user studies of dynamic hypertext system have been carried out at all, they were mainly focused on a particular user task or evaluating the learning effects of personalised hypertext (e.g., [6]). Some studies also measured the overall usability of the system but, to the best of our knowledge, no studies have attempted to elicit user preferences and acceptance of the different personalisation techniques used.

Subjects from the previous study were asked to rank the different page versions according to their own preferences and explain the reasons for their choice[5]. All subjects were experts in hypertext and Web browsers and have already interacted with a similar online system, so they could successfully simulate computer interaction using the paper mockups.

The mockup consisted of screen shots of the hypertext alternatives displayed in Netscape. The subjects were asked initially to customise the window and font size according to their preference and then the mockup material was produced to look exactly as it would on their screens. This was particularly important because, e.g., visually-impaired users use much larger fonts and their page ranking might have been affected if the experiment conditions did not match their everyday use.

The mockup experiment had to be performed on paper, because for most subjects it was not possible to show on the computer screen more than 2 windows in parallel. Most scenarios consisted of at least 4 alternatives, so we used a big table where the alternatives could be viewed simultaneously and compared. The order in which the alternatives were arranged was changed at random between the subjects.

The experiment differentiated two types of users with respect to content: – those who always preferred the most *concise texts* with links which they can explore further; and those who rated higher texts with *additional information* which might even lead them to material they did not initially intend to read. These preferences were consistent in all scenarios.

Due to space constraints, here we will only discuss the three scenarios which tested users' attitudes towards clarifying information (e.g., definitions of important unknown terms and familiar supertypes); a detailed discussion is available in [2].

One scenario tested the use of *clarifying information inside definitions*; the second tested its use in descriptions of *object parts*; the third one covered descriptions of *object subtypes*. All scenarios covered several alternatives:

1. provide *only a link* to the term (concise);
2. provide *a familiar supertype* in brackets and a link to the term;
3. provide *a definition* of the unknown term in brackets and a link to the term;
4. include a familiar supertype and no link to the term;
5. include the definition and no link to the term.

In all scenarios users always preferred to have the links included, because otherwise they would have to perform a search if they wanted to find further details about the term. Also, somewhat surprisingly, the experiments showed a connection between formatting and preferred alternatives. For example, definitions are acceptable when they are not too long (about 10 words), i.e., do not interfere with the flow of the main explanation. In parts and subtypes descriptions, definitions are preferred when list formatting is used because it makes it

[5] We chose to use the same subjects since they already had some experience with the electronic encyclopaedias and were familiar with the problems of using such systems.

easier to ignore them when skimming the page (see the generated example in Figure 2).

For the first scenario, half of the experts preferred the text with the definition in brackets (Figure 1, left), whereas the other half rated the one with the familiar supertype best (right). The difference comes from overall personal preference for concise versus more informative texts but there is also a connection with the user's familiarity with the words used in the definition.

For the second and third scenarios the most preferred version was the one that used lists to enumerate all the parts/subtypes and provided short definitions of them (see the generated example in Figure 2). For terms where the system had further information (e.g., properties), links were also provided. The preference for definitions is always the same, regardless of the user expertise in the domain, because the definitions can be easily ignored while skimming. In fact, one of the experts said he would rather have the short definitions there, rather than follow the link only to discover that the page contains just this information (i.e., is of no interest to him).

3 Summary

To summarise, the empirical studies revealed a lot of variation in the user preferences which motivated us to adopt a more individualistic approach, where users can customise the system adaptivity behaviour. The adaptivity techniques preferred by the majority of the mockup users are enabled by default in our implementation and the interface allows users to change them easily, including disabling all personalisation (see Table 3 for some examples).

Despite this variation, all users exhibited strong preference for well-formatted, concise explanations, where further detail and additional information can be obtained from links and the form interface.

The scenarios which tested different ways of providing extra material showed that some users always preferred the shortest text with links to further detail, while others always chose relatively concise, but more informative, explanations. Therefore the system interface was designed to allow easy selection of different levels of explanation detail with further finer-grained tuning available from the user preferences page.

A somewhat unexpected result was the strong desire of users to have control over the personalisation techniques applied by the system, i.e., customise the system behaviour with respect to both adaptivity and language use (e.g., use of phrases like As previously seen). Consequently, the system was designed and implemented with adaptivity control in mind. In order to make adaptivity control easily manageable, only adaptivity alternatives relevant to the current page are made available in the system top-level interface, with the full set of choices available from the preference page. For example, if the system has used definitions of unknown terms and links to related pages, only options related to these techniques (e.g., disable related links, switch to known superconcepts, disable all adaptivity) are displayed. The intuitive HTML forms interface with

Type of adaptivity	Default behaviour	User choices
Links to related pages	after the explanation grouped as More general, More specific, Similar	disable
Return to a visited page - using Back - using a link	 show same page modify page opening	 disable modification customise the page opening phrase
Already seen material	include with a cue phrase	disable
Clarify unknown terms - short, to-the-point text - more informative text	 known superconcept in brackets and short definitions of parts/subtypes short definitions	switch to links only switch to definitions switch to superconcepts
Related information	only as a link to a related page	include as section in current page

Table 1. Adaptivity techniques summarised

check boxes and radio buttons makes the whole adatability mechanism intuitive and fast to use.

4 The Implemented System in a Nutshell

Based on these results, we implemented an adaptive hypertext system which, similar to [8, 7, 1] , uses Natural Language Generation (NLG) techniques to create dynamically the hypertext nodes and links. HYLITE+ generates factual explanations of domain terminology which have been developed and evaluated in the domains of chemistry and computer hardware. The need for such explanations, for example in e-commerce, has been proven in practice by the increasing number of online computer shops that provide such information in reference guides and tutorials (see e.g. `www.action.co.uk`). Computer magazines like Ẃhat laptop' and Ẃhat PC' also have terminological glossaries, as part of their buyer's guides.

Following the distinctions made in [5], HYLITE can be classified as an *on-line information system* which provides *referential* information, without having educational goals as do, for example, intelligent tutoring environments. The information is requested by users with different knowledge and interests and typically each hypertext node corresponds to a domain concept.

The user interacts with the system in an ordinary Web browser (e.g., Netscape, Internet Explorer) by specifying a term she wants to look up. Further information about subtypes, parts, and other related concepts is obtained by following hypertext links or specifying another query.

Similar to all Web applications, HYLITE+ needed to *(i)* respond in *real-time*, i.e., avoid algorithms with associated high computational cost; and *(ii)* be *robust*,

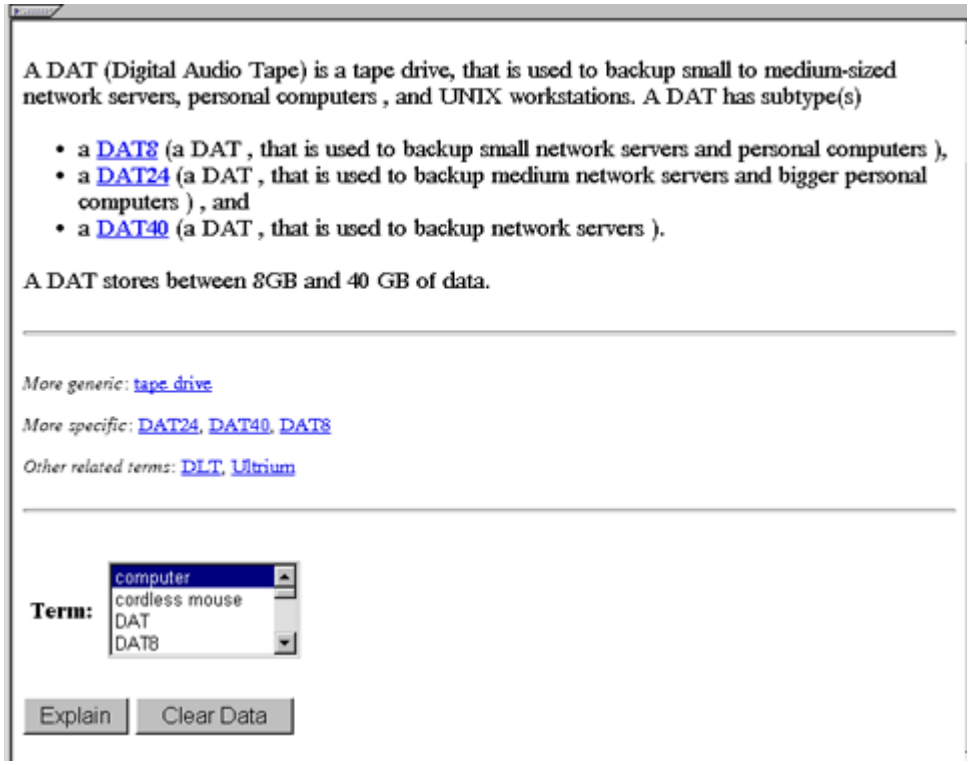

Fig. 2. An automatically generated text with added definitions and links to related material

i.e., always produce a response. Consequently the system uses some efficient and well-established applied NLG techniques such as text schemas and a phrasal lexicon (see [14, 8, 7]).

The adaptivity is implemented on the basis of a user and a discourse models. The user model is preserved between interactions and is updated dynamically, based on the user's interaction with the system.

A more detailed description of the system and the NLG and UM techniques is otside the scope of this paper, but see [2, 3].

5 Conclusion

The paper presented the *empirical studies* which were used to design the dynamic hypertext generation system HYLITE+. The results from a task-based empirical evaluation [2] of the implemented have shown that users prefer adapted texts to the neutral version where information about unknown terms is not automatically provided but has to be accessed by following a link. If relevant to their task, the additional information provided by the adaptive system is also used by the

users to minimise the number of visited irrelevant pages. In addition, the users' subjective opinion has shown that the adaptive system is easy to use and does not confuse users.

We are now about to perform the second stage of empirical evaluation, aimed at establishing the adequacy of the system's adaptability mechanism. The empirical studies of a system that adapts itself to the user and her task [11], have already shown that for interface adaptation systems, adaptivity and adaptability are not alternatives. Instead, better results are achieved when they are combined in a mixed-initiative system.

References

1. Liliana Ardissono and Anna Goy. Dynamic generation of adaptive web catalogs. In Peter Brusilovsky, Oliviero Stock, and Carlo Strapparava, editors, *Adaptive Hypermedia and Adaptive Web-Based Systems*, pages 5—16, Berlin Heidelberg, 2000. Springer Verlag. Lecture Notes in Computer Science 1892.
2. Kalina Bontcheva. *Generating Adaptive Hypertext Explanations with a Nested Agent Model.* PhD thesis, University of Sheffield, 2001. Forthcoming.
3. Kalina Bontcheva and Yorick Wilks. Generation of adaptive (hyper)text explanations with an agent model. In *Proceedings of the European Workshop on Natural Language Generation (EWNLG'99)*, Toulouse, France, May 1999.
4. Britannica Inc. *Encyclopaedia Britannica CD'99.* 1999. Multimedia Edition.
5. Peter Brusilovsky. Methods and techniques of adaptive hypermedia. *User Modelling and User Adapted Interaction*, 6(2-3), 1996. Special issue on adaptive hypertext and hypermedia.
6. Richard Cox, Mick O'Donnell, and Jon Oberlander. Dynamic versus static hypermedia in museum education: an evaluation of ILEX, the intelligent labelling explorer. In S.P. Lajoie and M. Vivet, editors, *Artificial Intelligence in Education: Open Learning Environment: New Computational Technologies to Support Learning, Exploration and Collaboration.* IOS Press, Amsterdam, 1999.
7. Alistair Knott, Chris Mellish, Jon Oberlander, and Mick O'Donnell. Sources of flexibility in dynamic hypertext generation. In *Proceedings of the 8th International Workshop on Natural Language Generation (INLG'96)*, 1996.
8. Maria Milosavljevic, Adrian Tulloch, and Robert Dale. Text Generation in a Dynamic Hypertext Environment. In *Proc. of 19th Australian Computer Science Conference*, Melbourne, 1996.
9. Jakob Nielsen. *Usability Engineering.* Morgan Kaufman, California, 1993.
10. Jakob Nielsen. *Designing Web Usability: The Practice of Simplicity.* New Riders Publishing, 2000.
11. Reinhard Opperman. Adaptively supported adaptability. *International Journal of Human-Computer Studies*, 40:455 – 472, 1994.
12. Jenny Preece, Yvonne Rogers, Helen Sharp, David Benyon, Simon Holland, and Tom Carey. *Human-Computer Interaction.* Addison-Wesley, Reading, Mass., 1994.
13. David Reinking and Robert Schreiner. The effects of computer-mediated text on measures of reading comprehension and reading behaviour. *Reading Research Quarterly*, Fall:536—551, 1985.
14. Ehud Reiter and Robert Dale. *Building Natural Language Generation Systems.* Cambridge University Press, Cambridge, U.K., 2000.