Active Alice: Using Real Paper to Interact with Electronic Text

Heather Brown[1], Robert Harding[2], Steven Lay[2], Peter Robinson[2], Dan Sheppard[2] and Richard Watts[2]

[1] University of Kent at Canterbury, England
[2] University of Cambridge, England

Abstract. Many documents exist in both paper and electronic forms. Paper has many well-known advantages, but electronic texts often contain useful information that is not easily accessible from printed paper versions. SGML texts, in particular, are rich sources of additional information. The Active Alice project shows how a reader can use a paper document to access information from its corresponding electronic version without having to manipulate the electronic version via a separate computer interface.
The project makes use of a DigitalDesk. This is an ordinary desk augmented with a video camera and computer-driven projector. The camera captures images of the pages on the desk and detects simple user actions such as pointing to specific words on a page. The images are used to associate the pages with their SGML counterparts. Information from the SGML versions can then be conveyed directly to the reader via information projected onto the page or onto other areas of the desk.
The project takes its name from the example text used—a version of *Alice's Adventures in Wonderland*.

'Curiouser and curiouser!' cried Alice.
Lewis Carroll

1 Introduction

The scope and scale of electronic documents has grown dramatically since the simple marked-up texts and word processor documents of the 1970s. At that time electronic texts were self-contained and were used almost exclusively for producing printed versions. They now come in a wide variety of forms, document content is often generated dynamically or retrieved on demand over a network, and many hypermedia and distributed documents are intended primarily for online viewing. Yet, in spite of a growing emphasis on online documents, the demand for paper remains and a great many electronic texts—particularly SGML [7] texts—are used to produce both paper and online versions. Sometimes these are published together, perhaps as a book with a supporting CD-ROM, but more often the two versions are regarded as separate items to be used in different ways for different purposes.

It is generally accepted that paper has many advantages in readability, portability, and convenience. Electronic documents, on the other hand, are acknowledged to provide significant enhancements in searching, cross-referencing, and the use of sound and moving images. Some also provide additional information for scholarly purposes or for specific applications. Much effort has been expended to improve the readability and convenience of electronic documents by making them behave more like their paper counterparts, but this work has nearly always assumed that the electronic document will be accessed purely through a conventional computer interface. An important motivation for the Active Alice project is to show that the current rigid separation of printed and electronic documents is unnecessary. In particular, it aims to demonstrate that a printed document can be used as a natural interface to electronic enhancements.

The key to this work is the *ubiquitous computing* concept [15] which aims to add unobtrusive computing power to everyday objects in a way that blurs the distinction between computers and other tools. The Active Alice project uses a DigitalDesk [16]: a computer-enhanced physical desk that can recognise real paper documents and use them as part of the overall system. When a paper document is placed on the DigitalDesk, the system attempts to read the visible text on it and to match this up with an electronic version held in its central document registry. If a match is found, the paper can then be used as a natural interface to the electronic version.

This paper explains how the Active Alice project uses the DigitalDesk to give direct access from paper documents to the information contained in the electronic texts of the British National Corpus (BNC) [2]. These are SGML texts, marked up in accordance with the Guidelines published by the Text Encoding Initiative (TEI) [10]. The text chosen to illustrate the possibilities is a simplified version of 'Alice's Adventures in Wonderland' by Lewis Carroll from the Oxford Bookworm series for young readers [6], but the principles are applicable to all BNC texts (amounting to approximately 90 million words) and, indeed, to a great many other documents which have SGML versions coded according to the TEI guidelines.

The paper is structured as follows. The next section introduces the DigitalDesk, explains how it works, and describes the main stages used in recognising pages and associating them with their electronic text. The following sections then outline relevant features of the Text Encoding Initiative and the British National Corpus, give details of the Alice example and its implementation, and discuss further further ways in which the information in TEI texts could be accessed via the DigitalDesk.

2 The DigitalDesk

In the 1970s, the *desktop metaphor* was developed at Xerox PARC. This made computers easier to use by presenting information on the screen in a way that looked and behaved rather like pieces of paper on a desk. Later developments attempted to make computers behave like intelligent diaries, intelligent whiteboards, and so on. Some recent research has taken the opposite view. Instead of simulating intelligent versions of other objects on conventional computers, work on *ubiquitous computing* and *computer-augmented environments* [12] has concentrated on adding computing facilities to the

other objects. The DigitalDesk [11] [13] is an attempt to produce an intelligent desk that can work with real paper, rather than simulating the paper on a computer screen.

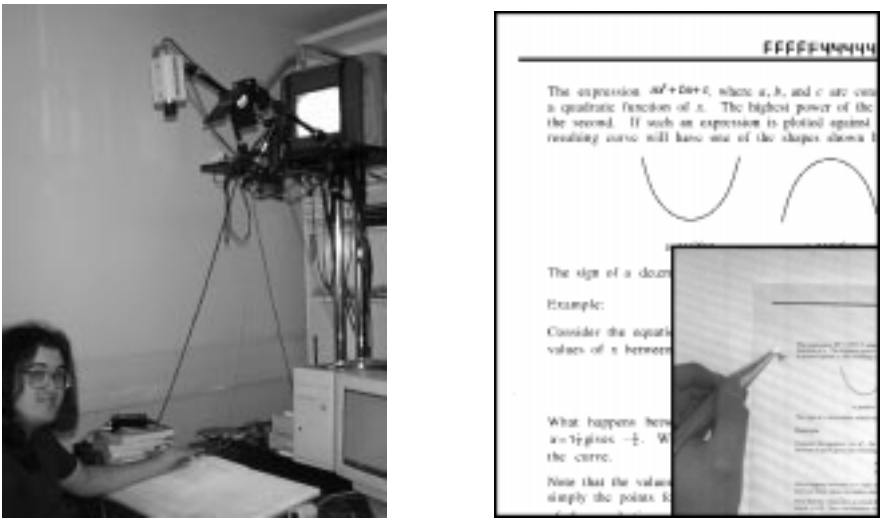

Fig. 1. Using the Digital Desk

The DigitalDesk, see Figure 1 (left), is a computer system built around an ordinary desk. It integrates the physical desktop into the system by means of a video camera and computer-driven projector. These are mounted above the desk, pointing down at the work surface. The video camera is used to recognise paper documents placed on the desk and to detect where the user is pointing. Figure 1 (right) shows part of an active mathematical textbook and (inset) the user pointing to it on the DigitalDesk with a LED-tipped pen. The projector enables electronic objects to be projected onto the work surface, either onto a paper document on the surface or onto other areas of the desk. Thus users can read and manipulate both the paper and electronic documents without having to divide their attention between the desk and a screen.

The current DigitalDesk prototype includes a single JVC TK-F7300 frame capture camera. It uses a prismatic arrangement to sub-sample a 756x576 image by a factor of 2, 3, or 6 in each direction. This gives it a maximum resolution of 4536x3456, or approximately 10 pixels/mm over an A3 working surface. Standard image processing methods, adapted to this low resolution, are used to recognise pages and lines of text within pages and to perform fairly crude OCR. A LED-tipped pen serves as a mouse. User actions are recognised by tracking a red LED on the end of the pen and noting pen clicks. The display capabilities are provided by a Proxima 1024x768 colour LCD panel attached to a high power overhead projector.

Previous projects using the DigitalDesk dealt with paper documents printed specifically for the application. These used unique identifiers on pages to aid in page recognition. Typically lines were printed at the top and bottom to aid registration of the page

image, and a unique identifier in an OCR font was placed at the top-right of the page. Figure 1 shows a page of this type from an active mathematical textbook designed to allow access to animations of mathematical functions. The page shown deals with general quadratic equations. If a student writes coefficients into a general quadratic printed on the page, the system captures the values and projects the resulting graph onto a space on the page.

For the Active Alice project, it was considered important to use real published documents. This meant developing a more general page recognition process that did not rely on convenient identifiers to match the page to a version in the central page registry. A real book does not lie completely flat on the desk, so special action is needed to recognise lines of text following unpredictable curves. It is also necessary to decipher enough of the line and word patterns to allow the page to be matched reliably to its electronic counterpart.

The recognition process works in two stages. In the first stage the overall page areas are detected by identifying vertical assemblies of lines from low-resolution (150dpi) images of the desktop. Higher-resolution images (300dpi) of the page areas are then used to determine the baseline and x-line of lines of text. Baselines are calculated by finding the bottoms of characters, smoothing, and then fitting a Bezier curve through a selection of the resulting points.

The lines in the image are matched to the electronic text by building a tree of page/line/word information from the image, finding the best match with similar trees built from the electronic text, and then creating appropriate links between the two trees. The matching process uses the pattern of ascenders, descenders, and word spaces in each line and allows a certain tolerance in the matching. Once a match for a page has been made in this way, the positions of the words and word spaces in the image can be established with reasonable accuracy.

Figure 2 shows the Alice book in use on the DigitalDesk during the development of the word recognition and matching system.

3 The Text Encoding Initiative

The Text Encoding Initiative (TEI) [1] [14] has published a comprehensive set of guidelines [10] for the preparation and interchange of electronic texts. Although these are aimed particularly at creating electronic versions of old texts in a form suitable for scholarly research, they have been used for a wide variety of old and new texts and for annotating speech. TEI projects cover archival and museum information, many types of literature, legal texts, and multilingual corpora.

The TEI is an application of SGML, so the Guidelines present an SGML Document Type Definition (DTD) together with a comprehensive set of recommendations for encoding texts using this DTD. Many different textual features may be encoded, from simple structural items such as sections and sentences, to complex linguistic and critical apparatus. Each TEI text begins with a header giving details of the original text and explaining which features have been encoded.

Because most texts use only a small subset of the available TEI features, the TEI DTD has been designed in a modular fashion that is often referred to as the Chicago

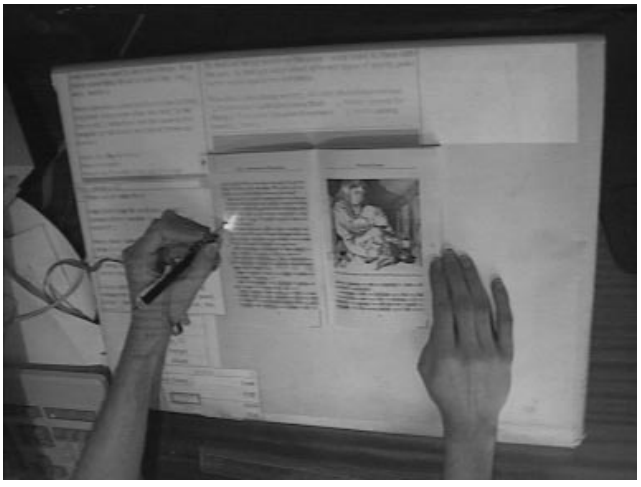

Fig. 2. The Alice text in use on the DigitalDesk

Pizza model [3]. Pizzas typically consist of a single base (thincrust or deep dish, say), plus some universal ingredients (tomato sauce and cheese) and any mixture of optional toppings. In accordance with this model, the TEI DTD requires users to pick one of a small number of *base tag sets* designed to cope with prose, verse, drama, speech, or dictionaries. The mandatory *core tag set* provides the universal ingredients: sections, paragraphs, highlighting, and simple versions of lists, references, names, links, editorial changes, and so on. These may be sufficient for many needs, but users may choose to add toppings in the form of any number of *additional tag sets*. These cover such things as advanced linking, analytical mechanisms, transcriptions of primary sources, critical apparatus, and detailed ways of identifying names and dates.

The TEI Guidelines amount to 1300 pages of text describing the use of more than 400 different SGML elements, so it is not feasible to provide a summary in a short paper. Instead, the following example attempts to give an idea of the richness of the facilities available by showing how names may be encoded.

The core tag set provides two elements for identifying names. The *referring string* element, `<rs>`, is used for a general-purpose name or name phrase and the *name* element, `<name>`, is used for a proper noun or noun phrase. Further information may be encoded using the attributes *type* (to indicate a person, place, country, etc.), *key* (to provide a unique key for this name), and *reg* to provide a 'regularised' form for the name. The following example illustrates a possible use of these elements.

```
<name type=person key=SMITH3 reg="Smith, Dr Ian">
Smith</name> lived in <name type=county>Somerset</name>.
. . . . . In 1890 the
<rs type=person key=SMITH3 reg="Smith, Dr Ian">Doctor</rs>
```

When information of this type is encoded, the electronic text is useful for many computer-based applications. As a simple example, all references to a particular person

or place could be identified by searching for all <name> and <rs> tags with a given *key* value.

An additional tag set for names and dates provides further elements and allows significantly more information about names to be encoded. The <persName> element, for example, may contain any number of the following subelements

<surname>	inherited family name
<forename>	given or baptismal name
<roleName>	official title or rank
<addName>	additional name, such as a nickname or alias
<nameLink>	connecting phrase within a name, such as 'van de'
<genName>	general information, such as 'Junior' or 'IV'

Each of these has several attributes that can be used to give further information. Thus the full coding of a name using the additional tag set might be

```
<persName key=JONES2>
<roleName type=military>General</roleName>
<forename>Peter</forename>
<forename full=init reg=Marcus>M</forename>
<addName type=epithet>Ironsides</addName>
<surname>Jones</surname>
</persName>
```

4 The British National Corpus

The British National Corpus [8] [4] [5] is a large collection of modern English text and speech marked-up using the TEI Guidelines. The Corpus contains over 3000 written texts (90 million words) and about 900 spoken texts (10 million words). It was created by a consortium of publishers, Universities, and the British Library. Some of the texts are freely available, others are only available for research purposes or in certain countries.

Every sentence, word and punctuation mark in every BNC text is tagged. In particular, each word is tagged with its part of speech. The part of speech encoding was performed using the CLAWS [9] system developed at Lancaster University. The additional tag set for simple analytic mechanisms provides six elements to represent the traditional linguistic categories of sentence, clause, phrase, word, morpheme, and character. The BNC uses three of these

<s> for sentences
<w> for words
<c> for punctuation characters

The part of speech for each word is given by the value of the *type* attribute of the <w> element, using a set of 57 *word class codes* based on the CLAWS annotation. For example

AJ0	adjective (e.g. good, green)
AJC	comparative adjective (e.g. better, smaller)
AV0	adverb (e.g. angrily, quickly)
DPS	possessive determiner (e.g. their, my)
NN0	common noun, neutral for number (e.g. aircraft)
NN1	singular common noun (e.g. pen, dream)
NP0	proper noun (e.g. Mexico, Linda)
PNP	personal pronoun (e.g. you, he)
VBD	past tense of verb to be (e.g. was, were)
VHG	-ing form of verb to have (e.g. having)
VVB	base form of lexical verbs (e.g. live, think)
VVD	past form of lexical verbs (e.g. lived, thought)

In addition to the 57 main class codes, 15 *portmanteau* class codes are used to indicate uncertainty between two different parts of speech. For example, AJ0-AV0 indicates that the word might be either an adjective or adverb. The following is a small sample of a BNC text

```
<w NP0>Alice <w VBD>was <w AV0>very <w AV0>nearly
<w AJ0>asleep <w AVQ-CJS>when<c PUN>, <w AV0>suddenly
<c PUN>, <w PNP>she <w VBD>was <w VVG>sitting <w PRP>on
<w AT0>the <w NN1>ground<c PUN>.
```

5 Active Alice

We chose to illustrate the possibilities of accessing electronic texts from paper documents by taking a child's text from the BNC and using it to provide a simple grammar lesson. This exploits the word class information in the BNC texts and has the double benefit of providing an application for a potentially large set of users and of avoiding one designed for computer professionals.

The following subsections outline briefly how the Active Alice system appears to a user, explain the small number of changes that we made to the BNC text, and discuss some practical difficulties encountered. The details apply specifically to the Alice text [6] but the principles apply to many other BNC texts.

5.1 The User Interface

The printed Alice text is a small book containing approximately 40 pages (see Figure 2). The BNC text covers the main running text but not the title pages, glossary, running headers, or the illustrations and their captions. When the open book is placed on the DigitalDesk, the running text areas are recognised and matched to the appropriate parts of the electronic text as outlined earlier. At this point, two small windows of information are projected onto the desk, one either side of the book. The area to the right of the book is used for information and menus designed to help the user find out about particular word types. The area to the left is used for a simple quiz on word types and to provide

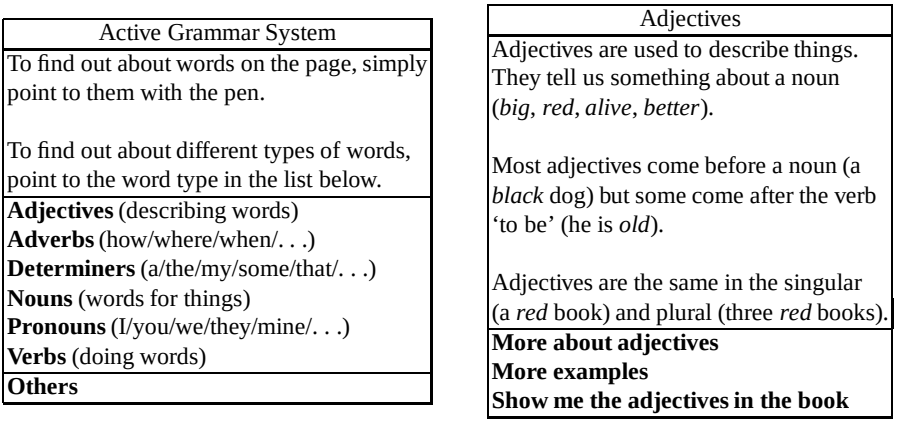

Fig. 3. Initial information (left) and information about adjectives

feedback when the user points to a word on the page. The quiz is designed to make the system fun to use and to encourage users to explore the text and try their skill.

Figure 3 shows the initial window that appears to the right of the book and the additional information that appears immediately beneath it if the user points to the *Adjectives* menu item. If the user points to **More about adjectives**, a third small window appears below these two giving information about ordinary, comparative, and superlative adjectives. If the user points to the **Show me . . .** item, immediate feedback is provided by highlighting each of the adjectives on the open pages of the book (by projecting coloured rectangles onto the physical book).

Altogether, about 30 different windows of information may be accessed to appear on the right of the book. These cover prepositions, conjunctions, numbers, interjections, negatives, infinitives, and possessive forms as well as the common word types shown in the menu in Figure 3. In all cases the user can choose a *Show me . . .* item to see all the words of the appropriate type on the open pages.

At any time the user may also point to a word on the open pages. Feedback is given by highlighting the word on the page (with a coloured rectangle as before) and providing information about the word in an area to the left of the book. Figure 4 shows the result of pointing to the word 'smaller'.

The quiz is simply a variation on the word information theme that invites the user to find certain types of words or to answer simple questions. Quiz problems always involve one or more words on the open pages. They are chosen at random from an internal list, with safeguards to ensure that a quiz contains 6 varied problems and that every problem can be solved using the currently open pages. Some typical problems are

- Find 2 adjectives
- Find a comparative adjective
- Find 3 different pronouns
- Find 3 words from the verb 'to be'
- Are the highlighted words adjectives, nouns or verbs?

Adjective — comparative
An adjective gives information about a noun (the *tiny* mouse, a *bigger* house, *green* fields). '*smaller*' is a comparative adjective – one that compares nouns (*taller*, *better*, *noisier*) without picking out the most extreme (*best*, *darkest*).

Fig. 4. Feedback on 'smaller'

- Is the highlighted word an ordinary, comparative, or superlative adjective?

Words of the same type are always highlighted in the same colour, and care is taken to provide a consistent level of feedback, whether the user is attempting a quiz or simply exploring the grammar lesson via the menus on the right of the pages.

5.2 Using the BNC Text

For this project two significant sets of changes were made to the BNC text. The first was to add `<pb>` and `<lb>` tags to indicate page and line breaks. These 'milestone' tags are in accordance with the TEI Guidelines and the BNC DTD. They were needed to allow the trees of page/lines/word information to be built to help the page recognition on the DigitalDesk.

The second was to resolve the portmanteau word types and, in a very few cases, to correct the word types. As only 220 words out of the 6400 in the entire text had portmanteau types, this was not a major job. *Alice's Adventures in Wonderland* is full of names like 'the White Rabbit'. One of the common causes of uncertainty or error came from this use of adjectives and nouns as names. 'Rabbit', 'March', and 'Hatter', for example, were sometimes classified as either singular or proper nouns (NN1-NP0), whereas 'Mock Turtle' was always classified as an adjective (AJ0) followed by a singular noun (NN1), and 'Cheshire Cat' was always classified as a proper noun (NP0) followed by a singular noun (NN1). Another understandable problem occurred with Alice's well-known but ungrammatical exclamation 'Curiouser and curiouser!'. This appeared in the text as

```
<w NN1-NP0>Curiouser <w CJC>and <w NN1>curiouser<c PUN>!
```

The most frequent portmanteau type to occur was VVD-VVN, indicating uncertainty between the normal past tense form of a verb (VVD) and its past participle (VVN). Most of these were resolved to VVD.

A few additional problems concerning the layout of the printed text had to be addressed during implementation. These were mainly caused by line breaks in the middle of words. Although the text did not hyphenate single words across lines, it did occasionally split hyphenated words such as *rabbit-hole* and *bread-and-butter*. The BNC text also contained a few multiword items that were split across lines in the printed

version (for example, the multiword prepositions *up to* and *on top of*). Several layout details like these had to be taken into account during page recognition and when highlighting items on a page.

Almost the opposite problem occurred when dealing with contractions such as *She'll*, *that's*, *doesn't*, and *won't*. These represent two separate words and are identified as such in the BNC text (as `<w PNP>She<w VM0>'ll` and `C<w VDZ>does<w XX0>n't`, for example), but the two words run into each other on the page. This means that the page recognition algorithm has to remember that not all words are separated by convenient word spaces. There is also a problem of accuracy when highlighting part of a contraction on a page. There is a similar, but less acute, problem of accuracy when highlighting very short words such as *I* or *a*.

With the relatively minor exceptions noted above, using the word information in the BNC text was straightforward. Internally, it was convenient to group the word types into a number of 'supertypes'. The supertype for adjectives, for example, includes the three BNC types for ordinary (AJ0), comparative (AJC), and superlative (AJS) adjectives. As the earlier examples showed, it is sometimes appropriate to use the supertype to pick out all adjectives and sometimes appropriate to provide separate information for the three different BNC types. For this application it was not appropriate to distinguish all the 57 BNC word types, so some of the supertypes were never broken down into their constituents. The system is designed so that is is easy to provide different levels of information for children at different stages by plugging in different lists of menus, quiz problems, and supertypes.

6 Exploiting the Text Encoding Initiative via the DigitalDesk

The Active Alice project was designed to exploit one particular type of information in one TEI corpus. This section considers other ways in which information in TEI texts could be accessed via the DigitalDesk.

One obvious alternative for accessing BNC information would be to provide an advanced interface for linguists. Instead of hiding some of the complications of the word types from the user, an application for linguists would allow scholars direct access to the full information and might provide additional tools for studying selected word types and for producing statistics. Some preliminary work has been undertaken on an interface of this type. It allows highlighting of several word types at once and provides simple facilities for finding specified combinations of word types.

However, it is more interesting to consider what other aspects of the TEI facilities could usefully be accessed from a paper document. The following are three possible applications based on different TEI additional tag sets.

1. **Simple Analytical Mechanisms.** The BNC uses only a small proportion of the syntactic encoding allowed by the tagset for simple analytical mechanisms. Further detailed syntactic information about clauses (`<cl>`) and phrases (`<phr>`) may be given. Example tags are

```
<cl type="finite relative" function="adjectival">
<phr type=V function="main verb">
```

Semantic encoding of 'spans' of text can also be used to identify themes, characters, type of discourse, images used, and allusions to other passages or texts. Access to this type of information could readily be provided using an extended version of the 'menus and highlighting' interface developed for the Active Alice project.

2. **Transcription of Primary Resources.** This tagset is designed to record interpretations of old manuscripts. Facilities include
   - annotation of abbreviations and their expansions
   - recording of features in the original manuscript (additions, deletions, corrections, gaps, and information on the ink and 'hand' of the writer)
   - additional information provided by the encoder (text supplied or conjectured from other sources to fill in gaps, for example)

   This type of information is usually recorded after extensive scholarly investigation. It is unlikely that users would have access to an original manuscript (and equally unlikely that the DigitalDesk would be able to recognise it), but printed versions of the corrected and updated texts would be useful and easy-to-read versions in their own right and could also be used on the DigitalDesk to provide access to the detailed information in the TEI version.
3. **Critical Apparatus.** This tagset provides facilities for recording variations of old texts. In particular, different *witnesses* (early editions, translations, or quotations of the work in other texts) may be recorded. This allows the history and evolution of a work to be encoded in a single text. A common way for different versions to be encoded is as a single *lemma* (or base version) and a number of alternative *readings* attributed to other witnesses. A simple example (taken from the TEI Guidelines) showing three variations of a word is

```
<app>
<lem wit="El Hg">Experience</>
<rdg wit="La" type="substantive">Experiment</>
<rdg wit="Ra2" type="substantive">Eryment</>
</app>
```

   A useful application would allow users to take one version of a text to the DigitalDesk in order to access the information on other versions.

7 Summary

The Active Alice project has demonstrated the feasibility of accessing additional information in a TEI text from a paper version. Two main problems need to be overcome before the Alice demonstration could be turned into a general application for BNC texts: limitations in the current DigitalDesk technology and the need to make changes to the BNC text. Fortunately, both these problems should be eased in the near future. The current limitations of the DigitalDesk are the poor OCR and the small size of the active area of the desktop. The cost of cameras and projectors is falling to the stage where it should soon be feasible to use several cameras and projectors to cover a larger area of the desk at better resolution. This would allow a more robust matching process and

should eventually eliminate the need to add the <pb> and <lb> tags to the BNC text. The work involved in checking and updating the word information in the BNC texts should also be eliminated when the texts are reissued using word information from an improved version of the CLAWS system.

The user interface was constrained by the small area of the desktop available for projection of information, but few problems were encountered in designing an interface to the BNC information. Although good interfaces to more diverse TEI texts will demand careful design, the two basic mechanisms used in the Active Alice project

1. providing information from the TEI text when the user points to words on the page
2. highlighting relevant parts of the page in colour in response to user questions or actions

appear to provide a suitable starting point for many paper-to-TEI applications.

Much more work is needed before a satisfactory and widely-available interface between paper and electronic documents is created. The Active Alice project is one step in this process. We believe it is also a significant step towards a new concept in parallel paper and electronic document publishing.

8 Acknowledgements

The original DigitalDesk was developed by Pierre Wellner, a research student at the Computer Laboratory at the University of Cambridge who was sponsored by Rank Xerox. Work on animated documents using the DigitalDesk is currently supported by a grant from the EPSRC. Heather Brown would also like to thank the Computer Laboratory at the University of Cambridge for its hospitality while the work desribed here was undertaken.

References

1. David Barnard, Lou Burnard, Jean-Pierre Gaspart, Lynne A. Price, C. M. Sperberg-McQueen, and Giovanni Battista Varile, 'Hierarchical Encoding of Text: Technical Problems and SGML Solutions', *Computers and the Humanities*, 29(3), pp. 211–231, 1995.
2. Gavin Burnage and Dominic Dunlop, 'Encoding the British National Corpus', in *English Language Corpora: Design, Analysis and Exploitation*, Jan Aarts, Pieter de Haan and Nelleke Oostdijk (eds), pp. 79–95, Amsterdam and Atlanta: Editions Rodopi, 1993.
3. Lou Burnard, 'Text Encoding for Information Interchange: An Introduction to the Text Encoding Initiative', TEI Document no TEI J31, Oxford University Computing Services, July 1995. (Available at http://www.uic.edu/orgs/tei/intros/ via the TEI home page.)
4. Lou Burnard (editor), *Users Reference Guide for the British National Corpus*, Version 1.0, Oxford University Computing Services, May 1995.
5. Lou Burnard, 'Using SGML for Linguistic Analysis: The Case of the BNC', *SGML '96 Conference Proceedings: Celebrating a Decade of SGML*, pp. 95–106, SGML '96 Conference, Boston, MA, November, 1996.
6. Lewis Carroll (retold by Jennifer Bassett), *Alice's Adventures in Wonderland*, Oxford Bookworms, Oxford University Press, 1994.

7. Charles F. Goldfarb, *The SGML Handbook*, Oxford University Press, 1990.
8. G. Leech, '100 million words of English', *English Today*, 9(1), 1993.
9. G. Leech, R. Garside, and M. Bryant, 'CLAWS4: The tagging of the British National Corpus', *Proceedings of the 15th International Conference on Computational Linguistics (COLING 94)*, pp. 622–628, Kyoto, Japan, 1994.
10. C. M. Sperberg-McQueen and Lou Burnard (editors), *Guidelines for Electronic Text Encoding and Interchange (TEI P3)*, ACH/ACL/ALLC (Association for Computers and the Humanities, Association for Computational Linguistics, Association for Literary and Linguistic Computing), Chicago/Oxford, 1994.
11. William Newman and Pierre Wellner, 'A desk that supports computer-based interaction with paper documents', *Proceedings of the ACM Conference on Human Factors in Computing Systems*, pp. 587–592, Monterey, May 1992.
12. Peter Robinson, 'Virtual offices', *Proceedings of Royal Society discussion meeting on Virtual Reality in society, science and engineering*, BT Publication SRD/R5/1, July 1995.
13. Peter Robinson, Dan Sheppard, Richard Watts, Robert Harding and Steve Lay, 'Animated Paper Documents', *Proceedings 7th International Conference on Human-Computer Interaction*, HCI'97, San Francisco, August 1997.
14. Text Encoding Initiative home page: http://www-tei.uic.edu/orgs/tei/index.html.
15. Mark Weiser, 'Some computer science issues in ubiquitous computing', *Communications of the ACM*, 36(7), pp. 74–83, (special issue on 'Computer-Augmented Environments'), July 1993.
16. Pierre Wellner, 'Interacting with Paper on the DigitalDesk', *Communications of the ACM*, 36(7), pp. 87–96, July 1993. (COLING 94), pp. 622–628, Kyoto, Japan, 1994.