

Citation Linking: Improving Access to Online Journals

S. Hitchcock*, L. Carr, S. Harris, J. M. N. Hey and W. Hall

Open Journal Project[†]
Multimedia Research Group
Department of Electronics and Computer Science
University of Southampton
Southampton SO17 1BJ, United Kingdom.

*Contact for correspondence: email sh94r@ecs.soton.ac.uk

[†]The Open Journal project is funded in the UK by JISC's Electronic Libraries Programme (eLib) award ELP2/35.
More information about the project is available at <http://journals.ecs.soton.ac.uk/>

Abstract

The most innovative online journals are maturing rapidly and distinctive new features are emerging. Foremost among these features is the hypertext link, popularised by the World Wide Web and which will form the basis of a new, highly integrated scholarly literature. Journal integration in this instance seeks to recognise, extend and exploit relationships at the level of journal content—the papers—while maintaining some of the familiar contexts, in some cases journal identities, that define the content hierarchy and inform decision-making by readers. Links are a powerful tool for journal integration, most immediately in the form of citation linking. The paper reviews examples of citation linking in practice, and describes a new system, a link service, which is being developed to support novel and flexible linking mechanisms on the Web. One application of this link service is the Open Journal project, which is working with journal publishers to investigate the most effective ways of applying these powerful link types to enhance online journals.

Keywords: electronic journals, hypertext, hypermedia, citation linking, link services

1 Introduction

To say that the number of online journals is growing dramatically is becoming something of a cliché, but not everything is changing so fast. As established publishers bring existing journals to the World Wide Web we are witnessing not so much an evolution but, in many cases, direct replication, even down to the actual appearance of the original printed page when it is viewed online. If the journal article is the basic building block of the scholarly literature, then the journals are the walls. By leaving these walls in exactly the same arrangement as in the print environment, however, the limitations of which are becoming increasingly apparent, the online user is just as enclosed in his or her subject domain, within his or her library, as before. Simply, this is not sustainable.

By embracing the Web, as journal publishers are (Hitchcock *et al.* 1996, Walter 1997), the assumption must be that everything, not just certain things, will change. This includes the structure of

journals publishing and the journals themselves, even down to the form of individual articles and the ways and styles in which they are authored. The authoring issue is for the longer term rather than this paper. It is a culture shock that awaits us, and is well anticipated by Landow (1992).

Instead, we want to examine why significant change in journal publishing is imminent and some of the ways in which it will be effected. In particular, the paper will consider the impact of hypertext links, and a novel way of applying, or *publishing*, links on the way we build and use the literature; how we already expect the journal content, in the future, to be integrated in new, purposeful ways, for example, permitting users to navigate freely among secondary and primary resources (Olvey 1995).

2 Accessing the journal literature

2.1 Current limitations

The most fundamental problem facing journal readers, library users and, it follows, librarians, is physically to get hold of all the journal articles they need when they need them. Experience varies enormously of course, but the days in which every essential journal was held in all major academic libraries have passed. The reason most commonly cited is economic, that journal price inflation has outstripped library (and personal subscriber) budgets, with rationalisation the inevitable result. There has also been growth in the number of journal titles, which reflects changes within traditional academic disciplines: broadening and subsequent fragmentation within fields, and greater emphasis on cross-disciplinary work. These are complex academic, socio-economic and political issues which have been widely discussed. (Cummings, A. M. *et al.* 1992) With the emergence of the Web, conventional, largely failing, print-world responses to these issues may no longer be relevant.

Cutting through this complexity, some points stand out. In a much debated article Tenopir and King (1996), using substantial but possibly dated survey data, revealed that university scientists are reading more papers than in the past, adding that: 'Scientists who read more are more productive and perform their work better.' Crucially, the paper also showed that the modern journals environment is not serving these scientists adequately: 'What has changed dramatically is that scientists read far more from library-provided journals, while at the same time libraries are reducing their journal collections.'

Permission to make digital/hard copies of all or part of this material for personal or classroom use is granted without fee provided that the copies are not made or distributed for profit or commercial advantage, the copyright notice, the title of the publication and its date appear, and notice is given that copyright is by permission of the ACM, Inc. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires specific permission and/or fee

DL 97 Philadelphia PA, USA

Copyright 1997 ACM 0-89791-868-1/97/7...\$3.50

In other words, as the number of published papers grows and the demand to read articles increases, our ability to access the literature is deteriorating.

Rightly or wrongly, the dominant motivation for scholarly journals publishing today, whether by commercial publishers, learned societies or university presses, is profitability (Lynch 1994). Commercial publishing exists to serve markets, so users, librarians too, have an influence. The above suggests that journals markets are not being served, however. While the long-running arguments over journal prices, the so-called 'serials crisis', have failed to resolve this problem, indeed have exacerbated it, the likelihood is that this issue of access will be a turning point. It affects most users more directly than do journal prices. The Los Alamos physics e-print archive, for example, is successful not because it gives access to papers more cost-effectively than do print journals (which it does), but because it fulfils a latent need that was not otherwise being served by publishers or which could not have been served by earlier technology.

The critical new factor is the Web, a massively popular platform that not only offers the chance to prevent further deterioration, but to transform access to the literature. The technology of the Web can be used for different purposes to suit different users in different disciplines, but as soon as that new functionality is fully realised so the purpose of journals publishing in all disciplines will also, ultimately, be transformed.

2.2 Creating information currents

This issue can be viewed from another angle. In principle academic research is progressive, one of its central tenets being that work should not be unnecessarily duplicated. Naturally this is reflected in the structure of the published literature: originality of work is a prerequisite for submission to journals, and the published works themselves form a record that informs, and might even direct, later research. Scholarly articles are not intended to be used in isolation, and one way in which this is reflected is the ubiquity of reference lists within modern papers. Although not referring specifically to scholarly works, Barlow (1994) adds a new perspective: 'Information that isn't moving ceases to exist as anything but potential ... at least until it is allowed to move again.' For a scholarly work, what characterises 'moving' is its ability to influence other work, that is, to be read, commented, acted upon or referenced by others in the field. While it can never be guaranteed that a work will be acted on or responded to, its influence will certainly be seriously proscribed if reader access to the publishing journals is difficult.

Networked electronic, or online, publishing represents a profound change from existing practice that has become commonly delineated in this way as a switch from static (print) to dynamic or 'moving' (online) information. Okerson (1991) anticipated the significance: '... the moment information becomes mobile, rather than static, this transformation fundamentally alters the way in which information is used, shared, and eventually created.' This is not universally accepted. Peek (1994) and Barron (1997), for example, are more cautious about claims for technological change and its impact, but in this case the evidence for sustainable change—current experience and the widespread acceptance of the Web—is compelling. There are persuasive historical visions (Engelbart 1963, Nelson 1987). Licklider and Taylor (1968) believed even then that they were 'entering a technological age in which we will be able to interact with the richness of living

information—not merely in the passive way that we have become accustomed to using books and libraries.' Most notable of all was Bush's (1945) famed vision of a 'web of trails'. Importantly, Bush envisaged being able to read and follow the literature at a speed that begins to match that of our thought processes—literature and life converging in a new dimension.

2.3 Improving access through online integration: anticipating user demands

The key to realising a dynamic journal literature is integration of information resources. It is not through the imperative of this cultural shift that integration is on the agenda, but basic market forces. The revelation of the Web, to commercial publishers at least, is that they can no longer be sure what has value, or what people are prepared to pay for. The volume of material freely available on the Web, and its improving quality in parts, suggests that it is not textual content that obtains financial value automatically as it does in print (Dyson 1994). Instead, it is widely suggested, services are the key. According to a European Commission (1996) study on electronic publishing developments, 'electronic publishers must focus on integrating their content with services, customer-driven product tailoring (and) the brokering of information.'

When print articles and journal issues were simply discrete entities, libraries and secondary publishers provided some integration through services such as bibliographic cataloguing, indexing and abstracting. As more materials become available electronically, far from abandoning these services, Kling and Covi (1995) argue: 'The segregation of e-journals into an electronic space that isn't (yet) integrated into the scholarly document systems of libraries, indices, abstracting services, is a formula for continued marginality.'

The introduction of Web-based library catalogues will begin to speed up access to individual journal articles where electronic versions are available, but there are still many problems. Critics such as Dempsey (1995) have highlighted the gap in providing detail about journal articles actually held by an individual library. On the other hand, digital libraries such as the Perseus Project (Crane 1996) have shown that a critical mass of multimedia, multivalent materials can be a catalyst for the transformation of practice.

Indirectly, there is evidence that users are ready to accept dramatic changes in practice. A major stimulus for improved access to electronic articles is increased electronic collaboration between researchers. Wulf (Kouzes *et al.* 1996) defined this environment as the 'collaboratory', merging 'collaboration' and 'laboratory' to describe a '.. center without walls, in which researchers can perform their research without regard to geographical location—interacting with colleagues, accessing instrumentation, sharing data and computational resource, and accessing information in digital libraries.' The recently completed Distributed, Collaboratory Experiment Environments (DCEE) in the USA involved seven testbed projects representing scientific environments of various scales and involving from tens to hundreds of scientists and students. (Johnston and Sachs 1997)

Elsevier's Tulip project attempted to deliver familiar journals to the user's desktop in electronic form and involved one of the most extensive evaluations to date of user responses. The final report

on the project (Borghuis *et al.* 1996) summarised users' requirements and included:

- access to all information from one source;
- effective search capabilities;
- high publishing speed (timeliness of the information);
- sufficient journal and time coverage;
- *linking of information.*

We can begin to see that the limited impact of online journals alluded to by Kling and Covi (1995) should not be allowed to obscure the importance of new integrating services which the online medium uniquely can support. Search engines are widely used on the Web, and increasing use is being made of custom alerting services to bring newly published material to attention promptly.

To these means of integration must be added the hypertext link. The link is the single most important, characteristic feature of Web pages. The link allows us to approach Bush's vision of an information resource in which any item of information is, potentially, an instant of a 'click' away, but to exploit this potential requires ingenuity and more flexible systems to be able to design and apply links more effectively than we do now.

A way in which this feature of the Web can be used immediately to enhance journal papers is to build on established practice by linking citations to the cited articles.

2.4 Some initiatives towards citation linking

Just as it is rare to find links in online journal papers, especially those reproduced from a print original, so too citations were absent from the earliest journals (Schaffner 1994). Given that citations are an integral feature of the modern scholarly paper, it is not hard to anticipate how citation links will proliferate in Web-based journals.

The likely effect of citation linking can be gauged by recognising a direct parallel with citation indexing, which is sometimes referred to as 'forward' referencing (i.e. for a given article, all the subsequent papers that cite it), developed by Garfield (1955). Garfield's description of citation indexing as an 'association of ideas' bears remarkable similarity to Bush's 'association of thoughts' which anticipated modern hypertext. Citation linking combines the two approaches, mapping both reference data and citation index data on to the text in the form of links. Adding electronic links to the literature is introducing a new culture in many respects, but citation linking is likely to be acceptable to the academic community because it builds on practices established in other forms such as print. It also allows the community to exploit its own intellectual input in this process, recognised in Garfield's original rationale that 'by using authors' references in compiling the citation index, we are in reality utilizing an army of indexers'.

Unsurprisingly, those publishers first to produce innovative online journals or to bring large journal programmes online are most alert to the possibilities of citation links. Not least of the problems to be solved if this is to be achieved, the articles must be available online, and that includes the current and historical literature. Further, citation linking between journal articles will only be effective if links are direct to the referenced article, and this immediately presents difficulties, not just technical

difficulties, but the need to pass through commercial walls, between other journals or other publishers, that are a legacy of print journals. New applications are highlighting this problem, but interim solutions are emerging.

With the substantial Los Alamos e-print archive to serve it, the field of physics is ideally placed to exploit the new facilities, as was revealed in an electronic publishing special issue of the *APS News* insert, *APS Online*. Articles posted in the e-print archive have identifiers: 'Certain non-APS journals have begun to accept the archive identifier as the electronic submission itself, and conduct their editor/refereeing interactions as well by means of the version retrieved from the archive.' (Ginsparg 1996) In some of these cases the publisher encourages 'preprint' papers, that is papers which have been accepted for journal publication but have not yet been 'printed' or published in the paper edition, to be placed in the archive, with a link established from the journal's Web site to the archive. Ginsparg continued: 'Physical Review D adds such links to its Web pages, uploading directly to the archive information concerning papers "to appear", and later their published status. The information is then available whenever users search the archive listings or browse abstracts.'

Another example, Physical Review Online Archives (PROLA) posts two lists on each document information page: the first provides a link to all PROLA articles that are referenced by the current article, and the second lists links to all the later papers that reference the current paper, including any errata that may have appeared. 'This hyperlinked cross-reference system allows users to locate and traverse a thread of related articles.' (Thomas 1996) PROLA offers hyperlinks to articles within the PROLA system, will automatically locate and list all references to archived articles from the 1996 online APS journals, and there are plans to offer links to other, non-APS, online journals that reference *Physical Review*.

Institute of Physics Publishing (IoPP) in the UK was one of the first journal publishers to make all of its titles available online. Its latest development is the Hypercite Project, through which a subscriber can link from a reference in any online IoPP paper to an abstract (if it exists) from an Inspec-provided database or can link directly to the full-text article within another IoPP journal. Online versions of IoPP journals can be purchased as a site licence on terms, in the UK at least, which encourage purchase of the whole journals catalogue, thereby making it possible to link between full-text papers within IoPP's site. The wider world of full-text physics papers beyond IoPP is mimicked by substituting the abstracts service. For users, even this should prove to be a significant step forward.

Similarly, Electronic Press, through its BioMedNet service for scientists in biology and medicine, links citations to full-text articles in journals it owns, and other citations are linked to abstracts in its Evaluated Medline database (Hitchcock *et al.* 1997). HighWire Press, a producer of 'high impact' biomedical and science online journals, is another to have begun exploring the use of links between full-text articles available through its services or to abstracts in a Medline database.

Astronomy, a smaller discipline, shows most promise in providing full-text linking between different publishers. According to Peter Boyce, Senior Associate in charge of electronic publishing at the American Astronomical Society (AAS), interlinking 'is easy in astronomy where 90 per cent of the peer-reviewed literature is contained in about six core journals, with very few references

which go outside this group... Most astronomical institutions subscribe to all the relevant astronomical journals, and all the journals are choosing to provide institution-wide licenses based on IP addresses, so in practise there should be no impediment.' Boyce's preference for making the system fast, however, eliminating passwords and unnecessary barriers between journals, at least for a couple of years, seems to be too much for other publishers to agree to: 'The financial uncertainty is so large that any suggestion of foregoing even a small percentage of revenue is too much.'

Instead, astronomy can call on an extensive archive of bibliographical data and abstracts held by the Astrophysics Data System (ADS), which is funded by NASA and used as a clearinghouse for references and citations by 'everyone' in the field. Now a group of journal publishers and data centers, in collaboration with the ADS, have announced a system, Urania, that uses this data for linking, providing 'references to 20 years of historical texts of journal articles (in the ADS) and all of it linked both backward (references) and forward (citations to the article being read)', Boyce says.

How useful is all of this? From the special issue insert in *APS News*, this view of an editor and user: 'When information about essentially *all* references in an article is available, with forward links to other articles, we will truly have added value to the paper form. Finally, when the entire text of a referenced article can be accessed by a mouse click, we will have reached a new level of information access.' (Austin 1996)

3 The Open Journal project and link services

The Open Journal project likewise is investigating the interlinking of journals and journal papers, in this case applying a new tool called a 'link service' that complements and enhances the capabilities of the Web to support links by introducing powerful link types (Carr *et al.* 1995). Specifically the project is using the Distributed Link Service (DLS) being developed at Southampton University, although there are other link services (Carr *et al.* 1996). Hitchcock *et al.* (1997) discussed the application of this tool to linking journal papers and compared it with the methods used for citation linking in some of the publishing examples described above.

Broadly, the project, funded by the UK Electronic Libraries programme (eLib), addresses the issue of information coherence in the information environment currently served by academic libraries. An increasing proportion of the diverse information assets in this environment are becoming electronic, and from a single 'terminal' users can access journals, databases and articles but, as we have seen, are required to navigate a complicated path through many providers' information gateways to locate any particular piece of information.

The goal of the project is to develop a framework of information retrieval technologies (based around the DLS) and to establish electronic publishing practices to be used by information providers (especially journal publishers, but also by librarians, perhaps) which will allow them to make publications available not as isolated resources, but as cooperating assets within an information delivery environment: the 'open' journal.

The facilities of the DLS, which allow authors to create links between arbitrary Web documents and then to publish those links for the benefit of other users, provide a basis for reintegrating the

Overlaying links on documents as a server-side process

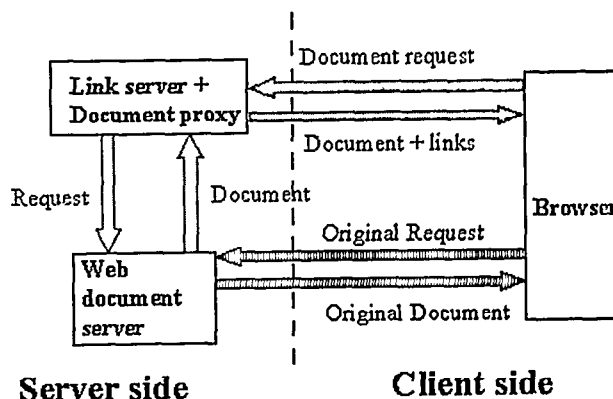


Figure 1: WWW proxy-based document enhancement

online literature. In hypertext terms, they allow authors to turn a bounded subset of the Web from a *hyperbase*, an undistinguished database of information nodes, into a *hyperdocument*, a collection of documents united by authorial intent. (Stotts 1991)

One of the major features of the DLS is that, instead of inserting link data within the HTML (or other) coding of a source document, the data are held separately in a link database, or linkbase. The links can be viewed superimposed on a Web page, as though a transparency were overlaid on it, by retrieving the document from a url via a proxy link server (Figure 1). This is easily set up using the standard menu facilities on popular Web browsers. The links appear to act as conventional Web links, so the application of the link service is effectively transparent, while at the same time it provides the link author and user with significantly improved link flexibility.

It is this capability of separating links from the authored content which is perhaps the key to allowing journals to interoperate, because it allows the creation of links between two third-party resources, not just to another publisher's documents, but *from* them as well.

3.1 Why a link service will aid the Web

Imagine an institutional library has introduced a new rule governing the use of its books and journals. To be allowed to take out a book, users need to present the issue desk not with the book itself, but some other book which contains a citation for the book actually required. Only when the librarian sees a valid citation for a book or journal will the user be allowed to read it. Further, the staff choose to keep all books behind the counter, inaccessible to library users but directly available to qualified librarians. In this scenario there is no way to idly look through stacks of new periodicals to keep abreast of the latest developments, nor can users select the particular book required on a topic by inspecting the whole range of books available on that topic.

This is almost the situation that users of the Web find themselves in, an environment where a private collection of documents can be requested individually, but only given a valid reference to the document (in the form of a url). In both situations, neither the

identity of a document, nor even its existence, can be deduced or guessed.

Under the new rules there are two approaches that will enable users to have access to all the related information they want from a collection of disparate books and journals from a range of publishers:

1. produce comprehensive 'meta-books'
2. provide links leading from every article, book and journal

Much activity has been directed at the former—producing lists of content, and lists of lists, making catalogues and indexes of subject material on the Web—but less in authoring and publishing links, towards which the link service is directed. Despite the above examples of citation linking, the scarcity of links in the majority of online versions of established journals is partly a reflection of this situation.

One likely effect of using a link service in a professional publishing environment, such as journals publishing, will be to transform link authoring into a link *publishing* task, possibly handled by link editors. In this way links are added at the moment of publication rather than during the authoring stage, an important feature in a dynamically changing environment such as the Web. The link maintenance problem also becomes more manageable. In addition the Open Journal project has extended the core functions of the DLS to work not just with html documents but also portable document formats viewed with Adobe Acrobat.

3.2 Building Open Journals

To demonstrate the concept the project is producing three Open Journals in the areas of biology, cognitive science and computer science. Each consists of original journals from a number of different primary and secondary publishers, served from a number of different sites in a number of data formats (principally html and pdf). For more general information on the project see the Web site: <http://journals.ecs.soton.ac.uk/>

Dictated by differing cultures, academic disciplines show various degrees of acceptance of online publishing, adopting different approaches. This is borne out by clear distinctions between the three Open Journals. Biologists have access to a diverse range of online materials, from journals to books, databases of gene sequences and other molecular structures, libraries of graphical images and teaching resources. The information is invariably well catalogued and often freely accessible. The databases in particular form a focal point for the field, lending it an online coherence that few other disciplines can match and attracting a higher proportion of the field's prestigious journals online than other fields. This area also provides the project with its longest list of participating journals, one of the project objectives in this case to begin to link these journals into the non-journal data resources.

In contrast, computer science offers large volumes of disparate online resources with little organisation apart from some loosely catalogued indexes, although there is an attempt to build a coherent 'library' of networked computer science technical reports (Davis 1995). Few established computer science journals are available online, and the odd new online journals in this field, with one or two exceptions, have been noticeably slow to grow. In this case the project aims to identify pockets of coherence that can be extended through the application of links, starting with

specialised journals provided by publishers collaborating in the project.

Cognitive science, likewise, seems ill-served with online journals. It is ironic, then, that the field delivered one of the first online journals of all, *Psychology*, edited by Stevan Harnad and launched in 1990. This journal and its extensive archives, together with its print counterpart *Behavioural and Brain Sciences*, which recently began an online preprints service, are the nucleus of the project's cognitive science Open Journal. This example forms the basis of the project's most extensive demonstration of the use of a link service to link journal papers from reference lists, and is described in more detail below.

3.3 Citation links and the Distributed Link Service

As has been argued, most journal papers available on the Web are unlinked. The DLS can be used to overcome the lack of links on most Web journals by dynamically matching a fixed database of pre-defined links against the contents of an article, inserting a link whenever an explicit match occurs.

More general than a programming language GOTO statement, a link is the specification of a relationship between a data source and destination where both the source and destination may expand to one of several places in a particular set of documents. These flexible relationships are usually coded explicitly by the DLS as a generic link (Carr *et al.* 1995), but they may be the by-product of a more complicated processing arrangement. For the purpose of citation linking this involves matching citations and bibliography lists in a document against a database of bibliographic information.

This is a complex process which requires an understanding of the house style of the publication, an understanding of the document coding format and the ability to parse complex language structures. Since this is a knowledge-based activity being undertaken on behalf of the user, we call it a *citation agent*. When the DLS processes a document which is defined to be in the Cognitive Science Open Journal, it launches the agent to act on the document before adding other links in the usual way.

As well as recognising the citation, the agent has the problem of retrieving the cited paper, *i.e.* where to link the citation to. In this project we have chosen to link the citation in the body of the document to the bibliographic reference at the end of the document (an internal link) and then to link that reference to an external bibliography database. This database, although not linking to the full-text articles, provides a complete set of metadata for the cited article including an abstract and the citations given in that cited document. Making that record available, with each citation linked back into the database, allows the user, in theory, to browse the complete literature (in summary form), with each reference instantly available. (Fig. 2)

Extracting the citation is a more or less difficult task for the agent depending on the format in which the paper is held. Highly structured document formats based on SGML are in principle the simplest as all the separate components of the bibliography data are marked explicitly. At the other end of the scale, simple ASCII provides problems in recognising both the existence of a citation and the boundaries of its component parts (for example, where the author names end and the title starts). HTML, the format of the

IV. PRIMING POSITIVITY

5. When words are repeated in a task where occasional targets (e.g., non-words) have to be detected, then the potentials evoked by repeated words are positively shifted (Bentin & Peled, 1990; Rugg, Pearl, Walker, Roberts & Holdstock, 1994). This broad positivity (between 250 ms and 700 ms) has been interpreted as an attenuation of the N400, followed by an enhancement of P3 (Rugg et al., 1994). A relationship of this broad positive shift to theta is not obvious. Further, there is no obvious relationship to the hippocampus, since the effect does not differ between lobectomized and other epileptic patients (Rugg, Roberts, Potter, Pickles & Nagy, 1991) nor between Alzheimer patients and controls (Friedman, Hamberger, Stern & Marder, 1992; Rugg et al., 1994).

V. RECOGNITION POSITIVITY

6. In this paradigm, two lists of words are presented, with the second list consisting of "old" items (i.e., members of the first list) and "new" items. ERPs are recorded during the second list. Correctly detected old items evoke a larger positivity than other items (Sanguist, Roubenagh, Syndulko & Lindsley, 1980; Karis, Fabiani & Donchin, 1984). This positivity starts later than priming positivity (Rugg & Nagy, 1989) and is larger the more clearly the word is remembered (Smith, 1993). It has been interpreted as enhancement of the very P3 due to great confidence in the decision (Rugg & Nagy, 1989) which would agree with a general regularity in P3's behavior (Johnson, 1986). But on the other hand, the topography of this effect might differ from P3's (Smith & Guster, 1993).

Figure 2a: Original article without links

Rugg, M.D., Pearl, S., Walker, P., Roberts, R.C. & Holdstock, J.S. (1994). *Word repetition effects on event-related potentials in healthy young and old subjects, and in patients with Alzheimer-type dementia*. *Neuropsychologia*, 32, 381-398.

Rugg, M.D., Roberts, R.C., Potter, D.D., Pickles, C.D. & Nagy, M.E. (1991). *Event-related potentials related to recognition memory: Effects of unilateral temporal lobectomy and temporal lobe epilepsy*. *Brain*, 114, 2313-2332.

Sanquist, T.F., Roubenagh, J.W., Syndulko, K. & Lindsley, D.B. (1980). *Electrocortical signs of levels of processing: Perceptual analysis and recognition memory*. *Psychophysiology*, 17, 568-576.

Smith, M.E. (1993). *Neurophysiological manifestations of recollective experience during recognition memory judgments*. *Journal of Cognitive Neuroscience*, 5, 1-13.

Smith, M.E. & Guster, K. (1993). *Decomposition of recognition memory event-related potentials yields target repetition and retrieval effects*. *Electroencephalography and Clinical Neurophysiology*, 86, 335-343.

Smith, M.E. & Halgren, E. (1989). *Dissociation of recognition memory components following temporal lobe lesions*. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 15, 50-60.

Smith, M.E., Halgren, E., Sokolik, M., Baudena, P., Masolino, A., Liegeois-Chauvel, C. & Chauvel, P. (1990). *The intracarotid topography of the P3 event-related potential elicited during auditory oddball*. *Electroencephalography and Clinical Neurophysiology*, 74, 725-740.

Figure 2c: Citation link followed to article's bibliography section

articles used for this Open Journal, can provide extra clues implicit in the formatting markup (for example, titles may be rendered in italics).

The project is using a 500 Mb Tagged Data Format (TDF) file supplied by the Institute for Scientific Information (ISI), with abstracts of 200 000 papers taken from 7000 journal issues from 300 journal titles covering the years 1991-1995 in the field of cognitive science.

Various ways of indexing and matching entries in a database of abstracts have been described. (Garfield 1955, Hitchcock *et al.* 1997) In this instance the citations are linked to the abstracts by performing a fuzzy database look-up, comparing the document's reference data (as located by the citation agent) with the TDF file. Due to the size of the TDF file the most practical way of locating an abstract in the data is using two levels of indices. The inner level has the surname and initials of one of the authors along with

IV. PRIMING POSITIVITY

5. When words are repeated in a task where occasional targets (e.g., non-words) have to be detected, then the potentials evoked by repeated words are positively shifted (Bentin & Peled, 1990; Rugg, Pearl, Walker, Roberts & Holdstock, 1994). This broad positivity (between 250 ms and 700 ms) has been interpreted as an attenuation of the N400, followed by an enhancement of P3 (Rugg et al., 1994). A relationship of this broad positive shift to theta is not obvious. Further, there is no obvious relationship to the hippocampus, since the effect does not differ between lobectomized and other epileptic patients (Rugg, Roberts, Potter, Pickles & Nagy, 1991) nor between Alzheimer patients and controls (Friedman, Hamberger, Stern & Marder, 1992; Rugg et al., 1994).

V. RECOGNITION POSITIVITY

6. In this paradigm, two lists of words are presented, with the second list consisting of "old" items (i.e., members of the first list) and "new" items. ERPs are recorded during the second list. Correctly detected old items evoke a larger positivity than other items (Sanguist, Roubenagh, Syndulko & Lindsley, 1980; Karis, Fabiani & Donchin, 1984). This positivity starts later than priming positivity (Rugg & Nagy, 1989) and is larger the more clearly the word is remembered (Smith, 1993). It has been interpreted as enhancement of the very P3 due to great confidence in the decision (Rugg & Nagy, 1989) which would agree with a general regularity in P3's behavior (Johnson, 1986). But on the other hand, the topography of this effect might differ from P3's (Smith & Guster, 1993).

Figure 2b: DLS adds citation links

Rugg, M.D., Pearl, S., Walker, P., Roberts, R.C., Holdstock, J.S. *WORD REPETITION EFFECTS ON EVENT-RELATED POTENTIALS IN HEALTHY-YOUNG AND OLD SUBJECTS, AND IN PATIENTS WITH ALZHEIMER-TYPE DEMENTIA.*

Event-related potentials (ERPs) were recorded from 16 healthy young (mean age 21 years) and 16 healthy old subjects (mean age 64 years), and from 11 subjects with a diagnosis of Dementia of Alzheimer Type (DAT). The task requirement was to attend to a series of visually presented words so as to respond to occasional animal names. Non-animal names repeated either a single or six intervening items. In the young subjects ERPs evoked by repeated words displayed a widespread, sustained positive-going shift relative to ERPs evoked by first presentations (the ERP repetition effect). This effect onset around 220 msec and did not differ as a function of inter-item lag. Other than for a delay in onset of approximately 60 msec, the ERP repetition effect in the healthy old group was in all respects equivalent to that of the young subjects. The ERP repetition effects in the DAT patients were statistically indistinguishable from those of an appropriately matched sub-set of the healthy old subjects.

Articles cited in this paper are:

- Bentin, S (1992). *J EXP PSYCHOL LEARN*, 18-1272
- Bentin, S (1990). *MEM COGNITION*, 18-358
- Craik, F.I.M (1992). *HDB AGING COGNITION*, 51
- Folstein, M.F (1975). *J PSYCHIAT RES*, 12-189
- Friedman, D (1992). *ANN NY ACAD SCI*, 658-33
- Friedman, D (1990). *BIOL PSYCHOL*, 30-61
- Friedman, D (1992). *J CLIN EXP NEUROPSYCH*, 14-448
- Friedman, D (1993). *PSYCHOL AGING*, 8-120

Figure 2d: Bibliography link followed to ISI bibliography database

the year of publication and a pointer into the main database, with one entry per author. e.g.

NICHOLSON,IR:1993:1023
NEUFELD,RWJ:1993:1023

The outer index holds starting points for each two-letter surname prefix in the alphanumerically-sorted inner index.

Citations may then be searched for by each author individually, for example, searching for "Nicholson, IR & Neufeld, RWJ (1993)" would require searching for "Nicholson, IR 1993" and "Neufeld, RWJ 1993". Both of these will, assuming the paper is present in the database and spelled correctly, return one or more pointers into the main database, e.g.

Looking for: NICHOLSON,IR (1993)
Found: NICHOLSON,IR:1993:1023

Looking for: NEUFELD,RWJ (1993)
Found: NEUFELD,RWJ:1993:1023
Found: NEUFELD,RWJ:1993:995

The resulting list of possible pages is then retrieved in order of likelihood and compared with other data from the citation. This comparison is chiefly done by looking at the titles of the papers and the title given in the citation. A straightforward string comparison is insufficient for this task due to differences in punctuation and spelling or typographical errors, so a text differencing algorithm is used:

- All punctuation is homogenised: hyphens and apostrophes joining parts of a word are removed, hyphens and other punctuation (commas, colons, semi-colons *etc.*) are converted to dots.
- All vowels which do not start words are removed
- Words are de-pluralised
- Strings of repeated characters are reduced to a single character.
- Lower-case characters are translated to upper-case.

For example, "Citation Linking and Meta-data: The Way Forward" would become "CITN LNKNG AND MTD. TH WY FWRD". A delta value of $|A \text{ xor } B|$ (*i.e.* the number of words in A but not in B plus the number of words in B but not in A), is calculated, where A is the set of words in the citation and B is the set of words in the database title. This value is compared with a threshold and the most likely paper (if any) is returned, taking the number of matched authors names into account. If necessary the match can be verified by comparing the journal name and issue numbers in the same manner, although this data is harder to extract and compare.

This process matches a citation to an entry in the database, and was originally performed at the time the user looks up the citation. In other words, the software identified the places in the document where the citations occurred and put the links in as queries to a database search script. For example, the reference 'Manis, M., Dovalina, I., Avis, N.E. & Cardose, S. (1980)' would have been automatically linked to the query URL `http://server/lookup.cgi?manis,m:dovalina,i:avis,ne:cardose,s:1980`. The advantage of this approach was that it delayed the relatively slow database look-up until the user had explicitly requested it by following the link. The strategy of adding links blindly, however, often resulted in the user being presented with an apology from the database instead of the document summary, because the domain of the database is limited principally in time but also in breadth of journal coverage.

For this reason, it was decided to verify the links before they were added to the document, and each citation is checked individually for a unique resolution in the database before the link is created. Also, the link is created directly to the correct record for the article, instead of indirectly to a query processor as before. This extra processing must now be performed as the document is delivered to the user. This slows down the delivery, but since HTML documents can be streamed (the user can see the start of a document while the DLS is still processing its end), and since the bibliography section is invariably at the end of a article, this is not too much of a disadvantage.

Current indications are that some 60% of the citations in the nucleus of the Cognitive Science Open Journal are outside the time-range of the database and hence are ignored. About half of the remainder are successfully recognised and linked, with the rest accounted for by citations to books or citations out of the core of the discipline. This is perhaps not surprising for such an interdisciplinary subject, with patterns of success rates emerging for different authors and different themes within the journals. Further work is being conducted to improve the hit-rate.

4 Conclusion: the way forward

With the arrival and, in some fields, broad acceptance of a core mass of scholarly online journals, the most progressive of these journals are evidently heading towards a new phase in which they are being enhanced to take advantage of the capabilities of the Web. Tools to support this new functionality are proliferating. Among them, link services offer the potential to transform the use of the dominating feature of the Web, the hypertext link.

Journal publishers of all types, whether with enthusiasm or by indirect coercion, are embracing the Web. The next challenge for these publishers is not how they make their journals available online, but how they can participate in this emerging, highly integrated environment, to the benefit of journal authors and users as well as themselves, without in the short term harming their most valuable assets, the journal titles. The broad sweep of journal titles is not a simple hierarchy but reflects both the wider and specialised interests within the academic community. Established titles also confer important recognition on contributing authors, the loss of which cannot be immediately replaced, while the existing journals structure has become the foundation for archiving and recording, another preciously guarded asset.

The wedge that threatens this arrangement will not be, as has long been regarded, purely economic, but the increasing gulf between the deteriorating accessibility of print journals and the irresistible improvements offered by online resources. To avoid an accelerating erosion of journal titles the onus is therefore on publishers to commit these titles to the online community in order that they can then exploit the new opportunities this will produce.

The Open Journal project demonstrates a tool, the Distributed Link Service, which has the potential to add new value to online journals through the use of flexible and cost-effective link 'publishing' strategies that are consistent with the new dynamic of the Web. In addition, by working directly with publishers the project is able to develop these strategies in ways which might preserve the important features of existing journals, the journal identities, while encouraging their deployment to the advantage of the whole online community.

Acknowledgements

The citation database for the Cognitive Science open journal has been provided by the Institute for Scientific Information. We would like to acknowledge our thanks to Robert Kimberley and Helen Atkins from ISI for their help and suggestions and to Stevan Harnad for the use of the journal *Psychology*.

References

1. Austin, S. (1996) An editor/user perspective. *APS News: APS Online* (insert), November, 7

2. Barlow, J. P. (1994) The economy of ideas. *Wired*, 2.03, March
3. Barron, D.W. (1997) Electronic paper: a model for electronic journals. Submitted to *Journal of Digital Information*
4. Borghuis, M. et al. (1996) The TULIP project final report <http://www.elsevier.nl/info/projects/trmenu.htm>
5. Bush, V. (1945) As we may think. *Atlantic Monthly*, July, 101-108
<http://www.isg.sfu.ca/~duchier/misc/vbush/>
6. Carr, L., De Roure, D., Hill, G., and Hall, W. (1995) The Distributed Link Service: a tool for publishers, authors and readers. *Proceedings of the Fourth World Wide Web conference*, Boston, MA, USA
http://wwwcosm.ecs.soton.ac.uk/dls/link_service.html
7. Carr, L., Davis, H., De Roure, D., Hall, W., and Hill, G. (1996) Open information systems. *Computer Networks and ISDN Systems*, 28, 1027-1036
http://www5conf.inria.fr/fich_html/papers/P12/Overview.html
8. Crane, G. (1996) Building a digital library: the Perseus Project as a case study in the humanities. *Digital Libraries 1996, Proceedings of the 1st ACM international conference on Digital Libraries*, edited by E.A.Fox and G. Marchionini (New York: ACM), pp.3-10
9. Cummings, A. M. et al. (1992) *University Libraries and Scholarly Communication: A Study Prepared for The Andrew W. Mellon Foundation* (Washington, D.C.: Association of Research Libraries)
<http://www.lib.virginia.edu/mellon/>
10. Davis, J. R. (1995) Creating a networked computer science technical report library., September <http://www.dlib.org/dlib/D-Lib Magazine september95/09davis.html>
11. Dempsey, L. (1995) The scandal of serials holding data. *Catalogue & Index*, 118, 9
12. Dyson, E. (1994) Intellectual property on the Net. *Release 1.0 report*, December
<http://www.edventure.com/release1/1294.html>
13. Engelbart, D. C. (1963) A conceptual framework for the augmentation of man's intellect. In *Vistas of Information Handling*, Vol. 1 (London: Spartan Books)
14. European Commission DG XIII/E (1996) *Strategic Developments for the European Publishing Industry Towards the Year 2000* (Brussels, Belgium: European Commission). An information note on this study is available at <http://www2.echo.lu/elpub2/en/infonote.html>
15. Garfield, E. (1955) Citation indexes for science: a new dimension in documentation through association of ideas. *Science*, 122, 15 July, 108-111
16. Ginsparg, P. (1996) Los Alamos XXX APS ONLINE. *APS News: APS Online* (insert), November, 8
17. Hitchcock, S., Carr, L. and Hall, W. (1996) A survey of STM online journals: the calm before the storm. In *Directory of Electronic Journals, Newsletters and Academic Discussion Lists*, 6th edition (Washington, D.C.: Association of Research Libraries)
<http://journals.ecs.soton.ac.uk/survey/survey.html>
18. Hitchcock, S. et al. (1997) Linking everything to everything: journal publishing myth or reality? *ICCC/IFIP Conference on Electronic Publishing*, Canterbury, UK, April
19. Johnston, W. E. and Sachs, S. (1997) Distributed, Collaboratory Experiment Environments (DCEE) Program Overview and Final Report (draft, version 3)
http://www-itg.lbl.gov/DCEEpage/DCEE_Overview.html
20. Kling, R. and Covi, L. (1995) Electronic journals and legitimate media in the systems of scholarly communication. *The Information Society*, 11 (4), 261-271
<http://www.ics.uci.edu/~kling/klinge2.html>
21. Kouzes, R.T., Meyers, J.D. and Wulf, W.A. (1996) Collaboratories—doing science on the Internet. *Computer*, 29 (8), 40-46
22. Landow, G. (1992) *Hypertext: The Convergence of Contemporary Critical Theory and Technology* (Baltimore, MD: Johns Hopkins University Press)
23. Licklider, J. C. R. and Taylor, R. W. (1968) The computer as a communication device. *Science and Technology*, April, 21-31
24. Lynch, C. A. (1994) Scholarly communication in the networked environment: reconsidering economics and organizational missions. *Serials Review*, 20 (3), Fall, 23-30
25. Nelson, T. H. (1987) *Literary Machines*, edition 87.1
26. Okerson, A. (1991) The electronic journal: what, whence, and when? *Public-Access Computer Systems Review*, 2 (1), 5-24
<http://info.lib.uh.edu/pr/v2/n1/okerson.2n1>
27. Olvey, L. D. (1995) Library networks and electronic publishing. *Information Services and Use*, 15, 39-47
28. Peek, R. P. (1994) Where is publishing going? A perspective on change. *Journal of the American Society for Information Science*, 45 (10), December, 730-738
29. Schaffner, A. C. (1994) The future of scientific journals: lessons from the past. *Information Technology and Libraries*, December, 239-247
30. Stotts, P. and Furuta, R. (1991) Hypertext 2000: databases or documents? *Electronic Publishing: Origination, Dissemination & Design*, 4 (2), 119-121
31. Tenopir, C. and King, D.W. (1996) Setting the record straight on journal publishing: myth vs. reality. *Library Journal*, March 15, 32-35
32. Thomas, T. (1996) PROLA—Physical Review Online Archives. *APS News: APS Online* (insert), November, 3
33. Walter, M. (1997) Online journals: print publishers move from pilot to full rollout. *The Seybold Report on Internet Publishing*, 1 (6), February, 10-20