The New ADONIS

Barrie T. Stern
ADONIS

Introduction

Ever since its inception, ADONIS has enjoyed a high and favourable exposure in the library and publishing world. Numerous discussions on the pros and cons of information technology in the information dissemination process have been triggered by the concept of automated document delivery systems.

After more than ten years of technical and market research, a workable automated document delivery system was finally put to the test during a two-year trial period, in which various components and assumptions of the application of information technology in document delivery were tested.

The results were encouraging and the possibility of a publisher-backed document delivery service, operating on a commercial basis was spurred by the request from a representative number of information professionals in the pharmaceutical industry. They wanted a tailor-made document delivery system. The publishers in the ADONIS consortium responded favourably to this as the pharmaceutical industry is a major consumer of STM information and drives much of the demand for document delivery.

Launched commercially in 1991, the new commercial service has been received with great interest. However, although there have been a number of papers on the results of the trial and the ADONIS concept, we feel that an update is required on the developments that have taken place since, as well as where ADONIS is today.

The Development of the New Service

To determine the information needs of the prime target group, the pharmaceutical industry, an extensive market survey was needed. It was found that a commercial document delivery service could succeed, provided that a number of criteria were met:

- The list of titles has to closely match the users' information needs
- It must be competitive compared with photocopying and ILL services
- The print-outs must meet the quality requirements of end-users
- The rapid conversion from paper to CD-ROM (less then three weeks) was required by the customers
- The service should run on 'standard' Personal Computers, commonly found in libraries, eliminating both the costly development of special hardware as well as the additional investments by users such hardware would require
- Special compression techniques had to be developed to store an ever-increasing number of pages to keep down the number of CD-ROMs produced,as there is a limit to the number of CD-ROM discs a librarian is willing to handle manually
- The system should offer an array of benefits, such as library management tools
- The service should offer a mechanism for the collection of royalties and provide market feedback to publishers

A number of drug companies were contacted, who agreed to supply ADONIS with the lists of periodicals they subscribed to. Merging and purging their lists resulted in a 'core' list of about 1,000 titles published by more than 200 publishers, primarily in the biomedical and biochemical disciplines.

Taking the existing physical limitations (storage capacity of pages) of CD-ROM into account, as well as the time needed to negotiate the terms and conditions with each individual publisher ADONIS decided to start with around 350 titles with a short-term target (2-3 years) of 700.

The suitability of these titles has been discussed with more than forty pharmaceutical companies, mainly in Europe and the United States.

Handling of Source material

Even with 385 journals covered by the Service, the amount of paper pages processed is impressive; starting at 450-500,000 annually with an expected rapid increase to 1,000,000+ pages per year, as new titles are constantly added to the service.

This volume and the quality requirements (at least as good as a photocopy) by the end-user, necessitated the selection of an experienced scanning operation with advanced high resolution scanning equipment.

A specialized software house, with a proven track record in the CD-ROM industry and experience of working with the selected scanning company, was selected to develop the search and retrieval software, the necessary production software for the scanning bureau and the user interfaces.

To retrieve articles from CD-ROM, an index is needed to point to the information each disc contains as well as to provide searching facilities. Full Boolean logic, simultaneous left and right truncation, character string searching, inverted file format for rapid retrieval, conventional bibliographic fields and a weekly cumulation of the index which can be held on a mainframe for networking access are featured. The retrieval software operates using icons and a mouse.

Problems & Solutions

ADONIS production involves a relatively complex process, apart from the effort to make totally transparent the specifications which link all the steps in the production process in three different European nations, by four companies. The main problems stemmed from the need to develop software to run on standard hardware yet meet the required level of performance.

Software Problems: compression

The standard compression techniques usually used to store scanned images on CD-ROM allow the capture of about 5-6000 pages on a single disc in image (as opposed to ASCII) mode. To meet the maximum of 50 discs per year, a compression algorithm was needed to at least double this amount. Compression involves a number of processes, starting with 'removing' the white space in scanned documents and encoding the remaining black dots using 'dithering' techniques.

The advanced software the system was to use would incorporate all existing techniques and the resulting images would then have to be 'super-compressed', by a special technique which ADONIS developed. Although the initial goal of 15,000 pages on a CD-ROM could be achieved, the print quality of the pages did not meet ADONIS' requirements; the quality, although readable, was not found to be good enough for the illustrations and smaller type sizes. Eventually a hybrid technique, offering good text and excellent halftone quality gave storage of 10-12,000 pages on a single CD.

ADONIS is giving high priority to the on-going development of even more sophisticated compression techniques, which will further increase the number of pages that can be stored on a CD-ROM disc.

Production Software

ADONIS converts paper documents to CD-ROM and the volume of paper involved is huge. To meet the required turnaround of less than three weeks the production process had to be automated. The nature of print

products, however, does not really make an intelligent automated approach easy; articles are continued on pages out of the normal sequence, advertisements break up the page order, etc.

Just scanning all pages in sequential order would therefore result in a chaotic jumble of images and sorting these out by visual recognition, in view of the sheer volume, would be extremely error prone. To get around this problem the software produces an intelligent scanner control mechanism to differentiate between text only pages and those containing halftone illustrations. This is matched with the index data (which allows the retrieval of pages from CD-ROM). Errors will be identified by software with matching capabilities and, if need be, a manual correction is made to the file. The printed pages are linked to the software by a sheet containing a bar-code.

The compression techniques developed by ADONIS require very powerful and dedicated equipment; testing this at the scanning bureau resulted in an average production time of more than a minute and a half per page which was clearly much too slow. The solution was to take the magnetic tapes containing the scanned page images and to 'supercompress' the data. The resulting 'supercompressed' images are then sent for disc manufacture.

Errata handling

When a publisher issues an erratum the page images of the original article are retrieved and the reference in the index is deleted. The erratum slip is then scanned and the image combined with the original article images and both sets are written to the next disc. A new index entry is also prepared. In this way a time path is created between the original article and the erratum.

Software limitations: The DOS limit

Virtually every company and library make use of PC's (the 'standard' computer) which operate under MicroSoft's DOS, which stands for Disk Operating System.

Although DOS is by far the most commonly used PC system, its restrictions in memory, its relative slowness, and its inability to simultaneously perform different tasks make it less than an ideal system for running optical image systems. The amount of optically stored data to be manipulated by a computer is by its nature vast.

DOS requires various additional programs to handle large files and CD-ROM discs which, together with other software needed for the monitor and CD-ROM player etc., occupies a considerable amount of the PC's memory. The DOS so-called 640 Kbytes base memory limit (DOS cannot address more than 640 Kbytes) proved to be a formidable barrier to overcome. By very structured programming and making use of all the tricks in the DOS book, sophisticated software has been produced, running within the 640 KBytes base memory limit. This software even allows users to search and retrieve articles during the printing of documents. However, it does have the trade-off that when a user performs searches whilst printing documents the system will slow down slightly.

Scanned images of pages require a large amount of memory and therefore need to be compressed. To access or print the physical contents the image has to be decompressed. The single processor environment in which DOS operates requires an extremely heavy use of the Central Processing Unit (CPU), the 'heart' of a Personal Computer. Only very fast CPU's are really suited for the complex task of displaying and printing scanned (bitmapped) images. The service can run on both fast and slower PC's, although performance limitations forced ADONIS to abandon support of the first generation of PC's.

Hardware

A more serious problem was found, unfortunately inherent in the common computer architecture; the speed of printing. A normal laser printer is

connected with a PC over what is known as a parallel or serial port. These ports are configured to 'communicate' with a printer at a set transfer speed, using a thin cable. Laser printers can accept data quite fast and the optimal setting is usually more than adequate for regular text to be printed out. However, ADONIS page images are enormous, compared to normal wordprocessor files, and printing files with a normal printer setup can take anything from 50 seconds to 3 minutes for a single page.

Clearly this is not acceptable for normal document delivery usage and although ADONIS' aim was to develop an entirely software driven service, this limitation is embedded in the computer topology, a hardware problem which cannot be solved by software; a concession had to be made.

Various printer 'boosters' were tested and a selection was made. An interface known as a video card (Tall Tree Systems' type JLaser 5 or 6 was selected) allows the software to obtain printing speeds of between 4 pages a minute depending on the PC (the more powerful, the faster) and the laser printer used.

Current Developments

The end of 1991 hailed two major breakthroughs for the ADONIS Document Delivery Service software; Network Support and Jukebox capabilities, both of which are major requirements at a large number of (potential) ADONIS users. The list of titles is growing as more and more titles have been and are still being added: 385 to date.

Networking

A multi disc CD-ROM product has to meet certain criteria; access to the information should be easy and guaranteed. Currently many libraries are connecting computer and information resources over networks; software and services manufacturers are developing products to support these networks. For ADONIS this was another reason for keeping the software as compact as possible. To accommodate digitised page images of articles, delivered on CD-ROM, a network must be fast enough to handle large chunks of data. ADONIS supports the almost de facto standard Novell network products, and will run on most networks with so-called NetBios compatibility.

Fast printing

One of the positive side-effects of networking was that ADONIS could finally meet the urgent requirement by its high volume users to speed up printing. Although most users find 4-7 pages per minute is satisfactory, the large institutions making tens of thousands of photocopies per year wanted ADONIS to support high-speed laser printers. Even high speed laser printers have difficulties in coping with the vast data flow which is transferred from the PC if a print is made from the System. However, certain high-speed printers can be configured to act as so-called network printers. With the necessary hardware modifications (interfaces based on the Ethernet protocol), a printer can be equipped with an Ethernet card. This means in practice that high-speed laser printers, such as the Hewlett Packard Laserjet IIIsi can print ADONIS articles at a rate of 10-12 pages per minute. This is fast enough to accommodate the speed requirements of the high volume users.

Finally a jukebox-that-works

45-50 CD-ROM discs per year require a changing device, or jukebox, in a network or even in a stand-alone environment. It is very inconvenient for a human operator to change discs every time the system asks for a disc change. In view of the relatively large number of discs per year, today's commonly used 'stacked' 4-8 drives will not do.

Although in the past a number of jukeboxes were introduced, the manufacturers met great difficulties and after a few initial tests they usually disappeared from the market. This year,

finally, a few jukeboxes have either entered the market, or are in an advanced state of testing.

ADONIS is already supporting the INCOM CDR-100 jukebox, a small and relatively inexpensive but sturdy changing device, capable of holding up to 100 CD-ROM discs. One can even link 7 of these jukeboxes with each other, allowing for a total storage capacity of 700 discs ! Other suitable machines will be supported if the quality and durability requirements are met.

Functionality

Subscribers to the ADONIS Service have repeatedly requested the addition of keyword and subject searching to the functionality of the system. Although primarily a document delivery service, it was felt that the functionality would benefit from added search facilities. The costs of adding these features are now being assessed and it is expected that subject and keyword searching possibilities will be added.

Naturally a far-reaching concept such as ADONIS leads to other possibilities and the members of the ADONIS group are implementing studies to fully integrate current awareness (on-line databases etc.) with the services offered by ADONIS. This is seen as an essential step towards a total current awareness service coupled to the individual article supply environment.

Benefits of Using ADONIS

Given the improvements in software functionality, the service can now be integrated in virtually any library environment. Indeed a number of ADONIS subscribers are currently implementing the service in their local library networks, providing access to the indexed information to their users. Others, such as the British Library Document Supply Centre are integrating the service in their automated document order fulfilment systems.

Although the majority of subscribers still use the ADONIS system as a stand-alone service from the library, the latest features introduced have convinced them of the possibilities of providing on-line access to the service, enhancing their library's role as information provider offering the combined possibilities of access to and retrieval of scientific information.

Other advantages can be identified: the cost efficiency, the advanced library management tools the service offers and its adaptability to the demands of library end-users, which may differ from organisation to organisation, but which can be steered by the information professional, something that was already apparent during the trial period.

Cost efficiency

Locating an issue of a journal containing the requested item and making photocopies when found, are time consuming tasks. If a library does not subscribe to the journal the author can be approached for a reprint or, more commonly these days, can order a photocopy from a third source and put up with the time interval from request to actual delivery.

Both situations are not only time-consuming but costly; automated user-friendly systems save staff time (up to 50% compared with photocopying), which will generally increase productivity.

Other elements adding to general library cost are storage space and binding: the hard copy of the — to this date 385 — journal issues occupy more than 100 metres of precious shelf space, the same data captured on 40 discs requires only 0.35 metres (35 centimetres), including their slim-line jewel casings. Naturally the workstation itself takes up space, but no more than a photocopier and with the prices charged for renting or buying real estate, cost savings are considerable. Although binding costs differ from location to location, they do contribute to the pressure on budgets. Binding can, however, be replaced with an automated system: ADONIS.

Library Management Tools

The use of the service is monitored by a statistics program, which updates a file each time an item is printed or transmitted over a network. The resulting encoded file, which has to be submitted to ADONIS for decoding and charging purposes, can be linked to the index database. A list of items printed-out is made available to users. It contains information on the number of copies printed per journal, per subscriber, in a given quarter. For those libraries who use the cost centre principle for charging back, a field in the user interface can be completed with the relevant information. This allows an even further subdivision: per journal, per subscriber and per cost centre within that subscriber.

An analysis of this information will greatly help in determining the use of information at the journal level, providing a reliable selection mechanism for subscriptions management.

Library Integrity

The direct access to (archival and current) information greatly enhances the overall performance of a library. The ideal of instantaneous access to the actual articles and a concept referred to as CAS/IAS or Current Awareness Services (CAS) linked to a form of automated Individual Article Supply (IAS) is emerging through the use of this new technology. In this concept a seamless integration of on/off-line reference database and articles in digital form, retaining their original format, should offer a one-stop shopping possibility to search, access and retrieve the document(s) required. This concept is now finding a widespread acceptance, especially since the individual article supply component is now established. Admittedly, the titles covered by the ADONIS Service do not satisfy all requests, but a substantial number can be satisfied. It is estimated that the journals already in the ADONIS Service would have satisfied 2% of the total request traffic of the British Library at Boston Spa on an annual basis. With the addition of more titles the overall number of requests satisfied will improve. We feel that the service is a leap towards the Digital Library, as it will assist publishers and librarians with the development of electronic (customized) journals.

Benefits to Publishers

To make a concept work publishers must naturally benefit equally. The statistical information obtained through ADONIS is already improving their insight into the use of their material. Bibliometric research often indicates where a competitive edge may be and in this respect the rapid monitoring of the actual use of scientific publications on an article basis is a valuable tool for enhancing the quality of journals. New trends and demands by the end-user of the information can be identified. The pay-on-demand principle of ADONIS will create a revenue stream for publishers, who feel the pressure mounting from third party use of their material. Eventually the experience gained from participating in ADONIS will assist the evolution of the traditional paper product to other, electronic formats.

Although paper journals will be here for a long time, the repackaging and customising of the information they contain is becoming more important and technically possible. The last step, offering articles in a variety of additional formats, in view of the funding of the development costs involved is not easily taken by individual publishers, but becomes possible through the open-ended nature of ADONIS.

Future Technical Developments

The scanning of articles is a pragmatic approach; accepting the contents of a journal in computer readable format would allow ADONIS to increase dramatically the number of pages on a CD disc, as well as adding valuable tools for literature researchers to be able to search full text.

Unfortunately publishers use different typesetters, who in turn use their own, often proprietary, software and text coding mechanisms.

For obvious reasons, the conversion of all different formats to produce a common code which on displaying and printing should exactly represent the pages as they are or will be published, is not yet logistically and financially viable. The ever-growing demand for the rapid publication of papers has triggered the development of a new computer language: the Standardized Graphics Mark-up Language (SGML). This language allows the coding of a paper with the necessary lay-out and other instructions at the time of input (often by the author).

Eliminating a few steps in the production process of a journal, including keyboarding, will speed up publication.

If enough publishers adopt this standard ADONIS can start accepting tapes, containing articles in SGML format together with software embedded codes to simulate fonts etc. at printing. Scanning will be needed only for (halftone) graphics. The physical page make up and the problems of columns, hyphenation and other, more general aesthetic properties of pages can be handled by software. The end result, at printing, will be a true regeneration of the article as it (will) appear(ed)s in the print journal issue. Initially this will result in a hybrid ADONIS Service, part of the articles on CD-ROM will be in scanned format, part in computer readable (SGML) form.

What the near future will bring

ADONIS currently concentrates on covering the subject area where most copies are demanded, namely biomedicine. In response to the demand from customers other areas such as chemistry will be added shortly, in line with the library holdings of our initial audience: researchers in the pharmaceutical industry and larger academic libraries.

ADONIS maintains close contact with the various parties in the scientific community to assess the need for additional periodicals and to monitor the validity of the core list.

Conclusion

The ongoing publisher financed developments in making information more accessible and retrievable indicate their commitment to further, and support, the true digital library. It is felt that although paper journals will continue to play a major role in the foreseeable future, better accessibility and retrievability of information is becoming more important. Ultimately, repackaging information in different combinations at the time of retrieval and leading to knowledge engineering by demand/supply packaging will become available.

End-users are fully aware of the possibilities information technology offers today and are demanding to be kept fully aware of what is happening in their fields. They wish for an ever faster, on demand and instantaneous delivery of copies of articles which they need for their research and study purposes.

For an increasing number of librarians and end-users ADONIS is the perfect solution.