Experiences with ExtraMED, an electronic full-text biomedical journal collection on CDROM

Dirk Schoonbaert

Assistant Librarian, Institute of Tropical Medicine Prince Leopold, Nationalestraat 155, 2000 Antwerpen, Belgium

Abstract: ExtraMED is an electronic full-text library distributed on CDROM, incorporating over 200 biomedical and health-related journals. Unlike Adonis, for example, ExtraMED does not specialise in high-impact journals, but focuses on local or regional biomedical journals from developing countries, most of which are not indexed by the major indexing services. Especially for specialties such as tropical or traditional medicine, this constitutes a valuable source of additional information. Also, as a bibliographic database ExtraMED is a worthwhile supplement to MEDLINE, but obviously not an equivalent alternative. The annual subscription fee is £2000 and includes 12 monthly CDROMs. Each disc contains about 8000 pages, so ExtraMED provides an estimated 100,000 pages a year. The practical advantages of this type of electronic document delivery are discussed, as well as the actual coverage and relevance. Though ExtraMED is certainly a laudable initiative, there are ample indications that the producers may have underestimated the problems of its practical realisation. The project suffers considerable delay, and as only six CDROMs were published in over 20 months, the contents still present a somewhat chaotic picture with lots of gaps. Fortunately things are getting gradually better. The retrieval software is generally adequate and includes a few remarkable features such as powerful truncation, proximity searching and automatic synonym look-up. Display and printing capabilities are good, though somewhat cumbersome in the latter case.

1. Introduction

ExtraMED is an electronic full-text library distributed on CDROM, incorporating over 200 biomedical and health-related journals. Unlike Adonis, for example, ExtraMED does not specialise in high-impact journals, but focuses on local or regional biomedical journals from developing countries, most of which are not indexed by the major indexing services. Especially for specialties such as tropical or traditional medicine, this constitutes an interesting source of additional information. The practical advantages of this type of electronic document delivery are obvious: a vast number of references can easily be retrieved (bibliographic database) and the articles themselves, which in their printed form are generally not easy to come by, are now immediately available (full-text images). Because of the CDROM format, the information is far less bulky than the paper versions and the management overhead is minimal, compared to subscribing to each individual journal separately (and claiming missing issues). ExtraMED is produced by the London based company Informania Ltd. It originated as a project of the World Health Organization (WHO), which for many years has been active producing regional supplements to MEDLINE (WHO 1994). ExtraMED is also supported by other international bodies such as UNICEF. The annual subscription fee is £2000, for which you receive 12 monthly CDROMs. Each disc contains about 8000 pages, so ExtraMED provides an estimated 100,000 pages a year. extraSCI and AgROM Extra, comparable series in cooperation with UNESCO and FAO, have already been announced.

2. Coverage

As librarians at an institute of tropical medicine, my colleagues and I were obviously interested in ExtraMED, and subscribed to it when it was first announced. Although it is essentially a great initiative, I have mixed feelings concerning its practical and technical realisation.

After initial discussions, ExtraMED started processing its first CDROM early in 1994. More than 20 months after subscribing, I had received only six discs. Publication pace has recently started to improve, so ExtraMED may soon reach cruising speed.

A survey of the number of megabytes and articles on each disc is included in Appendix 1. One obvious conclusion is that the CDROM discs are not filled to their potential. Undoubtedly this is partially due to the delay in the reception of original material and the pressure to publish more or less regularly. As each year 12 CDROMs are scheduled to be added, and disc swapping should be avoided as much as possible, I hope this filling ratio will soon be optimised.

At the start of the project, some 200 journals had agreed to join the ExtraMED Consortium, with many others to follow, including 90 newsletters re-
"ExtraMED can only aim to supplement the bibliographic MEDLINE information and can certainly not replace it"
This obviously does not improve the reliability of searches, nor does it inspire a great deal of confidence.

3. Software

ExtraMED uses Idealist retrieval software, developed by Blackwell Scientific Publishers. I am using the Windows version; a Windows NT version also exists and a Macintosh version was announced. In order to run smoothly, a 32-bit Windows expansion, which is provided with the program, needs to be installed first. During the actual software installation, four different icons are created: ExtraMED (the database program proper), CD-Manager (to install new CDROM discs), User registration (to allow the system administrator to define a number of user names, which may be linked to a password and a 'cost centre'), and User statistics (to log the actual use of each journal, user and cost centre). This valuable log information is a fine feature that would not be out of place in any sophisticated database program.

During installation of new discs, all bibliographic information, indexes and abstracts are copied to the hard disk and as such this can be used as a database by itself. The actual page images remain on the CDROMs, and only need to be retrieved for actually viewing the full-text articles. This splitting up of bibliographic and full-text information is a good option, allowing fast retrieval, but it should also be borne in mind that as the indexes copied to hard disk for the first six CDROMs take up some 12 Mb of hard disk space, probably over 20 Mb will be needed for each year of ExtraMED.

Though the system is now working fine, installation of the ExtraMED software was not without problems. Originally, most of the full-text images could not be displayed on screen, resulting in an error message. Contrary to earlier specifications, the 4 Mb RAM of my installation proved insufficient. This appeared to be the fault of Microsoft, which had underestimated the memory needed to run its own 32 bit Windows expansion. Adonis is reported to have encountered the same problem in the beginning. Upgrading to 8 Mb RAM solved this problem. For one thing, this clearly implies that ExtraMED needs a relatively powerful PC to run smoothly, which may partially compromise its suitability for efficient widespread distribution of biomedical information in developing countries. Of course, yesterday's high-end PCs are today's entry-level machines, and most present-day software gets ever more memory hungry, so this may not be an objection after all, provided that developing countries can and will follow this dazzling pace.

After installing a program upgrade, the CD-Manager failed to recognise the ExtraMED discs in the 'D:' drive. Trying to remedy this problem led to many hours of frustration, at the end of which I was ready to give up. Eventually, Informia found out the cause: in my Windows installation 'English (International)' was not installed as language in the international settings of the control panel. Changing this single (external) parameter solved the problem.

4. Retrieval

As ExtraMED is a Windows application, most functions are available via buttons, ALT menus and shortcut (function) key combinations. Online help is available from any screen.

Retrieval is fast and can be highly differentiated: practically all fields can be searched, including volume number, page ranges, number of pages and so on. The ExtraMED interface offers two basic templates: 'searching all fields' and 'searching by field'. In practice, both allow global and field-specific searching, including a number of fields in groups, such as 'text fields' or 'numeric fields'. The major difference is that the first incorporates natural language phrases (for example 'containing the word . . .'), while the second uses the more abstract standard Boolean operators AND, OR and NOT in a step by step approach. Straightforward retrieval using a command mode is also available in both (for example 'YPB=1993' to limit the search to 1993 as publication year), but the Boolean operators are recognised as such only if they are entered between brackets (for example 'africa [AND aids]'). Comparative searching with operators such as 'greater than', 'greater than or equal to', 'less than' and 'less than or equal to' is also available. Interestingly, fields can be selected on whether they actually contain data or are empty. Various degrees of proximity searching can be specified: search terms appearing within the same paragraph, within the same sentence, or a specific (maximum) number of words apart. Rangings can be both numerical and alphabetical. Searching is case independent and truncation capabilities are impressive: left hand and right hand truncation pose no problem, and even masking within a word is executed very fast.

Different retrieval templates can be defined with the 'Search-U-Like' feature. The two standard templates feature automatic index browsing with the possibility to select index items as search terms. This index itself is not field-specific (though the 'searching by field' template highlights the index terms if they are present in the specific field demanded) and includes only individual words. For some fields, such as 'journal name', an index of full field contents would have been more helpful than just the constituent words.

The software is generally adequate to find the information 'as-it-is' in the database. Finding a specific article is fast and easy. Searching for subject information, however, is basically limited. Since no keywords are included, only terms explicitly included in the title or abstract field can be retrieved. This lack of keywords also precludes a good thesaurus with powerful

"The actual page images remain on the CDROMs, and only need to be retrieved for actually viewing the full-text articles"

"Probably over 20 Mb will be needed for each year of ExtraMED"
“The ‘synonym’ list can be customised and the feature can be activated as a system default or turned off”

‘explode’ facilities (cf. MEDLINE’S MESH), and obviously reduces precision and recall. This lack of structured keyword searching is partly overcome by a few interesting additional techniques: a synonym list activates supplementary search terms. For example, ‘malaria*’ yields additional hits which do not contain the term ‘malaria’, but are triggered off by terms like ‘mosquito’, ‘falciparum’ or ‘chloroquine’. The ‘synonym’ list can be customised and the feature can be activated as a system default or turned off.

A comparable retrieval option is called ‘lookalikes’ or ‘soundalikes’ and is meant to neutralise spelling errors. In practice, this implies that ‘today’ and ‘toothpaste’, for example, are closely related. It is not clear if and how results from this function can be modified. This feature may prove helpful at times, but it is a good thing it is optional and not a system default.

The system uses a stopword list. This list can be modified but only from outside the program, for example using Windows Notepad. This is not a very user-friendly procedure and, more seriously, the changes made are not reflected in subsequent sessions. Just as before, the shortlisted words keep popping up in the retrieval results. Obviously, the index files themselves cannot be modified, as they are copied from the CDROM discs. Yet it might still be possible that the software ignored stoplisted items from this index, but this does not appear to be the case.

The active hit list can be modified (using the Boolean operators OR, AND, NOT or their natural language equivalents ‘widen’, ‘narrow’, ‘exclude’), but the subsequent search formulations are not kept as separate sets. So it is not possible to get a survey of all search formulations used in the present session, or to combine a set 3 with a set 6. On the other hand, the most recent search formulation can always be recalled, even if it was executed weeks before. It is possible to save search histories for use in later sessions. Individual hit lists can also be saved.

5. Display

ExtraMED features several display levels. First, the ‘overview hit list’ shows one line per record, consisting of author(s), year of publication and title. For the titles, upper and lower case is used indiscriminately, which gives a rather sloppy appearance. The actual combination of fields can be customised, but only from outside the system, changing an initialisation file in the Windows directory. According to the comprehensive manual, the order between individual records can be self-defined, but I could not find how to achieve this. It appears that this Idealist standard feature was deliberately omitted from the ExtraMED package, but may return in a future software upgrade (Trimmer 1995). At the bottom of the ‘overview hit list’, the search expression is listed in command mode. The screen layout can be customised: for example you can define what bars are included on the screen. Records can be marked (or dropped) while viewing the ‘overview hit list’. This subset can then be selected as an active ‘hit list’.

Full-record display features the complete bibliographic information of each record, including an abstract. The fields included and their relative order can be changed by defining one or more alternative display sheets. Hit terms are highlighted automatically. Unlike the overview hit list and the full-record display, full-text display does not use machine readable text but displays images (in TIFF format) of the articles themselves, which are retrieved individually from the CDROM. This works very well, although the result is ultimately dependent on the quality of the scanning of the original. At first a full-page view is displayed in the window (which itself can be sized up to fill the whole screen). From this full-page image, at least two subsequent enlargements can be activated. If the text required is not on the active disc, in most cases a message prompts you to load the right disc (e.g. ‘cannot open d:'ad940001.img’), but this message could certainly be formulated in a more user-friendly fashion.

6. Printing and exporting

Hit lists and bibliographic references can be printed during retrieval or display. Printing the full-text articles is somewhat more complicated, as the articles selected are first stored in a print queue from which actual printing is generated afterwards. I experienced some problems with this. When I did not specify individual printer settings for the headers, the hit lists and the text itself and relied on the (correct) default values, the system invariably claimed it could not print because of ‘insufficient printer information’. The headers — which include bibliographic information, credits and copyright indications, but also management information such as to which centre costs should be accounted — are certainly helpful in larger libraries or network environments. But it should be equally possible to ignore them, as in standalone situations they can be a waste of time and paper. Hit lists, full records and even TIFF images of full-text articles can also be downloaded to a file on disc. For libraries, it is interesting that ExtraMED works on a flat fee principle: no royalties or other extra costs are involved for printing and downloading.

7. Conclusion

ExtraMED features the full-text images of over 200 biomedical journals from developing countries at a average price of
£10 per journal. As such, it is certainly an interesting electronic library for the study of tropical medicine, though not necessarily an essential one, as the major international journals are not included and the actual selection is debatable. At this time ExtraMED still suffers substantial delay, and as only six CDROMs were published by the end of 1995, the contents still present a somewhat chaotic picture. The producers may have underestimated the problems of its practical realisation. As things are gradually getting better, these are probably growing pains which will soon be overcome. Also, as a bibliographic database ExtraMED is a worthwhile supplement to MEDLINE, but obviously not an equivalent alternative. The retrieval software could still be improved upon, and a few things that were promised have failed to materialise. Yet it is generally adequate for retrieval and includes a few remarkable features such as powerful truncation, proximity searching, and automatic synonym activation. Display and printing capabilities are good, though somewhat cumbersome in the latter case.

Post Scriptum

By the end of March 1996, a total of 10 discs, featuring nearly 10 000 articles, had been published.

References


Appendix 1: analysis of contents in ExtraMED discs 1–6

<table>
<thead>
<tr>
<th>Disc</th>
<th>94/1</th>
<th>94/2</th>
<th>95/3</th>
<th>95/4</th>
<th>95/5</th>
<th>95/6</th>
</tr>
</thead>
<tbody>
<tr>
<td>CDROM memory actually used:</td>
<td>237 Mb</td>
<td>430 Mb</td>
<td>363 Mb</td>
<td>370 Mb</td>
<td>279 Mb</td>
<td>314 Mb</td>
</tr>
<tr>
<td>Articles (cumulative)</td>
<td>1062</td>
<td>1683</td>
<td>2765</td>
<td>3685</td>
<td>4504</td>
<td>5455</td>
</tr>
</tbody>
</table>

Appendix 2: Distribution of publication years

<table>
<thead>
<tr>
<th>Disc</th>
<th>94/1</th>
<th>94/2</th>
<th>95/3</th>
<th>95/4</th>
<th>95/5</th>
<th>95/6</th>
</tr>
</thead>
<tbody>
<tr>
<td>1995</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1994</td>
<td>0</td>
<td>608</td>
<td>490</td>
<td>679</td>
<td>795</td>
<td>816</td>
</tr>
<tr>
<td>1993</td>
<td>405</td>
<td>420</td>
<td>826</td>
<td>1256</td>
<td>1660</td>
<td>1964</td>
</tr>
<tr>
<td>1992</td>
<td>497</td>
<td>495</td>
<td>785</td>
<td>988</td>
<td>1188</td>
<td>1667</td>
</tr>
<tr>
<td>1991</td>
<td>54</td>
<td>54</td>
<td>94</td>
<td>117</td>
<td>180</td>
<td>221</td>
</tr>
<tr>
<td>1990</td>
<td>12</td>
<td>12</td>
<td>97</td>
<td>121</td>
<td>121</td>
<td>167</td>
</tr>
<tr>
<td>1989</td>
<td>15</td>
<td>15</td>
<td>46</td>
<td>69</td>
<td>78</td>
<td>91</td>
</tr>
<tr>
<td>1988</td>
<td>42</td>
<td>42</td>
<td>56</td>
<td>69</td>
<td>86</td>
<td>86</td>
</tr>
<tr>
<td>1987</td>
<td>29</td>
<td>29</td>
<td>43</td>
<td>58</td>
<td>68</td>
<td>115</td>
</tr>
<tr>
<td>1986</td>
<td>8</td>
<td>8</td>
<td>34</td>
<td>34</td>
<td>34</td>
<td>34</td>
</tr>
</tbody>
</table>

No publication year

| | 0 | 0 | 294 | 0 | 0 | 0 |

Appendix 3: Informania Ltd

ExtraMED
Informania Limited
PO Box 1359
London W5 3ZP
UK
Fax: +44 (0)1730 265398
The author

Dirk Schoonbaert

Dirk Schoonbaert graduated from the University of Antwerp (Germanic Philology 1980, and Library and Documentation Science 1985), and has worked in several scientific libraries. Since 1986 he has been Assistant Librarian at the Prince Leopold Institute of Tropical Medicine in Antwerp. His professional interests include the electronic aspects of information management and retrieval.