

Pierre-Marie Belbenoit-Avich

Bibliothèque de l'Université Lyon I

LES BASES PLEIN-TEXTE BIOMÉDICALES ET LA FOURNITURE DE DOCUMENTS

C'EST DEVENU un lieu commun de dire que les nouvelles technologies ont profondément modifié notre profession. Les bases en ligne avaient complètement transformé le processus de recherche bibliographique, tout autant que le font aujourd'hui les CD-ROM¹. Et nous vivons actuellement la même révolution que celle qui avait, il y a peu, accompagné l'informatisation. Mais s'il est un domaine qui est également en profonde mutation, c'est celui de la fourniture de documents. On peut se poser la question de savoir si les bases plein-texte ne vont pas constituer une nouvelle voie d'accès au document primaire, permettant à l'utilisateur de ne plus recourir aux bibliothèques.

Dans une première partie, on étudiera les bases plein-texte générales puis celles qui se rapportent au domaine biomédical. Enfin, on essaiera de voir si, à l'avenir, les bibliothèques seront abandonnées au profit de ces bases, ou quelle part elles pourront conserver.

Les bases plein-texte générales

Nous allons d'abord tenter une définition de ces bases plein-texte en expliquant leurs avantages, avant d'étudier pour les bases générales leur multiplication, leur diversification et enfin leurs améliorations.

Définition

Nature : les bases plein-texte, comme leur nom l'indique, ne donnent pas simplement des références bibliographiques, même si elles sont parfois couplées aux bases bibliographiques ; ainsi sur BRS², le fichier CCML (Comprehensive Core Medical Library) est-il lié soit à *Medline*, soit à *Excerpta Medica* (30). Certes, on peut inclure sous ce terme les bases factuelles type RTECS, ou plus récemment GENE-BANK créé par la National Library of Medicine pour tout ce qui concerne les séquences génétiques (38). Mais, d'une manière précise,

une base plein-texte fournit le texte intégral des documents. Elle peut correspondre seulement à un livre, telle celle développée par la National Agricultural Library des Etats-Unis, et c'est encore le cas de tous les dictionnaires et encyclopédies ; il peut aussi s'agir d'une collection complète, tel l'ensemble de vingt-cinq CD-ROM élaboré par l'International Agricultural Research de la Banque mondiale et qui correspond à 5 500 livres ou rapports. Il peut enfin s'agir de périodiques, tel le fichier en ligne de TEXTLINE qui contient le texte d'articles parus depuis 1980 dans un certain nombre de périodiques internationaux majeurs.

Matériellement, ces bases peuvent être accessibles en ligne. Nous ne citerons parmi beaucoup d'autres que le périodique *News Analyst* qui insère les nouvelles dès leur arrivée, ou surtout l'ensemble des sources plein-texte accessibles sur Dialog, au nombre de 2 000. D'un autre côté, beaucoup de ces bases ne sont disponibles que sur CD-ROM, pour des raisons de coût évidentes, tels le US History CD-ROM, qui contient le plein-texte de plus de cent ouvrages sur l'histoire des Etats-Unis. Quelques autres sont également

1. CD-ROM : Compact-Disc Read Only Memory

2. BRS : Bibliographical Retrieval Services

accessibles sur disquettes, en particulier le London Dysmorphology Database ou le London Neurogenetics Database, qui identifient chacune 2 000 pathologies et possèdent chacune plus de 10 000 références bibliographiques.

Avantage : les bases plein-texte, qu'elles soient ou non du domaine médical, présentent beaucoup d'avantages, d'abord en ce qui

s'alimentant à plusieurs sources. Sur ces 2 000 sources, 549, soit 28 %, avaient été ajoutées en un an. Et les bases plein-texte représentent un pourcentage de plus en plus important de l'ensemble des bases. En 1988, il avait été calculé que dans le Cuadra Elsevier Directory of Online Databases, 32,4 % étaient de type plein-texte (5) ; récemment, en 1992, lors d'un simple sondage local, 10 %

en ligne (18, 28), et il peut y avoir interactions ou interrétions de lecteurs autorisés (18).

On le voit les bases plein-texte se développent dans l'ensemble des domaines.

Les bases plein-texte biomédicales

Les bases plein-texte médicales sont parmi les premières à être apparues au milieu d'un foisonnement d'espoirs variés. En 1984, le serveur BRS a créé CCML qui fournit le texte de plusieurs centaines de produits analysés dans l'*Index Medicus* et maintenant *Excerpta Medica* (33). Le serveur MEDIS a vu le jour peu après (49). En 1985, l'American Medical Association lui a donné licence pour tous ses périodiques scientifiques (24), deux ans après que l'éditeur IRCS ait commencé à faire paraître ses périodiques simultanément sous forme papier et en ligne (27).

Aujourd'hui les bases plein-texte couvrent plusieurs types de documents puisque cela va des fichiers de Dialog, consacrés à un périodique uniquement, tel le *New England Journal of Medicine* ou le *Morbidity and Mortality Weekly Report* ou à une collection de plusieurs titres, ainsi le fichier 442 qui représente l'ensemble des revues de l'American Medical Association. Mais, naturellement, il faut citer MEDTEXT, qui – sur Dialog – donne, comme son nom l'indique, le texte intégral d'un certain nombre de périodiques médicaux depuis 1982, avec mises à jour mensuelles, soit 55 000 références. On peut même y faire une recherche sur les références citées depuis 1990. Nombre de titres ont été mis sur CD : le *Journal of the American Medical Association* et autres titres du même éditeur, ou encore par exemple *Pediatrics in review*...

D'autres périodiques sont également concernés, en particulier ceux du domaine de la recherche, tout simplement parce que la publication électronique veut réduire les délais d'impression toujours longs dans la forme papier. Dans ce créneau des journaux de recherche, on peut don-

Donner aux lecteurs l'information qu'ils désirent immédiatement, plutôt que de les envoyer vers un service de prêt entre bibliothèques

concerne le prix puisque, d'après Barwick et Cornish du Document Supply Centre, une étude a montré que les journaux électroniques sont moins chers que la forme papier (4). D'autre part, et c'est bien évident, les bases sur CD économisent de la place. Un seul CD contient cinq années du *Journal of the American Medical Association*, qui occupe pour la même période deux mètres de rayonnages, etc. Enfin, et nous y reviendrons lorsque nous parlerons du lien des bases plein-texte et de la fourniture de documents primaires, ces bases s'insèrent dans un mouvement qui veut donner aux lecteurs l'information qu'ils désirent immédiatement, plutôt que de les envoyer vers un service de prêt entre bibliothèques (43). D'autant que, dans tous les cas, les logiciels ont été améliorés soit pour être rendus compatibles (22), soit pour être utilisés aussi bien par le novice que par l'utilisateur confirmé (23).

Evolution

Multiplication : apparues aux Etats-Unis vers 1984, les bases plein-texte ont vu leur nombre s'accroître de plus en plus vite (14, 17, 24, 26). Ces dernières années, l'augmentation a même été exponentielle. En 1991, Dialog proposait près de 2 000 sources plein-texte, ce qui ne signifie bien sûr pas autant de bases, une base

des utilisateurs d'une seule bibliothèque ont dit utiliser des bases plein-texte.

Diversification : non seulement le nombre de ces réalisations augmente sans arrêt, mais de plus elles se diversifient et il s'en crée dans tous les domaines. En 1992, par exemple, le producteur News Bank a lancé sur CD-ROM 34 périodiques des Etats-Unis d'importance nationale ou régionale et le répertoire *Full Text Sources Online Newspapers*, édition 1992, liste 125 périodiques quotidiens d'Amérique du Nord et au total 3 000 périodiques de toutes fréquences offerts en plein-texte. Dans le domaine littéraire, l'ensemble des CD intitulés *Cumulative Contents Index Humanities* offrira pour 600 périodiques leur sommaire ou, ce qui est mieux, leur texte intégral (13). Et on a déjà cité l'agriculture, ce qui touche à la banque mondiale, ou encore à la propriété intellectuelle ou aux mathématiques (15). C'est une vague en plein développement, qui ne fait que toucher l'Europe, alors qu'elle se répand aux Etats-Unis depuis près de dix ans (4, 9, 17, 21, 24, 27). Il s'agit donc d'un phénomène dont nous devons être très conscients.

Améliorations : celles-ci sont constantes. Le logiciel devient de plus en plus convivial, le délai de publication est parfois proche de cinq semaines (26), on peut se créer des profils d'intérêt dans le cas de bases

La recherche sur les bases plein-texte amène plus d'items pertinents mais en même temps moins de précision

ner les noms des publications éditées par Prous, en Espagne, comme *Drugs of the Future, Drugs of Today, Drug News and Perspectives, Drug Data Report* qui couvriront sous forme de CD la période postérieure à 1988, voire à 1985. Et on ne peut s'empêcher de citer le titre *On-line Journal of Clinical Trials*, lancé début 1992, qui ne sera disponible que sous forme électronique (28). Ce titre a été choisi précisément pour que les malades puissent bénéficier plus tôt des recherches et des découvertes. Les articles, d'après l'éditeur, seront de grande qualité graphique, avec courbes, illustrations, équations.

Bien sûr, on trouve dans ces bases plein-texte médicales, des ouvrages ou des ensembles d'ouvrages : par exemple *Drug Information Full Text*, l'une des bases plein-texte de pharmacie les plus importantes dans le monde. Elle fournit une information complète sur 50 000 produits commercialisés. Elle correspond à la forme papier de *American Hospital Formulary Service Drug Information* et du *Hanbook on Injectable Drugs* avec ajouts de mises à jour spécifiques. On peut aussi citer le CD-MAXX (Maximum Access to Diagnosis and Therapy) qui regroupe l'ensemble des textes d'ouvrages de références de la collection Little Brown manuals et d'informations de l'US-Pharmacopoeia Drug Information (33). Parlons enfin du Physician Data Query (PDQ). Créée par le National Cancer Institute des Etats-Unis, elle fournit une information courante, en plein-texte, sur le diagnostic et le traitement du cancer. Les auteurs de cette réalisation ont émis l'hypothèse qu'un accès direct et immédiat à PDQ briserait beaucoup de barrières pour l'accès à l'information en assurant une disponibilité continue (22).

Inconvénients

Ces bases sont très simples ; elles sont devenues très vite très faciles à utiliser. En ce qui concerne PDQ, pour lequel une étude complète a été réalisée, 87 % des utilisateurs ont trouvé le système simple et plus de la moitié étaient assez à leur aise après une demi-heure de maniement. Quant au gain de temps, il était souligné par 73 % des utilisateurs ; enfin, la disponibilité d'une telle offre est remarquable : 44 % des usagers se sont servis de ce système entre 18 heures et 6 heures du matin, soit quand la bibliothèque était fermée, cela grâce à des réseaux locaux.

Toutefois ces bases plein-texte peuvent poser des problèmes. Il est vrai que les bases plein-texte semblent avoir deux inconvénients sur lesquels on peut dire rapidement quelques mots. D'une part, lorsqu'elles sont en ligne, elles ne sont pas toujours mises à jour aussi vite que les bases bibliographiques correspondantes (30). En particulier, si on compare BRS / CCML ou MEDIS avec Medline, BRS est plus rapide que Medline seulement dans 18 % des cas et a une vitesse équivalente dans 52 % des cas ; MEDIS est plus rapide que Medline dans 28 % des cas et aussi rapide dans 36 % des cas. Si les bases bibliographiques sont parfois plus rapides (respectivement dans 30 et 36 % des cas), cela tient à la difficulté de saisir l'intégralité du texte d'un article.

Un autre inconvénient est à ne pas sous-estimer : dans le cas précédent, où une base plein-texte offre les mêmes ressources qu'une base bibliographique, sur laquelle des deux la recherche bibliographique est-elle plus pertinente ? Beaucoup ont écrit sur ce point (40, 41). Certains pensent que l'approche du plein-texte est différente selon que l'on est utilisateur final ou intermédiaire (11). Il semble que la

recherche sur les bases plein-texte amène plus d'items pertinents mais en même temps moins de précision, parce que les bases bibliographiques possèdent un thésaurus hiérarchisé que n'ont pas les autres (41).

Les bases plein-texte et la fourniture de documents

En fait, une double question se pose : pourquoi les bases plein-texte intéressent-elles l'utilisateur final, et que peuvent faire les bibliothèques ?

Intérêt pour les usagers

Les bases plein-texte sont un moyen important pour obtenir le document final, d'une part, parce que les périodiques qu'elles offrent sont les plus utiles. Les titres de BRS / CCML représentent pour notre bibliothèque environ 40 % du prêt-inter. D'autre part, les bibliothèques ne répondent pas toujours aux besoins des utilisateurs (43). L'inadaptation des collections de périodiques est soulignée par l'exhaustivité des bases bibliographiques. « *Nous avons fait de grands progrès dans l'accès à la littérature mais nous n'en n'avons pas fait autant en ce qui concerne l'accès aux documents, ce qui rend les lecteurs frustrés de devoir attendre* ».

D'autre part, les bases plein-texte sont disponibles continuellement. Ce qui est vrai de l'exemple de Physician Data Query, l'est aussi ailleurs (48). « *Pour le médecin qui n'a pas un accès facile à une bibliothèque médicale importante, l'accès immédiat à une source plein-texte est un avantage évident* ». C'est particulièrement sensible dans le domaine biomédical, et le plein-texte offre une réponse en temps réel.

De plus les bases plein-texte sont moins chères que la forme papier (4) ou que l'utilisation d'un service de prêt entre bibliothèques (34). Une étude a été menée à la Hong-Kong Polytechnic Library : il apparaît que les prix sont égaux ou inférieurs à ceux de la forme classique et qu'ils décroissent (25).

Le rôle des bibliothèques

Que peuvent faire les bibliothèques ? Les bases plein-texte sont-elles un danger ? Même si leur nombre augmente, tous les documents ne sont pas encore disponibles sur ce support et subsiste le problème des graphiques, des illustrations. Cependant le problème est important. L'Association of Computing Machinery, qui est spécialisée en informatique scientifique et universitaire, va mettre sur disque optique les dix dernières années de onze périodiques et 45 congrès annuels. Elle pense satisfaire 80 000 demandes au départ. Ce n'est qu'un exemple (46).

Si les lecteurs trouvent, dans BRS ou ailleurs, ce qu'ils veulent, ils ne vont plus utiliser les bibliothèques et le volume du prêt-inter va décroître. Déjà en France, beaucoup de laboratoires privés ont pris un abonnement à cette sorte de serveur. Partout où le prêt-inter induit des revenus, existe une menace. Bien plus, si les lecteurs trouvent le plein-texte, en ligne ou sur CD, ils ne vont plus utiliser nos collections de périodiques, sauf si – comme à la Tennessee University Knoxville Library – nous créons un système qui offre seulement le plein-texte des périodiques auxquels la bibliothèque n'est pas abonnée (20). Mais, soyons-en sûr, les bases plein-texte sont un des problèmes majeurs des années à venir. Si nous ne le résolvons pas, nos bibliothèques ne seront que les dépositaires des publications pré-électroniques.

Juillet 1992

BIBLIOGRAPHIE

1. **André, P.**, « Full text access and laser video discs : the National Agricultural Library System », *Library High Tech*, 1986, 4 (1), p. 13-21.
2. **Angier, J.J.**, « Use of the PDQ system in a Clinical setting », *Bulletin of the Medical Library Association*, 1990, 78 (1), p. 15-22.
3. **Barwick, M.M.**, « Interlending and Document supply : a review of the recent literature 19 », *Interlending and Document Supply*, 1991, 19 (1), p. 17-23.
4. **Barwick, M.M. ; Cornish, G.**, « Interlending and Document Supply : a Review of the recent literature 20 », *Interlending and Document Supply*, 1990, 18 (3), p. 106-110.
5. **Bianchi, G.**, *Biomedical libraries and full-text databases*, EAHIL 2nd Congress, Bologna 1988, p. 415-420.
6. **Bjorner, S.**, « Full text document delivery on-line : it makes sense », *On-line*, 1990, 14 (5), p. 109-112.
7. **Butler, M.**, « Full-text CD-ROM for international development », *Microcomputers for information management*, 1990, 7 (4), p. 273-291.
8. « CD-ROM newspapers not for the libraries », *Information World Review*, 1991, Oct., p. 21.
9. **Chandra, P.**, *Access to the full-text of nonfiction books : Design considerations from a study of public library users*, ASIS 1987, Proceedings of the 50th ASIS Annual meeting, p. 196-200.
10. **Collen, M.F.**, « Full-text medical literature retrieval by computer », *Journal of the American Medical Association*, 1985, 254 (19), p. 2768-2773.
11. **Consales, J.**, *End user access to full-text medical information : an Examination of searching behavior*, National on-line meetings Proceedings, New York 9-11 May 1989, p. 101-104.
12. **Desmarais, N.**, « Bureau of electronic publishing makes history », *CD-ROM Librarian*, 1991, 6 (2), p. 24-28.
13. « Electronic index to humanities journals », *Information World Review*, 1992 May, p. 10.
14. **Figley, M.M.**, « Publication in radiology : challenges to tradition », *Radiology*, 1985, 157 (3), p. 633-636.
15. **Flanders, B.**, « General periodicals on disc », *CD-ROM Librarian*, 1991, 6 (1), p. 27-32.
16. « From book to disc », *Information World Review*, 1991, Oct., p. 24.
17. **Franklin, J.**, *Biomedical Journals in an on-line full-text database : a Review of reaction to ESPL*, 7th International On-line information Meeting 1983, London 6-8 Dec. 1983, p. 407-410.
18. « Full-text meets real time », *Information World Review*, 1992, May, p. 9.
19. **Gallacher, K.E.**, « Full-text and bibliographic database searching in the health sciences : an exploratory study comparing CCML and Medline », *Medical References Services Quarterly*, 1990, 9 (4), p. 17-25.
20. **Gillikin, D.P.**, « Document delivery from full-text on-line files : a pilot project », *On-line*, 1990, 14 (3), p. 27-30.
21. **Gordon, D.W.**, *Acquiring full-text documents : the Information specialists ongoing problem*, National on-line meeting Proceedings 1984, New York, April 10-12, 1984, p. 67-76.
22. **Hall, S.**, « Many changes for newspapers discs », *Information World Review*, 1992, May, p. 10.

23. « Handbook equations come on-line », *Information World Review*, 1991, March, p. 14.
24. **Harris, D.K.**, « MEDIS : a new strategic option for acquiring medical information electronically », *Journal of the American Medical Association*, 1985, 254 (19), p. 2 801.
25. **Hawkin, (D.T.)**, « Document delivery gets personal », *On-line*, 1992, May, p. 106-108.
26. « Hypertext and images meet research needs », *Information World Review*, 1992, April, p. 30.
27. « IRCS speeds publication in powerful database links », *IRCS Medical Science*, 1983, 11 (3), p. 189-192.
28. **Kassirer, J.P.**, « Journals in bits and bytes », *New England Journal of Medicine*, 1992, 326 (3), p. 195-196.
29. **Markinson, O.F.**, « The electronic journal for dentistry », *International Dental Journal*, 1989, 39 (1), p. 17-24.
30. **McKinin, E.**, « Currency of full-text medical journals : CCML and MEDIS vs Medline », *Bulletin of the Medical Association*, 1991, 79 (3), p. 282-287.
31. **McKini, (E.)**, *A comparison of full-text and abstracts for information retrieval in clinical medicine*, National on-line meeting Proceedings 1989, New York, 9-11 May 1989, p. 295-303.
32. **McKinin, E.**, « The Medline full-text research project », *Journal of the American Society for Information Science*, 1991, 42 (4), p. 297-307.
33. **Miller, N.**, « Optical product review : MAXX, maximum access to diagnosis and therapy », *CD-ROM Librarian*, 1991, 6 (1), p. 34-37.
34. **Mok, W.M.**, *Using on-line full-text databases : the Future of interlibrary loan services*, On-line information 90. Proceedings of the 14th international on-line information meeting. London Dec. 11-12, 1990, p. 497-506.
35. **Nesbit, K.**, « BRS / Links to the future : an on-line hypertext is born », *On-line*, 1990, May, p. 34-36.
36. **Ojala, M.**, « The dollar sign : research into full-text retrieval », *Database*, 1990, 13 (4), p. 78-80.
37. « Optimal access in the electronic age », *ESP World*, 1992, January, p. 10-11.
38. **Pratt, G.E.**, « On-line retrieval of nucleic acid sequences from Genbank », *Database*, 1991, 14 (2), p. 102-104.
39. **Renschler, H.E.**, « Rationale arztliche fortbildung », *Schweizerische rundschau medizine praxix*, 1991, 80 (19), p. 515-523.
40. **Sievert, M.E.**, *A comparison of indexing and full-text for the retrieval of clinical medical literature*, ASIS 88 information technology : Planning for the next fifty years. Proceedings of the 51th annual meeting of the American society for information science, Atlanta, Oct. 23-27 1988, p. 143-146.
41. **Sievert, M.E.**, *Why full-text misses some relevant documents : an analysis of documents non retrieved by CCML or MEDIS*, ASIS 89 : Managing information and technology. Proceedings of the 52nd annual meeting of the American Society for Information Science, Washington DC, 1989, Oct. 30-Nov., 2nd, p. 34-39.
42. **Snow, B.**, « Keeping up to date with full-text FDC Reports », *On-line*, 1990, May, p. 94-97.
43. **Tenopir, C.**, « The Impact of Electronic References on references librarians », *On-line*, 1992, May, p. 54-60.
44. **Tenopir, C.**, *Users and uses of full-text databases*, On-line information meeting London, Dec. 6-8, 1988, p. 263-270.
45. **Tenopir, C.**, *Search strategies for full-text bases*, ASIS 88 Information technology : Planning for the next fifty years. Proceedings of the 51th Annual meeting of the American Society for Information Science, Atlanta, Oct. 23-27, 1988, p. 80-86.
46. « Ten years image library planned », *Information World Review*, 1992, April, p. 10.
47. « US papers discs proliferate », *Information World Review*, 1992, April, p. 5.
48. **Veenstra, R.J.**, « Access to information about AIDS », *Annals of Internal Medicine*, 1991, 114 (4), p. 320-323.
49. **Wertz, A.K.**, « CD-ROM : a new advance in medical information retrieval », *Journal of the American Medical Association*, 1986, 256 (24), p. 3 376-3 378.