

Les banques de données en 2004 : la génération Star Trek

Qu'est-ce qu'une banque de données en 2004 peut apporter par rapport aux informations trouvées sur Internet ? Sont-elles encore utiles ? Qu'y trouve-t-on ? Ces questions peuvent apparaître incongrues à la majorité des professionnels de l'information dont des générations successives ont appris - et apprennent toujours - le bien fondé d'une information structurée et organisée. Un rapide sondage auprès d'étudiants ou même d'enseignants montre leur méconnaissance du sujet : les moteurs de recherche Google et Yahoo, notamment, sont pour eux les moyens de trouver l'information qu'ils souhaitent. Peu importe la pertinence ou la fiabilité des informations trouvées, le plus important étant d'arriver au plus vite à un résultat ; seuls les chercheurs et les scientifiques, pourtant les premiers utilisateurs d'Internet, ont compris l'intérêt des résultats apportés par une banque de données professionnelle à leurs recherches. En 2004, que peut-on dire sur les banques de données ? Comment ont-elles évolué ? Quel est leur futur ?

La définition la plus simple du terme est qu'une banque de données est un entrepôt de données archivées. Un système de base de données permet, lui, d'accéder aux données entreposées. Si l'on applique cette définition à la lettre, il existe un grand nombre de banques de données : un serveur de mails, un document Word, une base de données relationnelles, un livre (mieux encore une encyclopédie)... Internet, pour sa part, peut être considéré comme un réseau complexe de bases de données hétérogènes.

Initialement, les banques de données n'étaient pas, bien sûr, électroniques. Quand les hommes ont eu besoin de conserver et d'archiver leurs connaissances, celles-ci étaient enregistrées sur des supports papier (le plus souvent un livre), puis retrouvées ensuite grâce à des systèmes complexes de classification, de catalogage, d'indexation mis en place par les bibliothécaires, les archivistes et les documentalistes. Aujourd'hui, quand on pense banque de données, cela inclut automatiquement l'idée de banque de données électroniques. Mais, papier ou électronique, la finalité était et reste la même.

Pourquoi les banques de données sont utiles ? Les raisons sont nombreuses :

- la mémoire humaine étant limitée, il est nécessaire de la conserver, pour la réutiliser ensuite.
- le cerveau humain est lui-même limité et ne peut engranger que quelques éléments d'information à la fois. Une banque de données donne la possibilité de consulter rapidement un grand nombre de données éparses sur un sujet : celles-ci sont triées selon les critères de recherche émis au départ, puis rassemblées sous la forme adéquate (le plus souvent une liste de références). Leur réutilisation, leur mise en forme sont alors possibles.
- elles permettent, à partir de données et de faits réunis et analysés par la machine, de faire des déductions, de récupérer les informations produites : c'est le cas, par exemple, du « *textmining* » qui va chercher dans les données entreposées celles qui se recoupent. La corrélation est rendue possible par le logiciel de recherche entre des données qui, a priori, n'ont rien à voir entre elles.

Il existe aujourd'hui un grand nombre de banques de données, hormis les banques de données bibliographiques. Il s'agit :

- des banques de données relationnelles qui, grâce à un langage spécifique, SQL ou autre, mettent en relation certaines données entre elles dans une structure à deux dimensions ;

- des banques de données objet : elles compensent les faiblesses des précédentes et permettent d'effectuer des équations de recherche plus complexes. Leur langage est le langage OSQL (object sql) ;
- des banques OLAP (On-Line Analytical Processing Databases) qui analysent les informations. L'intelligence artificielle (IA) a permis leur développement ;
- des banques de données spatiales pour les données géographiques (cartes, plans...) ;
- des banques XML : dernières arrivées, elles utilisent le langage XML qui a une aptitude à traiter n'importe quel type de document, grâce à la spécification de *définitions de type de document* (DTD) adaptées. Le projet Xperanto d'IBM initie un mouvement qui va certainement aller en s'amplifiant : avec XQuery, langage d'interrogation basé sur XML pour permettre l'accès à tous les types de données, structurées comme non structurées, il est possible d'obtenir des sources de contenus tels des documents (feuilles Excel), des bases d'e-mails en plus des traditionnelles bases de données relationnelles. Cette approche fédératrice permet aux données de rester dans leurs formats d'origine et à leurs emplacements d'origine. La capacité à interroger des données dans ce format va devenir incontournable.

Très différentes les unes des autres, ces banques ont des usages et des usagers multiples, mais avec un point commun : elles contiennent toutes des données qui sont structurées et organisées. Leur développement a été rendu possible grâce notamment aux recherches de scientifiques, de chercheurs, d'informaticiens et de sociétés informatiques telle IBM qui développa le langage SQL. Certaines banques de données pénètrent peu à peu notre vie quotidienne sans même que nous nous en apercevions, notamment quand une recherche est effectuée sur un plan de ville pour trouver le plus court chemin d'un point à un autre. D'autres sont appelées banques de données distribuées ou fédératives : leur particularité est d'être interrogeables en une seule fois même si elles sont localisées sur des sites différents, ce qui est transparent pour l'utilisateur. D'autres banques de données se développent avec des applications importantes dans le domaine de la médecine et de la recherche scientifique : c'est le cas, par exemple, des banques de données scientifiques sur l'ADN, le génome humain ou le séquençage. Certains prévoient même la possibilité de répliquer digitalement un corps humain dans une base de données. Est-ce de la science-fiction ou est-ce pour demain ? La génération Star Trek des banques de données est donc née.

Qu'est-ce qui caractérise une banque de données en 2004 ?

- c'est une banque de données bibliographique.
- elle contient du texte intégral.
- son contenu est disponible en format HTML ou PDF.
- elle propose des liens vers d'autres sources d'information.
- sa couverture est d'au moins 25 ans.
- elle donne la possibilité aux utilisateurs de télécharger ou d'imprimer des documents.
- elle propose une aide en ligne.

Les banques de données bibliographiques subissent quant à elles de plein fouet les contraintes économiques qui touchent également les bibliothèques et les services de documentation. Les fournisseurs d'information, conscients de cet état de fait, proposent dorénavant des liens vers des sources d'information multiples, généralement vers du texte intégral d'articles de périodiques : c'est notamment le cas de ProQuest. La possibilité donnée de lier les références contenues dans une base de données avec du texte intégral présente une valeur ajoutée certaine. Les banques de données interconnectées représentent une voie pour le futur.

L'utilisateur averti et exigeant a besoin de beaucoup plus que ce que proposent les moteurs de recherche actuels, aussi puissants soient-ils. Et seules les banques de données bibliographiques, avec des références sélectionnées par des professionnels, peuvent répondre à cette exigence. Une tendance se dessine d'ailleurs dans les bibliothèques pour un retour vers les banques de données bibliographiques car elles fournissent des réponses hautement plus fiables qu'Internet actuellement. Il est donc nécessaire, sinon impératif, d'éduquer et d'orienter les utilisateurs vers ces sources d'information. Cela suppose de prendre en compte les différents besoins de catégories d'utilisateurs hétérogènes (du grand public à l'étudiant et au professeur) et d'adapter l'offre bibliothéconomique et documentaire en la matière.

Jean-Philippe Accart
jean-philippe.accart@rero.ch

Auteur de l'ouvrage « Le Métier de documentaliste », en collaboration avec MP Réthy, 2^{ème} éd., Paris, Ed. Electre-Le Cercle de la Librairie, 2003. ISBN 2-7654-0872-6.
Sommaire et préface consultables sur le site : <http://www.accart.nom.fr>