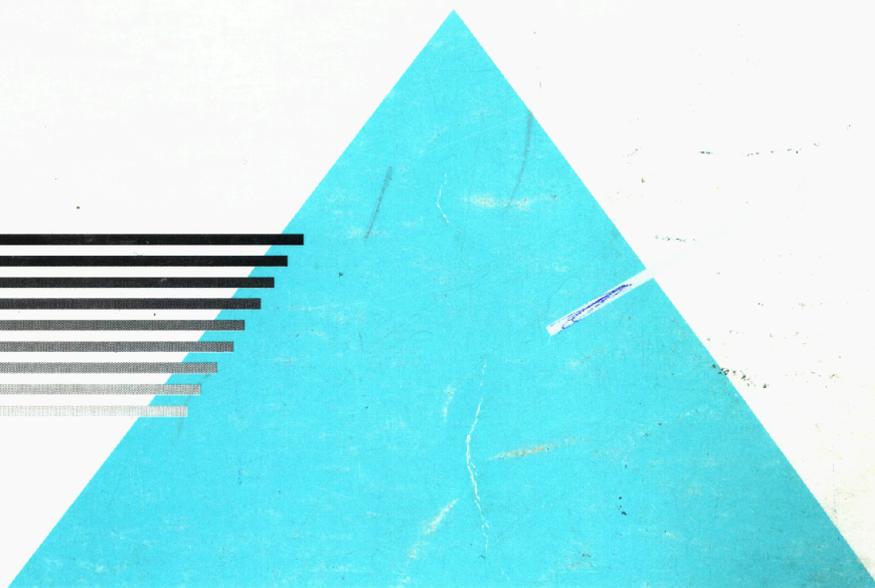


JEAN-PIERRE LARDY

# LES ACCÈS ÉLECTRONIQUES A L'INFORMATION

Etat de l'offre

BIBLIOTHEQUE DU CERIST



ADDBS  
EDITIONS

**LES ACCES ELECTRONIQUES  
A L'INFORMATION**

**ETAT DE L'OFFRE**

Décembre 1992

Jean-Pierre LARDY

**ADBS EDITIONS, 1993**

## TABLE DES MATIERES

INTRODUCTION	4
HISTORIQUE	5
L'évolution technique des ordinateurs	5
Les réseaux de télécommunication	8
Les pionniers de la recherche documentaire informatisée	9
LA SITUATION ACTUELLE	12
Des prévisions à la réalité	12
Les tendances	15
LES ACCES ELECTRONIQUES	18
LES ACCES CLASSIQUES	21
Les évolutions matérielles	21
Les évolutions logicielles	23
Les accès graphiques aux données chimiques	28
Les accès en temps réel aux informations financières	28
Des produits victimes de leur âge	29
Les coûts	29
L'accès aux documents primaires	30
LES ACCES GUIDES	36
Les caractéristiques	36
Les réalisations	36
Le VIDEOTEX	39
LES LOGICIELS FRONTAUX	44
Les caractéristiques	44
Quelques exemples récents	46
Les frontaux de l'avenir	47
Les frontaux graphiques	47
Le marché	48
LES PASSERELLES	51
Les caractéristiques	51
Quelques exemples	51
Un service à valeur ajoutée - CARTE EXPERT de Questel	55
LES "SUPERMARCHES" DE L'INFORMATION ELECTRONIQUE	58
Les exemples américains	59
La surabondance de sources	60

A travers l'histoire, de nombreuses avancées technologiques ont eu de grands impacts sur la société. Elles fournirent des possibilités inconnues jusque là, conduisant à des améliorations en efficacité dans nombre d'activités.

En sciences de l'information, il y a eu deux développements majeurs ; la presse à imprimer et les ordinateurs : le premier accrut fortement les possibilités de reproduction de l'écrit, le second donna un pouvoir d'organisation et de manipulation de l'information permettant la gestion de volumes de données en croissance constante.

Il y a 35 ans étaient créées les premières banques de données sur des principes restés identiques actuellement. Au même titre que les machines, les matières premières, les capitaux ou les hommes, l'information est une ressource avec un coût et une valeur. Mais, à la différence des autres, l'information est une notion multiforme et floue. Ces aspects à la fois formels et informels exigent des critères de jugement et d'exploitation autres que ceux employés pour la gestion des ressources classiques. L'information est maintenant reconnue comme un facteur déterminant de l'économie mais les moyens mis en oeuvre pour la gestion de son exploitation sont loin de répondre correctement aux besoins.

Cependant les travaux de recherche menés depuis laissent prévoir des retombées importantes que ce soit dans le multimédia ou la manipulation de la langue.

Le présent travail fait le point sur l'état de l'art en diffusion électronique, à distance ou localement, de l'information scientifique, technique, financière et autre.

Avec un âge de 25 ans, l'industrie des banques de données est arrivée à maturité et on considère qu'après une phase de consolidation elle se doit d'affronter de nouveaux défis pour mieux répondre et satisfaire la demande d'utilisateurs de plus en plus nombreux et avertis.

**HISTORIQUE  
DE LA RECHERCHE DOCUMENTAIRE  
INFORMATISEE**

La recherche documentaire informatisée est née aux Etats-Unis il y a presque 35 ans et ceci pour deux raisons :

- la première tient à la situation de la recherche scientifique avant la guerre. L'Europe en était le centre et les Etats-Unis importaient beaucoup de documentation. La victoire en 1945 a permis ainsi de mettre la main sur beaucoup de documentation d'origine allemande concernant la chimie, le développement de fusées à partir de documents trouvés à Peenemunde etc., sous la direction des services de renseignements. Il y eut donc la mise en place d'une politique de collecte, stockage et exploitation de l'information scientifique et technique.

- la seconde résulte du développement rapide de l'utilisation après la guerre de l'ordinateur, même si des machines de traitement automatique avaient été déjà développées en Europe, en particulier en Allemagne.

### L'évolution technique des ordinateurs

L'ordinateur a été conçu à l'origine en tant que machine à effectuer des calculs et non à manipuler de l'information textuelle. Mais dès les années 50 les caractères typographiques devenaient matière à traitement et ce dans deux voies différentes :

- d'un côté la **production** de textes qui connaît actuellement un développement énorme avec les logiciels de traitement de texte et leur prolongement, la publication assistée par ordinateur (PAO).

- de l'autre la **description** de fonds documentaires pour leur stockage et gestion en vue de les retrouver facilement.

On peut noter que la tendance actuelle tend vers un mariage des deux approches.

Au niveau stockage, la technologie de l'époque reposait sur les bandes magnétiques et était limitée à des enregistrements de longueurs fixes. Peter Luhn [LUH57], de la société IBM, fut le responsable du

développement d'un lecteur à cartes permettant d'utiliser un ordinateur pour manipuler du texte. Pour repérer les enregistrements répondant à une question, on entrait les critères de recherche sur une carte. L'ordinateur commençait à permettre le traitement de grandes quantités de données, certes encore en mode séquentiel, mais cela autorisait les tris et les éditions d'index.

Bien vite cependant le principal problème de la recherche d'information apparut : comment répondre à la question particulière d'un individu X à partir d'un ensemble varié de données issues de producteurs différents. Evidemment on ne parlait pas de rien : certains problèmes étaient étudiés et résolus depuis longtemps au niveau des index imprimés développés dès la fin du XIXe siècle.

C'est pourquoi, très naturellement, les techniques classiques furent reprises :

- stockage de l'information sous forme condensée et structurée en champs,
- représentation du contenu sous une forme compacte et manipulable par l'ordinateur : le texte intégral est remplacé par une liste de mots clés, attribués soit par l'auteur du document, soit automatiquement par l'ordinateur, soit le plus fréquemment par des indexeurs, accompagnés de codes de classement et quelquefois d'un résumé.

Le but de ces différentes méthodes est de rendre la recherche de l'information aussi rapide et précise que possible. Mais aucune n'est parfaite et même l'interrogation de la banque de données MEDLINE qui possède un des langages de description le plus approfondi, conduit à des erreurs, dues à une définition trop pauvre du contenu.

Pendant les années 60, on porta une attention soutenue à cet aspect et diverses propositions d'amélioration de la description du contenu virent le jour : pondération, termes qualifiant, indicateurs de rôle, liens. Mais tous ces éléments nécessitent une intervention humaine importante et les efforts se portèrent sur la possibilité d'automatiser l'indexation. Pour apprécier grâce à des indicateurs fiables la qualité des interrogations, on définit les notions de

"précision" et de "rappel". Ceux-ci sont d'un intérêt limité pour l'utilisateur final qui, en fait, n'a pas encore d'outil permettant d'apprécier la qualité de sa recherche et doit s'en remettre à son jugement après lecture des documents.

Les techniques utilisées pour la recherche dépendent à la fois de la représentation intellectuelle du contenu des documents et de la structure de la base. Contrairement aux formes papier qui ne permettent pas de répondre facilement à une question multicritères, la recherche informatique repose sur la logique booléenne qui permet l'association de descripteurs ou d'autres critères de recherche avec les opérateurs ET, OU, SAUF. La pondération de descripteurs peut être utilisée pour accentuer certains critères. Dans la recherche en langage non contrôlé (zones titre, résumé, texte intégral), c'est l'utilisation des opérateurs syntaxiques qui permet la précision dans la recherche. La préparation de la question nécessite la même attention que l'analyse de documents. Aussi c'est dans la grande majorité des cas un intermédiaire spécialisé qui effectue ce travail avec le concours du demandeur.

Ces dernières années une autre notion dépassant le contexte de la recherche documentaire actuelle apparaît : l'indexation intégrale de documents de tous types, sans structuration. La recherche est alors effectuée à partir d'une série de mots (qui devraient logiquement être dans les textes). Les apparitions de ces mots dans les différents documents sont comptabilisées. Ensuite un algorithme calcule un degré de pertinence des documents trouvés en fonction de la question. L'en-tête de chaque document est présenté à l'écran et l'utilisateur fait son choix. Les produits Topics, Wais fonctionnent selon ce modèle.

Ainsi au milieu des années 50, les principes actuels de la recherche documentaire informatisée étaient établis par PERRY et son équipe [PER56] à l'Institut Battelle :

- utilisation d'un vocabulaire contrôlé
- résumé de style télégraphique
- interrogation par combinaison de critères : ET OU SAUF
- structuration de l'information en phrases, paragraphes...
- évaluation d'une recherche : rappel, précision ...

Les évolutions technologiques successives de l'informatique furent rapidement intégrées au cas particulier de la recherche documentaire informatisée. Il en fut ainsi de la possibilité de faire travailler l'ordinateur sur plusieurs tâches en même temps (time sharing), et de l'utilisation de moyens de stockage importants des programmes et données des divers utilisateurs avec des temps d'accès courts. L'apparition des disques durs en remplacement des bandes magnétiques a permis l'exploitation de gros fichiers avec un accès direct à l'information. Ces problèmes furent résolus vers le début des années 60 en particulier pour la saisie et la mise au point de programmes informatiques.

Au niveau de l'organisation des données, c'est l'utilisation des index sous la forme de fichiers inverses qui fut retenue. Elle permet de donner rapidement le nombre de références répondant à une question et ainsi d'orienter la poursuite de l'interrogation. La consultation du fichier des références n'a lieu qu'à la visualisation des informations.

On peut noter que le développement accéléré des banques de données suivit de près le développement de la technologie des ordinateurs. C'est ainsi qu'en 1992 le serveur DIALOG commercialise plus de 450 banques de données représentant près de 2 téraoctets, interrogeables de n'importe quel point du globe à n'importe quelle heure.

### **Les réseaux de télécommunication**

Les premières liaisons à distance datent des années 60 et utilisaient le réseau téléphonique commuté.

Mais on s'orientait vite vers des réseaux de transport de données propres du fait d'une forte demande résultant du développement du télétraitement, demande elle-même confortée par les besoins des militaires en communication fiable permettant la mise en commun des moyens de calcul.

L'idée a consisté à découper et organiser les données en paquets indépendants, accompagnés de l'adresse du destinataire et de données de contrôle. Le réseau est maillé, plusieurs chemins sont possibles : ce sont les ordinateurs de commutation de paquets qui optimisent les transmissions par

le choix des chemins, gèrent les erreurs et reconstituent les données à l'arrivée. Contrairement au réseau téléphonique, il est possible de connecter des terminaux fonctionnant à des vitesses différentes.

La recherche documentaire en conversationnel était déjà une réalité, et le développement de tels réseaux a été un renfort puissant à l'essor de l'industrie de l'information. Alors débuta l'ère du "online". On sortait du terrain des pionniers pour entrer dans le domaine commercial.

Le premier réseau à transmission de paquets a été développé en 1969 aux Etats-Unis sous contrat de l'agence militaire ARPA (Advanced Research Projects Agency) et reçut le nom d'ARPANET. Il était accessible par ligne téléphonique. Tymshare Corporation, société de service vendant du temps partagé, installa un réseau privé du même type sous le nom de TYMNET qu'elle étendit rapidement hors des Etats-Unis [RAC73].

Le projet français CYCLADES a été lancé en 1972 et fut opérationnel en 1975. Il reliait une vingtaine d'ordinateurs situés dans des universités, centres de recherche ou centres informatiques, à la fois dans un but d'expérimentation et comme outil d'accès aux banques de données. Il permit le développement du réseau commercial TRANSPAC ouvert en 1978.

Au niveau européen, la CEE développa le projet EURONET engagé en 1975 par les divers PTT. La mise en service en 1980 s'appuyait sur un centre de gestion à Londres, quatre centres de commutation à Francfort, Londres, Paris et Rome et cinq autres points d'accès. Depuis l'Europe dispose d'un maillage généralisé.

La première réunion sur l'information en ligne en Europe s'est tenue en 1974 à Bruxelles sous l'égide de l'OTAN. Elle regroupait des spécialistes de la documentation et traitait des problèmes de réseaux pour les bibliothèques et centre de documentation [OTA75].

### **Les pionniers de la recherche documentaire informatisée**

La littérature sur l'histoire de la recherche documentaire informatisée est peu nombreuse. Elle est essentiellement américaine et on lira avec intérêt [NEU86] et [SAL76].

Ce fut l'ASM (American Society for Metals) qui, en finançant un projet sur cinq années et en s'attachant l'équipe de PERRY, démontra que les concepts étaient exploitables pratiquement et après une étude de faisabilité de deux ans, mit en service en 1960 le premier système vraiment opérationnel traitant environ 20 000 articles en métallurgie.

De même un programme pilote lancé en 1961 aboutit au développement du service d'information en éducation ERIC.

En 1964 la National Library of Medicine (NLM) aux Etats-Unis lançait le service MEDLARS fonctionnant en traitement par lots ; en 1965 Chemical Abstracts faisait de même à partir de travaux débutés en 1955, suivi en 1969 de BIOSIS en sciences de la vie et de la Bibliothèque du Congrès avec le système MARC. Ces services furent au début à l'usage exclusif de leurs producteurs : les questions étaient entrées en série puis un lot de recherches était lancé. Ce fonctionnement était cher, peu pratique et surtout peu satisfaisant au niveau des réponses car s'il était possible d'essayer diverses stratégies de recherche pour une même question, les résultats n'étaient obtenus qu'après plusieurs heures de traitement.

L'utilisation des disques durs comme mémoires de masse en complément des bandes magnétiques réduisit les temps de recherche à l'information (accès direct par rapport à l'accès séquentiel des bandes) et offrit la possibilité de créer des index élaborés. Ainsi on passa du traitement par lots au mode interactif permettant de poser une question et de modifier immédiatement la stratégie en fonction de la réponse.

En 1971 MEDLARS fut accessible en mode conversationnel et à distance grâce à la mise en place des premiers réseaux de télécommunication, successeurs du réseau ARPANET mis en place par les militaires. Parallèlement apparurent les premières banques en texte intégral (MDC LEXIS).

C'est aussi au milieu des années 60 que l'idée d'un ordinateur serveur de banques de données apparaît. En 1965 Lockheed fait la démonstration d'un système interactif de recherche d'information et en 1968

ce système, sous le nom de RECON (Remote Console Information Retrieval Service), est retenu par la NASA pour ses centres de recherche.

En 1972 débute la commercialisation de services en ligne par deux serveurs (DIALOG et SDC) offrant six banques de données tandis que les réseaux commerciaux américains TYMNET et TELENET offrent leur service à l'extérieur des Etats-Unis : ainsi les Etats-Unis deviennent le premier producteur et le premier fournisseur d'informations en ligne.

C'est cette ouverture sur l'extérieur qui va provoquer une prise de conscience des européens et décider les gouvernements à intervenir comme l'avait fait bien avant eux le gouvernement américain.

### **Conclusion**

Il ressort de cet historique que le rôle des pouvoirs publics a été déterminant aux Etats-Unis comme en Europe ainsi que le montre Serge CACALY [CAC90]. Cependant aux Etats-Unis très vite l'Etat se désengage après avoir investi dans l'infrastructure et laisse la gestion et la commercialisation au secteur concurrentiel. Ce désengagement n'est cependant jamais total en ce qui concerne les producteurs, les sociétés savantes comme l'American Chemical Society ou la National Library of Medicine en sont le meilleur exemple.