

CAHIERS DE L'ADBS

SCIENCES ET TECHNIQUES DE L'INFORMATION

MONOGRAPHIE N°2

BIBLIOTHEQUE DU CERIST

logiciels et systèmes documentaires

IST

418/1

adbs

CAHIERS DE L'ADBS

SCIENCES ET TECHNIQUES DE L'INFORMATION

MONOGRAPHIE N°2

BIBLIOTHEQUE DU CERIST

logiciels et systèmes documentaires

étude réalisée par le CXP

1975

ASSOCIATION FRANCAISE DES DOCUMENTALISTES ET BIBLIOTHECAIRES SPECIALISES
63 bis, rue du Cardinal Lemoine - 75005 - Paris téléphone: 033.33.38

L'étude ci-après a été financée par :

- le BNIST (convention n° 291-173)
- l'I R I A
- le C X P

*

Dans la même collection

N°1 LE DOCUMENTALISTE EN FRANCE ET SON SALAIRE
EN 1974
enquête réalisée sous la direction de Jean Thoumin

ISBN 2-901046-02-9

L'informatique documentaire est un sujet qui intéresse aujourd'hui de plus en plus de sociétés, d'administrations ou d'organismes nationaux et internationaux. Et le problème du choix du logiciel et/ou du matériel est loin d'être simple devant l'arsenal d'outils informatiques plus ou moins adaptés aux besoins des services d'information.

De façon à pouvoir répondre à de nombreuses demandes, en juin 1974 le BNIST (Bureau National de l'Information Scientifique et Technique) et la Délégation à l'Informatique décidaient de confier au CXP (Centre d'Expérimentation de Packages), la réalisation d'une étude sur les logiciels et systèmes documentaires disponibles en France. Le STI (Service Technique Informatique) de l'IRIA était chargé d'assurer, pour le compte de la Délégation, le suivi technique du contrat.

Cette étude doit permettre, dans un premier temps, d'avoir une vue globale des produits existants (qu'ils soient ou non commercialisés). Dans une deuxième étape, une dizaine de produits (commercialisés) seront étudiés d'une manière très approfondie et comparative.

L'étude a été suivie et orientée par un GROUPE DE PILOTAGE formé d'experts et d'utilisateurs. Le document final doit apporter des informations précises, sûres et représentatives.

Ce document ne représente qu'une partie de l'étude il sera suivi d'une présentation comparative de onze logiciels.

Jacques MICHEL
Secrétaire Permanent
du BNIST

Michel DELAMARRE
Chef du Service Technique Informatique
de l'IRIA

BIBLIOTHEQUE DU CERIST

S O M M A I R E

INTRODUCTION

| | |
|---|----|
| Remerciements | 5 |
| Réalisation de l'étude | 7 |
| Objet de l'étude | 9 |
| Présentation générale du marché | 11 |
| Les principales caractéristiques | 15 |
| Liste des produits : | |
| .par ordre alphabétique des noms des progiciels | 21 |
| .par ordre alphabétique des noms des distributeurs | 22 |
| .par matériel | 23 |

LISTE DES FICHES (entre parenthèses : nom du concepteur)

| | |
|---------------------------------|-----|
| CORA (ROUSSEL-UCLAF) | 27 |
| X CRISTAL (CNRS) | 31 |
| DAMAS (IREP) | 35 |
| DOCILIS (CEDIJ) | 39 |
| ECLAIR (IRIA-STERIA) | 45 |
| FABIUS (HEURTEY) | 51 |
| FIND II (ICL) | 57 |
| GOLEM II (SIEMENS-UNIDATA) | 63 |
| chaîne IFP (IFP) | 71 |
| IMDOC-T (INDUSTRI-MATEMATIK) | 75 |
| ISARD (IRT) | 85 |
| KWIC 360 (IBM) | 91 |
| KWOC version 72 (CDC) | 95 |
| MARABOU (ROUSSEL-UCLAF) | 93 |
| MISTRAL (CII) | 105 |
| X PASCAL 3 (CNRS) | 113 |
| RIRMS (IBM) | 119 |
| SABIR (Institut Gustave-Roussy) | 127 |
| SAFIR (CNET) | 133 |
| SAGEP (CREUSOT-LOIRE) | 141 |
| X SATIN 1 (URADCA) | 145 |
| SIC (SAMM) | 153 |
| SIDERAL+SID (ELF-ERAP) | 157 |
| SPHINX (INSEE) | 163 |

| | |
|---|-----|
| SPLEEN II VIRA (CNRS-CDSH) | 169 |
| STAIRS (IBM) | 175 |
| SYDONI (CRIDON) | 183 |
| SYGEBRE (M.B.P. de France) | 191 |
| TAC (SAMM) | 201 |
| TITUS II (I.T.F.) | 203 |
| VIRA-EPOS version PASCAL-VIRA (SINFOK-CNRS) | 211 |
| La recherche au CEA | 216 |

I N T R O D U C T I O N

| | |
|---|----|
| Remerciements | 5 |
| Réalisation de l'étude | 7 |
| Objet de l'étude | 9 |
| Présentation générale du marché | 11 |
| Les principales caractéristiques | 15 |
| Liste des produits : | |
| .par ordre alphabétique des noms des progriciels | 21 |
| .par ordre alphabétique des noms des distributeurs | 22 |
| .par matériel | 23 |

REMERCIEMENTS

Nous tenons à remercier toutes les personnes et les organisations qui ont permis la réalisation de cette étude. Nous remercions :

- . Monsieur MICHEL et Madame CORNIER (du BNIST) ainsi que Monsieur CARFANTAN de l'IRIA qui ont pris l'initiative de cette étude et en ont assumé le support financier.
- . Monsieur LEPIDI et les membres actifs du CXP qui ont accepté que cette étude figure parmi les travaux du Centre.
- . Les membres du Groupe de Pilotage :
Monsieur CARFANTAN, Madame CORNIER (déjà cités) Monsieur DELAUNEY (Institut Français du Pétrole), Monsieur DUCROT (Institut du Textile Français), Monsieur LEROUX (Institut National de Recherche Agronomique), Monsieur VIELLARD (Direction des Recherches et Moyens d'Essais), Madame WOLFF-TERROINE (Institut Gustave Roussy).
- . Nous remercions également :
 - Madame MOUREAU (IFP) qui a guidé nos premières investigations;
 - Madame CONNAT et toute l'équipe de SATIN qui nous a reçus à Marseille, ainsi que Monsieur DAUMIER qui nous avait présenté le système à PARIS.
 - Madame ALLAIRE (CNRS) qui nous a fourni de très nombreuses indications.
 - Monsieur COUDRON qui n'a pas hésité à sacrifier un samedi pour nous présenter à NAMUR le système GOLEM.
 - Monsieur RAAIJMAKERS et Monsieur WILLIAMSON qui sont venus de Hollande nous présenter IMDOC.

ainsi que tous les spécialistes qui nous ont reçu aidé et ouvert tous leurs dossiers...ils ont été si nombreux que la place nous manque pour les citer tous, qu'ils trouvent ici l'expression de nos sentiments reconnaissants.

PRESENTATION DU C.X.P.

... Qu'est-ce que le C.X.P. ?

L'article 2 des statuts de notre association le précise sans équivoque :

"Le C.X.P est un centre d'information et d'expérimentation créé par des utilisateurs et pour des utilisateurs de packages".

Le C.X.P. est une association sans but lucratif (loi de 1901) créé en 1973 par 8 importantes sociétés. Aujourd'hui le C.X.P. compte 8 membres ACTIFS, 7 membres ABONNES et une vingtaine de membres CORRESPONDANTS. Parmi les membres actifs et abonnés nous pouvons citer : Air France, Charbonnages de France, BRED, EDF-GDF, Société Générale, RATP, ELF ERAP, ORTF, Régie Renault, SNCF, GRADIA, ainsi que l'IRIA et le BNIST... qui ont demandé cette étude.

Le principe de fonctionnement du C.X.P. est simple : chaque MEMBRE ACTIF peut choisir une étude qui est réalisée par les ingénieurs du C.X.P. appuyés par des spécialistes si nécessaire. L'étude est envoyée à chaque société de service dont on compare les produits pour contrôle (et bien sûr entièrement gratuitement), puis elle est diffusée à tous les membres ACTIFS et ABONNES et, à la demande, aux membres CORRESPONDANTS.

Le C.X.P. réalise entre 8 et 10 études par an. Les cotisations de membres ACTIFS et ABONNES sont respectivement de 35 000 et 20 000 F par an et donnent droit (sans aucun autre frais) à :

- toutes les études réalisées pendant l'année
- l'abonnement au bulletin mensuel C.X.P.i
- l'utilisation du FICHER CENTRAL DES PRODUITS LOGICIELS développé par le C.X.P. avec l'aide de l'IRIA
- les conférences et réunions réalisées par le C.X.P. pour ses membres.

Pour en savoir plus, contactez Monsieur Jean-Eric FORGE - Responsable des Etudes. 256-10-99 11, Avenue Delcassé 75008 - PARIS

REALISATION DE L'ETUDE

L'étude a été réalisée par des ingénieurs du CXP assistés de spécialistes sous contrats.

SPECIALISTES

- Contrôle technique Général : Mademoiselle RIMBERT. Institut Gustave Roussy.
- Etude et contrôles détaillés des produits : Monsieur MANSUY assisté de Monsieur FEDIDA (GETI)

INGENIEURS CXP

- Coordination, suivi des travaux, mise en forme finale : Monsieur DARAS
- Direction technique et premières spécifications : Monsieur FORGE
- Suivi juridique et financier : Monsieur JACQUET
- Assistance technique : Monsieur ROTMAN

L'étude a été suivie au fur et à mesure de sa réalisation par un GROUPE DE PILOTAGE comprenant les personnalités suivantes (par ordre alphabétique).

| | |
|-----------------------|---|
| Monsieur CARFANTAN | (IRIA) |
| Madame CORNIER | (BNIST) |
| Monsieur DELAUNEY | (IFP) |
| Monsieur DUCROT | (ITF) |
| Monsieur LEROUX | (INRA) |
| Monsieur VIELLARD | (DRME-BIS) Président du Groupe de Pilotage |
| Madame WOLFF-TERROINE | (I.G.R.) |

L'étude a commencé en novembre 1974 et a suivi un plan en trois phases, qui avait été défini dans le cahier des charges du contrat.

- 1 - Survol du marché et définition d'un MODÈLE permettant de définir les frontières de l'étude ainsi que le plan de présentation des différents systèmes.
- 2 - Etude détaillée de 11 systèmes retenus par le groupe de pilotage.
- 3 - Synthèse et spécifications générales d'une procédure expérimentale de tests.

Les différentes "fiches de survol" ont été réalisées à partir d'un questionnaire mis au point par le CXP et soumis aux concepteurs et distributeurs des produits cités. Ces fiches ont été renvoyées après leur mise au point par le CXP aux différents responsables des produits cités individuellement d'abord, puis globalement, chaque responsable recevant le premier document (SURVOL) en entier.

La même procédure a été suivie en ce qui concerne l'étude détaillée des produits.

Les premiers documents ont été disponibles en Juillet 1975.

Quelques chiffres

L'étude a demandé environ 2000 heures de travail aux ingénieurs chargés de sa réalisation. Une centaine de rendez-vous ont été nécessaires pour recueillir les informations. 15.000 pages de documentation ont été étudiées, 1000 pages ont été écrites (dont 600 correspondant au document final)

O B J E T D E L ' E T U D E

L'objet de cette étude réalisée par le C. X. P. est de permettre une meilleure connaissance des produits qui existent déjà dans le domaine de l'informatique documentaire.

L'étude complète comprend trois grandes parties :

première partie : Un SURVOL des produits disponibles sur le marché français. Une étude très rapide des grandes options qui se dégagent.

deuxième partie : Une présentation comparative de onze progiciels choisis par le Groupe de Pilotage. Un modèle fonctionnel a été conçu qui permet de comparer (du point de vue de l'utilisateur) les produits très différents et originaux qu'il convient d'étudier.

troisième partie : Une présentation détaillée des onze progiciels choisis. Conclusions générales

LE PRESENT DOCUMENT CORRESPOND A LA
PREMIERE PARTIE DE L'ETUDE QUI A ETE
REALISEE PAR LE C. X. P. DANS LE CADRE DE
CE CONTRAT PASSE AVEC LE BNIST ET L'IRIA

PRESENTATION GENERALE DU MARCHÉ

La première difficulté de ce survol a été de dresser la liste des produits auxquels il fallait s'intéresser. En effet, le marché est assez fourni en produits très différents les uns des autres et il n'est pas toujours facile de dire si un progiciel peut être considéré comme un logiciel (ou un système) documentaire ou non.

Il fallait donc se fixer des frontières. Et tout d'abord avec un voisin très proche : les systèmes de gestion de bases de données (S. G. B. D.) Or cette frontière à plusieurs points de vue n'est pas très précise et l'incorporation d'un produit dans la liste peut paraître souvent subjective. Nous avons pris comme critère essentiel le "savoir-faire " contenu dans le produit..

Un système de gestion de bases de données est conçu pour manipuler facilement des informations réparties dans une structure plus ou moins complexe mais toujours définie et explicite : en fonction de ses caractéristiques, une donnée va se placer dans une zone déterminée de la structure. Les éditions sont très générales et doivent être définies par l'utilisateur. Enfin un système de gestion de bases de données est logiquement conçu pour la mise-à-jour plus que pour l'addition d'informations.

Un système documentaire au contraire a une structure extrêmement mobile : chaque document entre avec ses caractéristiques et "s'accroche" aux autres plus ou moins directement. Ici la structure est très changeante, ce sont les "règles d'accrochage" qui sont fixes. Au niveau des éditions les paramétrages sont fait en fonction des besoins spécifiques des documentalistes. Enfin les systèmes sont conçus pour supporter de très gros volumes d'informations dont la volatilité est faible et qui augmentent par accroissement successif.

Cependant cette règle n'est pas non plus très précise, d'autant que nous avons recherché à présenter les possibilités du marché pour chaque catégorie d'utilisateurs, en tenant compte de l'environnement : ainsi SOCRATE ne fait pas partie des produits étudiés alors que FIND en fait partie. SOCRATE a comme concurrent pour les matériels sur lequel il fonctionne des produits typiquement documentaires comme MISTRAL à la C.I.I. et STAIRS ou RIRMS pour I.B.M. ICL par contre n'a pas de produit spécifiquement documentaire et FIND est susceptible de répondre à des besoins documentalistes.

Une autre frontière est la spécificité des produits. Frontière également assez floue derrière laquelle on trouve des produits de gestion de bibliothèque, de composition automatique des textes, etc...

Il est intéressant de regarder l'origine de tous ces produits :

On trouve d'abord des produits figurant aux catalogues des constructeurs de matériel informatique :

RIRMS et STAIRS principalement chez IBM (D.P.S n'est plus maintenu) MISTRAL pour la CII, FIND chez ICL, GOLEM chez UNIDATA.

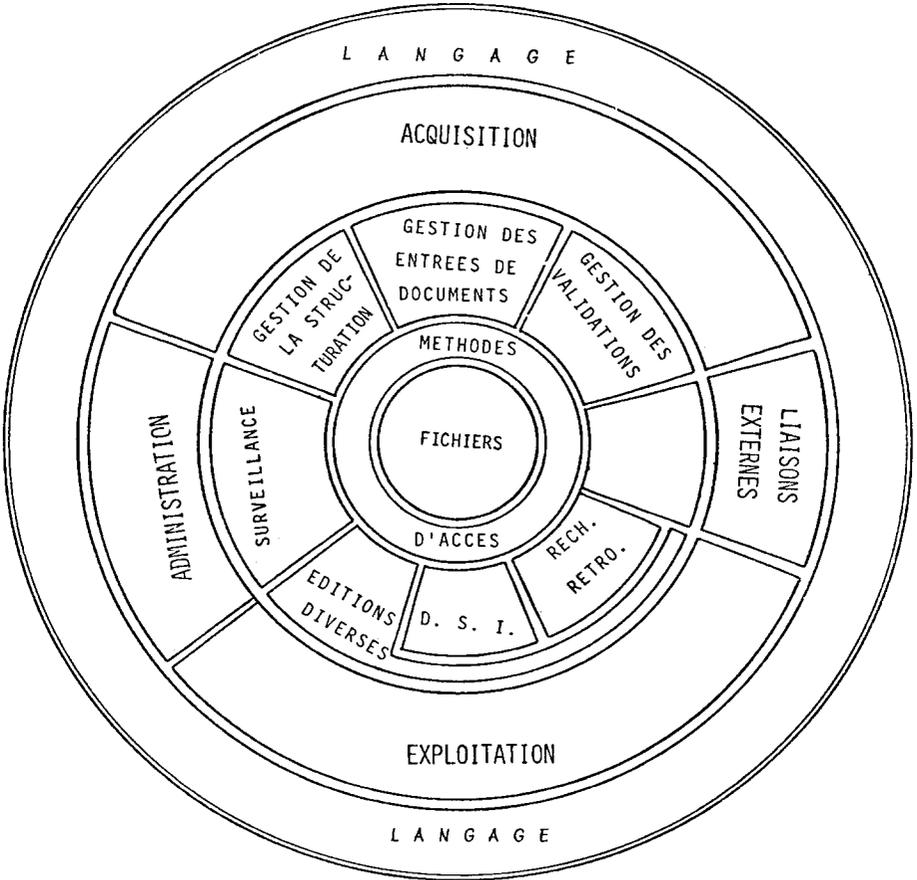
Autre catégorie : les gros produits-programmes écrits par des organismes désireux de se doter d'un matériel documentaire important : TITUS de l'Institut du Textile de France, SABIR de l'Institut Gustave Roussy, la chaîne de l'IFP, SAFIR pour le CNET, etc ... Ce sont en général des programmes écrits au début pour une application spécifique. Ils ont cependant des possibilités très étendues et sont susceptibles de répondre à des utilisateurs exigeants.

Viennent ensuite de nombreux produits dont le champ d'action commercial ainsi que les possibilités techniques sont en général très spécialisés. Ces produits sont intéressants à connaître parce qu'ils peuvent répondre parfaitement à un besoin, si celui-ci recoupe de très près celui des concepteurs du produit. Au niveau de notre étude cependant nous nous contenterons de les mentionner en les accompagnant d'une description sommaire : ils sont en général déjà connus des spécialistes oeuvrant dans le même champ et leurs spécifications sont souvent susceptibles d'évoluer.

Indiquons enfin que nous nous sommes limités à des produits effectivement diffusés en France. Nous avons cependant jugé utile de mentionner quelques produits qui, bien que n'ayant pas à l'heure actuelle d'utilisateurs, ont déjà une base solide de contacts et (surtout) des moyens d'assurances suffisants dans notre pays.

LES PRINCIPALES CARACTERISTIQUES

Nous donnons ci-dessous le schéma du "modèle" que nous avons mis au point. L'explication détaillée de ce modèle est donnée dans la seconde partie de cette étude.



BIBLIOTHEQUE DU CERIST

Les langages informatiques dans lesquels ont été écrits les programmes sont variés mais on y retrouve très souvent le langage Assembleur ce qui est d'ailleurs un obstacle à la portabilité du produit mais qui est souvent un gage d'optimisation des procédures. Les langages évolués utilisés sont FORTRAN, COBOL et PL/1.

Les constructeurs ont édité une documentation assez importante sur leurs produits. Certains autres produits ont une documentation d'aussi bonne qualité (manuels d'utilisation, description, mise en place, etc...). Il faut cependant reconnaître qu'en règle générale il y a une nette insuffisance à ce niveau.

Voyons maintenant les différentes fonctions :

La constitution du stock est en général faite par le produit lui-même mais l'importance des validations des informations à rentrer est très variable. Elles sont par exemple très développées dans PASCAL dont c'est le but principal. La plupart des produits ont choisi de faire cette constitution par lot mais certains comme TITUS, IMDOC, et STAIRS (par l'intermédiaire de ATMS), RIRMS offrent la possibilité de rentrer les documents à l'aide de consoles de visualisation.

L'indexation des documents est également envisagée différemment. Certains s'en tiennent à une indexation manuelle à l'aide d'un thésaurus prédéterminé, d'autres font de l'indexation plus ou moins automatique. Les plus avancés permettent d'incorporer des descripteurs nouveaux dans le lexique. Les systèmes fonctionnant en texte libre ne sont pas très développés et les essais d'utilisation de texte libre en langage naturel font encore partie du domaine de la recherche.

La façon dont est traitée la recherche rétrospective (RR) est plus homogène. Le système de recherche se sert également de descripteurs et se fait à l'aide d'un fichier inverse. Les opérateurs sont toujours un peu les mêmes, chaque produit se caractérisant par quelques perfectionnements supplémentaires. Bien souvent cette recherche est possible à partir de consoles qui permettent soit de simuler un traitement par lot (la console faisant alors office de lecteur de cartes et d'imprimante, soit d'utiliser un système plus élaboré où l'utilisateur dialogue véritablement avec le système. Il existe parfois de plus, des procédures d'aides à la recherche qui guident l'utilisateur dans la formulation de ses questions.

La diffusion sélective (DSI) est parfois considérée comme une recherche rétrospective où les profils sont de simples questions; mais généralement les procédures sont nettement différentes, adaptées à la taille du fichier sur laquelle on travaille et au grand nombre de questions posées. Les moyens d'édition en sortie de la DSI sont divers mais il faut noter l'évolution des moyens modernes tels que la photocomposition et l'édition sous microforme. Le terme de profil cache des contenus bien différents allant de la question de type recherche rétrospective cataloguée jusqu'à un ensemble d'informations comportant évidemment des renseignements pour la sélection des documents mais aussi des renseignements personnalisés destinés à améliorer le service de DSI.

Les éditions diverses sont assez importantes et permettent de tirer du fond documentaire des informations sous des formes adaptées aux différents besoins (éditions de thésaurus sous diverses formes, bulletins; listés de références, index...). Là aussi les supports d'édition sont variés et tournés vers les techniques modernes.

Bien des systèmes fournissent au gestionnaire de la base, des renseignements de types statistiques sur l'entrée des documents (fréquence d'utilisation des descripteurs...), l'exploitation (temps d'utilisation CPU par question,...), la gestion informatique de la base (occupation et besoins de restructuration des fichiers,...), etc... Ces renseignements sont loin d'être négligeables car ils permettent de bien gérer la base (du point de vue documentaire et informatique) afin de rendre son utilisation plus souple, et ce à un coût minimum.

Les possibilités multilingues existent sur peu de produits et sont encore assez limitées : TITUS et SABIR. Le congrès ADBS de décembre 1974 l'a souligné. Le multilinguisme est encore du domaine de la recherche mais le nombre de personnes qui s'y intéressent permet d'espérer l'intégration de fonctions de traduction dans les logiciels documentaires de demain.

Les problèmes d'accès au fonds documentaire ont été également examinés. Si certains concepteurs ont refusé un contrôle d'accès, d'autres ont conçu un système de secret assez élaboré, justifié souvent par l'utilisation : Les personnes ayant accès à la base ont un code qui leur permet soit de consulter uniquement la base (avec ou non possibilité d'édition) soit d'intervenir sur la base (création, suppression, mise à jour). Les accès peuvent avoir des portées à différents niveaux : le fonds documentaire, le document, une partie du document.

Les fichiers sur lesquels le système travaille sont sensiblement les mêmes, tout au moins dans leur nature : fichier bibliographique, fichier de recherche, fichier thésaurus, fichier des profils. Mais on trouve également des fichiers des mots vides, fichiers des synonymes, fichiers des questions catalogués. Le type d'organisation et leurs méthodes d'accès informatique sont fonction des traitements que l'on a à effectuer sur eux : séquentiel, séquentiel indexé, direct.

BIBLIOTHEQUE DU CERIST

Le descriptif des différents éléments composants les logiciels fait apparaître une disparité assez marquée dans bien des domaines, allant de la non existence de certaines fonctions à des possibilités très développées. Les logiciels présentés ci-après sont destinés à faciliter et à améliorer les travaux des documentalistes et peuvent être considérés comme un ensemble d'outils où chacun pourra y trouver des solutions plus ou moins adaptées à son problème.

§-§-§-§-§-§-§-§-§-§

Cette présentation préliminaire est développée de façon beaucoup plus importante et plus précise dans le deuxième document : ETUDE COMPARATIVE.