

Ecole Nationale
Supérieure de
Bibliothécaires

Université
Claude Bernard
LYON 1

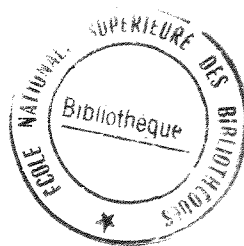
Diplôme supérieur
de Bibliothécaires

DESS Informatique
Documentaire

Note de Synthèse

Utilisation des systèmes experts
pour l'interrogation des bases
de données

Isabelle MANEVY



Sous la direction de :

Jean-Pierre LARDY

Unité Régionale de Formation en Information
Scientifique et Technique
Université LYON 1

SOMMAIRE

<u>SOMMAIRE</u>	p. 2
<u>RESUME</u>	p. 4
<u>ABSTRACT</u>	p. 5
<u>I. INTRODUCTION</u>	p. 7
<u>II. RECHERCHE BIBLIOGRAPHIQUE</u>	p. 9
1. La recherche manuelle	p. 9
2. La recherche automatisée	p. 10
2.1. Les bases utilisées	p. 10
2.2.1. La base PASCAL	p. 10
2.1.2. La base INSPEC	p. 11
2.2. La stratégie de recherche	p. 11
2.2.1. Interrogation sur PASCAL	p. 11
2.2.2. Interrogation d'INSPEC	p. 13
2.3. Analyse des résultats de la recherche	p. 13
3. Sélection et obtention des articles	p. 15
4. Conclusion	p. 15
<u>III. SYNTHESE</u>	p. 18
1. Qu'est-ce qu'un système expert ?	p. 18
1.1. Définition	p. 18
1.2. Evolution	p. 19

2. Intérêt des systèmes experts dans l'aide à l'interrogation des bases de données p. 20

- 2.1. Les problèmes rencontrés par l'utilisateur p. 20
- 2.2. Pourquoi utiliser les systèmes experts ? p. 21
- 2.3. Les utilisateurs potentiels p. 21

3. Les grandes fonctions que peuvent remplir les systèmes experts p. 22

- 3.1. La sélection de bases de données p. 23
- 3.2. La connexion p. 23
- 3.3. Le "Téléchargement" p. 24
- 3.4. L'élaboration de la stratégie de recherche p. 24
- 3.5. La traduction de la stratégie en langage de commande, et sa transmission p. 24
- 3.6. Le télédéchargement et les traitements ultérieurs p. 25

4. Les produits existants p. 25

- 4.1. Passage des interfaces simples aux systèmes experts p. 25
- 4.2. Vue générale du marché p. 26
- 4.3. Caractéristiques précises de quelques produits p. 28
 - 4.3.1. TOME SEARCHER p. 28
 - 4.3.2. MOSS : un système expert d'aide à l'amélioration de la stratégie de recherche p. 29
 - 4.3.3. REX : un système expert pour la mise en page de l'information p. 30

5. Evaluation des performances d'un système expert p. 30

IV. CONCLUSION p. 33

ANNEXES p. 34

V. BIBLIOGRAPHIE p. 40

RESUME

Mots-clés : Système expert, Base de données, Interrogation de base de données, Intelligence artificielle, IA, Recherche en ligne, Stratégie de recherche.

Au cours des vingt dernières années, le nombre de bases de données servant à la recherche documentaire informatisée a incroyablement augmenté, de même que le nombre de leurs utilisateurs.

Cet état des choses, ajouté à la difficulté qu'entraîne l'élaboration d'une stratégie de recherche pour un utilisateur inexpérimenté, a conditionné le développement d'interfaces d'aide à la recherche sur bases de données.

Les plus sophistiquées de ces interfaces sont les systèmes experts, qui sont des outils capables d'imiter le raisonnement d'experts dans un domaine donné. Les systèmes experts existant actuellement sont pour la plupart encore à l'état de prototypes. Ils résolvent des problèmes tels que le choix de la base de donnée appropriée à la recherche, la connexion, la formulation de la stratégie et son amélioration, sa transmission dans le langage de commande du serveur, et la mise en page des résultats obtenus.

ABSTRACT

Keywords : Expert system, Database, Database retrieval, Artificial intelligence, AI, Online, Search strategy.

During the last two decades, the number of databases for online information retrieval has greatly increased, as well as the number of their users.

This fact, added to the difficulties that may result from the unskilled user's work toward the research strategy, conditioned the development of interfaces to computer retrieval systems.

Among these interfaces, the most sophisticated are the expert systems, they are tools able to emulate the ability of an expert in a given area. The expert systems, up to this day, are, for most of them, still at a prototype stage. They solve problems such as database selection, logging on, search strategy formulation and improvement, transmission of the strategy, downloading and post-processing.

INTRODUCTION

I. INTRODUCTION

En 1972, la commercialisation de l'information en ligne se limitait à l'existence de deux serveurs et six bases de données bibliographiques en science et technologie. Actuellement il y a plus de 4000 bases de données sur environ 600 serveurs qui couvrent toutes les disciplines.

Cette augmentation massive des sources d'information pose d'importants problèmes aux utilisateurs des bases de données, aussi bien en ce qui concerne le choix de la base de données appropriée qu'en ce qui concerne la stratégie de recherche et sa transmission dans le langage de commande correspondant. Et ce, d'autant plus que cette prolifération des bases de données s'est accompagnée d'une augmentation de leur fréquentation par des utilisateurs novices pour qui la traduction d'une question en stratégie de recherche est loin d'être facile.

Ces différents motifs ont conduit les professionnels de l'information à élaborer des interfaces entre l'utilisateur et les serveurs, qui tentent de résoudre les problèmes liés entre autres au choix de la base de données et à la formulation de la stratégie de recherche.

Les plus sophistiquées de ces interfaces sont des systèmes experts, ils apportent de façon automatique les compétences d'un spécialiste de l'information.

Nous nous proposons donc dans ce rapport de déterminer le rôle que peuvent jouer les systèmes experts dans l'aide à la recherche documentaire informatisée, de définir leur fonctions exactes, et de présenter les produits qui sont actuellement développés.

RECHERCHE BIBLIOGRAPHIQUE

II. RECHERCHE BIBLIOGRAPHIQUE

1. La recherche manuelle

Elle s'est orientée le long de trois grands axes :

- la consultation d'encyclopédies pour une définition précise des systèmes experts, et en particulier la Mc Graw-Hill Encyclopaedia of science and technology (Mc Graw-Hill, 1982), qui comprend une définition assez décevante des systèmes experts car trop théorique, et pas assez technique. Il aurait été souhaitable, si cela avait été possible de consulter aussi l'Encyclopaedia of computer science and engineering (RALSTON A. et REILLY E.D., 1983).

- la lecture de traités généraux :

- . Online retrieval analysis and strategy (VIGIL P.J., 1988), notamment le chapitre 9, qui traite plus particulièrement des systèmes experts ;

- . Expert systems in libraries (Taylor GRAHAM, 1986) intéressant dans sa totalité ;

- . Online information retrieval in practice (Taylor GRAHAM, 1986).

- le dépouillement systématique de la revue Online.

La recherche manuelle n'avait pas pour but d'être exhaustive, mais de permettre une première approche du sujet. Elle a permis de comprendre comment on en était arrivé à développer des systèmes experts dans le domaine de la recherche informatisée, elle a fourni un premier aperçu de la littérature concernant le sujet (principalement américaine et

anglaise) et elle a apporté les premières références bibliographiques (en fait peu ont été utilisées, car ces références étaient trop nombreuses et souvent anciennes).

Mais elle a surtout servi de préambule à la recherche automatisée en fournissant une première idée du vocabulaire à utiliser.

2. La recherche automatisée

2.1. Les bases utilisées

La consultation de répertoires de bases de données a permis la sélection de bases correspondant au sujet. Elles sont au nombre de six : PASCAL, INSPEC, LISA, ISA (Information Science Abstract), COMPUTER DATABASE, COMPUTER AND MATHEMATICS SEARCH.

Pour des raisons de coût et de facilité d'accès, seules deux de ces bases ont été interrogées : PASCAL et INSPEC.

2.2.1. La base PASCAL

PASCAL est une base de données multidisciplinaire à caractère scientifique. Son domaine de compétence regroupe la physique, l'ingénierie, la chimie, les sciences de la terre, les sciences de la vie, la médecine, et les sciences de l'information. Ce dernier domaine étant particulièrement approprié à notre sujet.

Elle est produite par l'INIST-CNRS depuis 1973 à partir du Bulletin signalétique du CNRS. Elle a été interrogée sur le serveur Télésystèmes avec le logiciel QUESTEL+, et sur CD-ROM.

2.1.2. La base INSPEC

INSPEC est une base de données produite par The Institution of Electrical Engineers. Elle comprend tous les documents indexés depuis 1969 dans les bibliographies imprimées : Physics Abstracts, Computer and Control Abstracts, Electrical and Electronics Abstracts. Son domaine d'application s'étend donc à l'électronique, la physique, l'informatique, et les sciences de l'information.

Elle a été interrogée sur DIRS avec le logiciel DIALOG.

La base de données Téléthèse sur CD-ROM a aussi été interrogée, mais le résultat de la recherche ne fut que de trois références dont une seule intéressante.

2.2. La stratégie de recherche

La stratégie de recherche a été assez similaire au niveau des deux bases. La différence est que l'interrogation s'est déroulée en deux temps pour ce qui concerne PASCAL, puisque la première stratégie utilisée avait ramené trop peu de références.

2.2.1. Interrogation sur PASCAL

Les termes choisis pour l'équation ont été recherchés dans le Basic Index (I.e.. dans les champs titre, résumé, descripteurs et termes génériques français, anglais, allemands, espagnols). Ce qui permettait de récupérer davantage de références concernant le sujet, mais faisait encourir le risque de ramener plus de bruit qu'avec une interrogation dans le seul champ descripteur.

La stratégie de recherche qui avait été initialement choisie était la suivante :

1. ..LIM SCIENCE INFORMATION/FG
2. SYSTEME? AV EXPERT?
3. (INTERROGATION? OU CONSULTATION?) 2AV (BANQUE? OU BASE?) 1AV DONNEE?
4. 2 ET 3

Mais cette stratégie ne permit de ramener que 9 références. Il fallait donc remplacer un terme ou en ajouter un dans l'étape 3. C'est lors de l'interrogation d'INSPEC que l'idée d'utiliser le terme "online" est apparue, puisque ce terme revenait continuellement dans les articles traitant de la recherche automatisée.

Une seconde stratégie a alors été établie, et elle a fait l'objet d'une seconde interrogation de PASCAL, mais cette fois sur CD-ROM. L'équation s'écrit alors (LI est le basic index) :

1. LI=ONLINE OU ((LI=DATABA* ET (LI=RETRIEVAL OU LI=INTERROGATION*))
2. LI=SYSTEM* ET LI=EXPERT*
- 1 ET 2

Le choix des termes de l'étape 1 s'explique d'une part par le fait que les termes anglais ont souvent une signification plus précise que les termes français, et d'autre part du fait que les opérateurs de proximité ne sont pas utilisables sur CD-ROM, et l'anglais offre l'avantage d'exprimer en un mot une notion qui en requière trois en français. (DATABA* ramène les termes database et databank).

Cette recherche s'est effectuée sur les CD-ROM 87, 88, 89, et 90 de PASCAL.

2.2.2. Interrogation d'INSPEC

L'interrogation s'est déroulée de la façon suivante :

S1 CC=C72?	(ref=36372)
S2 EXPERT()SYSTEM? ?	(ref=17419)
S3 S1 AND S2	(ref=608)
S4 S3 AND (DATABASE? ? OR DATABANK? ?)	(ref=231)
S5 S5 AND ONLINE	(ref=59)
S6 S5/1985:1991	(ref=55)

S1...S6 sont les noms d'étapes

() est un opérateur de proximité

? est une troncature illimitée

? ? est une troncature limitée.

CC=C72? permet de se limiter au domaine des sciences de l'information (dont les concept codes commencent par C72).

L'étape 6 permet de restreindre les références à la période 1985-1991.

2.3. Analyse des résultats de la recherche

L'interrogation d'INSPEC a donc permit de récupérer 55 références. L'analyse de ces références a fait ressortir que 38 étaient pertinentes, mais ce nombre pourrait être plus important si il comprenait toutes les notices de congrès complets. En effet, chaque article de congrès est catalogué, mais le congrès complet l'est aussi, ce qui fait qu'étant donné le nombre de sujets traités, les notices des congrès complets sortent systématiquement à l'interrogation (en doublons des articles de congrès), et elles ne sont pas réellement utilisables.

En ce qui concerne PASCAL, 30 références ont été obtenues (8 pour 1987, 10 pour 1988, 3 pour 1989, 9 pour 1990), mais il

faut tenir compte du fait que la recherche ne s'est effectuée que sur quatre années. Sur ces 30 références, 22 se sont révélées pertinentes.

En ne prenant en compte que les références pertinentes des deux bases, on peut constater que 9 références sont communes, ce qui est important puisque pour PASCAL cela revient à dire 40% des références.

La comparaison entre les deux bases fait ressortir que les notices d'INSPEC sont beaucoup plus détaillées et agréables à lire que celles de PASCAL :

- les références sont numérotées ;
- les ISSN, ISBN, CODEN, sont spécifiés ;
- les résumés sont beaucoup plus volumineux (ils peuvent être inexistant dans PASCAL) ;
- les champs sont plus éclatés (pour les conférences, le seul champ "source" dans PASCAL regroupe le contenu des champs "titre", "lieu", "date", "éditeur", et "source" d'INSPEC).

Par contre, dans PASCAL, les descripteurs sont dans quatre langues différentes alors qu'INSPEC se limite à l'anglais.

Le tableau suivant récapitule les résultats obtenus dans les deux bases :

	Nb de Ref	Nb de Ref pertinentes	Taux de pertinence
PASCAL	30	22	73,33%
INSPEC	55	38	69,1%

3. Sélection et obtention des articles

La sélection des articles a été assez aisée dans INSPEC, et plus difficile et aléatoire dans PASCAL puisqu'elle s'est appuyée sur les résumés (qui étaient courts, voir inexistants dans PASCAL).

La facilité d'obtention des documents a aussi influencé leur sélection, puisqu'il est très difficile par exemple de se procurer des actes de congrès étrangers (à part les congrès ONLINE).

L'obtention des documents s'est fait selon un schéma très classique : localisation des documents à l'aide du CD-ROM MYRIADE (Catalogue Collectif National), et demande de prêt-inter lorsqu'il n'était pas possible de se les procurer sur Lyon.

D'une façon générale, les demandes de photocopies d'articles ont abouti puisque seul un article n'est pas arrivé à temps (il en a été de même pour la seule thèse commandée).

4. Conclusion

L'analyse des résultats de la recherche bibliographique montre d'une part que le sujet est assez récent puisque les articles sont pratiquement tous postérieurs à 1985, et d'autre part que la littérature qui le concerne est presque exclusivement anglo-saxonne.

Si d'une façon générale le résultat de la recherche est assez satisfaisant, puisque le taux de pertinence global est de 70,6%, la qualité des documents obtenus s'est parfois

révélée décevante. En effet, le thème des systèmes experts a visiblement inspiré beaucoup d'auteurs, sans qu'ils aient pour cela à parler d'un produit précis, simplement, ils dissertent de façon très générale sur le sujet. De plus, certains articles, bien qu'indexés avec le terme "système expert" n'abordent le sujet qu'en conclusion en expliquant que ce serait une application future intéressante.

SYNTHESE

III. SYNTHESE

1. Qu'est-ce qu'un système expert ?

1.1. Définition

Un système expert c'est tout simplement un programme qui simule les compétences d'un expert humain dans un domaine donné (Peter J. VIGIL).

C'est à dire par exemple un programme qui est capable de fournir un diagnostic en médecine (MYCIN).

Les systèmes experts sont donc utilisés comme des consultants pour résoudre des problèmes complexes.

Le développement de systèmes experts représente une part importante du domaine de l'intelligence artificielle (IA), science qui se donne pour but de simuler par le biais de machines les capacités de raisonnement de l'être humain (ex: reconnaissance de la voix).

Du point de vue technique, un système expert se compose :

- d'une base de connaissance, qui comprend des faits (données factuelles) et des règles qui lient ces faits entre eux (généralement de la forme (IF...THEN...)).

- d'un moteur d'inférence qui permet de contrôler l'utilisation des règles en fonction de la situation, et qui définit donc toute la stratégie de résolution du problème posé.

- d'une interface homme-machine, qui permet la communication avec l'utilisateur et qui comporte souvent un analyseur de langage naturel.

Grossièrement, le fonctionnement de cet ensemble serait le suivant : la question de l'utilisateur est lue par le système expert et traduite par l'interface de traduction de langage naturel. La signification de la question est interprétée par le moteur d'inférence grâce à l'utilisation de la base de connaissance. Une stratégie est alors mise en oeuvre pour aboutir à une ou plusieurs solutions.

La programmation des systèmes experts peut se faire de différentes façons :

- à l'aide de langages de programmation standards (procéduraux), type PASCAL ou C.

- à l'aide de langages de programmation orientés IA (PROLOG ou LISP).

- à l'aide d'une infrastructure de système expert existant : les shell (EMYCIN, STAGE, EXPERTEASE).

- à l'aide de langages spécialisés dans la représentation des connaissances (ROSIE, KRL, OPS51).

1.2. Evolution

Lors de leur apparition, les systèmes experts ont suscité une grande vague d'intérêt, puis l'enthousiasme a baissé, lorsque l'on s'est rendu compte que si effectivement ils pouvaient s'avérer très performants dans la résolution de problèmes complexes, c'était à condition de travailler dans un domaine précis avec des applications restreintes.

De plus les interfaces de traduction de langage naturel se sont souvent révélées décevantes, ayant parfois des difficultés à évaluer la signification d'un mot selon son contexte.

Enfin la complexité de ces systèmes fait que bien souvent ceux qui sont développés restent à l'état de prototypes et nécessitent encore des améliorations.

2. Intérêt des systèmes experts dans l'aide à l'interrogation des bases de données

2.1. Les problèmes rencontrés par l'utilisateur

Le développement massif de la recherche documentaire informatisée a amené de nombreux utilisateurs à s'y intéresser. Et ils connaissent des problèmes de trois types :

- la sélection de la base de données à interroger : qu'ils soient novices ou professionnels de l'information, le nombre de bases de données et de serveurs existants rend très difficile le choix de la base de données appropriée pour la recherche envisagée. Il existe bien des répertoires imprimés, mais ils sont chers, pénibles à utiliser, et ne concernent souvent qu'un seul domaine ou un seul serveur. D'autre part, il découle de cette prolifération des bases de données une certaine lourdeur des procédures de connexion.

- la formulation de la stratégie de recherche et sa traduction dans le langage du serveur : pour un utilisateur novice, il est particulièrement difficile de passer de sa formulation propre d'une question en langage naturel, à une équation booléenne qui nécessite des connaissances en indexation et en logique booléenne. Pour tous les utilisateurs il est de toute façon impossible de connaître parfaitement tous les langages d'interrogation qui existent.

- la récupération des références : les procédures de téléchargement ne sont en effet pas toujours évidentes pour l'utilisateur novice. De plus, les formats de sortie des références sont souvent peu satisfaisants, on ne peut pas choisir le type de classement, la présentation, etc.

2.2. Pourquoi utiliser les systèmes experts ?

Les systèmes experts sont des outils de décision dans le cadre de problèmes complexes, ils sont donc tout particulièrement adaptés à des problèmes tels que la sélection de bases de données. Ils peuvent stocker les caractéristiques des différentes bases de données (coût, domaine, type de documents, etc.) dans leur base de faits, et les critères de sélection dans leur base de règles.

Les systèmes experts comprennent généralement une interface de compréhension du langage naturel, ils sont donc particulièrement à même de pallier au problème de l'utilisateur novice qui ne parvient pas à formuler de façon logique sa question.

D'une façon générale, étant donné la complexité des problèmes rencontrés lors de la recherche (traduction de la stratégie dans le langage de commande, éventuellement repositionnement dans un autre langage à une autre base, reformulation, amélioration de la stratégie, etc.) ils peuvent avoir une grande utilité et abaisser de manière sensible le coût de la recherche.

2.3. Les utilisateurs potentiels

En fait, les systèmes experts peuvent s'adresser aussi bien aux professionnels de l'information qu'aux utilisateurs inexpérimentés.

Les premiers pourront grâce aux systèmes experts proposer leur stratégie sur plusieurs bases de données ou serveurs, l'améliorer facilement, et obtenir des résultats selon un format précis.

Les seconds pourront obtenir de manière efficace les références qu'ils souhaitent, et s'initier à la recherche documentaire informatisée, en étant guidés tout au long de leur interrogation.

De nombreux processus ont autant d'intérêt pour ces deux types d'utilisateurs potentiels :

- les processus de connexion automatique ;
- la recherche systématique de synonymes ou de termes associés dans des thésaurus ;
- la sélection des bases de données.

C'est au niveau de la recherche pure que la différence des besoins entre les deux niveaux d'utilisateurs se fait sentir. Il est donc important que les systèmes experts qui travaillent sur l'élaboration de la stratégie comprennent des modules d'évaluation de l'utilisateur. Ce qui peut éviter à l'utilisateur expérimenté de passer par une série de menus dont il n'a pas besoin pour élaborer sa stratégie.

3. Les grandes fonctions que peuvent remplir les systèmes experts

Les grandes utilisations possibles des systèmes experts dans la recherche d'information sur bases de données sont présentées ci-après. ces différents aspects de la recherche informatisée ne sont pas tous traités par un même système expert. Ou si c'est le cas, ce sont des systèmes très complets, mais moins performants ponctuellement.

3.1. La sélection de bases de données

Ils remplacent alors les répertoires imprimés et les manuels informatisés des producteurs.

- Ils peuvent simplement fournir des renseignements sur telle ou telle base, ou effectuer une véritable sélection.

- Ceci est possible en stockant toutes les informations relatives aux différentes bases de données et serveurs (coût, domaine, qualité, exhaustivité, types de documents, sources, etc.) dans la base de faits du système expert.

- De tels systèmes experts sont très performants lorsqu'ils se mettent à jour en télédéchargeant directement les renseignements des répertoires informatisés des producteurs.

3.2. La connexion

Pour des utilisateurs inexpérimentés, cette procédure peut être une véritable barrière à l'utilisation de certaines bases de données. Pour les spécialistes, cette procédure est particulièrement longue et fastidieuse.

- les systèmes experts permettent donc un gain de temps important lorsqu'ils effectuent la connexion.

- ils sont aussi d'une grande efficacité puisque si un réseau est saturé, ils peuvent en essayer un autre.

- ils fournissent généralement des explications en cas d'échec.

- l'inconvénient, c'est que la connexion requière souvent un équipement supplémentaire spécifique (modem Hayes).

3.3. Le "Téléchargement"

C'est une fonction très intéressante de certains programmes qui permet d'envoyer directement une stratégie qui a été enregistrée au préalable sur micro-ordinateur. (l'élaboration de la recherche se faisant hors-ligne).

- ceci est beaucoup plus rapide que de taper la recherche manuellement.

- c'est applicable à plusieurs bases de données.

- cela permet un gain de temps, et donc un coût moindre.

3.4. L'élaboration de la stratégie de recherche

- le système expert peut formuler ou améliorer la stratégie en cas d'échec ou de faible taux de pertinence.

- la base de connaissance du système comprend alors tous les termes se rapportant au domaine traité, et leurs relations (sous forme de thésaurus par exemple).

3.5. La traduction de la stratégie en langage de commande, et sa transmission

Ce qui implique de connaître les langages de commande des bases auxquelles il accède, le nom des différents champs.

3.6. Le télédéchargement et les traitements ultérieurs

- l'intérêt étant de pouvoir alors déterminer un format personnel et de le sauvegarder pour le réutiliser. La présentation du résultat est très importante pour les données numériques par exemple.

- à ceci s'ajoute la possibilité d'enlever des références non pertinentes, d'ajouter des titres, des cadres.

- et de se créer ainsi une petite base de données bibliographique personnelle.

L'apport des systèmes experts à la recherche informatisée à donc quatre conséquences :

- simplifier le processus de recherche
- abaisser le coût de la recherche
- gagner du temps
- améliorer le résultat de la recherche.

4. Les produits existants

4.1. Passage des interfaces simples aux systèmes experts

Les premières interfaces aidant à la recherche documentaire informatisée ont été commercialisées dès 1980. Elles ont donné naissance à deux types de logiciels, les logiciels gateways, et les front-end. Les premiers étant des logiciels élémentaires permettant la connexion avec le serveur (d'où leur nom qui signifie portail), les seconds étant plus évolués puisque permettant la traduction de commandes simples dans le langage d'interrogation de la base de données. Ces

interfaces sont souvent localisées chez le serveur et ont donc pour but premier de rendre son accès plus convivial. Ce sont par exemple : EASYNET, SCI-MATE, IN-SEARCH, PRO-SEARCH.

Puis, avec la complexité des recherches, la prolifération des bases de données, et le manque de standardisation, la nécessité de développer des logiciels plus sophistiqués s'est fait sentir. Ainsi sont apparus les premiers systèmes experts en recherche informatisée. Contrairement aux interfaces précédemment citées, les systèmes experts ne se situent pas chez un serveur ou chez un intermédiaire, ils s'implantent généralement directement sur ordinateurs personnels.

Le tableau de l'annexe 1 présente un comparatif entre les capacités des interfaces précédemment citées et un système expert.

4.2. Vue générale du marché

Il n'est pas vraiment possible ici de citer tous les systèmes experts actuellement développés. D'une part à cause de leur nombre et surtout en raison du fait que la plupart des systèmes experts existants sont encore à l'état de prototypes et n'ont pas encore obligatoirement une dénomination commerciale. Les systèmes experts cités ci-après sont d'ailleurs en majorité des prototypes.

. CANSEARCH

- c'est un spécialiste de la recherche documentaire dans un domaine spécifique : le traitement du cancer ;
- il s'utilise exclusivement sur MEDLINE ;
- il est écrit en PROLOG.

. PLEXUS

- il a été développé par le CIS (Central Information Service) de l'Université de Londres ;
- il travaille sur des références dans le domaine du jardinage, sur une seule base de données ;
- il aide à la formulation de la stratégie ;
- il est écrit en PASCAL.

. TOME SEARCHER

- il est inspiré de PLEXUS, et a été développé par TOME ASSOCIATES ;
- il permet des recherches sur des bases de données en physique, électronique, informatique ;
- il comprend une interface d'analyse du langage naturel et un module d'évaluation du niveau de l'utilisateur.

. EURISKO

- il n'a pas de domaine d'application spécifique ;
- il guide l'utilisateur dans sa recherche et peut proposer sa stratégie à plusieurs bases de données ;
- il comprend une interface d'analyse du langage naturel (en français) ;
- il accède à 40 bases de données sur Télésystèmes et Cedocar.

. ANSWERMAN

Il travaille sur les bases de données de texte intégral.

. SAFIR

- c'est un système complet qui s'applique à tous les domaines ;
- il sélectionne les bases de données, formule et transmet la question, et l'améliore.

. MOSS

- c'est un outil d'amélioration des stratégies de recherche dans tout domaine ;
- il est développé avec le shell Xi PLUS.

. REX

- c'est un outil permettant la mise en page de l'information récupérée et la production de bibliographies personnalisées ;
- il est écrit en PROLOG.

4.3. Caractéristiques précises de quelques produits

4.3.1. TOME SEARCHER

Le but de TOME SEARCHER est d'aider à la recherche documentaire informatisée en utilisant plusieurs bases de données sur différents serveurs (ESA, STN, ORBIT, DIALOG). Le domaine de compétence de TOME SEARCHER s'étend à la physique, l'électronique, l'informatique, les sciences de l'information. Il a un vocabulaire très étendu, et peut donc intéresser fortement des spécialistes. Il comprend un module d'évaluation du profile de l'utilisateur qui lui permet de proposer une aide sur mesure.

Du point de vue technique, il est écrit en PASCAL. Il accepte, analyse, traduit en langage booléen la question de

l'utilisateur. Il procède ensuite à la sélection de la base de donnée appropriée, se connecte, et transmet la stratégie dans le langage d'interrogation. Un algorithme détaille ces différentes opérations en Annexe 2.

4.3.2. MOSS : un système expert d'aide à l'amélioration de la stratégie de recherche

La modification et l'amélioration de la stratégie d'interrogation est un stade critique de la recherche, et peu de logiciels traitent efficacement ce problème, sauf dans des domaines précis (Grateful Med).

MOSS permet donc d'optimiser les stratégies de recherche. Il est développé sur le shell Xi PLUS mais sera bientôt implanté sur Léonardo (plus performant). MOSS peut proposer des stratégies à différentes bases de données.

Il fonctionne en distinguant deux cas de figure : celui où l'utilisateur a obtenu peu ou pas de références, et celui où il en a obtenu trop.

Dans le premier cas, il procède en :

- ajoutant des synonymes des termes utilisés
- ajoutant ou remplaçant ces termes par des termes au sens plus large
- soustrayant des concepts trop restrictifs

Dans le second cas :

- il crée des intersections avec de nouveaux concepts (opérateur ET)
- il utilise des termes équivalents mais ayant un sens plus précis
- il utilise du vocabulaire contrôlé
- il limite la recherche de termes à des champs précis

- il corrige les troncatures erronées
- il utilise l'opérateur SAUF pour faire des restrictions.

Le fonctionnement de ce système expert est détaillé dans le schéma de l'annexe 3.

4.3.3. REX : un système expert pour la mise en page de l'information

REX (Reference EXpert system) transforme de simples listings de références obtenus par téléchargement en sources d'information sophistiquées. Son grand intérêt est de créer une bibliographie unique à partir de plusieurs recherches dans des bases de données différentes. Il fonctionne en créant sa propre base de données, puis il travaille dessus grâce à sa base de connaissance, qui comprend des règles de classification, d'indexation, propres à chaque sujet. Il fournit ensuite une bibliographie et des index associés.

II est écrit en PROLOG.

Un exemple de la production du système REX est présenté en annexe 4.

5. Evaluation des performances d'un système expert

Plusieurs critères peuvent permettre d'évaluer la qualité d'un système expert :

- les performances de l'interface de compréhension du langage naturel :

- . capacité à accepter n'importe quelle entrée : phrase, mot, liste (la plupart des analyseurs ne travaillant que sur des phrases).

- . capacité à se débrouiller intelligemment avec des termes inconnus (listes de suffixes et de préfixes dans la base de connaissance).

- adaptation à tous les utilisateurs potentiels, convivialité :

- . pas de blocage en cas d'erreur, mais des messages d'erreur explicites.

- . pas de manuel d'utilisation nécessaire.

- . un nombre minimum de menus (lassant).

- . un domaine de connaissance important pour une utilisation par des spécialistes.

- . l'existence d'un module d'évaluation du profile de l'utilisateur.

- . explications sur la démarche utilisée.

- la performance de la recherche :

- . nécessité d'une bonne connaissance des techniques d'interrogation : opérateurs booléens, troncatures, opérateurs de proximité, utilisation des thésaurus pour modifier la recherche.

- . nécessité d'une bonne connaissance de la structure et de la terminologie employées dans le domaine en question.

CONCLUSION

IV. CONCLUSION

Dans le domaine de la recherche informatisée, les systèmes experts sont les premiers logiciels à prendre en compte tout ce que considérerait un spécialiste de l'information.

Si le développement de ces systèmes experts a débuté en 1983, ce n'est que récemment que sont apparus des produits vraiment performants. C'est pourquoi la plupart des systèmes experts ne sont encore que des prototypes, et qu'il existe peu de produits commerciaux.

Différentes raisons pourraient constituer des freins à leur commercialisation : d'une part le fait que de nombreux intermédiaires les considèrent non comme des outils complémentaires, mais comme des rivaux, et d'autre part, l'inquiétude des serveurs face à une baisse de leur revenus, du fait du gain de temps en ligne qui résulterait de leur utilisation.

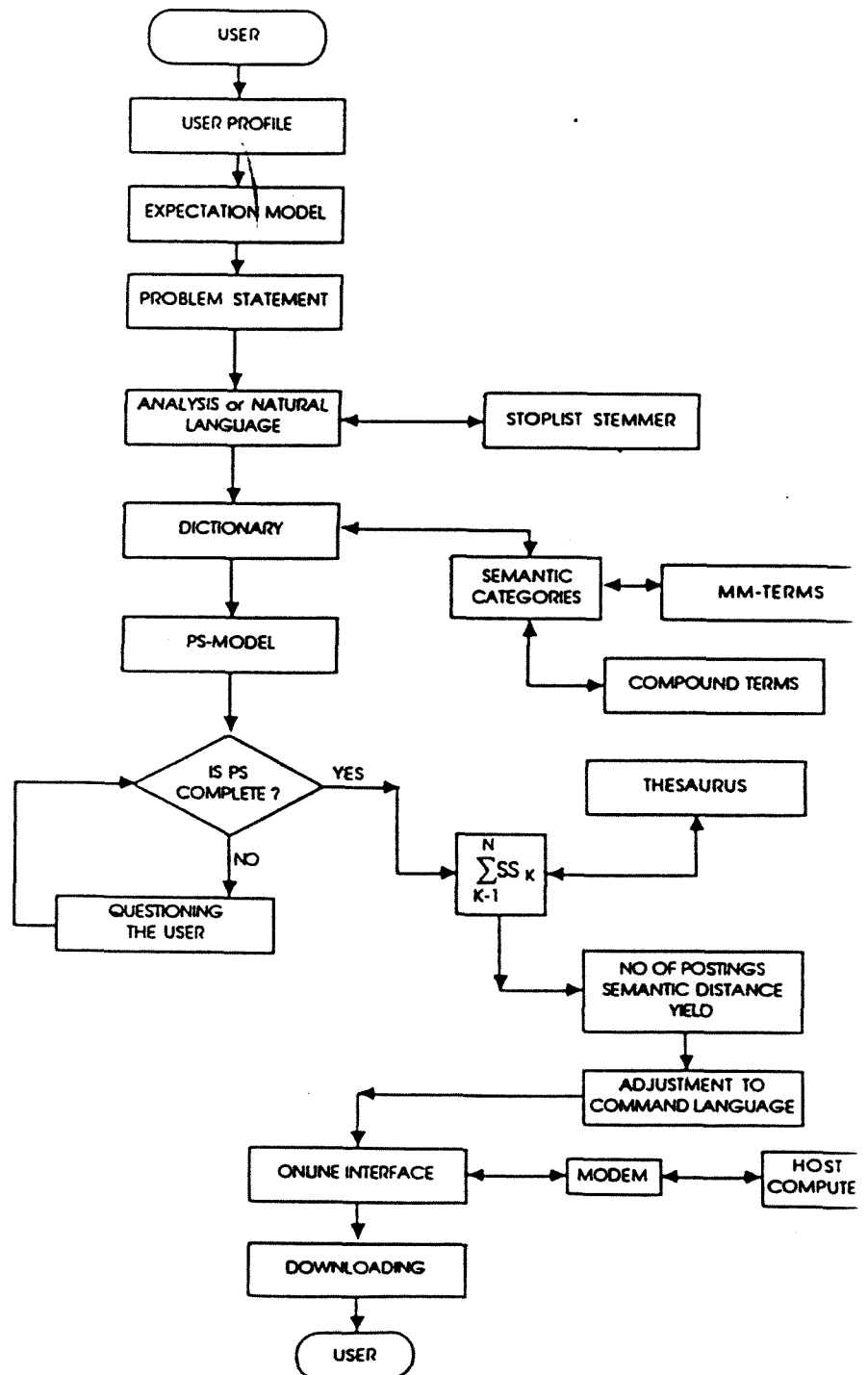
Quoi qu'il en soit, la majorité des systèmes experts actuellement développés nécessitent encore de nombreuses améliorations qui pourraient s'obtenir en consultant encore plus de spécialistes humains, et en essayant d'acquérir une meilleure connaissance des utilisateurs finaux.

ANNEXES

ANNEXE 1

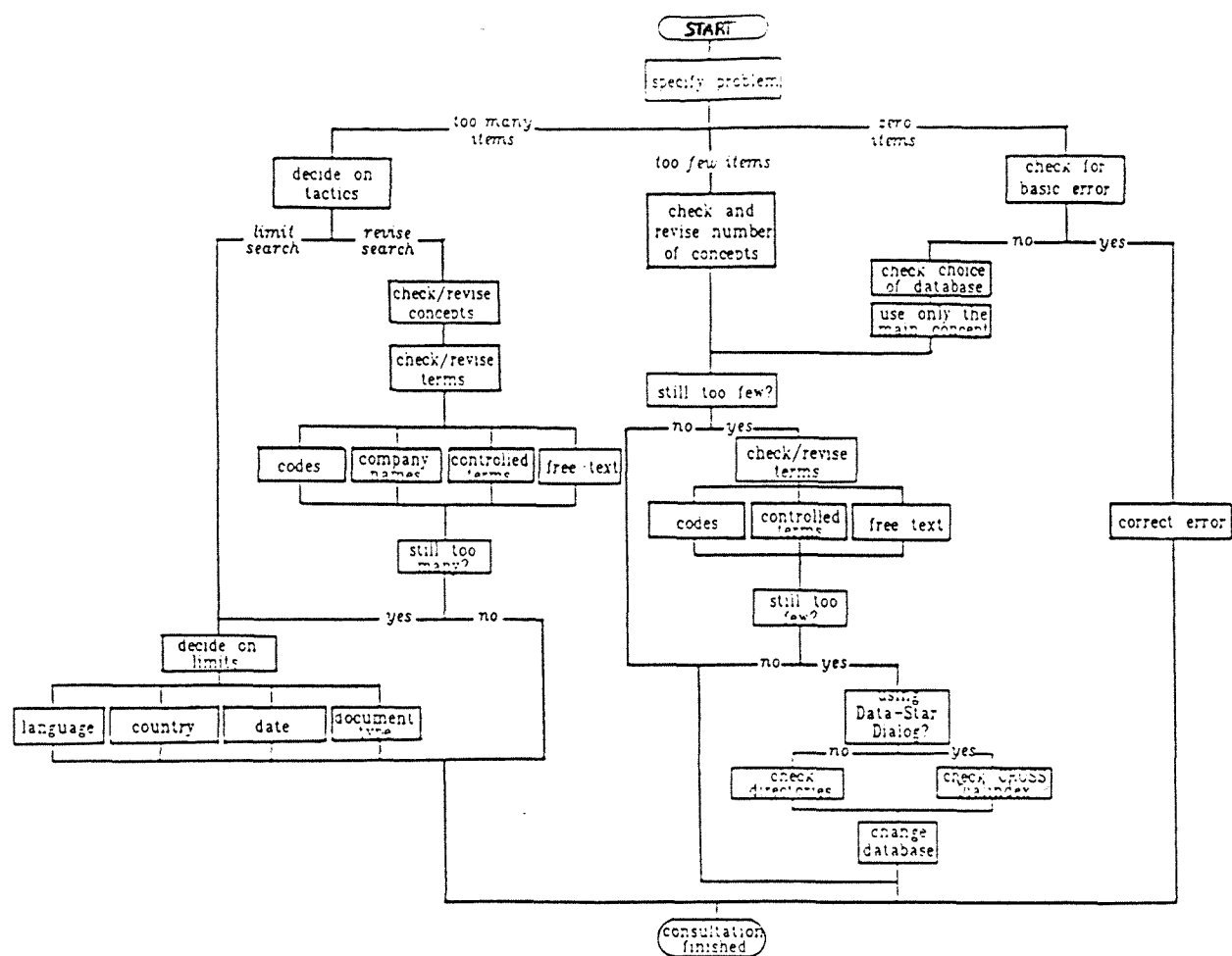
	EURISKO	SCI-MATE	PROSEARCH	EASYNET
request formulation with search steps	by EURISKO	by the user		
logic	known by EURISKO	known by the user		
database organisation	known by EURISKO	known by the user		
boolean operators	yes			
truncations	yes	no	yes	
proximity operator	yes			no
restriction to indexes and fields	yes			no
downloading	yes			
database selection	yes			no
extension, update	by user and producer	by the producer		
feedback process	in progress	no		

Comparatif entre différentes interfaces
et le système expert EURISKO

ANNEXE 2

Algorithme de traitement du système expert
TOME-SEARCHER

ANNEXE 3



Algorithme du système expert MOSS

ANNEXE 4

1, Aaby, A., SPECIFICATON OF DISTRIBUTED PROGRAMS USING TEMPORAL INTERVAL LOGIC, PH.D. THESIS, THE PENNSYLVANIA STATE UNIVERSITY 146pp., 1988

2, Adey, S., PLANNING AND EXECUTING PLANS WITHIN A NAVAL SIMULATOR, Alvey Expert Systems Research Theme, Report of the 6th Workshop of the Planning Special Interest Group, pp. 1-7,2-3 April 1987 Cambridge, UK, publ.: Alvey Directorate, London, UK. 128 pp., 1987

3, Alim, S., PROLOG-BASED EXPERT SYSTEM FOR ENCODING SEISMIC DESIGN PROVISIONS, Civil Engineering Systems, v4, n 1, Mar 1987, pp. 39-44, 1987

4, Antonov, A., Ratchev, B., Iwanov, S., Nikolov, N., Nikolov, W., Stojtshev, L., COMPUTER ARCHITECTURE AND SOFTWARE DESIGN OF A IS-TRIBUTED DIGITAL INTEGRATING MACHINE, Distributed Intelligence Systems: Methods and Applications, Selected papers from the IFAC/IMACS Symposium, pp. 221-6, 27 June-1 July 1988 Varna, Bulgaria, publ.: Pergamon, Oxford, UK. x+360 pp., 1989

5, Armstrong, J.L., Hibbert, D.B., REPRESENTATION AND MATCHING OF CHEMICAL STRUCTURES BY A PRO-LOG PROGRAM, Journal of Chemical Information and Computer Science, v 29, n 2, pp. 51-60, May 1989

Exemple de bilbiographie produite par
le système expert REX

Aaby, A.	1	Baba, M.F.	10
Aczel, M.	84	Bader, J.	11
Adams, F.	158	Baldwin, J.F.	12,13,29
Adey, S.	2	Baracos, P.	315
Agre, G.	276	Barcsak, J.	152
Akgiray, V.	293	Barker, H.A.	14,15
Alexandro, F.J.	235	Barnes, B.H.	30
Alim, Seema	3	Barry, J.	16
Alp, P.	75	Barthes, J.-P.A.	17
Ambrose, E.	31	Becker, B.-D.	18
Antaloczy, Z.	152	Becker, G.	6,7
Antonov, A.	4	Bell, M.G.H.	256
April, G-E.	33	Ben-Arieh, D.	19,20,21
Armstrong, J.L.	5	Bertin, B.	22
Armstrong, M.P.	204	Bharath, R.	23,24
Arnoux, M.	6,7	Bibel, W.	25
Arsentyeva, A.V.	8	Bihan, P.	26
Asthana, A.	177		
Auer, A.	9		

Exemple d'index auteur produit par
le système expert REX

BIBLIOGRAPHIE

V. BIBLIOGRAPHIE

La bibliographie comporte de nombreuses références. Ce sont pour l'essentiel des références d'actes de congrès et d'articles de périodiques. Ces références ont été classées selon deux critères de façon à fournir une bibliographie la plus lisible possible. Ce sont par ordre de priorité :

- par type de document : thèses, monographies, actes de congrès, articles de périodiques.
- par auteur.

THESES

1 - BARTHES C.- Spécificaton du contrôle d'un raisonnement en monde évolutif : application à la recherche documentaire en ligne.- Thèse de doctorat, Université de Toulouse, 1989, 240 p.- bibliogr. (216 ref).

2 - METAIS E.- Les mécanismes d'inférence et l'explication du raisonnement dans le système expert SECSI.- Thèse de doctorat d'état, Université de Paris 6, 1987.

3 - REYSSIER DANZART A.- SEREBRAL : un système expert pour l'aide à l'interrogation d'une base de données.- Thèse de doctorat, Université de Paris 11, 1985.

MONOGRAPHS

4 - DORRINGTON L., ed.- Online information retrieval in practice.
UK Online User Group.- Londres (U.K.) : Taylor Graham, 1987,
158 p.

5 - FORBES G., ed.- Expert systems in libraries.- Londres
(U.K.) : Taylor Graham, 1986, 97 p.

6 - VIGIL P.J.- Online retrieval analysis and strategy.-
Chicago (U.S.A.) : John Wiley, 1988, 242 p.

COMPTE-RENDUS DE CONGRES

7 - BARKER F.H.- Does any one of the 'user friendly' systems deserve a prize?. 12th International Online Information Meeting, Londres (U.K.), 6-8 déc. 1988.- Oxford (U.K.) : Learned Inf, 1988, vol. 1, p.121-37.- bibliogr. (1 réf.).

8 - BARTHES C.; FRONTIN J.; GLIZE P.- EURISKO : an artificial intelligence tool for automatic online retrieval. 11th International Online Information Meeting, Londres (U.K.), 8-10 déc. 1987.- Oxford (U.K.) : Learned Inf, 1987, p.431-42.- bibliogr. (11 réf.).

9 - BARTHES C., et al.- La connaissance experte d'Eurisko pour la recherche d'information en ligne. Congrès sur l'information et la documentation, Strasbourg (France), 5 déc. 1987.- [S.L.] : [S.N.], 1987, p.183-187.- bibliogr. (4 réf.).

10 - BIANCHI G.; GIORGI M.- Towards an expert system as intelligent assistant for the design of an online documentation service. 10th International Online Information Meeting, Londres (U.K.), 2-4 déc. 1986.- Oxford (U.K.) : Learned Inf, 1986, p.199-208.

11 - BOYLE T.; MITCHELL B.- Developing effective dialogue management in an expert front end for online information retrieval. Proceeding of the 11th BCS IRSG Research Colloquium on Information Retrieval, Huddersfield (U.K.), 5-6 juil. 1989.- Huddersfield : Huddersfield Polytech, 1989, p.255-63.- bibliogr. (17 réf.).

12 - COYLE K.; GALLAHER-BROWN L.- Record matching : an expert algorithm. Proceeding of the 48th ASIS Annual Meeting, Las Vegas (U.S.A.), 20-24 oct. 1985.- White Plain (U.S.A.) : Knowledge Ind., 1985, p.77-80.- bibliogr. (4 réf.).

13 - GAUCH S.- Evaluation of an expert system for searching in full text. Proceeding of the 13th International Conference on

Research and development in Information Retrieval, Bruxelles (Belgique), 5-7 sept. 1990.- New-York (U.S.A.) : ACM, 1989, p.255-78.- bibliogr. (23 réf).

14 - GAUCH S.; SMITH J.B.- Query reformulation search strategies for an intelligence search intermediary. Proceedings of the Annual AI Systems in Government Conference, Washington (U.S.A.), 27-31 mars 1989.- Washington : IEEE Comput. Soc. Press., 1989, p.65-71.- bibliogr. (14 réf).

15 - GAUCH S.; SMITH J.B.- Intelligent search of full-text databases. RIAO 88 Program, Cambridge (U.S.A.), 21-24 Mars 1988.- Paris (France) : C.I.D., 1988, p.162-71.- bibliogr. (13 réf).

16 - JANCARIKOVA D.- Expert system as an online search intermediary in interactive retrieval from databases. Application of artificial intelligence '89, Proceeding, Pragues (Tchécoslovaquie), 24-26 avril 1989.- Pragues : Ustav pro Informacni v Kulture, 1989, p.345-52.- bibliogr. (15 réf).

17 - JENSEN R.B.- Intelligent interfaces to Nordic Databases. Knowledge engineering - Expert Systems and Information Retrieval, Copenhagen (Danemark), déc. 1986.- Londres (U.K.) : Taylor Graham, 1987, p.74-91.- bibliogr. (16 réf).

18 - KINNEL S.K.; CHIGNELL M.H.- Who's the expert? : conceptual representation of knowledge for end user searching. National Online Meeting, Proceedings - 1987, New-York (U.S.A.), 5-7 mai 1987.- Medford (U.S.A.) : Learned Inf, 1987, p.237-43.- bibliogr. (3 réf).

19 - MCALEESE R.; DUNCAN E.B.- The graphical representation of 'terrain' and 'street' knowledge in an interface to a database system. 11th International Online Information Meeting, Londres (U.K.), 8-10 déc. 1987.- Oxford (U.K.) : Learned Inf, 1987, p.443-56.- bibliogr. (13 réf.).

20 - MORRIS A.; TSENG G.; NEWHAM G.- The selection of online databases and hosts-an expert systems approach. 12th International Online Information Meeting, Londres (U.K.), 6-8 déc. 1988.- Oxford (U.K.) : Learned Inf, 1988, vol. 1, p.139-48.- bibliogr. (11 réf.).

21 - REITER M.- Improving online information retrieval with an intelligent front-end system. 13th International Online Information Meeting Proceedings, Londres (U.K.), 12-14 déc. 1989.- Oxford (U.K.) : Learned Inf, 1989, p.597-604.- bibliogr. (2 réf.).

22 - SIITONER L.- Research on online searching by end users : implications for design and development of bibliographic databases. Bibliographic Databases and Networks, New-Delhi (India), 22-25 fev. 1989.- New-Delhi : Tata McGraw-Hill, 1989, p.1-15.- bibliogr. (24 réf.).

23 - SORMUNEN E.- A knowledge-based intermediary system for information retrieval. Knowledge engineering - Expert Systems and Information Retrieval, Copenhagen (Danemark), déc. 1986.- Londres (U.K.) : Taylor Graham, 1987, p.59-73.- bibliogr. (23 réf.).

24 - VICKERY A.- The experience of building expert search systems. 12th International Online Information Meeting, Londres (U.K.), 6-8 déc. 1988.- Oxford (U.K.) : Learned Inf, 1988, vol. 1, p.301-13.- bibliogr. (3 réf.).

25 - WATERS S.T.- Answerman : expert systems for librarians. Online '88 Conference Proceedings, New-York (U.S.A.), 10-12 oct. 1988.- Weston (U.S.A.) : Online, 1988, p.148-52.

ARTICLES DE PERIODIQUES

26 - BARTHES C.; GLIZE P.- A case study of planning in information retrieval. Expert Systems for Information Management (U.K.), 1988, vol. 1, n°1, p.50-65.- bibliogr. (22 réf.).

27 - BROOKS H.M. et al .- Research of information interaction and intelligent information provision mechanisms. Journal of Information Science, Principles and Practice (PAYS-Bas), 1986, vol.12, n°1-2, p.37-44.- bibliogr. (16 réf.).

28 - CREMERS A.B. et al.- KONDOR a knowledge based intermediary system for individual optimization of online search strategies. Nachrichten fur Dokumentation (Allemagne), août 1988, vol.39, n°4, p.257-61.- bibliogr. (20 réf.).

29 - CUPPENS F.; DEMOLOMBE R.- How to recognise interesting topics to provide cooperative answering. Information systems (U.K.), 1989, vol. 14, n°2, p.163-173.- bibliogr. (9 réf.).

30 - DATTERO R. et al.- Derived relations with exceptions : decision support capabilities. Journal of Management Information Systems (U.S.A.), 1990, vol.6, n°4, p.83-101.- bibliogr. (21 réf.).

31 - DETEMPLE W.- Future enhancements for full text databases. Online review (U.K.), avril 1989, vol. 13, n°2, p.155-60.- bibliogr. (23 réf.).

32 - DIERICK J. et al.- Online information retrieval, weldra zonder problem voor iedereen, via een expert systeem. Bibliotheek- en archiefgids (Belgique), 1987, vol.63, n°3, p.202-211.- bibliogr. (7 réf.).

33 - DUNNING P.- How much does the end user need an intermediary. Online review (U.K.), juin 1989, vol. 13, n°3, p.207-17.- bibliogr. (8 réf.).

34 - ERCEGOVAC Z.- Augmented assistance in online catalog subject searching. Reference Librarian (U.S.A.), 1989, n°23, p.21-40.- bibliogr. (92 réf.).

35 - FENICHEL C.H.; MURPHY J.J.- Using a microcomputer to communicate. Microcomputers for Information Management (U.S.A.), sept. 1985, vol.2, n°3, p.155-70.-- bibliogr. (12 réf.).

36 - GAUCH S.; SMITH J.B.- An expert system for searching in full-text. Information Processing and Management (U.K.), 1989, vol.25, n°3, p.253-63.- bibliogr. (18 réf.).

37 - GOLDSMITH G.; WILLIAMS P.W.- Online searching made simple : a microcomputer interface for inexperienced users. Library and information research report (U.K.), 1986, n°41.- bibliogr.

38 - GROSS D.- Applications of AI technology in online database services. Online review (U.K.), juin 1989, vol. 13, n°3, p.207-17.- bibliogr. (8 réf.).

39 - GROSS D.- AI online : improving via expert systems. Information today (U.S.A.), 1987, vol.4, n°7, p.7-17.

40 - HAWKINS D.T. et al.- Knowledge gateways : the building blocks. Information Processing and Management (U.K.), 1988, vol.24, n°4, p.569-68.- bibliogr. (20 réf.).

41 - HAYNES D.- TOME-SEARCHER : intelligent interface for searching and retrieving information. Computers and Libraries (U.K.), août 1988, vol.2, n°1, p.2-3.

42 - KEHOE C.A.- Interfaces and expert systems for online retrieval. Online review (U.K.), déc. 1985, vol.9, n°6, p.499-92.

43 - KLAUSS H.G.- Online-Dienst und mehr : Neuigkeiten für den Informationsprofi. Nachrichten für Dokumentation (Allemagne), 1988, vol.39, n°1, p.44-46.

44 - LIROV Y.; LIROV V.- Online search + logic programming = subject bibliography : an expert system approach to bibliographic processing. Online review (U.K.), oct. 1988, vol. 12, n°5, p.283-9.

45 - MAGLIOLA M.; DRACOS A.- Trends in online research facilitated mechanisms. Contemporary topics in information transfert (Pays-Bas), 1987, vol.5, p.167-171.

46 - MICCO M.; SMITH I.- Designing a workstation for information seekers. Reference Librarian (U.S.A.), 1989, n°23, p.135-52.- bibliogr. (12 réf.).

47 - MILNE R.- Artificial intelligence for online diagnosis. IEEE Proceedings (U.K.), 1987, vol.134, n°4, p.238-244.- bibliogr. (21 réf.).

48 - MOUREAU M.- Cost and know-how : the 'Matthew effect' in information retrieval. Online review (U.K.), 1987, vol. 112, n°6, p.355-360.

49 - OPPENHEIM C.- Online information services : presents plans and future prospects. 'The Pergamon Infoline View'. International Journal of Micrographics and Video Technology (U.K.), 1985, vol.4, n°2, p.85-9.

50 - SCEMANA C.- La journée de l'Association des Documentalistes et Bibliothécaires Spécialisés sur les systèmes experts. Bases (France), 1987, n°17, p.1-3.

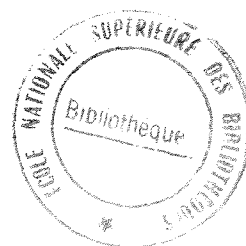
51 - STAUD J.L.- Das Expertensystem für Faktenretrieval ES-Fakt. Nachrichten für Dokumentation (Allemagne), 1987, vol.38, n°5, p.271-281.

52 - DE STRICKER U.- A menu interface to formulate Boolean logic can it be done?. Information Services and Use (Pays-Bas), 1988, vol.8, n°1, p.39-46.

53 - SUN QI-ZH.- Expert simulation for online production control. Computer integrated manufacturing systems (U.K.), 1989, vol.2, n°3, p.172-180.- bibliogr. (9 réf.).

54 - TRAUMAN R.; von Flittner S.- An expert system for microcomputers to aid selection of online databases. Reference Librarian (U.S.A.), 1989, n°23, p.207-38.- bibliogr. (25 réf.).

55 - WATERS S.T.- Answerman, the Expert Information Specialist an expert system for retrieval of information from library reference books. Information Technology and Libraries (U.S.A.), sept. 1986, vol.5, n°3, p.204-12.- bibliogr. (11 réf.).



BIBLIOTHEQUE DE L'ENSSIB



801613C