

E.N.S.S.I.B
ECOLE NATIONALE SUPERIEURE
DES SCIENCES DE L'INFORMATION
ET DES BIBLIOTHEQUES

UNIVERSITE
CLAUDE BERNARD
LYON I

DESS en INFORMATIQUE DOCUMENTAIRE

Rapport de recherche bibliographique

Les logiciels de veille technologique et stratégique

Michel CUEVAS

Sous la direction de
Madame Monique JOLY

A.R.I.S.T
Agence Régionale pour l'Information Scientifique et Technique

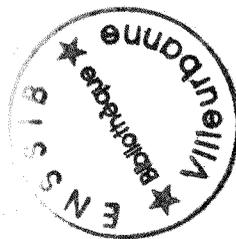
1995

E.N.S.S.I.B
ECOLE NATIONALE SUPERIEURE
DES SCIENCES DE L'INFORMATION
ET DES BIBLIOTHEQUES

UNIVERSITE
CLAUDE BERNARD
LYON I

DESS en INFORMATIQUE DOCUMENTAIRE

Rapport de recherche bibliographique



Les logiciels de veille technologique et stratégique

Michel CUEVAS

Sous la direction de
Madame Monique JOLY

A.R.I.S.T
Agence Régionale pour l'Information Scientifique et Technique

1995

1995

ID

22

LES LOGICIELS DE VEILLE TECHNOLOGIQUE ET STRATEGIQUE

MICHEL CUEVAS

RESUME : Les documentalistes ont longtemps assumé une fonction de centralisateur et de diffuseur d'information auprès des autres acteurs de l'entreprise. Mais de profonds changements tant au niveau des sources d'information que des technologies transforment leur rôle d'intermédiaire en organisateur mettant en oeuvre les solutions adéquates qui permettent des accès directs et rapides à l'information. Cette étude présente les principes de base de l'activité de veille et décrit la typologie de l'offre logicielle actuelle.

ABSTRACT : Information scientists have long enjoyed gatekeeper status in their organization providing service as intermediary between information and the people who need it. Changing information sources and technologies are transforming the role of intermediary to one of enabler, providing the mechanisms to supply quicker and more direct access to information. This article provides an overview of competitive intelligence system and of the software products available nowadays.

KEYWORDS : COMPETITIVE INTELLIGENCE / BUSINESS INTELLIGENCE

TABLE DES MATIERES

A) LA STRATEGIE DE RECHERCHE	1
I) Les objectifs.	2
II) La recherche documentaire.	2
II.1) La démarche adoptée.	2
II.2) Le choix des sources d'information.	2
II.3) La recherche exploratoire.	2
II.4) Les recherches ciblées.	3
III) L'analyse des résultats.	4
III.1) L'efficacité nuancée des descripteurs.	4
III.2) La qualité des réponses.	5
B) LA SYNTHÈSE.	6
I) Préambule.	7
II) Introduction.	8
III) Définition d'un système de veille.	9
III.1) La collecte de l'information.	9
III.1.1) Les sources.	9
III.1.2) La sélection.	10
III.1.3) Le stockage.	10
III.2) L'analyse de l'information.	10
III.2.1) L'interrogation ponctuelle.	11
III.2.2) La synthèse.	11
IV) L'offre logicielle.	12
IV.1) Les systèmes de base d'informations.	12
IV.1.1) AskSam 2.0 (de AskSam Systems).	12
IV.1.2) Folio Views (de Folio Corp).	13
IV.1.3) Wincite (de Quest Management System).	13
IV.2) Les gestionnaires de documents.	14
IV.2.1) SoftSolutions (de Novell Inc).	15
IV.2.2) PC DOCS Open (de PC DOCS Inc).	15
IV.3) Les agents intelligents.	16
IV.3.1) Hoover (de Sandpoint Corp).	16
IV.3.2) Topic (de Verity Inc).	17
V) Bibliographie.	18

LA STRATEGIE DE RECHERCHE

I) Les objectifs.	2
II) La recherche documentaire.	2
II.1) La démarche adoptée.	2
II.2) Le choix des sources d'information.	2
II.3) La recherche exploratoire.	2
II.4) Les recherches ciblées.	3
III) L'analyse des résultats.	4
III.1) L'efficacité nuancée des descripteurs	4
III.2) La qualité des réponses.	4

I) Les objectifs.

Cette étude vise à dresser un état de l'offre mondiale en matière de logiciels de veille technologique et stratégique en insistant particulièrement sur les Etats-Unis.

Les objectifs de cette recherche sont donc d'une part d'identifier les produits commercialisés à ce jour, et d'autre part de collecter des informations détaillées et critiques sur leurs caractéristiques et performances. Il ne s'agit donc pas d'établir une liste de références bibliographiques exhaustive mais d'isoler les documents les plus pertinents en vue du travail de synthèse à produire.

II) La recherche documentaire.

Cette étude étant orientée vers les produits anglo-saxons, ce sont les bases de données américaines qui sont logiquement choisies pour effectuer les recherches.

D'autre part, compte-tenu de l'obsolescence rapide que connaît la production logicielle c'est l'information récente disponible sur des bases de données on-line qui est privilégiée.

II.1) La démarche adoptée.

Une recherche préliminaire sera réalisée afin d'apprécier le volume d'informations disponible sur ce type de produit. Elle aura donc une visée exploratoire et sera menée d'une manière suffisamment large afin de récolter une masse substantielle de références en fonction de laquelle seront élaborées les étapes ultérieures de recherche.

II.2) Le choix des sources d'information.

Deux types de source sont visées; d'une part la presse spécialisée dans l'information professionnelle afin d'identifier les nouveaux produits commercialisés et d'autre part, les revues scientifiques afin de connaître les travaux de recherche récents.

Le serveur DIALOG fournit ces différentes informations :

- les bases *Business Software Database (256)* et *Computer Database (275)* sont des bases bibliographiques d'articles de presse qui proposent un résumé des articles référencés.

- les bases *Computer News Fulltext (674)*, *Computer ASAP (675)* et *Datapro Reports and Analysis (753)* sont des bases d'articles en texte intégral. Les plus intéressants dans le cadre de l'étude sont les analyses comparatives de produits.

- la base *Inspec (4)* est une base bibliographique scientifique qui propose également des résumés d'articles.

III.1) La recherche exploratoire.

Elle vise principalement deux objectifs : d'une part appréhender la largesse de l'offre et d'autre part vérifier la nature des descripteurs anglais associés à cette fonction.

Elle est menée sur les bases bibliographiques à vocation professionnelle :

Business Software Database (256) et *Computer Database (275)*.

La première requête cible résolument l'activité avec pour descripteurs :
COMPETITIVE(W)INTELLIGENCE et BUSINESS(W)INTELLIGENCE

Les résultats sont les suivants:

COMPETITIVE(W)INTELLIGENCE AND [5]
BUSINESS(W)INTELLIGENCE [2]

Le taux de réponse est peu concluant aussi une autre requête est préparée qui contrairement à la précédente se veut très large. Les mots-clés utilisés sont assez généraux mais conjugués ils permettent de cerner l'essentiel de l'activité de veille.

Le nombre de références trouvées pour chaque descripteur est indiqué entre crochets :

INFORMATION? AND ([15.469]
COLLECT? OR	[2.985]
GATHER?) AND ([430]
BUSINESS? OR	[8.917]
COMPETIT? OR	[2.363]
STRATEG? OR	[2.289]
MARKET?)	[5.254]

la réponse finale propose : [73] références.

A priori le nombre de références semble satisfaisant mais le dépouillement fait apparaître un faible taux de réponses pertinentes. L'analyse des résumés ne permet de retenir que neuf (9) références correspondant à l'activité étudiée. Le bruit de fond est donc très important, alimenté par des produits concernant des activités autres, soit financières (banque, assurance), soit commerciales ou comptables et il faut donc en conclure que la conjugaison des termes utilisés ne constitue pas un critère de recherche suffisamment spécifique. De plus il ressort que plusieurs produits ne traitent qu'en partie le sujet, ce qui fait ressortir une lacune au niveau de la requête qui omet de cibler la fonction de restitution de l'information.

L'analyse des descripteurs des quelques références pertinentes permet de repérer des concepts clés tels que : 'RETRIEVAL SYSTEM' ; 'DECISION SUPPORT' ; 'EXECUTIVE INFORMATION SYSTEM'. De nouvelles requêtes sont alors réalisées, complétées de ces descripteurs. Elles permettent une nette réduction du bruit de fond mais aboutissent aussi à une perte significative des références pertinentes isolées précédemment.

A l'issue de cette série de requêtes, il apparaît qu'il n'est pas possible, sur ce type de bases, d'isoler avec une seule interrogation un ensemble complet et homogène de références. Dans ces conditions, poursuivre dans cette voie supposerait de multiplier le nombre des interrogations en utilisant à chaque fois des descripteurs différents.

III.2) Les recherches ciblées.

Une autre voie est donc explorée qui consiste à rechercher dans les bases d'articles, les neuf (9) produits repérés durant l'étape précédente, avec pour objectif d'obtenir des documents d'analyses détaillées. A cette fin, le terme 'EVALUATION' est croisé à la liste des produits dans une requête ciblant la principale base bibliographique professionnelle :

Computer Database.

Les résultats sont les suivants:

EVALUATION AND ([93.592]
VIEWPOINT OR	[3]
ARRANGE OR	[6]
DATALOGIC OR	[4]
DATAPIVOT OR	[9]
TOPIC OR	[5]
ASKSAM OR	[9]
WINCITE OR	[0]
HOOVER OR	[21]
FOREST TREE)	[6]

la réponse finale propose : [18] références

Outre les éléments d'analyse qu'elles procurent, ces références sont intéressantes car elles mentionnent souvent des produits non encore identifiés, et qui à leur tour alimentent une nouvelle recherche ciblée.

L'analyse de ces références et ce cheminement progressif amènent ainsi à réviser la liste des produits pressentis et permettent d'appréhender la diversité de l'offre logicielle.

Enfin, un dernier type de recherche est réalisé qui vise des articles de recherche de façon à mesurer la présence des logiciels commercialisés dans des publications scientifiques. C'est donc la base *Inspec (4)* qui est ciblée avec les descripteurs les plus précis..

Les résultats sont les suivants:

COMPETITIVE(W)INTELLIGENCE	[51]
PY=1995 OR 1994	[222.340]

la réponse finale propose : [10] références

Le taux de réponse est tout à fait satisfaisant et reflète sans doute la plus grande efficacité du système d'indexation mis en oeuvre pour cette base mais s'explique aussi par la nature et la qualité des documents qui y sont référencés.

III) L'analyse des résultats.

III.1) L'efficacité nuancée des descripteurs.

Les descripteurs spécifiques à la veille, tels que COMPETITIVE INTELLIGENCE ou BUSINESS INTELLIGENCE sont quasiment inefficaces lorsqu'ils sont appliqués à la presse professionnelle. Il est possible d'invoquer la qualité assez médiocre de l'indexation des bases traitées mais l'explication essentielle réside probablement dans l'hétérogénéité d'une offre logicielle qu'il est donc très difficile de cataloguer.

En revanche cette confusion disparaît totalement dès lors que les articles référencés ne sont pas de nature professionnelle mais scientifique.

III.2) La qualité des réponses.

Les résultats des recherches s'apprécient différemment selon les bases interrogées.

a) Sur les bases professionnelles, le bruit de fond est toujours très important. Toutes les tentatives pour le réduire se sont soldées par une perte significative de l'information pertinente et ce, que la requête soit très spécifique ou au contraire très large. Il a donc fallu adopter une démarche plus pragmatique, de petits pas qui a consisté à réaliser des recherches ponctuelles au fil des lectures des références récupérées.

b) Sur la base scientifique *Inspec* en revanche, la qualité des réponses est importante et les informations obtenues sont d'une nature tout à fait complémentaires de celles issues de la presse professionnelle.

LA SYNTHÈSE

I) Préambule.	7
II) Introduction.	8
III) Définition d'un système de veille.	9
III.1) La collecte de l'information.	9
III.1.1) Les sources.	9
III.1.2) La sélection.	10
III.1.3) Le stockage.	10
III.2) La restitution de l'information.	10
III.2.1) L'interrogation ponctuelle.	10
III.2.2) Les mises en forme.	11
III.3) La diffusion de l'information.	11
IV) L'offre logicielle.	12
IV.1) Les systèmes de base d'informations.	12
IV.1.1) AskSam 2.0 (de AskSam Systems).	12
IV.1.2) Folio Views (de Folio Corp).	13
IV.1.3) Wincite (de Quest Management System).	13
IV.2) Les gestionnaires de documents.	14
IV.2.1) SoftSolutions (de Novell Inc).	15
IV.2.2) PC DOCS Open (de PC DOCS Inc).	15
IV.3) Les agents intelligents.	16
IV.3.1) Hoover (de Sandpoint)	16
IV.3.2) Topic (de Verity Incorporated).	17
V) Bibliographie.	18

I) Préambule.

L'objet de cette étude est de décrire l'offre logicielle en matière de veille technologique et stratégique aux Etats-Unis. Celle-ci présente la particularité d'être assez abondante et diverse, aussi ce document ne prétend pas constituer un catalogue de produits disponibles mais un guide typologique qui permet de positionner facilement dans le panorama proposé tout produit lié à cette activité.

II) Introduction.

Le concept de veille présente de multiples facettes. Il peut concerner aussi bien des domaines techniques (méthodes de production, normes, brevets, ...) que d'autres plus généraux, d'ordre économique, politique ou social.

Les structures de veille technologique sont d'abord apparues dans les grands groupes industriels. Elles ont utilisé la logistique des centres de documentation (qui ont pour fonction d'enrichir les bases de connaissances) afin de collecter puis de diriger vers des cellules d'analystes experts les flux d'informations à caractère technologique. Les premiers systèmes informatisés, inspirés de principes mis en place pour la gestion des fonds de bibliothèques, géraient en fait les références des documents en utilisant des descripteurs et/ou un thésaurus pour décrire le fonds documentaire et en permettre l'accès.

Plus récemment, depuis une quinzaine d'années, s'est développée la veille stratégique qui cible essentiellement l'environnement économique, politique et social de l'entreprise. Or cette activité est aujourd'hui en effervescence du fait de la multiplication des sources d'information disponibles sous une forme numérisée. Ce nouveau support de l'information en facilite l'introduction au sein du système de l'entreprise sous sa forme originelle (texte, dessin, son, image) mais nécessite la mise en place d'un nouveau type d'outils capables de manipuler des données sous une forme non structurée.

Cette diversité de nature de la matière stratégique reflète en fait la variété des contextes concurrentiels et industriels dans laquelle évoluent les entreprises. Les producteurs d'informations ainsi que les éditeurs de logiciels ont répondu à cette diversité en proposant des services et produits de plus en plus ciblés. Afin de mieux présenter cette offre il convient d'abord d'explicitier les besoins auxquels elle répond, et donc de définir les fonctions d'un système de veille.

III) Définition d'un système de veille ¹.

La veille technologique ou stratégique est un processus qui organise d'une façon systématique la collecte, l'analyse et la diffusion jusqu'aux décideurs, d'informations concernant l'environnement externe de l'entreprise. Les systèmes de veille sont bâtis autour d'outils informatiques qui permettent d'automatiser tout ou partie de ces fonctions.

III.1) La collecte de l'information.

C'est l'ensemble des opérations qui permet de sélectionner, parmi des sources variées, et de conserver au sein du système de l'entreprise, l'information susceptible de présenter un intérêt stratégique.

III.1.1) Les sources.

Outre les supports publiés (ouvrages, revues, journaux...) il existe une offre toujours croissante d'information sur support numérisé disponible notamment par l'intermédiaire du réseau Internet. ²

a) Les supports imprimés.

C'est encore le support le plus courant néanmoins il est à présent très facile de numériser tout document publié grâce à un scanner couplé à un système de reconnaissance optique de caractères. Il est ainsi possible de mémoriser le contenu intégral du document dans le système d'information qui devient alors directement accessible à l'utilisateur.

b) Les supports numérisés.

Outre les CD-Roms, les sources d'informations disponibles aux Etats-Unis sous une forme numérisée sont nombreuses (Nexis, CompuServe, Prodigy, Dialog, Dow-Jones, Financial Time). ³

L'accès à ces serveurs est possible soit à travers le réseau Internet ⁴ soit par liaison commutée classique mais le dialogue avec chacun d'eux s'appuie sur un langage propre que l'utilisateur doit donc apprendre. Il y avait donc là un besoin d'accès facilité à ces sources qui a été satisfait par l'apparition d'automates de connexion. ⁵

III.1.2) La sélection.

L'autre problème majeur auquel est confrontée aujourd'hui l'entreprise, est de maîtriser ces flux d'information ⁶ c'est-à-dire en premier lieu de filtrer à travers ces nombreuses sources toute mais uniquement l'information qui présente un réel intérêt. Les développements récents ont produit de nouveaux modules qualifiés d'agents filtrants ⁷ qui répondent en partie ⁸ à ce

¹ Developing Information Systems for Competitive Intelligence Support, Hohhof B

² Internet-sourced competitive intelligence, Cronin B; Overfelt K; Foucheraux K; Manzvanzike T; Mikyeong C; Sona E

³ Corporate services find new channel as businesses launch Internet invasion : (Microsoft Corp, MCI Communications Corp and others companies market on-line services), Patch K

⁴ The Internet and competitive intelligence : a survey of current practice, Cronin B; Overfelt K; Foucheraux K; Manzvanzike T; Mikyeong C; Sona E

⁵ Personalized news delivered to your desktop : (Internet news collection services), Frentzen J

⁶ Competitive intelligence gathering : A telecommunications nightmare, Castro C.R

besoin. Ils permettent de s'affranchir des syntaxes multiples de requêtes, de mémoriser celles qui présentent un caractère répétitif, et d'en programmer le déclenchement automatique avec l'envoi d'un message d'alerte au destinataire concerné après chaque collecte d'information. Tous ces outils concourent également à réduire les temps et les coûts de connexion.

III.1.3) Le stockage.

Il ne s'agit pas de s'étendre sur les aspects purement techniques de ce chapitre mais de mentionner l'évolution qui, de fichiers constitués d'enregistrements aux champs bien définis, a conduit à des bases dites déstructurées ou de format libre. Cette évolution est très importante car elle a permis le stockage de documents (texte, image, dessin...) dans leur intégralité. L'opération d'indexation qui consiste à leur associer des descripteurs n'est plus indispensable, puisque la recherche sur texte intégral est très performante, cependant elle reste possible et bien que contraignante, elle permet une réticulation de la base d'information qui en optimise l'exploitation.

III.2) La restitution de l'information.

C'est l'opération qui consiste à extraire du référentiel, à un instant donné, l'ensemble des éléments qui vérifient certains critères de sélection. Cette restitution peut résulter d'une interrogation ponctuelle de la base ou d'un traitement de mise en forme de l'information.

III.2.1) L'interrogation ponctuelle.

Elle regroupe un ensemble de fonctions qui permet la formulation d'une requête d'interrogation, la recherche des informations et la présentation des résultats.

L'interrogation s'appuie sur un langage de requête plus ou moins évolué (opérateurs booléens , de proximité, de troncation, reconnaissance de mots, de phrases, ...). La technologie objet permet de créer des agents de recherche qui sont des formes agrégées et complexes d'éléments du langage.

Le moteur de recherche est l'élément qui détermine la puissance du système. Il s'appuie sur une méthode de recherche de mots-clés, éventuellement couplée à un thésaurus.⁹ Il existe aussi d'autres systèmes plus complexes d'analyse basés sur des recherches morphologique ou sémantique.¹⁰

Enfin la fonction de restitution permet de présenter la liste des documents qui répondent aux critères de recherche. Certains systèmes pondèrent la pertinence relative de chacun d'eux à l'aide de méthodes d'analyse¹¹ statistique ou sémantique afin d'ordonner la liste des résultats proposée à l'écran et d'assister le décideur dans sa fonction.¹² D'autres permettent à partir de

⁷ Software Intelligence Filtering Technology (SIFT)

⁸ Wading through incoming data : (intelligent software agents incapable of using same judgement criteria as individuals), Gore A

⁹ Improving search quality using thesauri for query specification and the presentation of search results, Pollitt A.S, Ellis G.P

¹⁰ Retrieving information from full text using linguistic knowledge, Wacholder N, Ravin Y, Byrd R.J

¹¹ Term relevance feedback and mediated database searching : implications for information retrieval practice and systems design, Spink A

¹² SIADIC : an experimental system using relevance indexing to assist decision-making, Giraud E, Dou H

celle-ci et grâce à des liens hypertextes un accès direct au document par une simple action de la souris sur la ligne correspondante.

III.2.2) Les mises en forme.

C'est une fonction capable d'extraire des données de plusieurs documents ou fichiers de la base afin de produire des documents de synthèse prédéfinis. Ce type de fonctionnalité apporte une réelle valeur ajoutée à la restitution de l'information mais elle impose une structuration des données qui s'accomode mal des formats libres.

III.3) La diffusion de l'information.

Quelle que soit la valeur intrinsèque d'une information, elle n'acquiert un caractère stratégique que si elle est délivrée au décideur adéquat, au moment opportun. Il n'existe pas de logiciel expert qui soit capable d'apprécier à la fois, la pertinence d'une information et l'opportunité de l'instant. Aussi les systèmes existants mettent en oeuvre des solutions qui consistent à :

- automatiser la diffusion sélective du résultat d'une requête de recherche jusqu'à un usager prédéfini,
- offrir aux décideurs un accès direct à la base grâce à un système d'information ouvert.

IV) L'offre logicielle.

Elle reflète les évolutions technologiques du monde informatique si bien qu'aujourd'hui coexistent trois familles de produit qui reposent sur des architectures et concepts bien différents.¹³

La plus ancienne et la plus classique de ces familles correspond aux produits qui s'appuient sur une base propriétaire dans laquelle est conservé l'ensemble des documents ou références de documents. Ces systèmes constituent une base d'informations documentaires autour de laquelle sont développées les fonctions de gestion, de recherche et d'analyse.

Ensuite sont apparus les gestionnaires de documents qui ont permis d'étendre ces mêmes fonctions à un ensemble de ressources variées et dispersées au sein d'un réseau en s'appuyant sur des bases d'index identifiant et localisant les différents documents.

Enfin se développe la dernière génération de produits conçue autour du concept d'agent intelligent¹⁴ qui sont des objets exécutant une tâche très spécifique (recherche, alerte,...) et capable d'opérer sur des supports d'information très variés.

Tous ces produits répondent à une tendance marquée vers des architectures de plus en plus ouvertes autorisant un accès facilité à l'information tant interne qu'externe à l'entreprise. Il en résulte des bouleversements importants du schéma d'organisation classique qui autorisent des configurations tout à fait originales.

IV.1) Les systèmes de base d'informations.

Cette famille de produit répond à une pratique assez centralisée de l'activité comme par exemple au sein d'un centre de documentation. La constitution d'une base d'informations peut s'avérer fastidieuse en revanche elle permet une réelle maîtrise des données qui y sont conservées. Elle offre, de plus, des possibilités de structuration telles que les thésaurus, index ou dictionnaire de synonymes.

IV.1.1) AskSam 2.0 (de AskSam Systems).

AskSam fut parmi les premiers (dès 1986) à permettre la recherche sur texte intégral et demeure avec sa dernière version un produit innovant et très compétitif. C'est un système qui s'appuie sur une base propriétaire de format texte¹⁵ et qui présente un éventail de fonctions cohérent et efficace bien que peu sophistiqué.¹⁶

Ses points forts sont ses capacités de collecte et de diffusion de l'information.

AskSam propose des interfaces de connexion à des serveurs commerciaux (CompuServe et Lexis) et un module de numérisation (scanner et système de reconnaissance optique de caractères). Outre ces formats texte et image, il peut également importer des documents aux formats externes courants (Word, WordPerfect, dBase, RTF) dans sa base propre. Il fonctionne sans index ni thésaurus par contre il offre la possibilité d'établir des liens hypertextes entre les documents.

¹³ Searching for the last word: (text retrieval software) (includes related articles on case studies, basic functions of text retrieval), Shape R

¹⁴ Silent partners : (software 'agents'), Ditlea S

¹⁵ AskSam update fills small-office niche, Marshall P

¹⁶ AskSam 2.0 (AskSam Systems' DBMS), Spitzer T

AskSam autorise des recherches tant sur des champs clés que sur du texte intégral. Son langage de requête est complet (opérateurs numériques, booléens, de proximité, de troncation...). Il permet également l'utilisation de liens hypertextes mais ne sait pas réaliser de recherches progressives ni gérer un dictionnaire de synonymes. Malgré l'absence d'un index, le moteur de recherche s'avère aussi puissant lorsqu'il opère sur des champs de fichier que sur du texte.

Le traitement des résultats de recherche ne dispose d'aucune fonction d'analyse de pertinence mais la liste présentée à l'écran permet la navigation entre les documents grâce à des liens hypertextes.

Enfin une fonction de publication autorise la production et la diffusion à travers le réseau d'une mise en forme élaborée de l'information.

AskSam n'est disponible que dans les environnements DOS et Windows mais peut néanmoins constituer une alternative intéressante car très compétitive. Il est commercialisé en version monoposte (1500 \$) ou en réseau de 5 postes (2500 \$), de 25 postes (6000\$), de 50 postes (10.000\$) et même de 100 postes (16.500\$).

IV.1.2) Folio Views (de Folio Corp).

Folio Views ¹⁷ présente de larges similitudes avec AskSam. Ses points forts se situent pareillement autour de ses capacités de collecte et de diffusion de l'information.

Folio Views ¹⁸ propose aussi des interfaces de connexion à des serveurs commerciaux (CompuServe et Lexis) et peut également importer des documents aux formats externes courants (Word, WordPerfect, BMP, GIF, PIC ...). Contrairement à AskSam il intègre chaque mot du document importé dans un fichier d'index qui permet de meilleures performances de recherche.

Folio Views ¹⁹ fournit une panoplie de recherche plus complète et plus conviviale. Outre l'utilisation des opérateurs classiques il permet l'élaboration de requêtes complexes dans un mode de dialogue supporté par la constitution d'un arbre logique des résultats à chaque étape de la recherche. Il dispose également d'un dictionnaire de synonymes.

La fonction de publication est plus complète et plus puissante que celle proposée par AskSam et permet une diffusion plus variée de l'information.

IV.1.3) Wincite (de Quest Management System).

Wincite est un produit commercialisé depuis 1994 dont la force se situe autour de sa fonction d'analyse et de synthèse de l'information.

Il s'appuie sur un référentiel d'informations qui peut supporter aussi bien des fichiers structurés qu'au format libre mais contrairement aux produits précédents le seul support de constitution et d'enrichissement de cette base est la fonction d'importation. C'est au cours de cette opération qu'il est possible d'adjoindre des descripteurs aux documents.

¹⁷ Text data managers help untangle the mess: these products try to make sense out of unstructured data, Bobchin C

¹⁸ Folio Views create info databases : Producton kit provides easy control of corporate, commercial flat-file information repositories, Bethoney H

¹⁹ Intelligence gathering (Folio Corp's Folio Views), Powell J

A partir de ce référentiel et de liens définis entre les documents Wincite est capable de réaliser des synthèses, des recoupements qui aboutissent à la production de mises en forme sophistiquées de l'information.

Il dispose également d'un outil de publication qui permet un niveau supplémentaire de mise en valeur des données à travers la production en documents concis qui peuvent être diffusés à travers le réseau notamment jusqu'aux décideurs.

Wincite est un produit ambitieux, assez onéreux (à partir de 5.000\$) mais qu'il faut coupler à un système collecteur d'informations externes.

IV.2) Les gestionnaires de documents. ²⁰

Le gestionnaire de documents est une famille de logiciel dont l'essor est lié à celui des réseaux locaux. C'est en effet dans cet environnement qui conduit parfois à une prolifération mal maîtrisée de documents de tous types qu'il intervient comme un élément fédérateur et structurant. Il procure en effet une interface unique à partir de laquelle les usagers du réseau peuvent accéder à tout document, quelle que soit son application d'origine.

Sa fonction est donc d'organiser l'ensemble des fichiers puis d'en permettre des accès contrôlés en assurant la confidentialité et l'intégrité des données.

a) La base d'informations.

Le système s'appuie sur une base d'informations qui permet d'identifier, de localiser, de décrire et de gérer les différents documents dispersés à travers le réseau. Certaines de ces informations sont fournies par l'auteur lorsqu'il accède au document. Ainsi à chaque demande de sauvegarde le gestionnaire s'interpose pour demander les informations dont il a besoin afin de mettre à jour ses champs descripteurs et ses fichiers d'index, puis il rend la main à l'application d'origine.

Ces descripteurs permettent d'identifier et d'accéder aux documents mais également de regrouper au sein de dossiers, qui sont des entités virtuelles, tous ceux qui répondent à une communauté de critères définie. Il est ainsi possible de constituer des entités thématiques et de traiter un ensemble de documents hétérogènes quant à leur format et leur origine (images numérisées, texte ASCII, fichiers formatés, fax, images ou son).

b) L'indexation des documents.

Outre les descripteurs il est également possible de répertorier l'ensemble des mots du texte d'un document dans un fichier d'index. La puissance de cette fonction d'indexation se mesure par sa vitesse d'exécution ainsi que par la diversité des formats de fichiers qu'elle est capable de traiter.

c) Les fonctions réseau.

Elles contribuent à la fiabilisation du réseau en gérant les conditions d'accès des différents postes client au serveur ainsi que le routage des requêtes sur le réseau.

IV.2.1) SoftSolutions (de Novell Inc).

SoftSolutions ²¹ ²² est le produit leader du marché avec 54% des ventes mondiales. Il est disponible sur des plates-formes NetWare et Unix à l'extrémité serveur et Windows, Dos et Unix du côté client. Il dispose d'interfaces (DLL) vers les applications AmiPro, Excel, Lotus1-2-3, Word, WordPerfect, Office 4 et MicroSoft Mail qui autorisent le lancement automatique de l'application d'origine du document traité.

a) L'indexation.

L'opération d'indexation est souvent pénalisante en terme de temps d'exécution. Ce problème est ici contourné puisque cette opération peut être réalisée en parallèle et périodiquement par une autre machine dédiée à cette tâche, libérant ainsi le poste de l'utilisateur.

b) La recherche.

Le moteur d'indexation SpeedSearch est très performant parcequ'il s'appuie sur une base propriétaire. Il n'autorise que des recherches simples ou multiples sur des mots-clés alors que le module 'Intelligent Search' permet des recherches métamorphiques.

Enfin la base d'informations est également accessible grâce à une interface ODBC à des requêtes d'interrogation SQL formulées à partir d'Access, d'Excel ou de Visual Basic.

c) L'accès aux documents.

La version Suite avec le module 'Document Desktop' offre une interface graphique améliorée qui confère une grande souplesse d'utilisation. Elle permet l'utilisation d'icônes pour accéder facilement aux dossiers ou encore pour déclencher les requêtes de recherche préalablement mémorisées.

Le module de base est commercialisé à 495\$ par serveur et 295\$ par poste client. La version Suite avec le module 'Document Desktop' est proposée à 695\$ par serveur et 395\$ par poste client. Le module 'Server Enhancement Module' qui optimise la gestion du réseau majore le système de 1995\$

IV.2.2) PC DOCS Open (de PC DOCS Inc).

PC DOCS Open ²³ ²⁴ est souvent présenté comme le meilleur produit de sa catégorie et concurrence donc très sérieusement SoftSolutions. Il est disponible pour les serveurs Microsoft SQL, Sybase SQL, Oracle7 et en version Windows (Dos et Macintosh annoncées pour début 95).

²¹ SoftSolutions Suite 4.0 : a lending library for your files, White R
²² A power manager for network files : (SoftSolutions Technology's SoftSolutions 4.0 file manager software), O'Connor K
²³ Easier access to corporate data : (PC DOCS Open document management system), O'Connor K
²⁴ PC DOCS Open : (version 2.0 from PC DOCS Inc), Garris J

Contrairement à SoftSolutions, PC DOCS Open ne requiert pas une base propriétaire pour gérer le catalogue des documents mais il peut s'accommoder de la base disponible dans les serveurs tels que Microsoft SQL Server, Sybase, Oracle ou Watcom. Il dispose également d'interfaces vers les applications AmiPro, Excel, Word, WordPerfect, DocuComp et Compare Rite.

Ses fonctionnalités sont tout à fait équivalentes à celles de SoftSolutions. ²⁵

La version de base est commercialisée à 495\$ par serveur et 295\$ par poste client. Couplé au serveur Sybase ou Oracle il coûte 695\$ par serveur et 395\$ par poste client. Enfin il existe une version portable commercialisée à 445\$.

Les éditeurs ont compris la place que peut occuper ce type de produit dans un système de veille malgré sa faiblesse relative au niveau de la collecte d'information. L'offre est importante avec comme principaux autres produits Mezzanine (de Saros Corp) ²⁶ , PageKeeper (de Caere Corp) , ²⁷ ZyIndex ²⁸ ou encore Isys. ²⁹

IV.3) Les agents intelligents.

Les agents intelligents ³⁰ sont des modules qui s'intègrent parfaitement dans des architectures en réseau car ils satisfont à des exigences d'efficacité fonctionnelle et d'interopérabilité.

IV.3.1) Hoover (de Sandpoint)

Hoover est un produit récent sur le marché qui utilise des "agents filtrants" pour assurer la recherche et la collecte d'informations.

L'agent 'Business Topic Research'.

Cet agent présente des fonctions d'accès automatisé à des serveurs commerciaux (Wall Street Journal, Financial Times, ...) qui permettent de s'affranchir des procédures de connexion et des syntaxes d'interrogation spécifiques à chaque serveur. ³¹ Il peut aussi accéder à d'autres réseaux, à des CD-Roms ou à Internet pour explorer des sources externes d'information.

La formulation des requêtes de recherche s'appuie non plus uniquement sur des mots-clés mais sur des concepts qui sont une forme agrégée d'éléments du langage ou de sous-concepts. Outre cette capacité d'expression d'algorithmes de recherche élaborés, l'agent gère la séquence des accès aux différents serveurs cibles de la requête.

²⁵ The documents masters : PC DOCS, Saros and SoftSolutions : (high-end document management systems), Chen Steven C.M

²⁶ Mezzanine/Document Manager : (Mezzanine 3.1c and Document Manager 1.0 from Saros Corp), Garris J

²⁷ Caere delivers intelligent document management : from OCR to a desktop solution (PageKeeper), Berry M

²⁸ ZyImage adds scanning access to ZyIndex, Marshall P

²⁹ Text data managers help untangle the mess: these products try to make sense out of unstructured data, Bobchin C

³⁰ Agent Technology : the fine line between smart design and intelligence software : (upcoming innovations in intelligent software), Miley M

³¹ Hoover acts as your online scout. (SandPoint Corp. Hoover), Powers A

Enfin la liste proposée des documents sélectionnés découle d'une analyse statistique des différents documents qui permet un tri en fonction de leur degré de pertinence.

L'agent : 'NewsAlert'.

NewsAlert³² permet de paramétrer les déclenchement automatiques des requêtes d'interrogation (fréquence, bases ciblées, ...) et de retourner, en cas de collecte d'information, un message à son destinataire.

Hoover acquiert toute sa puissance couplé avec **Notes de Lotus** et met à profit les fonctions du groupware pour assurer les fonctions de stockage, d'indexation, de recherche sur texte intégral, d'annotation des documents et de diffusion aux usagers à travers le réseau.

NewsAlert est onéreux (à partir de 40.000 \$ pour 10 postes).

IV.3.2) Topic (de Verity Incorporated).

Verity est l'éditeur dominant sur le marché des systèmes de recherche sur texte intégral avec son produit Topic (5500 installations en juin 94).

Sa dernière version a intégré la technologie objet³³ pour définir des agents³⁴ qui assurent la recherche (objet Searcher), la surveillance des sources d'information (objet Watcher) et l'analyse de la pertinence des résultats de recherche (objet Analyst). Tous ces agents sollicitent le coeur du système qu'est le moteur de recherche.

L'agent : 'Searcher'.

Cet agent présente des fonctions analogues à celle du "Business Topic Research" de Hoover avec des modules supplémentaires d'accès à Mosaic, à Web et à des bases de données relationnelles.

L'agent: 'Analyst'.

L'agent d'analyse classe les documents sélectionnés par l'agent de recherche selon leur degré de pertinence déterminé en fonction de la fréquence de certains mots-clés dans chacun des documents.

L'agent : 'Watcher'.

Il surveille tous les flux d'information entrant dans le système et adresse des messages d'alerte d'une manière tout à fait sélective aux usagers qui sollicitent cette fonction.

Topic est commercialisé à partir de modules et donc le coût de la configuration du système est variable.

³² Competitive intelligence, software robots and the Internet : the NewsAlert prototype, King D; Jones K

³³ InfoAgent's Verity's new strategic direction, Berry M

³⁴ Verity's InfoAgents, Morgenstern D.

V) Bibliographie.

- A power manager for network files : (SoftSolutions Technology's SoftSolutions 4.0 file manager software), *O'Connor Kevin*,
PC User , N°. 230, pp. 71(2) ,March 9,1994, ISSN: 0263-5720
- Agent Technology : the fine line between smart design and intelligence software : (upcoming innovations in intelligent software), *Miley Michael*,
Mac WEEK, Vol. 7, No. 16, pp. 41(4), ISSN: 0892-8118
- AskSam 2.0 (AskSam Systems' DBMS), *Spitzer Tom*,
DBMS, Vol. 8, No. 1, pp. 74(2), Jan, 1995, ISSN: 1041-5173
- AskSam update fills small-office niche, *Marshall Patrick*,
InfoWorld, Vol. 16, No. 49, pp.98(4), Dec 5, 1994, ISSN: 0199-6649
- Caere delivers intelligent document management : from OCR to a desktop solution (PageKeeper),
Berry M,
Workgroup Computing Report, Vol. 17, No. 7, pp. 24(3), July 1994, ISSN: 1068-9699
- Competitive intelligence gathering : A telecommunications nightmare, *Castro C.R*,
Proceedings of 16th Annual Pacific Telecommunications Conference, 16-20 Jan 1994
- Competitive intelligence, software robots and the Internet : the NewsAlert prototype, *King D; Jones K*,
Proceedings of the Twenty-Eighth Annual Hawaii International Conference on System Sciences, 3-6 Jan 1995, ISBN: 0 8186 6945 4
- Corporate services find new channel as businesses launch Internet invasion : (Microsoft Corp, MCI Communications Corp and others companies market on-line services), *Patch Kimberly*,
PC Week, Vol. 11, No. 16, pp. 1(2), Apr 25, 1994, ISSN: 0740-1604
- Developing Information Systems for Competitive Intelligence Support, *Hohhof Bonnie*,
Library Trends, Vol . 43, No. 2, Fall 1994, pp. 226(12), ISSN: 0024-2594
- Delivery and retrieval technology. (Seybold Special Report: Seybold Seminar Boston 1994),
Seybold Report on Publishing Systems, Vol. 23, No.16, pp. 20(16), May 10, 1994, ISSN: 0736-7260
- Easier access to corporate data : (PC DOCS Open document management system), *O'Connor Kevin*,
PC User , N°. 227, pp. 47(2) , Jan 26, 1994, ISSN: 0263-5720
- Folio Views creates info databases : Producton kit provides easy control of corporate, commercial flat-file information repositories, *Bethoney Herb*,
PC Week, Vol. 11, No. 44, pp. 129(1), Nov 7, 1994, ISSN: 0740-1604
- Hoover acts as your online scout. (SandPoint Corp. Hoover), *Powers Andrea*,
PC World, Vol. 12, No. 8, pp. 9(1), August 1994, ISSN: 0737-8939
- Improving search quality using thesauri for query specification and the presentation of search results,
Pollitt A.S, Ellis G.P,
Proceedings of 3rd International Conference of the International Society of Knowledge Organization, 20-24 June 1994, ISBN: 3 88672 023 3
- InfoAgent's Verity's new strategic direction, *Berry Margaret D*,
Workgroup Computing Report, Vol. 17, No. 5, pp.34(3), May, 1994, ISSN: 1068-9699

- Intelligence gathering (Folio Corp's Folio Views), *Powell James*,
Windows Magazine, Vol. 5, No. 10, pp. 158(1), Oct 1994, ISSN: 1060-1066
- Internet-sourced competitive intelligence, *Cronin B; Overfelt K; Foucheraux K; Manzvanzvike T; Mikeyong C; Sona E*,
Proceedings of National Online Meeting, pp. 87(5), 10-12 May 1994, ISBN: 0 938734 84 9
- Mezzanine/Document Manager : (Mezzanine 3.1c and Document Manager 1.0 from Saros Corp),
Garris John,
PC Magazine, Vol. 13, N° 19, p. 5(2), Nov 8, 1994, ISSN: 0888-8507
- Mining for information gold, *Haskin David*,
PC Magazine, Vol. 13, No. 21, pp. 44(1), Dec 6, 1994, ISSN: 0888-8507
- On-target searches. (Target Systems Corp's Target-Smart Serach text search and retrieval software),
Brook Roseann,
DEC professional, Vol. 13 , No. 8, pp. 16(1), Aug 1994
- Personalized news delivered to your desktop : (Internet news collection services), *Frentzen Jeff*,
PC Week, Vol. 11, No. 44, pp. 19(1), Nov 7, 1994, ISSN: 0740-1604
- PC DOCS Open : (version 2.0 from PC DOCS Inc), *Garris John*,
PC Magazine, Vol. 13, N° 19, p. 8(2), Nov 8, 1994, ISSN: 0888-8507
- Retrieving information from full text using linguistic knowledge, *Wacholder N, Ravin Y, Byrd R.J*,
Proceedings of National Online Meeting, Conference Date: 10-12 May 1994, ISBN: 0 938734 84 9
- Searching for the last word: (text retrieval software) (includes related articles on case studies, basic functions of text retrieval), *Shape Richard*,
Computer Weekly, pp. 38(4), June 23, 1994, ISSN: 0010-4787
- SIADIC : an experimental system using relevance indexing to assist decision-making, *Giraud E, Dou H*,
Documentaliste-Sciences de l'Information, Vol. 31, No.2, pp.71(6), March 1994, ISSN: 0012-4508
- Silent partners : (software 'agents'), *Ditlea Steve*,
PC Computing, Vol. 7, N° 5, p. 160(9), May 1994, ISSN: 0899-1847
- SoftSolutions : (version 4.0 from Novell Inc), *Garris John*,
PC Magazine, Vol. 13, N° 19, p. 10(2), Nov 8, 1994, ISSN: 0888-8507
- SoftSolutions Suite 4.0 : a lending library for your files, *White Ron*,
PC Computing, Vol. 7, N° 11, p. 96(1), Nov 1994, ISSN: 0899-1847
- Term relevance feedback and mediated database searching : implications dor information retrieval practice and systems design, *Spink A*,
Information Processing & Management, Vol. 31, No. 2, p. 161(10), 1995, ISSN: 0306-4573
- Text data managers help untangle the mess: these products try to make sense out of unstructured data,
Bobchin Craig,
Data based Advisor, Vol. 12, No. 12, p. 68(6), Dec 1994, ISSN: 0740-5200
- The documents masters : PC DOCS, Saros and SoftSolutions : (high-end document management systems), *Chen Steven C.M*,
PC Magazine, Vol. 13, N° 11, p. 316(2), June 14, 1994, ISSN: 0888-8507

The Internet and competitive intelligence : a survey of current practice, *Cronin B; Overfelt K; Foucheraux K; Manzvanzvike T; Mikyeong C; Sona E*,
International Journal of Information Management, Vol. 14, No. 3, pp. 204(18), Jun 1994,
ISSN: 0268-4012

Verity's InfoAgents, *Morgenstern David*,
Which Computer ?, Vol.17, No. 7, pp. 36(1), July, 1994, ISSN: 0140-3435

Wading through incoming data : (intelligent software agents incapable of using same judgement
criteria as individuals), *Gore Andrew*,
Mac Week, Vol.8, No. 6, pp. 32(1), Feb 7, 1994

ZyImage adds scanning access to ZyIndex, *Marshall Patrick*,
InfoWorld, Vol. 16, No. 15, pp. 73(3), April 11, 1994, ISSN: 0199-6649

