



Pour être en mesure de choisir la stratégie et les outils les plus efficaces, il faut pouvoir s'appuyer sur trois préalables : bien définir ses besoins, connaître les critères d'évaluation des produits et services et comprendre la structure et les modes de recherche offerts.

savoir rechercher et interroger

Les repères méthodologiques

■ La première étape de toute recherche d'information consiste à bien préciser ses objectifs. Voici quelques points de repère et questions pour vous aider à les définir :

- S'agit-il de trouver une information ponctuelle, de mettre à jour ses connaissances, de réaliser un exposé, un mémoire, une thèse, un article scientifique, une bibliographie ?
- Quels sont les types de documents, d'informations et leurs niveaux ?
- Quels sont les services, lieux et personnes ressources susceptibles de produire, fournir cette information ? Sont-ils accessibles aisément ?
- Degré d'actualité de l'information nécessaire ?
- Degré d'exhaustivité ?
- Sous quelle forme les données seront-elles réutilisées ? Des questions de respect du droit d'auteur peuvent alors intervenir à cette étape.

- Dans quels délais est-il nécessaire de les obtenir ?
- A quel coût ?

■ Il est important de délimiter et situer son sujet pour choisir les sources et le ou les médias les plus adaptés. En effet, on peut trouver des produits, données, informations identiques d'un support à l'autre.

Cependant, le plus souvent, ils sont déclinés selon des logiques différentes en fonction des capacités du système, du public ciblé par le producteur ou encore des différents types d'usages auxquels ils correspondent le mieux. Bien étudier les écrans de présentation, les guides imprimés ou en ligne peut favoriser un gain de temps en permettant de prendre la mesure du domaine et de la période couverts, des types de documents analysés ou présents dans la base, des opérateurs disponibles, de la ou des langues dans lesquelles on peut interroger etc...

Les dépêches de l'AFP offrent une bonne illustration des déclinaisons possibles :

- On peut faire une recherche rétrospective importante sur les dépêches, en interrogeant la banque depuis 1985 sur un **serveur ASCII** comme Questel. Ex. p. 19 et 20.
- En revanche, l'interrogation des services offerts sur **Minitel** ne peut porter que sur les dépêches des trois derniers mois pour le service grand public (36 15 AFP) et des 12 derniers mois pour le service professionnel sur les hauts palliers (08 36 29 00 53 AFP-PRO). Ex. p. 19.
- Les documents de synthèse (fiches pays, gouvernements, biographies...) réalisés par l'AFP sont disponibles sur le **céderom** AFP DOC. Un autre produit thématique, le **céderom " Sciences, Mémoire de l'infor-**

mation scientifique de 1990 à nos jours " offre l'intégralité des informations scientifiques issues du bulletin Sciences. Une connexion directe au serveur Web de l'AFP est également proposée pour accéder aux bulletins scientifiques en cours.

- Enfin, sur le **serveur Web de l'AFP** (<http://www.afp.com>) de nouveaux services sont en cours de lancement : AFP à la Carte, Canal infos, AFP direct etc... Ils reposent sur le principe de personnalisation de la diffusion de l'information pour les entreprises ou les particuliers, Ex. p. 22.
- Les dépêches de l'AFP sont également consultables dans le cadre **d'autres services sur Internet** : annuaires thématiques comme Nomade (Ex p. 44), chaînes de Webcasting ...

Les critères de qualité

La qualité des produits et de l'interrogation dépend de plusieurs paramètres auxquels il faut se montrer très attentif.

- **le producteur, l'éditeur, l'auteur** : sa réputation scientifique ou sa notoriété, l'institution à laquelle il appartient, la reconnaissance de celle-ci dans son champ d'activité, son caractère officiel ou non, la richesse de la couverture documentaire (temporelle et géographique), les délais de mise à jour, la qualité de l'indexation et des outils de recherche (manuel, thesaurus).
- **le serveur et les éditeurs, distributeurs de céderoms** : la capacité et les performances de l'ordinateur et des logiciels d'interrogation, la structuration de la banque de données (nombre de champs interrogeables,

opérateurs utilisables), la convivialité de l'interface, le nombre de banques offertes, leur complémentarité, la périodicité des mises à jour, les services associés, les délais de la fourniture du document, la formation des utilisateurs, l'assistance commerciale.

- **Le réseau** : la fiabilité de la transmission des informations (en terme de sécurisation des données, coupures...), les débits offerts, la capacité (problème de saturation).
- **L'utilisateur** : son niveau de connaissance du domaine traité, du contenu et de la structuration de la banque, sa familiarité avec le langage d'interrogation, son aptitude à évaluer les réponses fournies et à modifier l'interrogation en fonction de ces dernières.

Le traitement du document

Avant de présenter les différentes étapes d'une recherche, il est nécessaire de rappeler quelles sont les pratiques des professionnels de l'information pour **décrire et caractériser**

le contenu du document (texte, images, etc...) avec des outils documentaires qui normalisent l'écriture des termes, structurent les relations entre eux et imposent l'usage de

Exemple d'utilisation du thésaurus du cédérom Delphes

Question : la publicité effectuée par le constructeur automobile RENAULT

termes univoques aussi bien dans la phase de description que d'interrogation.

• **L'indexation** est le premier mode de représentation du contenu : suite à l'**analyse** du document, des **mots-clés (keywords) ou descripteurs (descriptors)** sont choisis ou extraits pour **décrire** les principaux concepts. Ces mots-clés peuvent ensuite être organisés dans le cadre d'outils documentaires tels que la **liste alphabétique des mots-clés ou le thésaurus**.

Ce dernier permet d'établir entre les termes les **relations** suivantes :

- hiérarchiques (termes génériques - **TG** - Tourisme dans l'exemple ci-dessous et termes spécifiques - **TS** - Promotion du tourisme) ;
- d'équivalence (entre les synonymes, - **EP** - employé pour) ;
- associatives (relations de voisinage, - **VA** - voir aussi ou terme associé - **TA** - Loisirs terme associé de Tourisme).

Exemple de thésaurus

Extrait adapté du Thésaurus de DELPHES (banque de données du réseau des chambres de commerce et d'industrie couvrant plus de 900 sources françaises et européennes, sur les marchés, les produits, les entreprises...), Chambre de commerce et d'industrie de Paris, Ed. 1993, à partir du terme générique **TOURISME**.

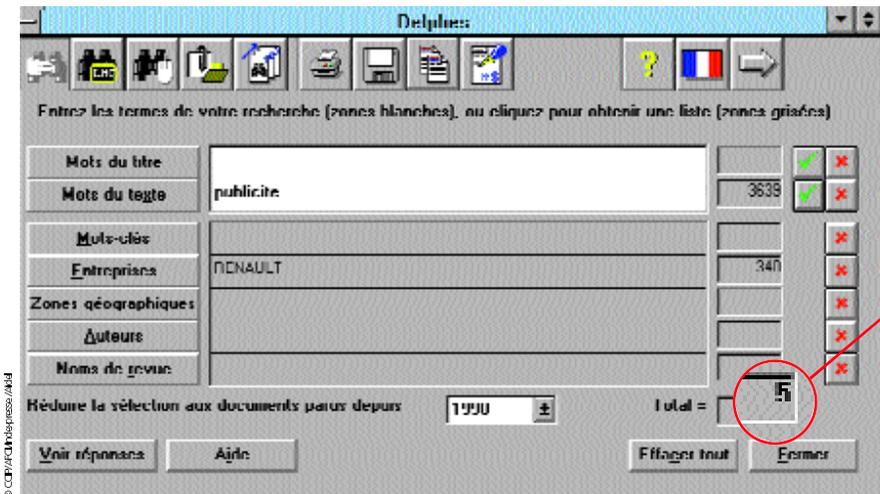
DESкриПТЕUR GÉNÉRIQUE	57 TOURISME	TERMES ASSOCIÉS LOISIRS 43
DESкриПТЕURSpécifiques DÉPENDANTDES PREMIERS	<ul style="list-style-type: none"> 57.1 PROMOTIONDU TOURISME <ul style="list-style-type: none"> 57.11 AGENCEDEVOYAGE 57.12 TOUROPERATEUR 57.2 FREQUENTATIONTOURISTIQUE <ul style="list-style-type: none"> 57.21 SAISONTOURISTIQUE 57.3 AMENAGEMENTTOURISTIQUE <ul style="list-style-type: none"> 57.31 HERBERGEMENTTOURISTIQUE <ul style="list-style-type: none"> 57.311 HOTELLERIE 57.3111 HOTEL 	<ul style="list-style-type: none"> LOCATIONSAISONNIERE 57-315 TRAVAILSAISONNIER 17-22657 EQUIPEMENTSSPORTIF 43-301 RESTAURATION 52
INDICESNUMÉRIQUES (COMPOSITIONDÉCIMALEPRÉCISANT LAHIERARCHIEDESTERMES).	<ul style="list-style-type: none"> [...] 57.4 PRODUITTOURISTIQUE <ul style="list-style-type: none"> 57.401 TOURISMEBALNEAIRE 57.402 TOURISMECULTUREL 57.403 TOURISME D'AFFAIRES 	<ul style="list-style-type: none"> STATIONBALNEAIRE 57-35 ACTIVITESCULTURELLESETARTISTIQUES 43-1 CONGRES 41-61 SEMINAIRE 41-612

La disponibilité croissante de **texte intégral** sous forme numérique - des articles de presse, des textes juridiques aux données produites par les organisations grâce à la généralisation de l'usage des logiciels de traitement de texte, en passant par tout ce qui est désormais diffusé sur Internet - donne un rôle de plus en plus important à l'**indexation automatique et aux outils de traitement linguistique**.

Les logiciels d'indexation automatique de type plein texte (full text) travaillent principalement au niveau de la forme en retenant toutes les chaînes de caractères séparées par des espaces à l'exception de mots vides (comme les articles) pour constituer des index. Les limites rencontrées par ces systèmes au regard

de la pertinence de l'information retrouvée, sont de plus en plus souvent compensées par les apports d'outils de traitement linguistique. Ils visent à réduire les ambiguïtés générées par la richesse du langage naturel en traitant les aspects **morphologiques, syntaxiques et sémantiques**.

• **Le résumé** (abstracts) est un autre mode de caractérisation du contenu par réduction de celui-ci. De plus en plus souvent, dans le cadre des services proposés via Internet, d'autres éléments comme les **tables des matières** (table of contents, TOC en abrégé) ou des **extraits** (sonores, de films...) sont offerts pour aider à mieux cerner le contenu du document.

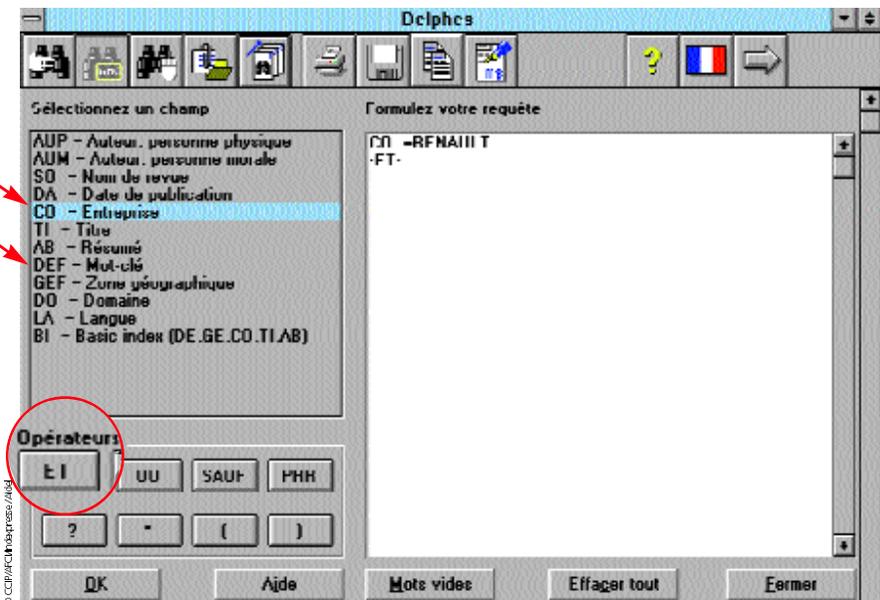


MODE ASSISTÉ

- l'utilisateur a saisi les termes **publicité** et **RENAULT** dans les zones correspondant aux champs proposés.

- 5 documents sont ainsi retrouvés.

Comparaison entre une recherche en mode assisté (écran supérieur) et en mode expert avec utilisation du thésaurus (écrans suivants).



MODE EXPERT

- le champ **CO** (correspondant à **Entreprise**) est sélectionné à gauche de l'écran et le terme **RENAULT** saisi.

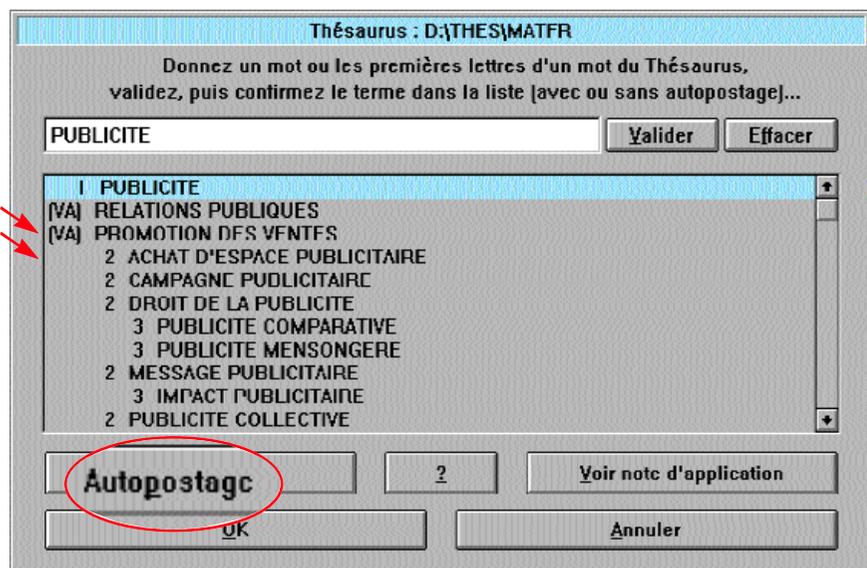
- opérateur booléen choisi pour combiner les deux critères : **ET**.

- la sélection du champ **DEF** permet d'ouvrir le thésaurus et d'aboutir au terme **publicité** (cf.écransuivants).

UTILISATION DU THESAURUS

- le mot clé **publicité** est affiché dans son environnement (termes associés précédés de **VA** et termes spécifiques précédés des numéros **2** ou **3** qui précisent leur niveau de spécificité).

- la fonction **AUTOPOSTAGE** permet de sélectionner en une seule fois tous les termes spécifiques de **publicité**. Cette fonction est activée dans le cadre de la présente requête.

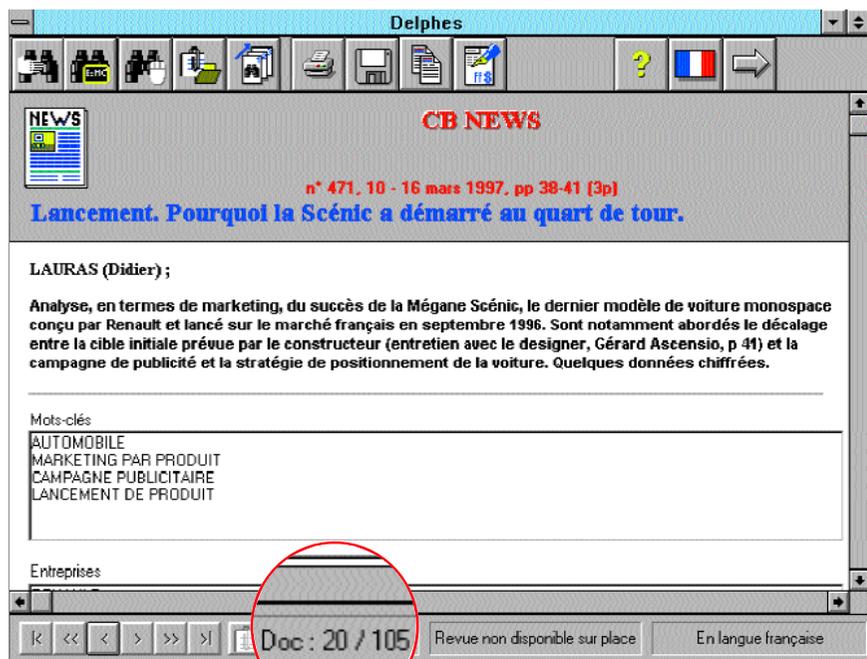


© CCB/FAC/Inspresse/AN&I

RÉSULTATS DE LA RECHERCHE EN MODE EXPERT

- 105 documents sont ainsi retrouvés contre 5 seulement lors de la recherche en mode assisté.

- la référence du 20^{ème} document est affichée.



© CCB/FAC/Inspresse/AN&I

La formulation de la requête

■ Analysez le sujet à traiter en le décomposant pour extraire ses **principaux concepts** et n'oubliez pas de formuler tous les aspects qui peuvent permettre de préciser le contexte (lieu, temps...).

■ Balayez pour chaque mot important de votre question les **alternatives** possibles. Il s'agit à cette étape de lever les sources d'ambiguïté :

- un même mot peut présenter des sens différents. Ex. : or, le métal ou la conjonction de coordination.

- Un concept peut être exprimé de différentes manières, il faut donc penser aux synonymes et également aux termes et expressions équivalents.

C'est bien sûr à cette étape que la consultation des outils décrits précédemment peut vous être utile.

L'usage des **troncatures** (voir tableau p.18) permet en remplaçant un ou plusieurs caractères de trouver les mots au singulier ou au pluriel, d'autres mots de la même famille. En effet, le logiciel lors de la recherche ne fait rien d'autre que de comparer les termes de la question entrés par l'utilisateur avec ceux qui ont été enregistrés pour décrire le document. Si vous avez tapé robot, il ne retrouvera pas robots, robotique...

■ Les références des documents sont structurées en diverses **zones** ou **champs** : auteur, titre, source, résumé, mots-clés... A chaque champ correspond un **index** (liste de tous les termes ou valeurs contenus dans chaque champ).

Généralement, on peut afficher les index dans des menus déroulants, le thésaurus (Ex. : cédérom DocThèses, index mots-clés p.29). Consulter et sélectionner les noms d'auteur à partir de leur index est particulièrement conseillé pour éviter les problèmes d'orthographe, d'initiales...

Les résultats de recherche de ses différents champs peuvent être **combinés** entre eux.

Lorsque vous saisissez vos critères dans des grilles de saisie (ou formulaires ou encore scripts) comme la majorité de produits et services le proposent aujourd'hui, certains champs sont combinés par défaut par le logiciel (le plus souvent avec OU). Il vaut donc mieux lire au préalable la présentation du mode de fonctionnement de celui-ci.

■ La recherche multicritère

Le principe de base de la plupart des recherches consiste à traiter une question par étapes successives en constituant des sous-ensembles de références pertinentes, combinés entre eux grâce aux **opérateurs booléens** : ET, OU, SAUF. Les résultats sont ainsi affinés progressivement. Vous pouvez également poser directement votre équation de recherche en utilisant des **parenthèses**.

Le tableau p. 18 résume leur usage ainsi que celui des autres opérateurs qui peuvent être à la disposition de l'utilisateur pour préciser sa question aussi bien pour interroger les banques de données commerciales que les moteurs de recherche offerts sur Internet (pour certaines particularités voir p. 39).

opérateurs	équivalents anglais	usages	exemple	commentaires
opérateurs booléens				
OU	OR, all of the terms	Union : sélectionne les documents où apparaissent le 1 ^{er} terme seul, le 2 ^e seul ou les deux à la fois.		1) La recherche porte sur tourisme ou loisirs ou les deux. OU permet d'élargir la recherche en prenant en compte les déclinaisons possibles d'un mot ou d'un concept.
ET	AND, any of the terms	Intersection : fournit les documents contenant les deux termes exclusivement.		2) Les résultats de l'étape précédente sont combinés avec station de sport d'hiver.
SAUF	NOT, NOT AND	Exclusion : élimine les notions non pertinentes.		3) La recherche va donc être faite sur les documents portant sur le tourisme ou les loisirs dans les stations de sport d'hiver sauf dans les Alpes.
opérateurs de proximité				
ADJ ADJ/n	NEAR NEAR/n WITH FOLLOWED BY	Ils permettent d'indiquer la distance entre deux mots, leur ordre, s'ils doivent figurer dans le même paragraphe... Ils sont particulièrement utiles pour interroger le texte intégral et permettent d'être plus précis qu'en employant l'opérateur ET. <i>Ces opérateurs et leur syntaxe varient selon les logiciels ou ne sont pas systématiquement disponibles.</i>		Dans l'exemple précédent ADJ était utilisé pour indiquer que station et sport doivent être adjacents.
troncatures				
* , ? , \$, ou +		Elles peuvent remplacer un ou plusieurs caractères. <i>Attention le signe employé change selon les logiciels.</i>		Dans l'exemple précédent ? remplaçait un caractère (\$ pour retrouver sports, + ou remplaçait généralement un nombre illimité de caractères (robot+ retrouve robot, robots, robotique, robotisé).
opérateurs numériques				
= , > , <		Ils permettent d'affiner l'équation de recherche sur les champs numériques (dates...).		

Exemples d'interrogations de produits de l'AFP

Déclinaison de l'information (Ascii, vidéotex, Internet) et évolution des interfaces.

■ Exemple d'interrogation en mode Ascii de la banque de données AGRA (AFP) sur le serveur QUESTEL (ensemble des dépêches françaises et internationales, tous pays, tous domaines depuis 1985).

— LANGAGE de COMMANDE : l'utilisateur saisit toutes les commandes et les codes —

La connexion

- Connexion sur le réseau Transpac
- Connexion sur le serveur Questel
- Choix de la version logicielle Questel+2
- Identification de l'utilisateur

La recherche

- Choix d'une banque de données
- Recherche de 2 termes adjacents dans le titre des dépêches
- Recherche sur une date avec l'opérateur numérique = (dépêches publiées en 1995)
- Combinaison des résultats obtenus aux questions 1 et 2

La visualisation

Visualise le document sélectionné en format maximum (tous les éléments de la dépêche)

La déconnexion

Arrêt de la consultation du serveur

DÉROULEMENT DE LA CONSULTATION

```

1042424009
COM
QUESTEL
Mode acti du serveur : 7 bits
Choice / Choix ? Q
Logon ? 11111 : 11
Mot de passe ? XXXX

..BASE / ..INFO / ..GUIDE
? ..BA AGRA
QUESTEL
Base sélectionnée : AGRA
Dépêches du Service Général de l'AFP
Question 1
? SCIENCE 1AV FETE/1
** Question 1, nombre de réponses 4
Question 2
? / DA = 1995
** Question 2, nombre de réponses 111 884
Question 3
? 1 et 2
** Question 3, nombre de réponses 1

Question 4
? ..VI QU 3 MAX
NUMERO AGR 19950612-0000272
SOURCE AFP (Service Général) ©AFP
TYPE DEPECHE
TITRE Science - Lancement d'un concours de nouvelles à partir du bulletin AFP Sciences pour la Science en Fête - Paris, 12 Juin 95 (277 mots)
TEXTE Un concours de nouvelles ayant pour thème dix sujets récents traités dans le bulletin hebdomadaire d'informations scientifiques de l'Agence France-Presse. AFP Sciences, vient d'être lancé par la Médiathèque du Centre Jean Renoir de Dieppe (Seine-Maritime) dans le cadre de la manifestation annuelle de la Science en fête.

Question 4
? ..ST FI
    
```

■ Sommaires de deux services vidéotex de l'AFP : grand public (les dépêches des 3 derniers mois) et professionnel (les dépêches des 12 derniers mois).

— MENUS GUIDÉS : l'utilisateur ne saisit que les codes correspondant à ses choix parmi les options proposées.



Sommaire du 36 15 AFP destiné au grand public.



Sommaire du 08 36 29 00 53

■ Interrogation identique à celle de l'exemple 1 (banque AGRA chez Questel) mais cette fois à l'aide du logiciel Imagination 2.

INTERFACE GRAPHIQUE : menus déroulants, ascenseurs, icônes etc...

Sélection du domaine d'intérêt

Sélection de la base de données

Mise en œuvre de la période sur laquelle portera la recherche: 1995

Saisie des termes de recherche dans le formulaire zone TITRE

Présentation du panel des résultats : - Historique de la recherche.

- nombre de résultats

- Choix du mode de visualisation : prévisualisation, document complet, focus

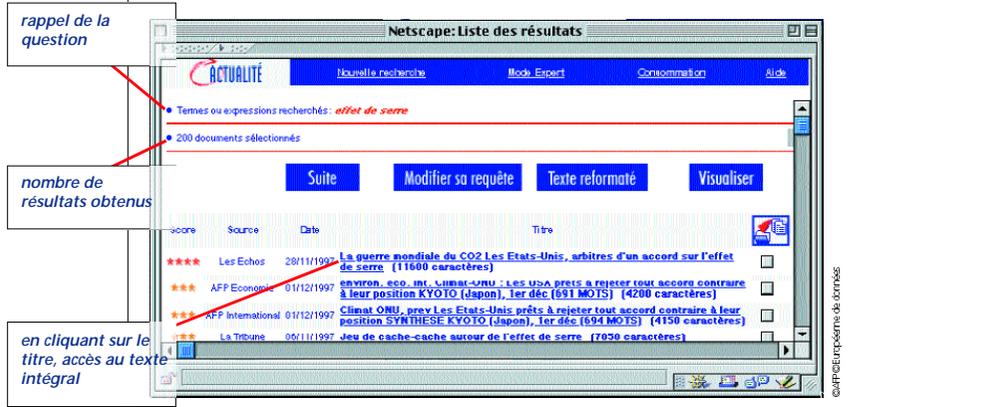
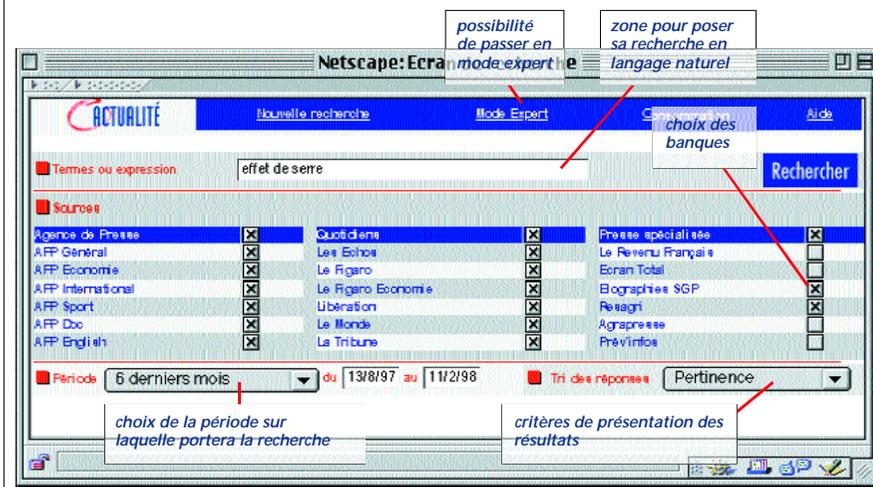
Option choisie "document complet" : visualisation en raccourci du résultat (il suffira de cliquer sur cette icône pour déplier le texte.)

The screenshots show the following steps:

- Sélection du domaine d'intérêt:** The user navigates through a tree view to 'Presse' and then 'Presse Française'.
- Sélection de la base de données:** A dialog box 'Sélection des bases de données' is shown with 'AGRA - AFP Service Général' selected.
- Mise en œuvre de la période:** The search criteria are set to '1995-01-01 1995-12-31 - 8'.
- Saisie des termes de recherche:** The search terms 'science et fête' are entered in the 'TITRE' field.
- Présentation du panel des résultats:** The search results are displayed in a list. A 'Prévisualisation' dialog box is shown with '18' pages.
- Option choisie "document complet":** A result is expanded to show a preview of the document.

■ Exemple d'interrogation des dépêches de l'AFP sur le serveur de l'Européenne de données

INTERFACE WEB, LANGAGE NATUREL : liens hypertextes, formulaires...



LE SERVEUR WEB DE L'AFP : Services personnalisés

■ Le serveur Web de l'AFP : page de présentation des différents services .

<http://afp.com>

