

Tricot, A. (1994). Recherche d'information dans des documents non-linéaires et récupération volontaire en mémoire. *Actes du 1<sup>o</sup> Colloque "Jeunes Chercheurs en Sciences Cognitives"*, La Motte d'Aveillans, 23-25 Mars (pp. 129-138).

## **Recherche d'information dans des documents non-linéaires et récupération volontaire en mémoire**

### **Introduction**

Dans les domaines bien distincts de la recherche d'information sur support informatique (RI) et de la récupération volontaire en mémoire humaine, le début des années 90 est marqué par des déclarations d'intentions collaboratives d'auteurs très réputés : van Rijsbergen écrit que la recherche consacrée à la RI "établit des liens solides entre l'IA, les systèmes de gestion de base de données (SGBD), les interfaces homme-machine (IHM) et les sciences cognitives" (van Rijsbergen & Agosti, 1992, p.193). Anderson (1990) affirme que l'activité de récupération volontaire est plus "naturelle" que l'activité de résolution de problème et que l'étude de cette activité mentale devrait marquer un tournant de la psychologie cognitive. Pour ce même auteur (voir notamment Anderson & Milson, 1989) c'est en se basant sur les principes de la RI que l'on doit élaborer des modèles généraux de la récupération en mémoire (intégrant la récupération volontaire). Son modèle adaptatif de la mémoire en est une première ébauche.

Le but de cette communication est de montrer l'intérêt théorique et surtout pratique de ces propositions, et d'essayer de préciser quelques modalités de cette probable collaboration. D'un côté la psychologie cognitive pourrait apporter à la RI des formats et des modes d'organisation des connaissances plus souples et plus riches, ainsi que des modèles "naturels" de la recherche d'information. D'un autre côté la RI pourrait fournir à la psychologie cognitive des situations de recherche d'informations souples, contrôlées expérimentalement, et partiellement organisées par le sujet.

Après avoir rappelé quelle était l'histoire et l'actualité des relations entre ces deux thèmes de recherche, j'évoque grossièrement les quelques concepts impliqués. Dans le chapitre suivant je fais un point sur les aspects (cognitifs) communs à la RI et à la récupération volontaire en mémoire. Enfin je tente de montrer comment et sous quelles conditions ces deux problématiques peuvent s'éclairer l'une l'autre, à travers les exemples de la modélisation des stratégies de recherche d'information et des tâches de recherche d'information. Ce dernier chapitre évoque largement les documents non-linéaires, qui concernent plus particulièrement

ma recherche et qui offrent un environnement dans lequel la recherche d'information s'apparente fortement à une résolution de problème.

## **Etat des relations entre RI et récupération volontaire en mémoire**

### *Historique des relations entre bases de données et mémoire*

La modélisation de l'organisation de la mémoire humaine est typiquement, depuis la fin des années 60, un thème où les liens entre psychologie cognitive et informatique sont très étroits. Les premiers modèles de mémoire sémantique (Collins & Quillian, 1969), en particulier, sont célèbres pour, d'un côté avoir fait avancer la psychologie cognitive, d'un autre côté être irréalistes tellement ils sont "informatiques".

L'étude de la récupération involontaire en mémoire est aussi une partie de la psychologie cognitive souvent calquée, jusqu'à la caricature, sur les progrès de l'informatique. On s'accorde aujourd'hui sur l'irréalisme des modèles d'exploration sérielle exhaustive (Sternberg, 1966) ; les modèles de génération-reconnaissance (Kintsch, 1970 ; Anderson & Bower, 1972), ont constitué un grand pas en avant, mais ils ont le désavantage de ne pas prendre en compte le fait selon lequel des items peuvent être rappelés même s'ils ne sont pas reconnus.

Plus généralement en psychologie cognitive, notre époque semble marquée par la relative perte de popularité de modèles procéduraux et quelques promesses intéressantes du connexionnisme : encore deux modèles empruntés à l'informatique ou aux mathématiques.

L'échange d'idées entre ces deux domaines de recherche se fait aussi dans l'autre sens : par exemple, c'est à partir du concept psychologique d'associationnisme, que Bush (1945) a tracé les grandes lignes des conceptions classiques (indexage) et modernes (base personnelle, hypertexte) des SGBD.

### *Les relations entre RI et récupération volontaire*

Des auteurs comme Baddeley (1990) ou Tulving (1983) reconnaissent que la récupération volontaire en mémoire est très mal connue, principalement à cause de difficultés méthodologiques. Par contre, Anderson (1990) affirme que l'activité de récupération volontaire doit être étudiée, en utilisant les principes de la RI.

D'un autre côté, avec les progrès et la souplesse introduite dans les SGBD depuis 10 ans, une problématique ergonomique est en train de se poser : l'étude de l'activité de recherche d'information. En effet, les apports de l'I.A. (en représentation des connaissances, notamment avec les modèles objets), de la logique floue et des hypermédias aux bases de données sont tels que :

- l'utilisation n'est plus réservée aux experts,
- l'utilisation ne se réduit plus à l'interrogation mais peut supporter des activités comme l'apprentissage,
- l'interrogation n'est plus une activité d'exécution mais de résolution de problème (voir la distinction posée par Hoc, 1987).

Autrement dit, d'un côté comme de l'autre, on dispose d'une situation nouvelle qui ne pourra apparemment se développer que dans le cadre de collaborations. Si on ne dispose pas de résultats nouveaux sur la collaboration entre les deux disciplines depuis quelques années, on se trouve bien à un moment où les relations entre ces deux thèmes de recherche sont sur le point de se développer.

## **Les concepts utilisés**

### *Mémoire humaine et récupération*

Le modèle de Tulving (1983), toujours d'actualité, est un modèle de la récupération involontaire. En voici une simplification : il y a préalablement une phase de "création d'un engramme" où un événement est encodé (phase d'apprentissage) ; cet événement a alors une place, une trace. La deuxième phase est la "question" : dans un contexte quelconque (résolution de problème, question de l'expérimentateur), le sujet doit récupérer l'information à partir d'un indice (ce que Tulving appelle l'ecphorie). Cet indice :

- peut être similaire à l'événement encodé : c'est cas de la reconnaissance,
- il peut avoir un lien avec l'événement encodé (le plus classiquement ce lien a été induit lors de la phase d'apprentissage) : c'est le cas du rappel indicé,
- il peut être "temporel" : le sujet doit récupérer l'ensemble des informations stockées lors d'une phase identifiée comme étant "la phase d'apprentissage" : cas du rappel.

L'étude expérimentale de la récupération involontaire se consacre donc à un mécanisme d'appariement item / indice. Les situations du genre "à quelle place se trouve ma voiture que j'ai laissé samedi dernier sur le parking de l'aéroport" peuvent être traitées par ce type de mécanisme... ou par des raisonnements inférentiels beaucoup plus complexes. D'autre part, ces situations complexes sont difficiles à étudier expérimentalement.

Par exemple, Rippoll (1993) a montré dans le cas du raisonnement par analogie, combien il était difficile de modéliser la phase de récupération en mémoire du problème source, autrement que comme un processus peu contrôlé et laissant une large place au hasard. Autrement dit, il y a deux problèmes : comment étudier un processus complexe, de haut niveau mais peu contrôlé ? comment étudier un processus dont le déclenchement ne semble pas directement lié à des caractéristiques de la situation stimulus ?

Avant de passer à la partie suivante, il faut rappeler que les modèles de la récupération involontaire sont toujours sous-tendus, plus ou moins explicitement, par un modèle de la mémoire à long terme (MLT). Comme je vais parler de recherche d'information dans les bases de données et tenter un parallèle avec la récupération volontaire en mémoire, il faut que je pose une limite théorique à tous les travaux sur la récupération : les difficultés des recherches sur la récupération sont peut-être dues aux modèles de l'organisation des connaissances en MLT.

### *SGBD et RI*

#### Les SGBD

Un SGBD est composé :

- d'une organisation : soit un ensemble d'éléments, caractérisé par sa structure, son extensibilité (capacité à s'étendre) et sa dynamique (capacité à évoluer),
- d'un (de) mécanisme(s) de recherche dont la finalité est de sélectionner des éléments en fonction d'une requête de l'utilisateur,

- d'objets indexés, de taille et de formats différents, parfois variables,
- de relations entre objets, qui peuvent être logiques, sémantiques, ou simplement d'appartenance.

## La RI

Le but général d'une RI est de sélectionner les items pertinents (rappel) et de ne pas sélectionner les items non-pertinents (précision). Salton & Mc Gill (1983) ont décrit le principe rationnel qui guide une RI :

C désigne le coût engagé (par les réponses non-pertinentes, le temps passé à la recherche...)

G le gain obtenu par une réponse pertinente

$p[A]$  la probabilité pour que l'item A soit pertinent

Un système rationnel d'interrogation obéit à :

$$p[A]G < C$$

Le problème principal est l'estimation de  $p[A]$ .

Un deuxième niveau de problème est l'adéquation entre le "besoin d'information" de l'utilisateur et les requêtes qu'il formule au système. Il y a globalement trois aspects :

- la connaissance que l'utilisateur a du système, soit son degré d'expertise dans le langage de requête, sa connaissance du contenu, de la structure et du fonctionnement du système,
- le besoin d'information de l'utilisateur, qu'il peut clairement identifier ou pas, qui peut évoluer ou pas,
- le retour de l'utilisateur sur ses sélections, sa possibilité de reformuler une requête.

Le constat habituel est que moins les connaissances en (a) sont importantes, moins les "réponses" du système seront satisfaisantes. Une des solutions est "d'aider" voire de "suggérer des pistes" à l'utilisateur. Or, une fois les aspects techniques réglés, il reste le niveau sémantique (Kok, 1991) : que signifie telle donnée, dans tel contexte, pour tel utilisateur ? Soit un problème de pertinence.

## Les supports classiques

Les SGBD de 1° génération (hiérarchiques, réseaux) et de 2° génération (relationnels) ont pu atteindre de hauts niveaux de subtilité et ont pu être dotés de fortes puissances de calcul. La seule raison qui justifie qu'on les occulte ici est l'intermédiaire obligatoire et souvent fastidieux d'un langage de requête (d'ailleurs le problème de l'interface langage de programmation / SGBD se pose aussi dans ce type d'environnement).

## L'apport des modèles objets

Les langages orientés objets sont des langages informatiques basés sur une idée commune à l'IA et à la psychologie cognitive des années 70 : l'idée d'objets structurés, qui est une façon de représenter des connaissances non pas comme objets élémentaires, mais comme objets complexes, munis de caractéristiques et de variables. Les quatre caractéristiques des SGBD objets sont (Gardarin & Valduriez, 1991) : la possibilité de gérer des objets complexes (y compris des images), l'interface programmation (les langages objets sont plus "au format" des données), la dynamique, et l'extensibilité. Même si, dans la pratique, l'utilisation n'est pas toujours aisée, les modèles objets accroissent le rôle de l'utilisateur : ce dernier peut effectuer une opération sur un objet et rendre cette opération exécutable par n'importe quel autre utilisateur. Les "nouveau-tés" apportées par l'utilisateur peuvent aussi être des "types d'objets"

ou des "type d'opération" (ainsi l'utilisateur peut définir des nouveautés au niveau local ou au niveau global).

#### L'apport de la logique floue

FIRST (Guardalben & Lucarella, 1993) est un système basé sur la logique floue (l'argument est que la pertinence n'est pas une variable binaire mais continue) qui détermine des attributs aux objets, attributs pouvant prendre des valeurs différentes selon les différents contextes (requêtes). Les objets sont reliés entre eux par des liens typés, qui précisent la nature de la relation (sémantique) et qui ont une valeur de vérité qui reflète la force de l'association entre les concepts. Ainsi :

- si la requête  $q$  porte sur le sujet  $c$  et que le document  $A$  porte sur le sujet  $c$ , alors  $A$  est pertinent
- si la requête  $q$  porte sur le sujet  $c$  et que le document  $B$  a un lien avec le sujet  $c$ , alors  $B$  est pertinent selon la valeur du lien qui l'associe à  $c$ ,
- et ainsi de suite pour l'ensemble des documents qui ont un lien avec  $c$  (ensemble que l'on nomme agrégat),
- on peut optimiser l'agrégat, faire peser sur lui des contraintes, ainsi
- à partir de la requête  $q$ , on détermine un sous ensemble flou  $R_q$  de documents
- et un sous ensemble limité  $R_q(a)$ , sur lequel a porté une contrainte  $a$  qui détermine une valeur seuil des liens.

On peut présenter à l'utilisateur les ensembles  $R$  par valeur décroissante des liens et indiquer ces valeurs. Les auteurs ont conduit une expérience sur une base de données de 300 documents (175 concepts). Ils ont enregistré 30 requêtes et mesuré la précision et le rappel. Ces deux indices varient inversement ; Guardalben et Lucarella ont identifié un optimum (soit un équilibre entre pertinence estimée par l'utilisateur et nombre de suggestions du système).

On voit que ce type de système ne change pas fondamentalement le mécanisme d'interrogation mais que peu à peu la notion de pertinence est "intelligemment" intégrée (rationnellement pris en compte).

#### L'apport des hypermédias

Un hypermédia est un (grand) ensemble de données multi-média, généralement facile d'accès, et structuré en réseau : un ensemble de noeuds connectés par des liens. Chaque lien part d'un ancrage (mot, zone d'écran, icône) dans le noeud d'origine, cet ancrage étant manifesté par un bouton (mot en gras, surligné, partie encadrée, icône). Ce type de système constitue un ensemble de configurations virtuelles : chaque utilisateur choisit son trajet parmi les données. La "philosophie" des hypertextes est la libre association entre les éléments de connaissance (Bush, 1945) et la possibilité de faire des ajouts personnels, voire de tout stocker (Nelson, 1981).

D'un point de vue cognitif, l'argument des pionniers dans le domaine est toujours le même : c'est plus "naturel" ou "intuitif". Pour ces non-psychologues, "naturel" semble concrètement vouloir dire :

- limitation du nombre de procédures, ou,
- intégration des procédures, comme dans les modèles des schémas,
- fonctionnement de la mémoire sur un mode associatif "peu logique", comme dans le modèle de Collins & Loftus (1975).

Pour les bases de données, ce type d'interface facilite une démarche de "browsing", (to browse = explorer, butiner, en quelque sorte "flâner parmi les données"), qui présente trois avantages principaux (Thomson & Croft, 1989) :

- il n'est pas besoin de formuler précisément par avance une recherche,
- la procédure d'interrogation est plus facile,
- il y a un "feed-back" direct de l'interrogateur sur ses sélections (selon la pertinence ou non des sélections).

La démarche de browsing dans des environnements linéaires a aussi un grand intérêt pour les tâches d'apprentissage. L'argument classique est qu'un hypermédia n'imposant pas de parcours, l'apprenant pourra choisir ce qui l'intéresse, ce qui lui semble pertinent, en utilisant le "mode d'apprentissage" qui lui convient. Cet argument est séduisant, mais très vague. Mais on peut trouver, sur des problèmes très précis, des études comme celle de Pierce & al. (1993), qui ont montré qu'une exploration relativement libre de l'espace problème, sans contrainte de cheminement vers le but, produit un schéma de résolution (exportable vers d'autres situations) de meilleure qualité qu'une exploration contrainte (cheminement dirigé vers le but).

### Conclusion

Avec les modèles objets, les hypermédiats et la logique floue, la RI se fait dans une base structurée de façon associative, qui pourrait être un modèle rudimentaire mais acceptable de la mémoire sémantique. L'intérêt de l'étude de la mémoire à travers des modèles SGBD peut être discuté. Les limites les plus évidentes sont liées aux aspects sémantiques et de pertinence, qui semble-t-il, sont de mieux en mieux pris en compte et sont dépendants d'un contexte. D'autre part on voit, depuis l'introduction des modèles objet, l'accroissement du rôle de l'utilisateur dans l'interrogation, parfois même dans la structuration et le fonctionnement de la base, et dans la création d'objets, ainsi, et c'est probablement le plus important, un contrôle direct et continu sur la pertinence des items sélectionnés.

### Aspects (cognitifs) communs à la RI et la récupération volontaire

Ce qui il y a en commun entre la recherche et la récupération d'information c'est :

- un problème à résoudre (avec les nouveaux SGBD, on est bien devant *une tâche à effectuer sans savoir immédiatement comment* -ce qui pourrait être une définition de "problème").
- une base organisée de connaissances,
- un (des) mécanisme(s) de recherche.

A chacun des trois niveaux, il y a d'importantes différences :

- **au niveau du "problème"** : dans la récupération volontaire en mémoire il y a surtout une difficulté méthodologique (comment étudier un problème dont on ne sait s'il va être résolu par un simple mécanisme d'appariement ou par un mécanisme plus complexe ? quels sont les observables ?) ; du côté de la recherche d'information, on a à faire à des situations de résolution de problème de plus en plus "classiques" ; il semble qu'un des points sur lesquels les deux environnements puissent être comparables, c'est dans la possibilité de changer de problématique (mais là encore, quid des observables ?).
- **au niveau de la base** : une base de données est connue et logique ; la mémoire est moins connue et moins logique. Il semble néanmoins qu'une base du type FIRST, c'est à dire à la fois "orientée objet" et gérée par la logique floue, constitue un modèle acceptable de la

MLT. La prise en compte systématique du contexte de la connaissance est un grand pas en avant. Il devrait être possible, comme cela est développé à Marseille par Jean-Paul Coste, de proposer des bases constituées par l'utilisateur lors d'une tâche d'apprentissage. Des bases partiellement initiées par l'utilisateur existent notamment dans des contextes d'enseignement (ARCADE à Grenoble, MACWEB à Montpellier ou AQUANET à Palo Alto). Il pourrait être très intéressant de proposer des tâches de recherche d'information dans une base constituée par le sujet lui-même.

- **au niveau des mécanismes de recherche** : nos connaissances sur les activités mentales impliquées lors de la recherche d'information ou de la récupération volontaire en mémoire sont très limitées. Une première avancée peut être faite si l'on admet que chacun des deux niveaux ci-dessus (la base et le problème) est descriptible et relativement isomorphe en environnement naturel ou artificiel.

A un niveau plus général, une des limites de cette proposition et de la recherche interdisciplinaire en sciences cognitive est la suivante : qu'est ce que la signification et qu'est-ce que la pertinence ?

### **Perspectives de recherche**

Dans cette partie, j'évoque deux recherches en cours qui, s'inscrivant dans le cadre de la recherche sur la RI, pourraient fournir des pistes de travail à la recherche sur la récupération volontaire en mémoire.

#### *La modélisation des stratégies de recherche d'information*

Dans les systèmes classiques, la recherche d'information pouvait être résumée à : (a) formuler une requête, (b) lire les documents sélectionnés. Il n'y avait pas d'activité cognitive spécifique, mais une alternative : on savait formuler une requête ou pas. D'autre part, Kulthau & al. (1988) ont montré que des SGBD plus élaborés étaient sous-utilisés par les utilisateurs occasionnels, notamment à cause d'une représentation très proche d'une "procédure linéaire" du processus.

Une première ébauche de modélisation du traitement en situation de recherche d'information (horaires d'avion, feuille de paie) a été proposée par Guthrie (1988) : (a) formation d'un but cognitif ; (b) sélection d'une catégorie d'information ; (c) extraction d'information ; (d) intégration à l'information préalablement extraite ; (e) recommencer jusqu'à ce que le but soit atteint.

Là où la situation devient différente grâce aux nouveaux SGBD c'est surtout au niveau de (b) : le système fait des suggestions, les critères ne sont plus seulement croisés. Les relations entre thèmes et items ne sont plus strictement hiérarchiques.

Rouet (1992) a repris le modèle de Guthrie et l'a adapté aux hypermédia : la phase (a) du modèle de Guthrie correspond au traitement de la question : pour que le processus de recherche puisse être initié, il faut qu'une représentation cognitive de l'objectif ait été construite. A ce niveau intervient la complexité des questions, qui rend plus ou moins difficile la construction de cette représentation.

La phase (b) correspond à la sélection d'un thème. La formulation des questions affecte le déroulement de cette phase (...). Cependant, comme le montre l'évolution des résultats avec l'entraînement, il faut à nos sujets une certaine expertise pour différencier leurs stratégies de sélection selon le type de question.

L'extraction de l'information (c) est ici un processus plus complexe que dans les documents utilisés par Guthrie. Dans un cas il s'agit de localiser une valeur dans une table, alors que dans

l'hypertexte il s'agit de construire une représentation sémantique d'un passage de texte. Rouet (1990) suggère que l'exécution de cette phase requiert une transition entre différents niveaux de processus : la compréhension d'un passage spécifique suppose un ensemble de processus locaux, alors que les autres phases de la recherche requièrent une approche "globale" de la tâche.

Les phases (d) et (e) correspondent respectivement à la mise en relation de plusieurs passages, et à la décision de prolonger ou de cesser la recherche. Ces phases ne concernent en principe que les questions complexes, qui demandent plus d'une sélection. Cependant le sujet est amené à recycler dès lors que le premier passage sélectionné ne contient pas d'information utile (sélection d'un thème non-cible), ou bien s'il considère ne pas avoir obtenu assez d'information pour répondre à la question. Evaluer si le but est atteint pose parfois un problème : il se peut notamment que le sujet oublie en cours de recherche l'objectif défini initialement.

On voit, plus généralement, avec ce type de système, que la boucle vers (a) peut se faire à chaque sélection de fichiers, voire dès l'étape (b). Du coup, on admet plus facilement que (a) peut se modifier à chaque étape. La boucle après (c) est particulièrement intéressante :

- boucler au dessus de (c) = balayage,
- boucler au dessus de (b) = essayer un autre thème,
- boucler au dessus de (a) = changer de but (peu probable : on change de but quand on a intégré une information).

D'un autre côté, on peut, avec un but stable, sélectionner différentes catégories d'information (phase (b) de Guthrie) : exemple, information de niveau supérieur, information de niveau inférieur, information de même niveau (contextuelle par exemple).

La figure 1 représente une modélisation de l'activité de recherche d'information dans un support non-linéaire, où j'ai distingué la phase (b) de Rouet de la phase (b) de Guthrie.

Ainsi, selon la forme du réseau, le type de tâche et la connaissance que le sujet a du système, on pourra décrire des stratégies de recherche d'information à partir de la forme des déplacements du sujet : par exemple, l'ouverture successive d'au moins 4 cartes de niveaux de détail de plus en plus fins pourra être interprétée comme stratégie de focalisation.

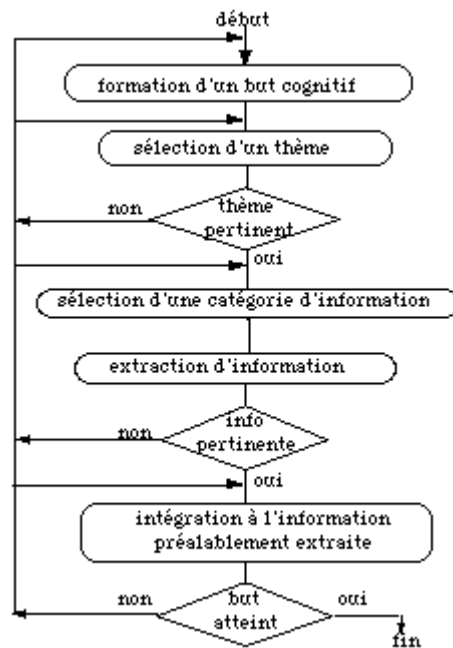
*La modélisation des tâches de recherche d'information*

La description de l'activité de traitement en situation de recherche d'information doit aussi passer par une description précise des différents types de tâche que l'on peut effectuer. Une telle description permettrait : de déterminer des variables dépendantes pour des expérimentations, de décrire des stratégies de navigation et des environnements particulièrement adéquats à tel ou tel type de tâche (cette dernière phrase peu sembler "évidente", mais dans le domaine des hypermédias les descriptions de tâche et les définitions de VD sont très peu avancées : cf. Tricot, 1993).

Figure 1. *La recherche d'information dans un support non-linéaire*

		Formulation de la question	
		précises	non-précise
Données ciblées répondre	ponctuelle	chercher un renseignement	explorer
	nombreuses	chercher	agréger





Par exemple, pour la recherche d'information, j'ai défini quatre tâches par deux critères croisés (Tricot, 1994) :

- "chercher un renseignement" : l'utilisateur a une question précise qu'il sait formuler et qui porte sur une donnée ponctuelle de la base (notons au passage que ce type de tâche, prise isolément, n'est pas du tout favorisée par les environnements hypermédia : un simple index alphabétique est bien meilleur),
- "explorer" : l'utilisateur n'a pas de question précise et/ou ne sait pas la formuler de façon satisfaisante, mais il pense qu'il va trouver quelque chose d'intéressant pour lui dans la base,
- "chercher" : l'utilisateur a une demande qu'il sait formuler et qui porte sur un ensemble de données de la base (cet ensemble pouvant être identifié comme tel ou non dans la base),
- "agréger" : l'utilisateur n'a pas de question précise et/ou ne sait pas la formuler de façon satisfaisante, mais il pense qu'il va trouver de nombreuses données intéressantes pour lui dans la base. Un cas particulier : "ré-agrégé" : l'utilisateur a préalablement des connaissances qui figurent dans la base (sous forme de données) et la consultation lui permet de "faire de nouvelles associations", "compléter une connaissance". Il est possible que ce type de tâche puisse rendre telle ou telle connaissance "fonctionnelle" dans un contexte nouveau, en favorisant des contextualisations différentes (cf. Bastien, 1992) : il est possible de montrer à un sujet que telle connaissance (qu'il possède déjà) peut être utilisée efficacement dans tel contexte (qu'il connaît peu).

Une consultation met en jeu alternativement plusieurs stratégies et/ou des stratégies hybrides. Par exemple, la stratégie correspondant à une tâche d'exploration quand elle se poursuit sur plusieurs objets devient stratégie "browsing". Elle peut aussi évoluer vers une stratégie "chercher" ou une stratégie "agréger".

#### *Structure des bases de données et sémantique cognitive*

D'autres travaux impliquant la RI et la psychologie cognitive pourraient, me semble-t-il, être conduits. Par exemple, une des limites des bases de données actuelles concerne le concept d'agrégat qui n'est qu'un ensemble d'items, alors que les travaux sur la catégorisation et en sémantique cognitive proposent des modèles munis des notions intéressantes comme la

hiérarchie et les représentations privilégiées (voir la synthèse de Cordier, 1993). La relation de "représentation" d'un item pour un autre est beaucoup plus souple et riche dans ces modèles que dans les bases de données, y compris "objet".

## **Conclusion**

Il me semble que l'on peut maintenant poser sérieusement deux questions :

- a) en quoi les mécanismes de recherche d'informations peuvent nous aider à étudier la récupération volontaire d'information ?
- b) en quoi les travaux sur la récupération, sur la MLT et sur la catégorisation peuvent nous aider à concevoir des SGBD ?

### *Réponse à la question (a)*

Des avancées dans l'étude des stratégies de recherche d'information pourront nous permettre de décrire précisément dans quel contexte et pour quel problème le sujet utilise telle stratégie (de balayage, de focalisation, d'exploration systématique, etc). Une formalisation de six stratégies de recherche d'information est en cours à Aix (équipe CREPCO) et Marseille (équipe Hermès) .

### *Réponse à la question (b)*

Les travaux sur la mémoire et sur la catégorisation essaient, entre autres, de décrire la nature et le fonctionnement des liens qu'il y a entre au moins deux items stockés. Cette description souligne le rôle de "représentant" (indice, prototype, "exemplar model") qu'un item peut avoir pour au moins un autre. La richesse des modèles dans ce domaine dépasse de loin les bases de données.

Il est encore assez rare de trouver des conceptions de SGBD qui veulent "correspondre" à la mémoire humaine. Il est encore plus rare de lire des travaux dans lesquels la récupération volontaire en mémoire est étudiée. Vouloir rapprocher deux champs aussi peu avancés est sans doute farfelu. Pourtant, à la fin de la 2<sup>e</sup> guerre mondiale est paru un article encore plus farfelu (Bush, 1945), qui proposait la même démarche en se basant... sur l'associationnisme de Locke. Ce texte est un exemple cruel pour tous les rationalistes : l'argumentation n'est basée sur aucun fait avéré (c'est d'après l'auteur lui-même un discours d'"anticipation") ; mais Bush y invente une technologie qui est entrain de révolutionner les SGBD, l'EAO, et plus largement, le génie logiciel : les hypertextes. 15 ans plus tôt, le même auteur, avec la même idée de base, eut une inspiration encore plus délirante : il inventa l'architecture parallèle des ordinateurs.

Peut-être sommes-nous armés aujourd'hui, non-pas pour avoir une autre idée géniale, mais pour étudier rationnellement les stratégies de recherche d'information et la récupération volontaire en mémoire.

Ce travail est soutenu par une allocation MRT (dans le cadre d'une thèse sur les hypermédias) et par un contrat entre les équipes Hermès, CREPCO et la DRED. Je remercie Jean-Paul Coste ainsi qu'un lecteur anonyme pour leurs commentaires sur une version antérieure du manuscrit.

## **Bibliographie**

- Anderson, J.R., & Bower, G.H. (1972). Configurational properties in sentence memory. *Journal of Verbal Learning and Verbal Behavior*, 11, 594-605.
- Anderson, J.R., & Milson, R. (1989). Human memory, an adaptive perspective. *Psychological Review*, 96 (4), 703-719.

- Anderson, J.R. (1990). *The adaptative character of thought*. Hillsdale, NJ: LEA.
- Baddeley, A. (1990). *Human memory. Theory and practice*. (trad. fr. "La mémoire humaine. Théorie et pratique", Grenoble: PUG, 1992).
- Bastien, C. (1992). Le décalage entre logique et connaissances. *Le Courrier du CNRS*, 79 "Les sciences cognitives", 38.
- Bush, V. (1945). As we may think. In J.M. Nyce & P. Kahn (Eds.) (1991). *From Memex to Hypertext, Vannevar Bush and the mind's machine*. Boston, MA: Academic Press.
- Collins, A.M., & Quillian, M.R. (1969). Retrieval time for semantic memory. *Journal of verbal learning and verbal behavior*, 8, 240-247.
- Collins, A.M., & Loftus, E.F. (1975). A spreading activation theory of semantic processing. *Psychological Review*, 82, 407-428.
- Cordier, F. (1993). *Les représentations cognitives privilégiées*, Lille: PUL.
- Guardalben, G., & Lucarella, D. (1993). Information retrieval based on fuzzy reasoning. *Data & Knowledge Engineering*, 10, 29-44.
- Gardarin, G., & Valduriez, P. (1990). *SGBD avancés*. Paris: Eyrolles.
- Guthrie, J.T. (1988). Locating information in documents : examination of a cognitive model. *Reading Research Quarterly*, 23 (2), 178-199.
- Hoc, J.-M. (1987). *Psychologie cognitive de la planification*. Grenoble: PUG.
- Kintsch, W. (1970). *Learning, memory and conceptual process*. New York, NY: Wiley.
- Kok, A.J. (1991). A formal approach to user models in data retrieval. *International Journal of Man-Machine Studies*, 35, 675-693.
- Kuhlthau, C.C., Belvin, R.J., & George, M.W. (1989). *Flowcharting the information search, a method for eliciting user's mental maps*. School of Communication, Information and Library Studies, Rutgers, NJ: The State University.
- Nelson, T.H. (1981). *Literary machines*. Swathmore, PA: Nelson (Sausalito, CA: Mindfull Press, 1987).
- Pierce, K.A., Duncan, M.K., Gholsn, B., Ray, G.E., & Kambi, A.G. (1993). Cognitive load, schema acquisition, and procedural adaptation in nonisomorphic analogical transfer. *Journal of Educational Psychology*, 85 (1), 66-74.
- Rippoll, T. (1993). *Recherche en mémoire d'un problème analogue*. Thèse de l'Université de Provence, UFR de Psychologie.
- Rouet, J.-F. (1990). Interactive text processing in inexperienced (hyper-) readers. In A. Rizk, N. Streitz & J. André (Ed.s) *Hypertext, Concepts, systems and applications*, Proceedings of the European conference on Hypertext, Versailles, Cambridge: CUP.
- Rouet, J.-F. (1992). Apprendre à lire un hypertexte - Une étude expérimentale. à paraître dans les Cahiers de Linguistique Sociale
- Salton, G., & Mc Gill, M.J. (1983). *Introduction to modern information retrieval*. New York, NY: McGraw Hill.
- Simon, H.A. (1991). Cognitive architectures and rational analysis, comment. In K. van Lehn (Ed.) *Architectures for intelligence*, Hillsdale, NJ: LEA.
- Sternberg, S. (1966). High speed scanning in human memory. *Science*, 153, 652-654.

- Thompson, R.H., & Croft, W.B. (1989). Support for browsing in an intelligent text retrieval system. *International Journal of Man-Machine Studies*, 30, 639-668.
- Tricot, A. (1993). Ergonomie cognitive des systèmes hypermédia. Actes du Colloque de prospective "Recherches pour l'Ergonomie", CNRS (PIR Cognisciences), Toulouse.
- Tricot, A. (1994). il s'agit de l'article : Tricot, A. (1995). Un point sur l'ergonomie des interfaces hypermédia. *Le Travail Humain*, 58 (1), 17-45.
- Tulving, E. (1983). *Elements of episodic memory*. Oxford: Oxford University Press.