

# L'analyse de références bibliographiques assistée par ordinateur

## The Analysis of Bibliographic References Using a Computer

### Análisis de referencias bibliográficas con asistencia de la computadora

Marcel Ouellet

Volume 38, numéro 2, avril-juin 1992

Analyse et gestion de l'information textuelle

URI : <https://id.erudit.org/iderudit/1028615ar>

DOI : <https://doi.org/10.7202/1028615ar>

[Aller au sommaire du numéro](#)

#### Éditeur(s)

Association pour l'avancement des sciences et des techniques de la documentation (ASTED)

#### ISSN

0315-2340 (imprimé)

2291-8949 (numérique)

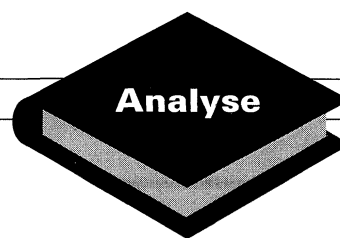
[Découvrir la revue](#)

#### Citer cet article

Ouellet, M. (1992). L'analyse de références bibliographiques assistée par ordinateur. *Documentation et bibliothèques*, 38(2), 103-109.  
<https://doi.org/10.7202/1028615ar>

#### Résumé de l'article

Comment traiter et analyser à l'aide de l'ordinateur des résultats d'interrogation de bases de données bibliographiques pour étudier les caractéristiques de la littérature sur un sujet donné ? C'est la question à laquelle cette recherche tente de répondre en prenant pour sujet les aspects cognitifs en sciences de l'information. La démarche réalisée se compose de quatre étapes principales soit : 1) la collecte des données (références bibliographiques extraites de LISA et de ERIC); 2) le traitement (tri, codification, uniformisation et fusion) des données recueillies; 3) l'analyse des données traitées pour en extraire les caractéristiques de la littérature; et 4) la présentation et l'utilisation des résultats de l'analyse. Lors de l'application de cette méthode, deux logiciels ont surtout été utilisés : un traitement de texte (WordPerfect) et un système d'analyse de texte (SATO).



## L'analyse de références bibliographiques assistée par ordinateur\*

Marcel Ouellet

Centre de documentation — Postes de répartition  
Hydro-Québec (Région Montmorency)

*Comment traiter et analyser à l'aide de l'ordinateur des résultats d'interrogation de bases de données bibliographiques pour étudier les caractéristiques de la littérature sur un sujet donné? C'est la question à laquelle cette recherche tente de répondre en prenant pour sujet les aspects cognitifs en sciences de l'information. La démarche réalisée se compose de quatre étapes principales soit: 1) la collecte des données (références bibliographiques extraites de LISA et de ERIC); 2) le traitement (tri, codification, uniformisation et fusion) des données recueillies; 3) l'analyse des données traitées pour en extraire les caractéristiques de la littérature; et 4) la présentation et l'utilisation des résultats de l'analyse. Lors de l'application de cette méthode, deux logiciels ont surtout été utilisés: un traitement de texte (WordPerfect) et un système d'analyse de texte (SATO).*

### **The Analysis of Bibliographic References Using a Computer**

*How can the bibliographic references retrieved from data bases be analysed in order to study the characteristics of the literature of a given subject? The following research attempts to answer that question using the cognitive aspects of information science. The procedure consisted of four steps: 1) the retrieval of the data (bibliographic references found in LISA and ERIC); 2) the manipulation of this data (sorting, codification, standardisation and merging); 3) the analysis of the data in order to draw conclusions regarding the characteristics of the literature; and 4) the organisation and use of the results of the analysis. The methodology required two programmes: a word processing software (WordPerfect) and a text analysis software (SATO).*

### **Análisis de referencias bibliográficas con asistencia de la computadora**

*¿Cómo tratar y analizar con ayuda de la computadora, los resultados de la interrogación de bases de datos bibliográficos para estudiar las características de la literatura sobre un tema dado? Esta investigación trata de responder a esa pregunta, tomando como tema los aspectos cognitivos en las ciencias de la información. El estudio se compone de cuatro etapas principales: 1) la recopilación de datos (referencias bibliográficas extraídas de LISA y ERIC); 2) el tratamiento (selección, codificación, uniformización y fusión) de los datos recopilados; 3) el análisis de los datos tratados para extraer de los mismos las características de la literatura; y 4) la presentación y utilización de los resultados del análisis. Durante la aplicación de este método, se utilizaron sobre todo dos programas: uno de tratamiento de texto (WordPerfect) y un sistema de análisis de texto (SATO).*

Les bibliographies sont beaucoup plus qu'une simple liste de sources secondaires. Elles constituent d'ailleurs une des sources de données (auteurs, titres de périodiques, années de publication, descripteurs, mots des titres et résumés, ...) utilisées dans les études bibliométriques<sup>1</sup>. En effet, l'étude des caractéristiques de la littérature produite<sup>2</sup> dans un domaine<sup>3</sup> permet d'établir: la production de documents par variables de nature bibliographique: par auteur, par année, par pays, par langue, par titre de périodique; la production par variables faisant référence au contenu des documents: par descripteur, par les mots des titres et résumés; et la production d'une variable en fonction d'une autre, comme la

production des auteurs selon les années.

Dans les études bibliométriques, on constate un intérêt grandissant pour l'utilisation des résultats d'interrogation de bases de données<sup>4</sup>. Dans la présente recherche, nous voulons développer une méthode pour analyser les références bibliographiques extraites de bases de données sur support informatique en recourant aux possibilités de l'ordinateur à partir du repérage des notices jusqu'à la présentation des résultats de l'analyse. Par la même occasion, nous cherchons à connaître les caractéristiques de la littérature sur les aspects cognitifs en sciences de l'information.

\* Cette recherche a été effectuée dans le cadre d'un séminaire donné par madame Suzanne Bertrand-Gastaldy à l'École de bibliothéconomie et des sciences de l'information de l'Université de Montréal.

1. S.M. Lawani, «Bibliometrics: its Theoretical Foundations, Methods and Applications», *Libri*, vol. 31, no. 4 (1981), 294-315.
2. Il importe de ne pas les confondre avec les « analyses de citations » qui servent à cerner les caractéristiques de la littérature citée ou encore utilisée par les chercheurs d'un domaine.
3. S.M. Lawani, « Bibliometrics ... »
4. Paul McGhee et al., « Using Online Databases to Study Current Research Trends: an Online Bibliometric Study », *Library and Information Science Research*, vol. 9, no. 4 (1987), 285-291.

### Problématique (Revue de la littérature)

D'après A. E. Jackson<sup>5</sup>, puisque la recherche en ligne est une activité maintenant bien établie, le temps est venu pour les professionnels de l'information de développer de nouveaux services. En ce sens, Henri Dou et ses collègues<sup>6</sup> affirment que la réalisation « répétitive » d'interrogations en ligne ne constituera plus l'un des principaux aspects du travail documentaire.

Même si plusieurs utilisateurs se sont habitués à recevoir seulement une liste non organisée de données bibliographiques comme résultats de la recherche automatisée, d'autres désirent obtenir une information plus directement utilisable<sup>7</sup>. D'ailleurs, les besoins et les attentes des usagers varient beaucoup et cela amène le bibliothécaire à fournir des réponses différentes à des demandes semblables d'information. Ainsi, par exemple, si la plupart des gestionnaires désirent un produit édité, organisé, les chercheurs sont plutôt intéressés par l'obtention de résultats sur support magnétique afin de pouvoir les traiter ultérieurement<sup>8</sup>.

Les serveurs de bases de données offrent diverses possibilités de traitement des résultats d'une recherche automatisée. Toutefois, les opérations disponibles sont limitées et occasionnent des frais additionnels<sup>9</sup>. De plus, même si la plupart des serveurs permettent différents formats d'impression, ceux-ci diffèrent selon la base de données utilisée et sont souvent incompatibles avec les formats des autres serveurs<sup>10</sup>.

Nous savons par ailleurs que les nouvelles technologies offrent une capacité de stockage accrue, une plus grande facilité pour le télédownload, un transfert plus rapide, une capacité d'analyser de façon automatique des corpus de plus en plus grands<sup>11</sup>.

En réalité, les serveurs de bases de données sont réticents au traitement *a posteriori* des résultats<sup>12</sup>. Pourtant, d'après Judy K. Burnam, les serveurs ont peu à craindre des bibliothécaires, puisque ceux-ci n'ont pas comme mission d'exploiter les données réorganisées à des fins commerciales. Cette auteure pense également qu'il s'agit là d'une utilisation équitable de l'information qui a été payée<sup>13</sup>.

Même si le télédownload et le traitement des résultats de recherche présentent plusieurs avantages, il faut temps, argent et connaissances particulières pour accomplir ces opérations<sup>14</sup>.

Judy Hedden<sup>15</sup> évalue que le temps de réorganisation des résultats équivaut à quatre fois celui de la recherche automatisée. De plus, Judy K. Burnam affirme que les dépenses doivent être justifiées en termes de coût/bénéfice. En effet, toutes les opérations (télédownload, reformatage, édition, réorganisation et transfert des résultats) doivent permettre de livrer un produit dans un temps acceptable, dans une forme utilisable et ce, à un prix abordable<sup>16</sup>. A. E. Jackson fait remarquer que la réutilisation des données électroniques représente un investissement contrairement à un achat pour un usage unique.

Plusieurs logiciels peuvent servir à exploiter des données télédownloadées. Judy Hedden les a regroupés par catégories. Il y a d'abord les logiciels spécialement conçus pour traiter les résultats de recherche (par exemple, BiblioLink). Ils permettent, entre autres, d'éliminer les données inutiles tels les messages du serveur lors de l'édition des données télédownloadées.

Une autre catégorie regroupe les autres logiciels d'application. On y retrouve les traitements de texte avec lesquels on peut « nettoyer » les résultats, ajouter une introduction ou encore, fusionner plusieurs fichiers. D'ailleurs McPherson<sup>17</sup> estime que le traitement de texte est le moyen le plus courant et le plus flexible pour traiter des données télédownloadées. Cet auteur est toutefois conscient que certaines opérations sont difficiles, voire impossibles à effectuer avec ce seul type de logiciel, et qu'elles requièrent des logiciels spéciaux.

Judy Hedden inclut également les chiffriers et tableurs parmi les autres logiciels d'application. Avec eux, on peut analyser les données, réaliser des graphiques et calculer la distribution des fréquences. Les logiciels de gestion de bases de données font aussi partie de cette catégorie. Toutefois, la difficulté d'utiliser des champs de longueur variable et différents formats d'enregistrements en constitue le principal problème.

La dernière catégorie de logiciels décrits par Judy Hedden sont ceux qui intègrent plusieurs de ces applications.

Nous avons regroupé, dans la liste ci-après les types de traitement ou d'analyse des références bibliographiques sur support informatique qui sont suggérés par les auteurs consultés (Judy K. Burnam, Judy Hedden, A.E. Jackson et Malcolm G. McPherson):

- Le « nettoyage » des résultats de recherche, c'est-à-dire l'élimination des stratégies de recherche, des informations inutiles, des erreurs, des messages du système, des références non pertinentes, etc.;
- L'élimination des doubles et la fusion dans un même fichier des enregistrements provenant de différentes bases de données. La duplication est un problème important. Une étude réalisée par June E. Slach<sup>18</sup> conclut que, dans 31 recherches automatisées, le taux de duplication se situait entre 0 et 64 %;
- Une présentation plus conviviale des résultats, c'est-à-dire l'ajout d'une introduction, d'une page de titre,

5. A.E. Jackson, « After Downloading: Enhancing the Value of your Data », *Aslib Proceedings*, vol. 40, no. 4 (1988), 111-122.
6. Henri Dou et al., « Veille technologique et information documentaire: de l'usage de la bibliométrie dans les services de documentation », *Documentaliste*, vol. 27, no 3 (1990), 132-141.
7. A.E. Jackson, « After Downloading ... »
8. Malcolm G. McPherson et al., « Enhancing Search Results by Editing, Analysis and Packaging », *Information Services & Uses*, vol. 9, no. 1-2 (1989), 101-106.
9. *Ibid.*
10. A.E. Jackson, « After Downloading ... »
11. Henri Dou et al., « Veille technologique ... »
12. A.E. Jackson, « After Downloading ... »
13. Judy K. Burnam, « Users' Needs for Repackaging and Re-use of Information », *Information Services & Uses*, vol. 1 (1982), 359-366.
14. Malcolm G. McPherson et al., « Enhancing Search Results ... »
15. Judy Hedden, « Packaging and Transmitting Search Output », *Database End-user*, vol. 1, no. 3 (1985), 15-17, 20-22.
16. Judy K. Burnam, « Users' Needs ... »
17. Malcolm G. McPherson et al., « Enhancing Search Results ... »
18. June E. Slach, « Detection and Elimination of Duplicates from Multidatabase Searches », *Bulletin of the Medical Library Association*, vol. 73, no. 3 (1985), 235-237.

d'en-têtes, d'annotations, d'un texte explicatif, etc. On peut apporter les améliorations suivantes: la pagination du document, la standardisation des formats, la justification du texte et l'utilisation du caractère gras afin de faire ressortir les mots clés utilisés dans la recherche. Le bibliothécaire peut également identifier les documents disponibles à la bibliothèque;

- Le tri et l'indexation, c'est-à-dire l'utilisation de descripteurs pour un tri des enregistrements par sujet, la production d'un index des auteurs ou d'une liste permutée des titres, etc. A.E. Jackson attire l'attention sur un problème pouvant affecter le tri: le même contenu peut différer d'une base à l'autre dans sa forme. Il donne comme exemple les différentes façons de représenter la même date ou de citer un nom d'auteur (Jackson, A., Jackson AE, Jackson, A.E., etc.);
- L'analyse des stratégies de recherche ou du contenu des notices repérées. L'analyse des stratégies peut servir, par exemple, à identifier quelles bases de données procurent les meilleurs résultats. Une analyse du contenu peut fournir des indications sur les tendances de la recherche, la couverture des sujets dans un domaine, etc.

A.E. Jackson mentionne, au sujet des analyses bibliométriques, qu'elles peuvent faire ressortir des aspects qu'il serait très difficile à discerner autrement.

*There are a number of types of summary information which could be a great benefit to a user. A numerical breakdown of cited authors, journal coverage or corporate source for a data set gives insight into the important participants in a given field. An analysis of indexing or controlled terms may provide useful feedback to aid reformulation of SDI profiles, stored queries or the building of a thesaurus or domain dictionary. Still more informative can be cross-correlations. For example, a cross-correlation of authors will reveal major co-workers. Cross-correlating authors or corporate source with indexing terms may reveal the main areas of work of an individual or concern. When carried out over successive years it may be possible to chart the changes in emphasis<sup>19</sup>.*

Nous avons relevé un certain nombre d'études bibliométriques basées sur les résultats de recherches automatisées. C'est le cas d'une analyse réalisée par H.G. David et ses collègues portant sur l'évolution de la recherche dans le domaine du textile où les auteurs ont établi la fréquence des descripteurs selon les années<sup>20</sup>. Ils estiment que cette méthode est sans aucun doute la plus rapide et la plus facile parmi toutes celles qu'ils ont utilisées.

Dans leur investigation sur l'état actuel de la recherche concernant le développement cognitif, Paul McGhee *et al.*<sup>21</sup> ont produit des tableaux sur les occurrences de chaque descripteur en relation avec les périodiques et l'année de publication. Cette recherche montre comment les méthodes bibliométriques peuvent contribuer à déterminer les principaux champs de recherches d'une discipline. Ils concluent que des études semblables pourraient être menées pour évaluer l'apport de certains auteurs dans un champ d'étude spécifique.

Henri Dou *et al.*<sup>22</sup> ont procédé au traitement automatique de données provenant de deux bases de données. Au sujet de l'interrogation, ces auteurs font remarquer qu'il faut réaliser des ensembles larges en considérant qu'un « bruit » contrôlé est nécessaire, car la question trop précise prédétermine la réponse. Quant aux logiciels d'analyse automatique, les auteurs nous préviennent que ces logiciels fournissent des informations dont l'interprétation peut rapidement devenir affaire de spécialistes. Ils recommandent donc d'utiliser au départ des logiciels simples.

Enfin, Gisèle Vergnes *et al.*<sup>23</sup> soulignent les difficultés des analyses bibliométriques à partir de références provenant de différentes bases de données, en particulier les nombreuses références en double ou non pertinentes. Ils mettent l'accent sur l'importance d'utiliser plusieurs bases de données pour valider les données du corpus, la méthodologie et les résultats.

### Méthodologie

Dans notre propre expérimentation, la méthodologie se résume à quatre grandes étapes: 1) la collecte des données; 2) le traitement des données

recueillies; 3) l'analyse des données traitées; et 4) la présentation et l'utilisation des résultats de l'analyse. Avant de présenter les objectifs et les procédures spécifiques à chacune de ces étapes, nous décrivons succinctement les données et le matériel utilisés.

### Les données

Les notices bibliographiques recueillies réfèrent à des articles de périodiques sur les aspects cognitifs en sciences de l'information. Elles ont été extraites de deux bases de données bibliographiques sur CD-ROM<sup>24</sup>: une en sciences de l'information (LISA; 1969-1990) et une autre contenant à la fois des notices en bibliothéconomie et en sciences cognitives (ERIC; 1966-1990). Nous avons obtenu trois bases de données locales: une bibliographie pour chaque base et une cumulative (fusion des deux autres bases).

Dans la bibliographie créée à partir de LISA, on retrouve 81 références pertinentes et dans celle qui a été extraite d'ERIC, 62. La bibliographie cumulative contient 117 références bibliographiques.

Dans les notices, les zones qui nous intéressent d'une façon particulière sont les champs auteur, titre, source (périodique), année de publication, langue, résumé et descripteurs (et dans ERIC, la zone identificateurs).

### Le matériel

Nous rappelons que toutes les étapes de la présente recherche ont été

19. A.E. Jackson, « After Downloading ... », 120.

20. H.G. David *et al.*, « The Examination of Research Trends by Analysis of Publication Numbers », *Journal of Information Science*, vol. 3, no 6 (1981), 283-288.

21. Paul McGhee *et al.*, « Using Online Databases ... »

22. Henri Dou *et al.*, « Veille technologique ... »

23. Gisèle Vergnes *et al.*, « Analyse bibliométrique des collaborations internationales des laboratoires de l'IN2P3 en 1984 », *Documentaliste*, vol. 27, no 6 (1990), 280-282.

24. Comme les bases de données étaient disponibles sur CD-ROM, nous avons utilisé ce support plutôt que de recueillir nos données auprès d'un serveur de bases de données en ligne.

assistées par ordinateur. Selon l'étape réalisée, nous avons eu recours au matériel<sup>25</sup> suivant :

- Micro-ordinateur avec un lecteur de CD-ROM ;
- CD-ROM et logiciel d'interrogation (SilverPlatter Information Retrieval System ver. 1.5 pour les deux bases) ;
- Logiciel d'analyse de textes : la version 3.51 de SATO (Système d'Analyse de Textes par Ordinateur). Ce logiciel vise à faciliter les analyses de contenu. Il permet, entre autres, la génération du lexique des formes dans un corpus textuel ou dans une de ses parties avec calcul des occurrences des différentes formes ainsi que la définition de domaines (subdivisions de textes) et la catégorisation des mots hors contexte (dans le lexique) ou en contexte (dans les textes).
- Logiciel de traitement de texte : WordPerfect (version 5.0). Il permet, entre autres, la lecture de fichiers saisis avec d'autres logiciels ; la manipulation d'un document ; la création d'outils informatiques (par exemple, les macro-commandes) ; la conversion des fichiers WordPerfect en fichiers ASCII pour traitement avec d'autres logiciels ainsi que la mise en forme et l'impression des documents.

### La méthode

Nous présentons maintenant chacune des quatre étapes suivies en décomposant les principales démarches.

#### Étape 1 : Collecte des données (références bibliographiques sur le sujet)

A. Interrogation des bases de données (LISA et ERIC) afin de créer deux bibliographies (une par base de données) le plus exhaustives possible sur le thème : « les aspects cognitifs en sciences de l'information ».

Voici les énoncés de recherche :

LISA: Cogniti\* and PY = 198\*

ERIC: 1/ CH = IR

2/ 1 and DTN = 080

3/ 2 and PY = 198\*

4/ « Cogniti\* » and (« Librar\* » or « Information »)

5/ 3 and 4

Les énoncés comportent les paramètres suivants :

- Recherche en vocabulaire libre (termes génériques) dans toutes les zones recherchables par défaut.
- Dates de publication : entre 1980 et 1989 (Publication year = 198\*), ce qui correspond à une période de dix ans.
- Type de documents : articles de périodiques (ERIC: Document Type Number = 080 (journal article) ; LISA: ne peut être spécifié lors de la recherche).
- Pour la recherche dans ERIC, inclusion de la catégorie contenant les périodiques en sciences de l'information (Clearinghouse Number = Information Resources).

B. Téléchargement sur disquette pour constitution de bases de données locales contenant les références bibliographiques afin de pouvoir les manipuler et les analyser ultérieurement. Dans chaque fichier, nous avons veillé à conserver la stratégie de recherche (pour information) et les références avec tous leurs champs.

#### Étape 2 : Traitement des données téléchargées

A. À l'aide de WordPerfect, tri des références bibliographiques afin d'éliminer les notices qui réfèrent à des documents :

- autres que les articles de périodique (dans LISA)
- ne concernant pas les aspects cognitifs en sciences de l'information (dans les deux fichiers) et constituant du bruit.

B. Pré-traitement des deux bases de données triées (ERIC2 et LISA2), à l'aide de WordPerfect, afin de pouvoir les analyser avec SATO : déclaration de l'alphabet, remplacement des caractères réservés au logiciel, attribution de propriétés textuelles et lexicales, etc.

C. Pour chacune des bases, uniformisation des auteurs, des périodiques et des descripteurs, à l'aide de SATO et de WordPerfect. Cela s'applique dans le cas d'un périodique nommé dans certaines notices dans sa forme complète et dans d'autres, de façon abrégée.

D. À l'aide de WordPerfect, création d'une base de données cumulative contenant les références bibliographiques des deux bases traitées précédemment en prenant soin d'éliminer les doubles, plus uniformisation des

formes qui diffèrent selon la base de données interrogée.

#### Étape 3 : Analyse des données à l'aide du logiciel SATO : Étude des caractéristiques de la littérature

A. Analyse descriptive des références bibliographiques à l'aide du logiciel SATO afin d'établir :

Pour la base de données cumulative :

- les auteurs ayant produit le plus grand nombre d'articles sur le sujet ;
- les périodiques ayant publié le plus grand nombre d'articles ;
- les langues de publication des articles ;
- la production au cours des années.

Pour les bibliographies provenant de LISA et ERIC séparément :

- les descripteurs le plus souvent utilisés.

B. Analyse par croisements de variables des références bibliographiques à l'aide du logiciel SATO afin d'établir :

Pour la base de données cumulative :

- les auteurs selon les années (1980-84, 1985-89) ;
- les périodiques selon les années (1980-84, 1985-89).

Pour LISA et ERIC séparément :

- les descripteurs selon les années (1980-84, 1985-89) ;
- les auteurs selon les trois principaux descripteurs ;
- les périodiques selon les trois principaux descripteurs ;
- les descripteurs cooccurant avec les principaux descripteurs.

C. Analyse du vocabulaire employé dans les titres et les résumés dans le but d'identifier dans la base de données cumulative :

- les mots les plus fréquents ;
- les mots cooccurant avec les mots les plus fréquents ;
- les mots les plus fréquents entre 1980 et 1984 et entre 1985 et 1989 ;
- les formes présentes dans une seule des deux périodes (mots disparus, mots nouveaux).

25. Les équipements ont été fournis par le laboratoire d'informatique de l'EBSI (ordinateur, logiciels, CD-ROM LISA) et par la bibliothèque E.P.C. de l'Université de Montréal (CD-ROM ERIC).

## Étape 4: Présentation et utilisation des résultats de l'étude

A. Production, à l'aide de WordPerfect, de tableaux présentant les résultats les plus significatifs obtenus lors de l'analyse des données avec SATO. Ceci inclut les opérations suivantes :

- élimination des commandes de SATO ;
- ajout de titres ;
- remplacement des minuscules par des majuscules (noms des auteurs, etc.) ;
- élimination des traits d'union entre les noms des auteurs, etc. ;
- compilation des fréquences non présentes dans les tableaux (exemple : le nombre de formes ayant une occurrence unique).

Pour chaque tableau, rédaction d'un texte faisant ressortir les principales tendances.

B. Production, à l'aide de WordPerfect, d'une bibliographie sélective basée sur les résultats de l'étude des caractéristiques de la littérature sur les aspects cognitifs en sciences de l'information.

Présentation des références bibliographiques sélectionnées sous une forme agréable (exemple: format standard).

### Présentation des résultats

Dans cette section, nous rapportons les principaux résultats obtenus lors de l'analyse des références bibliographiques (étape 3). Il s'agit donc d'un exemple du produit résultant de l'étape 4-A. Les tableaux qui suivent nous permettent de mieux connaître les caractéristiques de la littérature sur les aspects cognitifs en sciences de l'information. C'est à partir de ces résultats qu'est réalisée la bibliographie sélective.

Les langues de publication	
Nbre de réfs	Langue
1	czech
2	danish
<b>106</b>	<b>english</b>
1	french
4	italian
1	japanese
1	portuguese
1	rumanian
1	russian

La quasi-totalité des articles de périodiques (106 sur 117) enregistrés dans la base de données cumulative sont en anglais. Il est donc essentiel de comprendre cette langue pour consulter les textes répertoriés.

Le nombre de références par année	
Nbre de réfs	Année
6	1980
<b>13</b>	<b>1981</b>
11	1982
9	1983
9	1984
7	1985
<b>13</b>	<b>1986</b>
<b>14</b>	<b>1987</b>
<b>16</b>	<b>1988</b>
<b>19</b>	<b>1989</b>

À partir de 1986, on note une augmentation graduelle du nombre d'articles sur le sujet. Pour une étude qui porterait également sur la période 1980-1984, les années 1981 et 1982 sont celles où l'on observe le plus grand nombre de textes.

Compte tenu du nombre relativement limité de publications par année (moyenne = 11,7), les compilations de variables par année ont été faites en regroupant les années en deux périodes de cinq ans, soit de 1980 à 1984 et de 1985 à 1989.

Les auteurs selon les années			
1980-89	1980-84	1985-89	Auteur
<b>5</b>	<b>1</b>	<b>4</b>	<b>Hannabuss, Stuart</b>
5	4	1	Ingwersen, Peter
<b>5</b>	<b>0</b>	<b>5</b>	<b>Kuhlthau, Carol C.</b>
4	4	0	Johnson, Kerry A.
4	4	0	White, Marilyn D.
3	1	2	Ford, Nigel
3	0	3	Jakobovits, Leon-A.
3	0	3	Nahl Jakobovits, D.
2	0	2	Berkowitz, Bob
2	0	2	Daniels, P. J.
2	1	1	Stein, Barbara L.
2	1	1	Totten, Herman L.

Il est intéressant de noter que pour l'ensemble de la période, la grande majorité des auteurs (126 sur 138) n'ont produit qu'un seul article. Toutefois, les auteurs les plus fréquemment mentionnés dans la bibliographie cumulative sont *Hannabuss*,

*Ingwersen* et *Kuhlthau*. Pour des références plus récentes, l'utilisateur devra d'abord consulter les articles de *Hannabuss* et *Kuhlthau* car la plupart des textes de *Ingwersen* ont été produits au cours de la période 1980-1984.

Signalons enfin que *Johnson* et *White* sont les deux autres auteurs ayant produit le plus grand nombre d'articles entre 1980 et 1984.

Les périodiques selon les années			
1980-89	1980-84	1985-89	Périodique
12	8	4	Journal of the American Society for Information Science
<b>7</b>	<b>0</b>	<b>7</b>	<b>College and Research Libraries</b>
6	3	3	RQ
5	2	3	Journal of Documentation
5	5	0	Journal of Education for Librarianship
<b>5</b>	<b>0</b>	<b>5</b>	<b>School Library Media Quarterly</b>
4	0	4	Information Processing and Management
3	2	1	Journal of Information Science
3	0	3	Library and Information Science Research
3	0	3	Reference Librarian
3	1	2	Research Strategies

Parmi les 62 périodiques mentionnés dans la base cumulative, nous présentons les titres des 11 périodiques cités plus d'une fois. On note que le *Journal of the American Society for Information Science* a publié le plus grand nombre d'articles mais sa production a baissé au cours des dernières années. Pour les cinq dernières années, c'est le *College and Research Library* qui a été le plus productif. Cette information nous laisse croire qu'il serait opportun de consulter les derniers numéros (1990-1991) de cette revue pour vérifier si de nouveaux articles ont été édités sur le sujet.

### Les descripteurs

L'analyse des descripteurs a été faite séparément pour les bases de données locales provenant de LISA et de ERIC, et non à partir de la base de données cumulative, car les langages documentaires utilisés par ces deux services sont différents.

LISA			
1980-89	1980-84	1985-89	Descripteur
19	9	10	reader services
19	9	10	services
<b>19</b>	<b>4</b>	<b>15</b>	<b>subject indexing</b>
<b>18</b>	<b>4</b>	<b>14</b>	<b>technical processes and services</b>
<b>17</b>	<b>3</b>	<b>14</b>	<b>information retrieval</b>
<b>17</b>	<b>3</b>	<b>14</b>	<b>information storage and retrieval</b>
16	8	8	information work
15	9	6	information science
15	9	6	librarianship
15	9	6	profession

Ce tableau donne la liste des 10 descripteurs (sur 133 formes) les plus souvent utilisés. En tête de liste, on retrouve: reader services, services et subject indexing. Tandis que les deux premiers font davantage référence aux articles publiés avant 1985, le troisième descripteur est nettement plus utilisé après cette date.

Dans la bibliographie extraite de LISA, *Ingwersen* est l'auteur qui a le plus produit sur ces trois aspects. Avec cet auteur, *Nigel* est celui qui a le plus écrit au sujet des « reader services » et des « services » et *Daniels*, au sujet du « subject indexing ».

C'est dans le *Journal of the American Society for Information Science* que l'on retrouve le plus grand nombre d'articles sur ces différents aspects, mais ce nombre est peu élevé (deux articles par sujet) et ceux-ci ne représentent que le sixième des articles provenant de ce périodique.

Les descripteurs avec lesquels ils sont le plus souvent associés sont: information work pour les descripteurs « reader services » et « services »; et information retrieval, information storage and retrieval, et technical processes and services pour le descripteur « subject indexing ».

ERIC			
1980-89	1980-84	1985-89	Descripteur
18	8	10	cognitive style
14	5	9	cognitive processes
13	4	9	information seeking
9	4	5	information retrieval
8	4	4	librarians
8	2	6	library instruction
7	2	5	decision making
7	3	4	man-machine systems
7	3	4	search strategies
<b>6</b>	<b>0</b>	<b>6</b>	<b>library skills</b>
<b>6</b>	<b>0</b>	<b>6</b>	<b>online searching</b>

Parmi les 11 descripteurs (sur 144 formes) les plus utilisés, Cognitive style demeure celui qui est le plus souvent utilisé pour représenter le contenu des articles et ce, autant entre 1980 et 1989 qu'avant ou après 1985. Dans la bibliographie extraite de ERIC, les auteurs qui ont le plus souvent écrit sur cet aspect sont: *Johnson* et *White*. Mentionnons également que c'est dans le *Journal of Education for Librarianship* que l'on retrouve le plus grand nombre d'articles traitant de cet aspect.

Si l'on désire étudier un aspect spécifique en relation avec le descripteur « cognitive style », notons que les descripteurs information seeking et librarians sont ceux que l'on retrouve le plus souvent associés, dans un même article, avec ce descripteur.

Les autres descripteurs les plus fréquents sont: cognitive processes et information seeking. Dans les deux cas, *Kuhlthau* est, parmi les différents auteurs, celui ayant produit le plus d'articles indexés par ces descripteurs.

Il existe une différence entre les descripteurs utilisés dans les deux bases de données. Il en est de même pour les thèmes abordés. Ces différences peuvent s'expliquer par les domaines couverts par ces deux services. Tandis qu'on répertorie dans LISA des articles en bibliothéconomie et en sciences de l'information, ERIC retrace des références relatives à d'autres domaines dont les sciences de l'éducation. Les sujets couverts dans la base auraient une influence sur la fréquence d'utilisation d'un descripteur. Ainsi, par exemple, même si « cognitive style » est le descripteur le plus fréquent dans ERIC, il se retrouve au 16<sup>e</sup> rang dans la liste faite à partir de LISA.

Le vocabulaire employé dans les titres et les résumés			
Titre	Résumé	1985-89	Forme lexicale
<b>3</b>	<b>12</b>	<b>15</b>	<b>expert</b>
3	6	9	critical
2	6	8	affective
2	6	8	perceptions
1	6	7	documents
<b>1</b>	<b>6</b>	<b>7</b>	<b>interfaces</b>
1	6	7	ways
<b>2</b>	<b>4</b>	<b>6</b>	<b>hypermedia</b>
0	6	6	stages
1	4	5	electronic

L'analyse du vocabulaire semble démontrer l'intérêt des chercheurs, au cours des dernières années, pour l'**informatique**. De plus, des termes faisant référence à des personnes ne sont plus utilisés après 1985 (exemple: professionals) alors que nous retrouvons parmi les mots nouveaux des termes davantage associés à la technologie (exemples: interfaces, hypermedia).

### Discussion

Au cours de cette recherche, nous avons mis au point une méthode d'analyse assistée par ordinateur de résultats de recherches bibliographiques automatisées dans le but d'étudier les caractéristiques de la littérature sur les aspects cognitifs en sciences de l'information.

Les résultats obtenus démontrent sans équivoque qu'il est avantageux de traiter et d'analyser, grâce aux outils informatiques actuels, les références bibliographiques pour en extraire des « informations » non apparentes à première vue mais pouvant être fort utiles pour l'utilisateur final. En fait, le produit résultant de ce type d'analyse peut, par exemple, guider l'utilisateur final au début d'une recherche sur un nouveau sujet, l'informer sur l'état de la recherche avant même qu'il aborde les textes, l'inviter à explorer certains thèmes peu traités, etc.

La méthodologie mise au point dans le cadre de cette recherche était pour le moins aussi importante que les résultats obtenus à la suite de son application. On peut dire, entre autres, que chacune des étapes réalisées avait une influence sur la suivante car des erreurs méthodologiques au cours d'une seule de ces étapes risquaient de compromettre les résultats escomptés.

Au sujet de l'interrogation, il est important de consulter le plus grand nombre de bases de données se rapportant au sujet pour obtenir un portrait des caractéristiques de la littérature le plus fidèle possible à la réalité. Cependant, même si nous

n'avons utilisé que deux bases (LISA et ERIC), celles-ci forment un échantillon représentatif des articles écrits sur un sujet comme le nôtre.

Nous trouvons aussi qu'il faut être exhaustif lors de la recherche automatisée. Il suffit de signaler que seulement deux des dix premières notices apparaissant dans la bibliographie sélective (non reproduite ici) contiennent le mot «cogniti\*» dans la zone descripteur. Ainsi, un documentaliste qui aurait limité sa recherche documentaire à cette zone aurait perdu des données fort pertinentes. Il est sûrement préférable de prendre le temps nécessaire pour faire le tri des notices plutôt que de perdre de l'information ou de limiter les résultats à cause d'un énoncé de recherche trop spécifique.

Nous croyons aussi avoir démontré, malgré certaines limites, que le recours à des bases de données bibliographiques pour effectuer une analyse bibliométrique permet d'analyser un nombre important de références avec beaucoup moins d'effort qu'à la main.

Cependant, le type de traitement exposé ici requiert un travail supplémentaire de la part du bibliothécaire. Surtout si l'on considère que notre tâche a été facilitée par l'utilisation de deux bases de données seulement, disponibles sur des CD-ROM développés par le même producteur (avec, par exemple, des traits d'union entre les mots dans les noms des auteurs, les descripteurs et les titres de périodique), que nous n'avons retenu qu'un seul type de documents (article de périodique assurant toujours la même structure de l'enregistrement), et que les notices étaient rédigées dans la même langue soit l'anglais (un élément important lors de l'étude des descripteurs et du vocabulaire).

Ce type d'étude demande aussi de la part du spécialiste de l'information de posséder diverses connaissances. Il doit, entre autres, avoir un savoir-faire en recherche documentaire automatisée, posséder des notions de base en bibliométrie, utiliser les logiciels de façon efficace et, de

préférence, connaître le sujet à l'étude. Ce dernier point est important surtout au moment du tri des références pour déterminer celles qui concernent le sujet de la recherche. À la rigueur, cette étape pourrait être déléguée à un expert du domaine. Quant à la connaissance des logiciels, elle est primordiale notamment dans le cas du logiciel d'analyse de texte utilisé (SATO), si on veut exploiter toutes les fonctionnalités utiles.

Si l'on évalue le temps pris pour chacune des étapes de la méthodologie, on s'aperçoit que la dernière étape (présentation et utilisation des résultats) requiert à elle seule près de la moitié du temps. Le traitement des données vient en deuxième place (environ le quart). Finalement, c'est l'analyse des données avec SATO et l'interrogation, demandant beaucoup moins d'intervention humaine lors de leur exécution, qui se font le plus rapidement. Il s'agit ici de la répartition du temps consacré à l'application de la méthodologie qui peut varier selon certains facteurs tels la connaissance des logiciels, la profondeur de l'analyse, le travail d'édition, etc. Une évidence s'impose: si la même analyse avait été faite sans l'assistance de l'ordinateur, elle aurait pris un temps beaucoup plus considérable.

Même si ce type d'analyse nécessite des démarches supplémentaires, il offre une solution de compromis intéressante entre la production d'une simple liste bibliographique et une analyse complète sur un sujet (un état de la question, par exemple).

Nous sommes bien conscient qu'il s'agit d'une étude quantitative avec les limites que cela impose. Ainsi, même si un auteur a écrit un très grand nombre d'articles sur un sujet, un autre peut n'avoir publié qu'un seul article mais sous un angle original. C'est pourquoi il importe de bien définir les objectifs poursuivis par l'analyse des références bibliographiques. Comme nous voulions faire une étude des caractéristiques de la littérature, une analyse quantitative correspondait bien aux buts visés et l'ordinateur s'est avéré un outil précieux pour ce type de travail.

D'autres traitements auraient pu être effectués pour pousser plus loin l'étude. Nous en suggérons un certain nombre en guise de conclusion:

- Établir les fréquences pour d'autres zones (exemples: code de classification, identificateurs);
- Inclure d'autres types de documents (monographies, thèses, ...) et comparer les résultats obtenus;
- Utiliser des bases de données en français. L'analyse du vocabulaire à l'aide de SATO pourrait être approfondie. Une liste de termes simples et complexes du domaine pourrait être établie à partir des titres et/ou des résumés;
- Comparer les résultats obtenus selon la base de données consultée (du même domaine: ISA vs LISA; de domaines différents: LISA vs ERIC);
- Étudier la possibilité et les avantages d'utiliser d'autres types de logiciels: chiffrier, gestionnaire de bases de données, analyseur statistique (les résultats de SATO peuvent être édités en format admissible directement admissible par SPSS).