

Périodique électronique étudiant

<CURSUS>

de l'École de bibliothéconomie
et des sciences de l'information
de l'Université de Montréal.

Volume 1, numéro 2 (printemps 1996)

ISSN 1201-7302

[Comité de rédaction](#)

[Comité de lecture](#)

Dans ce numéro

[Table des matières avec résumés](#)

[Luc Ainsley, "Politique d'acquisition: l'exemple des archives littéraires"](#)

[Caroline Archambault, "La communication dans un groupe de discussion scientifique: analyse du groupe de discussion Biomch-I"](#)

[Guylaine Beaudry, "La *Text Encoding Initiative*: Les moyens pour donner de la valeur à un texte numérisé"](#)

[Carole Brouillette, "Vers une définition de la lecture professionnelle"](#)

[Hélène Guertin, "Des Mini-centres d'information de réalité virtuelle"](#)

[Jean-François Hamel, "De l'accès sujet au principe de respect des fonds: pour une diffusion plus large de nos archives"](#)

[Jean-Noël Plourde, "Définition et application de critères d'évaluation d'outils de recherche dans Internet"](#)

[André Vigneau, "Les documents informatiques: pour une classification efficace"](#)

Page d'accueil de [Cursus](#)

Page d'accueil de l'[EBSI](#)

Dernière mise à jour: mai 1996

Cursus vol.1 no 2 (printemps 1996) - Comité de rédaction

Ève-Marie Beaudoin
Guylaine Beaudry
François Papik Bélanger
Geneviève Frédérick
Michel Giroux
Caroline Lessard
Christian Levac
Richard Pringle
Guylaine Roy
Valérie Simard
Chantal St-Louis

Page d'accueil de [Cursus](#)

Cursus vol.1 no 2 (printemps 1996) - Comité de lecture

Ève-Marie Beaudoin, étudiante
Guylaine Beaudry, étudiante
François Papik Bélanger, étudiant
Pierrette Bergeron, professeure adjointe
Suzanne Bertrand-Gastaldy, professeure titulaire
Carol Couture, professeur titulaire
Geneviève Frédérick, étudiante
Louise Gagnon-Arguin, professeure agrégée
Michel Giroux, étudiant
Diane Lanteigne, chargée de cours
Caroline Lessard, étudiante
Yves Marcoux, professeur adjoint
Dominique Maurel, responsable de la formation professionnelle
Gracia Pagola, chargée de cours
Johanne Perron, chargée de cours
Richard Pringle, étudiant
Guylaine Roy, étudiante
Valérie Simard, étudiante
Chantal St-Louis, étudiante
Albert N. Tabah, professeur adjoint
Guy Teasdale, étudiant
James Turner, professeur adjoint
Hélène Vien, auxiliaire de recherche

Périodique électronique étudiant

<CURSUS>

de l'École de bibliothéconomie
et des sciences de l'information
de l'Université de Montréal.

Volume 1, numéro 2 (printemps 1996)

ISSN 1201-7302

[Comité de rédaction](#)

[Comité de lecture](#)

Dans ce numéro

[Luc Ainsley, "Politique d'acquisition: l'exemple des archives littéraires"](#)

L'élaboration de politiques d'acquisition respectueuses des autres institutions d'archives et la création de réseaux délimitant les tranches de spécialisation et les territoires propres à chacun ne pourront être envisageables qu'à la condition explicite que soit instauré un dialogue sur les bases d'une coopération suivie entraînant la diminution des chevauchements, et par le fait même, une baisse de la compétition. Un tel dialogue est hautement souhaitable pour l'ensemble du milieu archivistique canadien et, notamment, pour les nombreuses institutions qui se sont données un mandat d'acquisition touchant aux archives littéraires. En effet, source d'un intérêt plus grand, les fonds de nos écrivains ont valeur d'objets précieux. Le prestige qui leur est associé (leur *plus-value culturelle*) s'est traduit par une forte demande, impossible à combler et, conséquemment, par une hausse considérable de leur *juste valeur marchande*, s'établissant au détriment de leur *valeur intrinsèque*. La compétition rattachée à l'acquisition de ce type d'archives entraîne une fragmentation des fonds et une génération indue de documents alors que les auteurs encouragent la surenchère entre les institutions d'archives majeures et l'escalade des prix.

Caroline Archambault, "La communication scientifique dans un groupe de discussion scientifique: analyse du groupe de discussion Biomch-I"

Cet article se penche sur les normes de communication dans un groupe de discussion en biomécanique (BIOMCH-L). Les messages du mois d'août 1994 ont été analysés à l'aide d'une grille développée dans le cadre de cette recherche. Cette grille s'intéresse aux caractéristiques de la structure des messages et aux caractéristiques de contenu des messages.

Les conclusions de l'étude dénotent une certaine tendance, de la part des scientifiques abonnés à la liste Biomch-L, à se conformer aux normes présentes sur les réseaux quant aux caractéristiques de la structure des messages. Pour ce qui est des caractéristiques de contenu, la nature des messages analysés n'étant pas très révélatrice, aucune conclusion sérieuse ne pourrait en être tirée. Nous pouvons cependant observer que les normes scientifiques sont encore solidement ancrées dans les habitudes de communication des scientifiques et qu'elles sont sérieusement confrontées aux normes de communication présentes sur les réseaux.

Cette recherche, effectuée sur un seul mois, est un projet pilote. Elle a donc ses limites compte tenu du nombre de messages retenus pour l'analyse, mais ouvre des pistes de recherche intéressantes pour une meilleure compréhension de la communication scientifique à l'ère électronique.

Guylaine Beaudry, "La Text Encoding Initiative: les moyens pour donner de la valeur à un texte numérisé"

La Text Encoding Initiative (TEI) a comme objectif de produire un modèle général et des lignes directrices pour l'encodage sur support électronique de tous les genres de textes littéraires. Ce projet a été entrepris suite à la reconnaissance de la nécessité d'une normalisation pour l'encodage et l'échange des textes numérisés. Après avoir identifié les raisons pour lesquelles le Standard Generalized Markup Language (SGML) a été choisi pour la conception du format d'encodage, on présente les principales composantes de la Définition de Type de Document (DTD) de la TEI. Une liste de projets de numérisation utilisant cette DTD est donnée ainsi que les objectifs de la Text Encoding Initiative pour les prochaines années.

Carole Brouillette, "Vers une définition de la lecture professionnelle"

État de la question sur les recherches en lecture pouvant mener à une définition de la lecture professionnelle. La lecture effectuée dans le cadre d'un travail est pressentie comme différente de celle faite dans une autre situation, dont par exemple celle de loisir qu'on lui oppose fréquemment. La lecture professionnelle aurait pour principale caractéristique distinctive le fait d'être intégrée à un processus productif. De ce fait, les lecteurs touchés par cette pratique,

n'auraient plus pour seule tâche d'interpréter le document, mais bien de le travailler, de l'élaborer. Cette façon d'articuler le discours, ou de fabriquer le texte, a des conséquences sur la recherche documentaire et sur les stratégies de lecture conviées par ce travail. C'est au travers des perceptions de la lecture exprimées dans divers domaines, dont les études littéraires et la psychologie, que l'esquisse d'une définition de la lecture professionnelle est devenue possible. Les textes retenus pour faire état d'une certaine *lecture-travail* sont américains, français et québécois et datent d'au plus vingt-cinq ans. Leur date de parution n'a pas été restreinte aux toutes dernières années, la fin des années '60 et le début des années '70 ayant vu naître plusieurs textes théoriques sur la lecture qui supportent en plusieurs points l'évolution de la lecture professionnelle, l'informatisation de la matière textuelle et leurs conséquences en milieu de travail.

Hélène Guertin, "Des Mini-centres d'informations de réalité virtuelle"

Le but de cet article est d'initier le futur spécialiste en sciences de l'information (SI) à l'application de la réalité virtuelle (RV) dans son champ d'action. Compte tenu de la popularité du partage de l'information via le réseau Internet, il n'est pas exclu de penser qu'un jour, des mini-centres d'information privés, construits en RV et gérés par des spécialistes en SI, feront concurrence aux catalogues en ligne des bibliothèques officielles. Les futurs diplômés en sciences de l'information, plus familiers avec la micro-informatique que leurs précécesseurs, auraient donc intérêt à se familiariser dès maintenant avec la réalité virtuelle.

Jean-François Hamel, "De l'accès sujet au principe de respect des fonds: pour une diffusion plus large de nos archives"

Dans cet article, l'auteur démontre que l'accès par sujet, dans un contexte d'informatisation des archives, devient un moyen important pour accroître la clientèle des services d'archives et l'étendre à de nouvelles catégories de chercheurs. Loin de remettre en cause la nécessité du principe de respect des fonds, cette approche utilise des outils informatiques nouveaux qui permettent de questionner directement un texte numérisé. Mais l'informatisation entière et complète des centres d'archives n'est pas pour demain. Avant d'en arriver là, beaucoup d'améliorations restent à apporter aux instruments de recherche traditionnels si l'on veut utiliser l'accès sujet de façon efficace. De plus, l'étape obligatoire de la normalisation doit nécessairement tenir compte des documents numériques qui se créent dans les administrations selon des normes souvent imposées par les informaticiens. Les archivistes doivent affirmer davantage leur présence au stade de la création des documents pour en faciliter leur accès par sujet.

Jean-Noël Plourde, "Critères et évaluation d'outils de recherche des ressources dans internet"

En raison de la croissance rapide du Web, plusieurs outils de recherche ont été développés pour

le repérage d'information. Choisir l'outil qui répond bien à ses besoins demeure un exercice difficile. Pour faire son choix, l'utilisateur devrait disposer de critères de sélection clairement définis et permettant de mettre en évidence les qualités respectives des différents outils.

La première partie de notre étude porte sur l'identification de tels critères. Les critères retenus, contrairement aux critères traditionnels basés sur la précision et la pertinence des résultats, reposent sur la typologie des outils et les services offerts par ceux-ci. Dans cette typologie, deux types sont examinés: répertoire (tel Yahoo) et index (tel Lycos). Les services offerts, quant à eux, sont analysés en fonction de trois catégories: les services de mise à jour, les services de repérage et les services de présentation des résultats.

Dans la deuxième partie de notre étude, nous évaluons treize outils de recherche selon les critères établis précédemment. Ces outils sont: Aliweb, World Wide Web Catalog, InfoSeek, Lycos, Open Text Web Index, WebCrawler, Harvest Broker, WWW Worm, Galaxy, GNN's Whole Internet Catalog, WWW Virtual Library, Yahoo, Clearinghouse for Subject-Oriented Internet Resource Guides [Un. de Mich.]. Cette étude s'est déroulée pendant les mois de septembre à décembre 1995.

[André Vigneau, "Les documents informatiques: pour une classification efficace"](#)

Avec l'arrivée de la micro-informatique dans les entreprises, les organismes et chez les particuliers, se pose le problème de la classification des documents créés par ordinateur et sauvegardés par l'utilisateur dans le système informatique. L'auteur rappelle d'abord les grands principes qui guident la classification et suggère l'utilisation d'un plan de classification dont les axes principaux sont déterminés par les fonctions de l'organisme. Puis, après les fonctions de l'organisme. Puis, après avoir établi qu'un document informatique est avant tout un document et qu'à ce titre il doit trouver place dans un plan de classification, l'auteur précise d'abord que ce document doit trouver place dans un plan de classification, l'auteur précise d'abord que ce document doit être inséré dans le même plan de classification utilisé pour les documents sur support papier et qu'ensuite ce plan de classification peut être réalisé grâce aux fonctionnalités mêmes de trois systèmes d'exploitation. Ainsi, en utilisant les capacités d'organisation logique de l'ordinateur, l'utilisateur est en mesure non seulement de reproduire à l'écran la structure du plan de classification de l'organisme mais aussi peut-il y classer les documents créés par ces mêmes systèmes.

Dernière mise à jour: 19 octobre 1995

Politique d'acquisition: l'exemple des archives littéraires

par

Luc Ainsley

Cursus vol.1 no 2 (printemps 1996)

Cursus est le périodique électronique étudiant de l'École de bibliothéconomie et des sciences de l'information (EBSI) de l'Université de Montréal. Ce nouveau périodique diffuse des textes produits dans le cadre des cours de l'EBSI.

ISSN 1201-7302

URL: cursus@ere.umontreal.ca

URL: <http://www.fas.umontreal.ca/ebsi/cursus/>

L'auteur

BIOGRAPHIE DE L'AUTEUR ICI

Politique d'acquisition: l'exemple des archives littéraires a été présenté dans le cadre du cours *Recherche en archivistique (BLT 6621)* donnée par le professeur Carol Couture.

Pour joindre l'auteur:

Droits d'auteur

Tout texte demeure la propriété de son auteur. La reproduction de ce texte est permise pour une utilisation individuelle. Tout usage commercial nécessite une permission écrite de l'auteur.

INTRODUCTION

Dans une société où la transcendance devient le fait de quelques individus, où l'avoir prime outrageusement sur l'être, le désir de possession oriente dangereusement nos choix jusque dans leur arbitraire. La possession du savoir est l'une de ces tentations et l'archiviste s'y trouve soumis de longue date. Sa responsabilité en regard des activités d'acquisition l'y rend plus particulièrement vulnérable. Car posséder un fonds ou une partie de fonds pour le seul fait de la chose, et cela indépendamment de sa nature et de son contenu, n'est plus acceptable. Le posséder mais, par le seul fait de sa possession, le soustraire à ceux pour qui il est vraiment signifiant en vertu d'un mandat institutionnel d'acquisition, en vertu de champs de compétence reconnus, l'est d'autant moins. À la source de l'irrespect d'autrui se profile l'absence de coopération, de dialogue.

L'élaboration de politiques d'acquisition respectueuses des autres institutions d'archives et la création de réseaux délimitant les tranches de spécialisation et les territoires propres à chacun ne pourront être envisageables qu'à la condition explicite que soit instauré un dialogue sur les bases d'une coopération suivie entraînant la diminution des chevauchements, et par le fait même, une baisse de la compétition. Un tel dialogue est hautement souhaitable pour l'ensemble du milieu archivistique canadien et, notamment, pour les nombreuses institutions qui se sont donné un mandat d'acquisition touchant aux archives littéraires. En effet, source d'un intérêt considérable, les fonds de nos écrivains ont valeur d'objets précieux, de marchandises. Le prestige qui leur est associé (leur *plus-value culturelle*) s'est traduit depuis quelques années par une forte demande, impossible à combler et, juste retour des choses, par une hausse considérable de leur valeur monétaire. Par conséquent, ce sont les achats qui caractérisent ce marché, du moins pour ce qui est de l'acquisition des fonds d'auteurs d'envergure. Des achats que seules les grandes institutions nationales peuvent se permettre et qui s'ajoutent à la quantité d'acquisitions effectuées via la formule du don pour fins de crédit d'impôt. Il en résulte un déséquilibre dans l'acquisition de ce type de fonds en défaveur des petits services d'archives qui, en fin de compte, ont le plus à gagner d'une concertation, d'un dialogue soutenu.

Dans cet article, nous définirons d'abord des éléments fondamentaux tels la politique d'acquisition et le partage par tranches de compétence, les notions de territorialité et de réseaux d'acquisition. En second lieu, il sera question de la *plus-value culturelle* associée aux fonds d'archives littéraires. Suivra une partie consacrée à la situation actuelle des archives littéraires au Canada. Enfin, nous traiterons des conséquences de la compétition en regard de ce type d'archives bien particulier.

1. LE PARTAGE DES TRANCHES DE COMPÉTENCE: UN

ÉLÉMENT PRIMORDIAL DE TOUTE POLITIQUE D'ACQUISITION

La politique représente l'expression concrète de la volonté, exprimée par l'organisation qui la met de l'avant, d'en arriver à une efficacité pour ce qui est de l'exécution de l'une ou de plusieurs de ses activités. Elle traduit donc un choix certain, une direction, le désir de ne rien laisser au hasard. James Lambert et Louis Côté la définissent plus précisément comme: "(...) une ligne directrice qui établit les paramètres et les limites de décisions ou d'actions répétitives" (Lambert et Côté, 1992, p.5).

Fidèle à cette définition, la politique d'acquisition tend vers une rationalisation des activités liées à la sphère des acquisitions. Une telle politique trouve sa justification dans la nécessité croissante pour les institutions d'archives d'être plus sélectives dans leurs choix en regard du volume et de la nature de leurs *collections* (1). Des choix souvent mal planifiés dont dépend le sort de quantité de fonds qu'une acquisition injustifiée (c'est-à-dire ne tenant pas compte des acquis documentaires passés) risque de reléguer dans un isolement autant physique qu'intellectuel, une quasi-destruction. La capacité de choisir implique donc une responsabilité de taille, celle d'orienter tout effort expansionniste vers l'amélioration et l'accroissement du capital documentaire déjà en place. Ainsi, une politique d'acquisition ne peut se vouloir largement permissive quant au volume de documents à acquérir sans se limiter à un ou plusieurs champs d'activité précis. À cet égard, se montrer trop gourmand correspondrait à occuper inutilement de précieux espaces d'entreposage, à dilapider des ressources de plus en plus difficiles à obtenir. Un gaspillage qui se ferait aux dépens de fonds qui auraient vraiment eu leur place parmi les collections déjà constituées ou en voie de l'être. Le rôle premier d'une politique d'acquisition, si on considère le problème sous cet angle, serait donc de délimiter précisément ce qu'on veut acquérir après s'être interrogé sur le pourquoi, la justification à la base de nos choix.

Toute politique d'acquisition qui se veut le moins sérieuse devrait comporter, à tout le moins, une délimitation par tranches de compétence des priorités d'acquisition de l'institution, une liste de directives et de contraintes en matière d'acquisition et, enfin, un suivi des types de coopération établis avec les autres institutions d'archives. La délimitation des champs d'excellence mérite plus particulièrement qu'on s'y attarde. En effet, on constate qu'un nombre sans cesse croissant de services se spécialisent, entraînant une vive compétition et la nécessité d'établir des balises territoriales clairement identifiées afin de diminuer les chevauchements et d'encourager la coopération entre institutions. Mais l'identification de ces balises n'est pas sans causer sa part de problèmes, puisqu'elle est largement sujette à interprétation. De fait, c'est la notion même de territorialité qui se trouve mise en cause ici. Dans un document produit par les Archives nationales du Québec, on dit de la territorialité des archives qu'elle implique: "que les archives soient conservées dans les services d'archives ayant juridiction sur le territoire dans lequel elles ont été produites" (ANQ, 1990, p.6).

Cependant, on ne mentionne pas comment le *territoire* doit être défini, ni même départagé entre les instances qui l'occupent simultanément, à des niveaux différents cependant. Cet *espace intellectuel*, nous le verrons plus loin, peut être revendiqué à la fois par des institutions d'archives au mandat d'acquisition à portée locale, régionale et même nationale.

Pour revenir à notre propos initial, on note que l'établissement et la reconnaissance de champs de spécialisation comportent bien d'autres avantages. Lambert et Côté les énumèrent ainsi: "l'amélioration des compétences du personnel, qui peut concentrer sa formation et son perfectionnement sur des domaines délimités; la concentration de fonds dans un même domaine, réduisant les déplacements des chercheurs; des liens plus évidents entre des fonds complémentaires; une réduction du fractionnement de fonds" (Lambert et Côté, 1992, p.17).

Mais à la source de toute reconnaissance des spécificités propres à chacun, il doit y avoir coopération. Une coopération inter-institutionnelle qui doit nécessairement passer par la politique d'acquisition qui officialise les paramètres et conditions des ententes et permet de communiquer à l'ensemble de la communauté archivistique les orientations de l'institution en matière d'acquisition et de domaines de spécialisation. De même, la politique contraint l'institution à tenir compte des mandats d'acquisition et des autres services d'archives officiellement reconnus. Ainsi, une connaissance détaillée de la teneur des collections et des *intérêts manifestes* des institutions ayant un mandat d'acquisition similaire s'avère nécessaire pour l'archiviste dans sa définition d'une politique d'acquisition qui se veut limitative, s'ajuste en fonction de la mosaïque de compétences dont il est fait étalage à l'échelle locale, régionale et nationale. Cette mosaïque aux mille visages de notre culture présente quantité de spécialisations s'articulant autour de critères comme la langue, les activités professionnelles, la compétence de l'établissement (limite territoriale) et la chronologie. Bref, un patrimoine archivistique riche quoique chaotique, sans ordre apparent, en grand besoin de planification, de concertation.

Cette concertation qu'il faudrait générale, étendue à l'ensemble du pays, elle passe par l'élaboration de réseaux d'acquisition visant à coordonner les activités des institutions d'archives. On parle ici d'une quantité de réseaux à rayon variable, trouvant leur existence dans l'impossibilité d'établir un réseau national au sein duquel les 700 services d'archives du pays agiraient de concert (2). Le type de coopération que les organismes adhérents y pratiqueraient, suivant le développement de stratégies d'acquisition clairement énoncées et le partage des domaines de compétence, pourrait se concrétiser par: a) l'initiative prise par un service d'archives d'informer une institution soeur de la disponibilité d'un fonds qui se situerait davantage dans ses champs de compétence; b) le fait de diriger un chercheur ou même un éventuel donateur vers un service d'archives qui pourrait mieux répondre à leur besoin spécifique; et c) un partage de ressources pouvant aller jusqu'au développement de programmes de microfilmage des fonds. Par surcroît, comme en fait état le rapport du Comité des acquisitions du *Conseil canadien des archives* (CCA): "Les membres demeureraient libres de définir leurs propres spécialités ou *zones d'excellence*, selon leur mandat, leurs fonds et collections et leur expertise particulière. Les services d'archives respecteraient l'intégrité des fonds, les mandats et les spécificités des autres établissements, la volonté des donateurs et les intérêts des chercheurs. **L'accent serait mis sur la coopération plutôt que sur la compétition** [Je souligne]" (CCA, 1993, p.11).

Ce dernier point nous paraît particulièrement important puisqu'il met en évidence le problème des conflits d'acquisition, une réalité bien de ce temps pour nombre d'institutions d'archives. (3) Toutefois, comme l'a démontré jusqu'ici l'exemple du *Group e d'archivistes de la région de Montréal* (GARM), une communication des priorités d'acquisition tend à diminuer cette compétition de façon significative. Le GARM est un regroupement qui a à coeur, depuis sa fondation en 1982, l'harmonisation des politiques

d'acquisition des services d'archives du Grand Montréal. Les quinze institutions qui y participent sur une base volontaire se sont dotées d'une politique d'acquisition qui trace les limites des champs d'acquisition propres à chacune. Évidemment, le chevauchement des compétences persiste mais à un moindre degré, puisqu'il y a des limites évidentes à la spécialisation. En effet, celle-ci ne saurait être trop détaillée, trop spécifique, au risque d'être discriminante à l'excès et de sombrer dans l'absurde. Chaque adhérent au groupe est de ce fait conscient de sa position par rapport à l'ensemble (en vertu de la délimitation de ses champs de compétence), en plus d'avoir une meilleure connaissance des ressources archivistiques offertes dans son territoire .

Le GARM, cependant, constitue un cas particulier, une exception dans la communauté archivistique. Par comparaison, il faut savoir qu'au Canada, peu d'établissements peuvent se vanter d'avoir une politique officielle d'acquisition et que rares sont les services d'archives qui collaborent entre eux. Il reste encore beaucoup à faire sur le plan de la rationalisation et de la coordination des activités en matière d'acquisition. Les résultats d'un sondage mené en 1992 par le CCA sur les fonds et collections de dépôts d'archives canadiens en constituent un indice révélateur. Dans une partie touchant à la planification, aux mandats et aux politiques d'acquisition des organismes parrains, on mentionne que: "Environ 75% des dépôts d'archives ont pu établir leur propre politique d'acquisition. Malgré tout, seuls 58% se sont dotés d'une politique écrite et approuvée. En outre, seuls 54% font état d'un plan ou d'une stratégie d'acquisition et seuls 32% disposent de normes de sélection et de critères d'évaluation écrits" (4) (CCA, 1992, p.2).

Autres données révélatrices, mais touchant cette fois aux catégories et sous-catégories de sujets que privilégient les services d'archives canadiens: on y constate que sur les 309 établissements ayant répondu à l'étude, une majorité planifie et effectue régulièrement des acquisitions de fonds non institutionnels touchant au domaine des *arts* et de la *culture*, arrivant au premier rang à ce titre. Ce même domaine se classe également en tête de liste dans les catégories *intérêt majeur* et *intérêt mineur*.

L'importance accordée à ce type de fonds est donc plus que manifeste. Il ne faut pas se surprendre alors si, de tous les domaines d'activités, celui des arts et de la culture est au coeur de tant de conflits d'acquisition. Certes, la catégorie est large, un peu à l'image des multiples manifestations des arts et de la culture qui enrichissent notre société. Mais ce phénomène de compétition n'en est pas moins actuel, particulièrement dans le cas des fonds d'archives littéraires et musicaux.

2. LA PLUS-VALUE CULTURELLE ASSOCIÉE AUX FONDS D'ARCHIVES LITTÉRAIRES

Les institutions d'archives publiques et privées s'intéressent depuis longtemps à l'acquisition d'archives non institutionnelles (5) s'appuyant sur des *dons sans compensation* de particuliers et sur une pratique d'achats, cette dernière de moindre importance cependant. Mais c'était bien avant que ne s'impose le concept de *valeur marchande*, renforcé par le biais des législations fédérales sur les dons et les biens culturels. La dernière en date, la *Loi sur l'exportation et l'importation des biens culturels*, propose l'achat indirect des fonds privés d'archives, financé par l'émission de reçus pour fins de déduction d'impôt. On

se trouve ainsi en présence d'une sorte de budget spécial de l'État dont l'octroi s'établit sur la base de la sauvegarde du patrimoine culturel national. Cette mesure, peu connue à l'origine, a bien vite gagné en popularité au point où l'on assiste maintenant à la disparition graduelle de la raison philanthropique alors que la plupart des donateurs se prévalent de leur droit légal à une évaluation (6). Et qui dit évaluation monétaire dit détermination de la *juste valeur marchande* et, implicitement, jeu de l'offre et de la demande. Ainsi, comme l'écrit R.S. Gordon: "La seule estimation stable et valable des documents réside dans le nombre toujours plus petit d'exemplaires disponibles sur le marché, ou dans le nombre toujours plus grand de personnes ou d'institutions cherchant à se procurer l'unique exemplaire de tel ou tel document" (Gordon, 1972, p.19).

Voilà qui montre bien quelle dynamique anime la *juste valeur marchande*, ou valeur monétaire, rendue vulnérable aux lois du marché dans sa dépendance envers de puissants leviers ayant pour noms *rareté* et *besoin*.

Ce concept de *juste valeur marchande* qui s'est imposé au quotidien de l'archiviste depuis les vingt dernières années semble bien vague, en vérité, et mérite qu'on s'y attarde quelque peu. Revenu Canada le définit de la façon suivante: "(...) en règle générale, on entend par là le prix qu'un bien rapporterait sur le marché libre dans une transaction entre un vendeur et un acheteur consentants qui seraient indépendants l'un de l'autre et qui agiraient en toute connaissance de cause. Quand il n'y a pas de lien de dépendance entre les personnes, la vente ou l'achat du bien à la date d'évaluation ou à une date rapprochée confirme généralement la valeur du bien à cette date " (Revenu Canada, 1991, p.7).

Il faut comprendre cependant que la *juste valeur marchande* n'est qu'une valeur parmi tant d'autres et que si elle s'impose en force dans le cas des archives littéraires, elle le fait au détriment de la *valeur intrinsèque*, ou *valeur de recherche*, constituant la seule véritable préoccupation des archivistes faisant l'acquisition de ce type de fonds dans le but d'en assurer la préservation. De fait, pour ce professionnel de l'information organique et consignée, la *valeur intrinsèque du fonds devrait être le seul critère à considérer au moment de l'évaluation pour fins d'achat ou de crédit d'impôt. La situation, néanmoins, se présente tout autrement: il n'existe aucun marché pour les fonds de recherche. Toute la place est laissée aux arts et à la culture en vertu du prestige qui s'y trouve associé. Ajoutons à cela que cette différence entre valeur de recherche et juste valeur marchande n'a pas bonne écoute auprès des marchands de manuscrits et de livres rares par qui les institutions doivent souvent passer pour acquérir les pièces et les fonds qui manquent à leur collection. Un différend profond qui ne manque pas de rejaillir sur un autre plan: celui de l'évaluation des archives qu'on exécute en vertu des exigences du ministère du Revenu et à la demande du dépôt bénéficiaire du don pour fins de crédit d'impôt. C'est le Conseil national d'évaluation des archives (CNÉA) qui a charge (7) de mener à bien les évaluations permettant de déterminer la juste valeur marchande des biens culturels à être versés. C'est ce même CNÉA que les marchands de manuscrits accusent de s'opposer au concept de valeur de vente, puisque les individus qui y siègent ont tendance à évaluer les fonds à la baisse. La juste valeur marchande alors attribuée se fait indépendamment et contre l'état du marché au moment de l'évaluation, ce qui n'est pas à l'avantage du vendeur.*

À la lumière de ces faits, on peut facilement s'expliquer pourquoi certains écrivains et autres donateurs

potentiels, bien avisés de la situation, engagent un expert indépendant (quoique reconnu par le CNÉA) pour évaluer leurs documents d'archives: ils en retiennent l'espoir de pouvoir contourner le système en vendant leur fonds à un service d'archives, ce qui leur éviterait l'arbitrage proposé par le Conseil. Au risque d'empiéter sur le propos de notre prochaine partie, il serait bon d'ajouter sur ce point qu'au Canada, les achats de fonds d'archives littéraires (8) sont devenus la chasse gardée des institutions nationales et des universités d'envergure, seules à pouvoir mobiliser les sommes nécessaires à pareilles transactions. Il apparaît donc clair que la juste valeur marchande diffère selon que le bien culturel est évalué pour fins de vente ou pour fins de don impliquant un déboursé indirect de l'État. Les marchands de manuscrits, de leur côté, auraient avantage à ce que le processus d'évaluation se fasse via la tenue d'enchères, ce qui constituerait, selon leur point de vue, un moyen plus juste et impartial d'établir la valeur marchande.

En résumé, il existe une quantité de valeurs qui peuvent être attribuées aux documents. Dans le cas bien précis des fonds d'archives littéraires, deux valeurs priment, s'affrontent même, soit la valeur monétaire et la valeur intrinsèque, la première ayant pris largement la mesure de la seconde en raison des demandes du marché, de la relative rareté des fonds et surtout de la plus-value culturelle qui est généralement associée aux papiers de nos écrivains nationaux. De fait, la juste valeur marchande de ce type de fonds est tellement liée au jeu de l'offre et de la demande que le marché pour cette denrée culturelle a atteint ces dernières années des sommets jamais égalés auparavant, allant même jusqu'à se comparer avantageusement à celui des livres rares. Dans la prochaine partie, nous examinerons la situation des archives littéraires au Canada et nous nous pencherons sur l'hypothèse d'un partage des champs de compétence entre institutions d'archives comme solution éventuelle à la compétition qui fait rage en regard de cette catégorie d'archives.

3. LES ARCHIVES LITTÉRAIRES AU CANADA OU LA NÉCESSITÉ D'UN DIALOGUE

La plus-value culturelle associée aux documents d'archives littéraires est à la source de la flambée inflationniste qui déferle présentement dans ce milieu qui est, somme toute, assez restreint. À vrai dire, il faudrait se questionner sur les raisons qui motivent les institutions d'archives, hormis l'habituel motif de l'inclusion dans les champs de compétence, à faire l'acquisition de tels fonds. Ces raisons sont sans nul doute nombreuses, mais la première qui nous vient à l'esprit est liée au prestige associé à l'acquisition de fonds de créateurs. En effet, un service d'archives qui peut faire état dans son bilan annuel de l'acquisition d'un fonds de contributeur à la littérature locale, régionale ou nationale, fait bonne impression et semble justifier la confiance que l'organisme parrain porte en lui. Dans le milieu des archives comme dans beaucoup d'autres, on le sait, les apparences tiennent une place importante.

Ajoutons à cette plus-value culturelle les quelques précédents où le donateur a reçu plus d'argent que la juste valeur marchande ne le prescrivait, et l'on se retrouve dans une situation critique de marché à la hausse. En faisant littéralement sauter la banque, Yusef Karsh, Michel Tremblay et tous les autres donateurs de grand renom ont pavé la voie aux revendications de leurs pairs qui exigent depuis lors une part égale de deniers publics. De fait, ils ont établi une nouvelle échelle de comparaison et c'est toute la

dynamique relative à l'acquisition des fonds d'archives littéraires qui s'en est trouvée bouleversée. Ainsi, une large proportion des centres d'archives ayant un mandat touchant à l'acquisition de ce type de fonds a été brusquement évincée de la course, faute de moyens financiers suffisants. Une situation qui se produit à un bien mauvais moment alors que le milieu des archives se voit contraint de sabrer dans les budgets d'acquisition afin de répondre aux exigences des organismes parrains aux prises avec un contexte économique difficile. Aussi, il semble contradictoire que la valeur de ces fonds s'établisse à la hausse alors que partout, il est question de coupures budgétaires importantes et de rationalisation des dépenses.

Telle est la situation du marché des archives littéraires en vertu des achats. Maintenant, celle qui prévaut en regard des dons présente une tout autre trame même si le résultat et le rapport de force restent inchangés: les grands, encore une fois, l'emportent sur les petits.

Alors que les achats mettent en cause les fonds d'auteurs d'envergure et nécessitent des déboursés que les grandes institutions nationales sont seules à pouvoir défrayer, les fonds secondaires, de moindre intérêt, restent accessibles à tous, puisque leur acquisition est conséquente à un don pour fins de crédit d'impôt. Il importe de préciser ici que ce mode de récompense permet d'éviter la surenchère puisqu'il étouffe toute forme de compétition ouverte en obligeant le donateur à passer, en principe, par une seule et même instance pour l'évaluation de son fonds, soit le CNÉA. Cependant, c'est un mode qui ne peut empêcher, à un tout autre niveau, les jeux de coulisse. On entend par là les faveurs octroyées et les avantages proposés qui vont influencer le choix du donateur quant à l'organisme qui accueillera ses papiers. Et à ce jeu, ce sont encore les grandes institutions qui sortent gagnantes puisque, naturellement, elles ont le plus à offrir.

Mais qui sont les grandes institutions canadiennes mises en cause ici? Pour les fonds d'archives de langue anglaise, outre de grandes universités comme l'Université de Toronto et l'Université McGill, on retrouve les Archives nationales du Canada (ANC) et la Bibliothèque nationale du Canada (BNC); et pour les archives de langue française, on dénombre, encore une fois, les deux grandes institutions nationales (ANC et BNC), auxquelles s'ajoute la Bibliothèque nationale du Québec (BNQ) qui se fait cependant plus discrète depuis quelques années sur le plan des acquisitions. Quant aux autres dépositaires de fonds d'archives littéraires, ils doivent composer avec des budgets d'acquisition infimes quand seulement ils en ont. À ce propos, David Walden livre des statistiques intéressantes bien qu'elles datent, empruntées au Rapport Wilson (1980): "... financial resources available to archives are extremely low. [...] while academic libraries spend, on the average, between 25% and 30% of their budget on acquisition, 80% of all archives spend less than 10% of their budgets on acquisition and 42% do not even have an acquisition budget. It is small wonder therefore that Canadian archives have little money available to purchase collections at the prices dealers believe are equitable" (Walden, 1980-1981, p.106).

La situation présente n'est donc guère brillante, et il est permis de se demander où va nous mener semblable compétition si elle perdure. Probablement vers une centralisation des fonds de grand et de moyen intérêts au profit des institutions canadiennes d'importance dont les principaux acteurs, ANC et BNC, concentrent le gros de leurs activités dans la région de la capitale nationale. Cependant, y a-t-il

lieu de croire que la situation pourra changer un jour, qu'on pourra reprendre un certain contrôle sur la façon dont se distribuent les fonds d'archives littéraires parmi les différents territoires d'acquisition? Et si on évoquait la solution plus tôt étudiée de la délimitation des tranches de compétences, composante essentielle de la politique d'acquisition? Si on imaginait l'implantation de réseaux sur la base desquels se coordonneraient les activités d'acquisition en concertation avec l'ensemble des services d'archives partageant un même intérêt pour les fonds littéraires? Certes, on ne résoudrait pas tous les conflits territoriaux, on n'abolirait pas l'ensemble des chevauchements, mais on pourrait à tout le moins espérer en réduire significativement le nombre et la portée. Tout cela à condition que soit instauré un dialogue durable basé sur une coopération franche et ouverte. Mais cette harmonisation des activités d'acquisition, si elle est souhaitable pour les petits services d'archives, l'est davantage pour les institutions nationales canadiennes au mandat trop large. En effet, la politique d'acquisition dont se sont pourvues les ANC, la BNC et la BNQ sanctionne actuellement tous les passe-droits, et cela au nom de l'importance nationale (9) ou de son équivalent, selon l'institution où l'on se trouve et le point de vue adopté. Ainsi, la BNQ applique le concept sous l'étiquette de la sauvegarde de la culture québécoise alors que les institutions nationales fédérales le font sous celle de la préservation du patrimoine culturel canadien. Au nom de l'importance nationale, on justifie ainsi les intrusions répétées sur le territoire d'acquisition de voisins de moindre stature. Ce vaste mandat devrait être interprété comme un droit qu'on se donne à acquérir non pas ce qui semble représentatif de la culture mais plutôt ce qui a une véritable portée nationale et/ou internationale. Il faudra, de toute évidence, penser à redéfinir le concept plus en détail (10) de manière à en limiter davantage la portée tout en le rendant plus efficace. Peut-être permettra-t-il alors de freiner, sinon de minimiser, la dure compétition que se livrent les trois grandes institutions nationales au mandat quasi identique. À cet effet, la rivalité qui surprend le plus oppose les Archives nationales du Canada à la Bibliothèque nationale du Canada (11), deux institutions relevant du gouvernement fédéral et occupant, par surcroît, un même bâtiment. En 1989, on a bien tenté de ressusciter le Comité consultatif mixte chargé des collections mandaté de présider à un dialogue et d'agir en qualité de médiateur dans des cas de conflits d'acquisition mettant en cause des archives littéraires. Malheureusement, l'initiative mise de l'avant par l'Archiviste national et le Directeur général de la Bibliothèque nationale a été vaine. La communication ne se fait toujours pas. Des chevauchements de mandats, tolérés on ne sait trop pourquoi par l'appareil étatique, qui coûtent cher. D'autant plus cher que ces grandes institutions se ressentent de plus en plus des compressions budgétaires qu'on leur impose année après année. Avec un pouvoir d'achat diminué se traduisant par une participation moins marquée aux activités d'acquisition, on risque d'assister à une fluctuation du marché des archives littéraires s'établissant à la baisse. Car s'il ne se trouve plus de grandes institutions pour acheter les fonds monétairement surévalués, qui d'autre prendra la relève? Il y a fort à parier que les prix redescendront à un niveau plus raisonnable, reflétant davantage la valeur intrinsèque des fonds. À moins que ne se manifeste la menace des spéculateurs en matière de manuscrits avec tout ce qu'elle laisse supposer.

4. LES CONSÉQUENCES DE LA COMPÉTITION

La question de la propriété des archives devrait être sérieusement étudiée avec l'avènement des collectionneurs, de l'arrivée sur le marché de spéculateurs en manuscrits. Pour beaucoup, leur venue

constitue une intrusion non fondée puisque les biens culturels, par leur nature même, devraient être la prérogative du groupe culturel à travers lequel ils sont significatifs, et non celle d'un particulier à la recherche de spéculations culturelles. Le phénomène n'est pas encore ressenti dans le milieu des archives littéraires au Canada mais cela ne saurait tarder, l'intérêt pour cette catégorie de biens culturels étant appelé à grandir. De fait, on mesure mal l'impact que ces spéculateurs auront sur le marché des archives d'écrivains quoique, si on se base sur un commentaire de David Walden, leur implication pourrait se traduire par des hausses additionnelles: "This individual, who has appeared at recent auctions, feels no emotion for the intrinsic worth of the material he purchases, but is merely seeking a concrete investment in much the same way some people purchase gold. Such inexperienced investors in manus cripts may be affected by the excitement and momentum of the auction, with the result that they pay exorbitant sums for individual items" (Walden, 1980-1981, p.107).

D'un strict point de vue archivistique, leur action est hautement condamnable puisqu'elle va à l'encontre du principe de respect des fonds et du code d'éthique de la profession. En éparpillant les documents d'un même fonds, en brisant l'unité de l'ensemble, on perd toute l'information qui a trait au contexte, en plus de brimer le droit des chercheurs à un libre accès. De fait, un manuscrit acquis par un collectionneur est un manuscrit dont l'accès devient conditionnel au bon vouloir du propriétaire; un document, en somme, dont la collectivité se trouve privée.

Cette menace de fractionnement des fonds n'est pas ressentie en raison seulement de la venue probable de ce nouveau groupe d'acheteurs, indépendants de toute institution, de toute loyauté, elle l'est également en regard de la vive compétition que se livrent actuellement les institutions d'archives impliquées activement dans l'acquisition de fonds d'archives littéraires. Selon l'éthique, il est de mauvaise pratique d'acquérir le fonds partiel d'un écrivain quand une autre institution en a déjà reçu une première partie. Semblable division diminue considérablement la valeur de recherche de l'ensemble. Pourtant, cela s'est déjà produit par le passé, et quelquefois même tout à fait par accident, l'archiviste en place négligeant de vérifier si le donateur avait déjà cédé ses archives ailleurs. Il faut comprendre que le donateur, sauf exception, se soucie peu que ses papiers soient réunis ou pas; le choix d'une institution d'accueil dépend le plus souvent de considérations monétaires.

À ce dernier problème se superpose celui de la création vertigineuse de documents, c'est-à-dire de la multiplication des acquisitions d'un même fonds. Un phénomène qui s'explique par un enthousiasme excessif envers la formule du don pour fins de crédit d'impôt mise de l'avant comme mesure incitative afin d'encourager les Canadiens à se départir de leurs archives. La mesure est si populaire qu'elle prend des proportions démesurées qui ne vont sûrement pas dans le sens de ce que ses initiateurs avaient souhaité. David Walden écrit à ce sujet: "What has happened with the demystification of the tax credit system (...) is that after donating a significant collection of papers to an archives and receiving a handsome tax rebate, some donors appear to be virtually creating documents on an annual basis with the specific intention of donating them to an archives for further tax relief" (Walden, 1984, p.88).

Rien n'est simple dans le cas présent. Sachant très bien que ces acquisitions saisonnières d'un même fonds ont une valeur intrinsèque minimale, les archivistes ont néanmoins l'obligation morale de s'en porter acquéreur, puisque les repousser équivaldrait à morceler l'ensemble. Même en resserrant leurs

critères de sélection, les institutions d'archives se trouvent alors prises en otage, à moins de refuser de se prêter au jeu.

Par ailleurs, alors qu'on assiste à une escalade des acquisitions de fonds d'archives privés empruntant cette voie royale, le pire est à craindre. Ainsi, les abus perpétrés sur le dos du système de crédit d'impôt pourraient forcer le gouvernement à mettre un frein à cette forme indirecte de financement. De toute façon, nombreux sont les archivistes qui mettent en doute le bien-fondé du système actuel. C'est qu'ils ont la certitude que les gens continueront de céder leurs fonds aux institutions d'archives, même sans être récompensés pour leur geste. Du même coup, on réhabilitera, dans le cas des fonds de moindre importance, la valeur de recherche comme valeur première et non soumise aux fluctuations du marché. La compétition, cependant, sera toujours présente, quoiqu'elle devrait se manifester sous une autre forme.

CONCLUSION

En 1980, le Rapport Symons soulignait les avantages d'une meilleure coopération entre institutions d'archives. Quinze ans plus tard, on ne remarque toujours aucune amélioration sur ce point si ce n'est de l'initiative du GARM. Conscient du problème, le Conseil canadien des archives en a fait l'une de ses priorités pour les années à venir, car il devient de plus en plus urgent de mettre fin à la compétition résultant du chevauchement des mandats. En guise de solution, on préconise la création de réseaux d'acquisition à l'échelle nationale, régionale et locale, ce qui favoriserait grandement l'établissement d'une politique d'acquisition écrite et officialisée étendue à l'ensemble des institutions participantes. Une telle politique aurait l'avantage de rationaliser les activités en matière d'acquisition en plus de permettre l'identification précise des champs de compétence.

Mais au-delà de ces idées sur la coopération et sur un dialogue concerté, il y a la réalité qui nous rappelle à l'ordre. L'application du principe de territorialité et le suivi d'une politique d'acquisition ne seront pas sans rencontrer des obstacles incontournables, des circonstances devant lesquelles ils devront forcément s'incliner. La volonté du donateur (12) est l'une de celles-là, et pour cause: on ne peut, en toute conscience, refuser d'acquérir le fonds d'un individu qui insiste pour choisir le lieu où seront versés ses documents, et cela sous prétexte que le don proposé n'entre pas dans le mandat d'acquisition de l'institution. Dans une certaine limite, il faut respecter ce choix. Comme il faut délimiter plus précisément le concept de territorialité, considérant, par exemple, que le territoire national revendiqué par les institutions nationales d'archives occupe le même espace que celui revendiqué par quantité d'autres services d'archives ayant un mandat régional ou même plus large. Comment ce partage va-t-il s'établir et quels en seront les enjeux?

Mais avant toute chose, il faudra bien que les archivistes se dissocient un jour de ce fétichisme malsain qui leur fait adorer l'objet archivistique - littéraire, notamment - au point de vouloir absolument et totalement le posséder. La véritable richesse, elle se trouve encore dans l'information contenue dans le document d'archives. Une information d'un type particulier qu'il importe de situer dans un contexte idéal, c'est-à-dire la rendant davantage significative en raison de la proximité d'informations de même

nature. Et le contexte ici, c'est la "collection" de fonds acquise en vertu des directives territoriales précises dont un service d'archives s'est doté. Ce qui peut nous porter à penser, à juste titre, que la clé de notre salut passe forcément par la politique d'acquisition en vertu de laquelle s'initiera le dialogue tant nécessaire.

BIBLIOGRAPHIE

Archives nationales du Canada. Acquisition Policy / Politique d'acquisition. Ottawa, 1988.

Archives nationales du Québec. Projet de politique d'acquisition d'archives privées du centre d'archives de Montréal. (Préparé par Louis Côté), 1990.

Boles, Frank. "Mix Two Parts Interest to One Part Information and Appraise Until Done: Understanding Contemporary Record Selection Processes." American Archivist 50 (Summer 1987): 356-368.

Boles, Frank. Archival Appraisal. New York: Neal-Schuman Publishers, 1991.

Cline, Scott. "Collection Policy or Documentation Strategy: Theory and Practice." American Archivist 54 (Winter 1991): 44-52.

Conseil canadien des archives. Directives concernant l'élaboration d'une politique d'acquisition. Ottawa, 1990.

Conseil canadien des archives. Rapport de l'enquête du CCA sur les fonds et collections des dépôts d'archives canadiens: résultats nationaux. (Préparé par le Comité des acquisitions du CCA) Ottawa, 1992.

Conseil canadien des archives. Rapport du Comité des acquisitions du CCA sur une stratégie nationale d'acquisition. Ottawa, 1993.

Ericson, Timothy L. "At the "rim of creative dissatisfaction": Archivists and Acquisition Development." Archivaria 33 (Winter 1991-1992): 66-77.

Gordon, R.S. "L'évaluation des documents historiques". (...), (1972): 1-31.

Grigg, Susan. "A World of Repositories, a World of Records: Redefining the Scope of a National Subject Collection." American Archivist 48, no 3 (Summer 1985): 286-295.

Ham, Gerald F. Selecting and Appraising Manuscripts. Chicago: The Society of American Archivists, 1993.

Houde, Michel. "La préservation du patrimoine documentaire: rôle de l'archiviste, responsabilité institutionnelle et stratégie nationale." Archives 21, no 3 (1990): 3-13.

Lambert, James et Louis Côté. "Les outils de travail en archivistique: la politique d'acquisition: pourquoi, comment, critères et exemples." Archives 23, no 3 (1992): 3-21.

Ouellet, Fernand. "Le Conseil national d'évaluation des archives: un séminaire d'orientation." Université d'Ottawa, 1976: 41-42.

Phillips, Faye. "Developing Collecting Policies for Manuscript Collections." American Archivist 47, no 1 (Winter 1984): 30-42.

Revenu Canada. Dons en nature. Ottawa, 1991.

Walden, David. "Stretching the Dollar: Monetary Appraisal of Manuscripts." Archivaria 11, (Winter 1980-1981): 101-113.

Walden, David. "The Tax Credit System: Blessing or Burden?" Archivaria 18, (Summer 1984): 84-90.

À ces quelques titres, il faut ajouter les entrevues données par Mme Andrée Lavoie des ANC, M. Claude LeMoine de la BNC et, enfin, M. Michel Brisebois, évaluateur de fonds d'archives littéraires et de livres rares, ayant déjà siégé au sein du Conseil national d'évaluation des archives. Nous tenons à remercier plus particulièrement M. Brisebois sans qui ce travail n'aurait pu être possible.

1. Nous nous permettons d'employer le terme "collection" dans ce contexte particulier, car nous croyons qu'il reflète bien la disparité caractérisant l'ensemble des fonds détenus par un service d'archives, et cela même si ces fonds sont assujettis à des champs de compétence. En effet, le choix d'un fonds est en soi un geste hautement discriminant, la marque d'un collectivisme auquel il est difficile d'échapper.

2. (CCA, 1993, p.10)

3. "L'enquête montre l'ampleur possible du problème, puisque 60% des dépôts font état d'un territoire d'acquisition régional, provincial, interprovincial, national ou international. Or plus ce territoire est vaste, plus le risque de conflits d'acquisition est élevé. Avec une politique d'acquisition imprécise, les conflits d'acquisition entre dépôts deviennent inévitables. De fait, 29% des répondants en ont déjà fait l'expérience" (CCA, 1992, p.3).

4. À noter que 309 établissements d'archives ont répondu au sondage et qu'il n'est fait aucunement mention du taux de réponse au questionnaire.

5. Il faut savoir que le concept de valeur marchande ne s'applique pas aux archives du domaine public

puisque ce type d'archives, de propriété collective justement, ne se trouve pas soumis au jeu de l'offre et de la demande.

6. Ceci est plus qu'évident dans le petit univers des archives littéraires.

7. Pour les fonds qu'on estime d'une valeur de 1 000 \$ ou moins, le ministère accepte une évaluation interne menée par un archiviste de l'institution qui en sera bénéficiaire.

8. À noter que les achats mettent toujours en cause des fonds de prestige, des fonds d'auteurs dont le rayonnement justifie une dépense directe. En comparaison, les fonds de moindre importance sont généralement acquis par don.

9. Les Archives nationales du Canada en donnent la définition suivante: "Les documents qui ont une importance nationale sont ceux qui nous renseignent sur la vie des Canadiens. Ils relatent les efforts et les expériences de particuliers, de groupes, d'établissements, de corps constitués et d'autres organismes qui sont reconnus à l'échelle nationale ou internationale. Ils renferment également de l'information sur le milieu physique canadien de même que sur les événements et les tendances (d'ordre culturel, politique, économique, social, démographique, scientifique et religieux) ayant une large portée nationale" (ANC, 1988, p.4).

10. Dans un document de travail daté d'octobre 1993, la Commission canadienne d'examen des exportations de biens culturels en donne une définition plus fouillée quoique insuffisante. Encore une fois, on omet de prendre en considération la notion de territorialité, de l'établissement d'un partage.

11. La BNC fait l'acquisition de manuscrits littéraires dans le but d'enrichir ses collections d'imprimés.

12. Le donateur prestigieux aura toujours la liberté de choisir où il veut céder - vendre - son fonds puisque toutes les portes lui sont ouvertes. Ce sera plus difficilement le cas pour le donateur de moindre envergure, plus susceptible d'être dirigé vers un service d'archives ayant une compétence correspondant à sa catégorie de fonds.

La communication dans un groupe de discussion scientifique: analyse du groupe de discussion Biomch-L

par

Caroline Archambault

Cursus vol.1 no 2 (printemps 1995)

Cursus est le périodique électronique étudiant de l'École de bibliothéconomie et des sciences de l'information (EBSI) de l'Université de Montréal. Ce nouveau périodique diffuse des textes produits dans le cadre des cours de l'EBSI.

ISSN 1201-7302

cursus@ere.umontreal.ca

URL: <http://www.fas.umontreal.ca/ebsi/cursus/>

L'auteure

Diplômée de l'UQAM en administration des affaires (1986) et de l'EBSI en 1995, **Caroline Archambault** a occupé pendant un an le poste d'auxiliaire d'enseignement aux laboratoires d'informatique documentaire de l'EBSI. Elle occupe depuis avril 1996 le poste de spécialiste de l'information chez KPMG, une firme de comptables et de fiscalistes.

La communication dans un groupe de discussion scientifique: analyse du groupe de discussion Biomch-L a été écrit en 1994 dans le cadre du cours BLT 6341 *Recherche en gestion des services et des ressources d'information* donné par le professeur Pierrette Bergeron.

Pour joindre l'auteure : carcham@odyssee.net

Droits d'auteur

Tout texte demeure la propriété de son auteur. La reproduction de ce texte est permise pour une utilisation individuelle. Tout usage commercial nécessite une permission écrite de l'auteur.

Table des matières

- [1. Introduction](#)
- [2. Problématique](#)
- [3. Revue de littérature](#)
 - [3.1 La communication scientifique](#)
 - [3.2 La communication par ordinateur](#)
- [4. Méthodologie](#)
- [5. Analyse](#)
 - [5.1. Les caractéristiques techniques](#)
 - [5.1.1 Le nombre de sujet par message, le destinataire du message, la fidélité de la ligne sujet et l'origine géographique du message.](#)
 - [5.1.2 Les éléments de la signature](#)
 - [5.1.3 Les citations](#)
 - [5.1.4 Référence à un autre message de la liste](#)
 - [5.2 Les caractéristiques du contenu](#)
 - [5.2.1 Demande](#)
 - [5.2.2 Sommaire](#)
 - [5.2.3 Les annonces](#)
 - [5.2.4 Biomch-L, Internet et autres](#)
- [6. Conclusion](#)
- [7. Bibliographie](#)

1. Introduction

La communication scientifique est en train de passer des canaux informels plus traditionnels, comme le téléphone et le courrier, vers les canaux moins traditionnels disponibles sur les réseaux informatiques. Cette étude pilote porte sur les normes de communication dans un groupe de discussion électronique en

biomécanique (Biomch-L) et essaiera de répondre aux questions suivantes: y a-t-il des normes, scientifiques ou sociales, qui se dégagent de ce groupe de discussion? et peut-on comparer les résultats avec ceux obtenus par d'autres études déjà faites?

Dans un premier temps, la problématique est présentée, suivie d'un état de la question abordant deux thèmes: la communication scientifique et la communication par ordinateur. Ces deux thèmes permettront une meilleure compréhension de la communication scientifique et serviront de base pour l'analyse qui suivra. Nous présenterons ensuite la méthodologie utilisée pour l'analyse de contenu des messages du groupe de discussion étudié, ainsi que l'analyse qui en découle. Une discussion suivra, basée sur la revue de littérature, sur l'analyse des messages et sur notre expérience personnelle avec ce moyen de communication.

2.Problématique

La communauté scientifique est reconnue pour avoir développé une infrastructure sociale et technologique large et sophistiquée qui supporte la communication scientifique et technique. Cette infrastructure définit le rôle que la communication joue dans le processus scientifique (Doty, 1992). La communication avec les collègues, entre autres, a toujours été une des sources importantes de la communication scientifique (Kovacs et Kovacs, 1991; Garvey, 1979). Rice et Tarin (1993) remarquent également que les recherches scientifiques s'effectuent de plus en plus en collaboration et sont de plus en plus interdisciplinaires. La dispersion géographique qui en découle rend important le besoin des chercheurs de communiquer rapidement et facilement (Doty et al., 1990).

Les nouvelles formes de communication, comme la communication par ordinateur, permettent cette efficacité en facilitant le transfert d'information entre les chercheurs. Certains prévoient, dans un futur proche, des changements fondamentaux dans la façon dont les recherches et la communication scientifiques, ainsi que d'autres activités des chercheurs, seront faites (McClure et al., 1991). Ces changements seront en grande partie dus à l'arrivée massive des réseaux et de leurs possibilités de communication.

Cependant, même si plusieurs auteurs ont écrit sur la communication ou la communauté scientifique ainsi que sur les nouvelles formes de communication par ordinateur, peu de recherches ont été faites sur les normes qui influencent et guident le comportement des scientifiques communiquant fréquemment par ordinateur (Rosenbaum et Snyder, 1991) ainsi que sur le profil de ces chercheurs (Doty et al., 1990). La plupart des auteurs qui ont écrit sur le sujet affirment que sans cette compréhension, il sera difficile d'établir de bonnes structures pour la mise en place éventuelle de nouveaux réseaux.

C'est donc dans cet esprit que nous nous sommes intéressés aux groupes de discussion, en faisant une analyse de contenu d'un de ces groupes. Deux types de normes seront abordés dans cette étude, les normes scientifiques, propres à la communauté qui nous intéresse, et les normes sociales qui découlent de l'utilisation populaire de la communication par ordinateur et que nous identifierons par la " netiquette ", terme employé par la littérature pour parler du comportement que devrait avoir tout bon navigateur de

l'Internet.

3. Revue de littérature

La revue de littérature qui suit sert de point de départ pour l'analyse des messages du groupe de discussion Biomch-L. Elle nous permet de mieux comprendre le processus de communication chez les scientifiques ainsi que les normes qui émergent de la communication par ordinateur, principalement des groupes de discussion. Nous aborderons donc, dans un premier temps, la communication scientifique avec ses normes, son système de récompense unique et son processus de publication. Ce sont trois éléments qui ont un lien très fort entre eux, mais qui commencent à être quelque peu bousculés par la vitesse à laquelle se développent les réseaux de communication.

L'autre point abordé est la communication par ordinateur, plus précisément celle de l'Internet et des groupes de discussion. Nous verrons les normes sociales qui émergent de ces groupes ainsi que les résultats de quelques recherches qui ont été faites sur des groupes de discussion dans le milieu scientifique.

3.1. La communication scientifique

La science est considérée comme une activité unique guidée par ses propres règles (Hagstrom, 1965). Son but fondamental est l'accumulation du savoir, savoir vérifié et empirique. Merton préconisait quatre normes importantes dans le milieu scientifique, constituant ce qu'il appelait " the classical ethos of science " (Merton, 1973): l'universalité; la mise en commun des résultats (basée sur la reconnaissance que la communication est essentielle à la science et que le secret est l'antithèse de cette norme); le désintéressement (c'est-à-dire ne pas agir pour des intérêts personnels); et le scepticisme (avoir un esprit critique développé). L'avancement de la science dépend donc de l'efficacité avec laquelle les idées sont communiquées. Le système de communication est le système nerveux de la science, le système qui reçoit et transmet les stimuli à ces différentes composantes (Cole and Cole, 1973).

À travers les années, la communauté des chercheurs a développé un système de reconnaissance dans lequel la communication, les normes scientifiques et les récompenses sont étroitement liées (Merton, 1973). Cette reconnaissance apporte au scientifique plus de ressources et plus de pouvoir à l'intérieur de la communauté scientifique. Certains deviennent éventuellement membres de groupes spécialisés communément appelés " collègues invisibles " (Price, 1963; Crane, 1972). Ces groupes sont réputés pour avoir des normes d'éthique assez rigoureuses. Cela engendre cependant un certain cloisonnement et les nouveaux chercheurs sont souvent exclus de l'échange qui se fait entre certains scientifiques dans ces fameux collègues invisibles. L'avènement de la communication par ordinateur, qui rend la communication beaucoup plus facile, changera-t-elle la nature même de la communication scientifique en permettant un accès plus facile au milieu scientifique?

Depuis près de 300 ans, la communication scientifique se fait entre autres par le biais des périodiques. Hiltz (1984) affirme que le processus de la communication scientifique n'a pas beaucoup changé depuis.

Sauf que pour plusieurs disciplines, la croissance exponentielle des recherches effectuées a ralenti le processus de publication et allongé la période entre la fin d'une recherche et sa publication dans un périodique (Rice et Tarin, 1993). Les articles de périodique prennent donc tellement de temps à paraître que souvent ils ne servent qu'à informer les chercheurs de ce qui se passe dans d'autres domaines, ou encore comme signe de reconnaissance pour l'accomplissement scientifique (Hiltz, 1984).

À cause de la lenteur du processus, Garvey et Griffith (1971) concluent, suite à une étude de la communication scientifique en psychologie, que les chercheurs utilisent beaucoup les discussions informelles, les petites réunions et l'échange de documents préliminaires (" drafts ") et prétrages (" preprints "). Ces échanges leur permettent de rester à jour dans leur domaine de recherche et de connaître les points de vue de la communauté sur la valeur des recherches effectuées dans des domaines connexes. La communication informelle prend donc une importance considérable puisqu'elle permet l'accès aux recherches privées, non publiées, mais qui restent des éléments vitaux pour l'avancement de la recherche (Ginsparg, 1994; Price, 1961).

On voit donc ici un premier impact concret que peut avoir la communication par réseau sur la communication scientifique et sur les normes de la communauté, puisqu'elle peut rendre ces échanges informels beaucoup plus facile.

La communication par ordinateur permettra-t-elle un accès plus facile à la communauté scientifique en réduisant l'élitisme qui prévaut? Chose sûre, il paraît évident que les réseaux bouleversent les mécanismes solidement établis de la communication scientifique, autant pour la communication informelle que pour la publication formelle (Doty et al., 1991). Les normes scientifiques jouent en effet un rôle dans l'utilisation des réseaux électroniques, mais la nature exacte de cette influence n'est pas claire quoique plusieurs tendances se dessinent depuis quelques années.

3.2. La communication par ordinateur

Chaque mois, des milliers de nouveaux usagers découvrent le pouvoir de l'Internet. Mais avec ce pouvoir vient aussi des responsabilités (Fraase, 1993). Le pouvoir de l'Internet vient de la communication qui s'effectue entre les gens, pas seulement du transfert brut de l'information d'un site à l'autre. De ces relations se bâtissent des communautés et, en parallèle, apparaissent des règles sociales (Hiltz, 1984; Vallée, 1975).

Comme n'importe quelle autre activité sociale, la communication adopte des standards de comportement, standards auxquels les participants sont tenus de se conformer (Rosenbaum et Snyder, 1991). Les groupes de discussion, d'un point de vue plus général, constituent un phénomène social. Il est donc tout à fait normal que des règles sociales s'y soient greffées avec le temps. Les principes de base sont simples, et peuvent s'appliquer à toute entrée dans une nouvelle " société " (Fraase, 1993): observer, apprendre, vivre et laisser vivre.

Ces règles semblent être de plus en plus respectées par les utilisateurs en général. Des guides sur

l'Internet (Caroll et Broadhead, 1994; Gibbs et Smith, 1993) nous ont permis d'en relever quelques unes, adressées à tout type d'utilisateur: ne jamais oublier qu'on s'adresse à un humain, donc adhérer aux mêmes normes de comportement en réseau que celles que l'on suit dans la vie de tous les jours; garder ses messages le plus clair possible (discuter d'un seul sujet par message, bien indiquer le sujet du message dans la ligne sujet, lors d'une réponse à un message précédent, inclure seulement les parties du message auxquelles on répond directement, signer avec adresse électronique); éviter le contenu commercial dans les messages (sauf si un groupe le permet explicitement); consulter le FAQ (Frequently Asked Questions) des groupes de discussion.

Nous insistons sur le fait que ces règles ne sont pas des " lois ", elles sont tout au plus des conventions qui permettent le respect de part et d'autre. À travers les études sur les groupes de discussion consultées, on observe que ces règles, ou normes, quoique largement respectées, le sont à des degrés différents. Il n'en ressort pas de normalisation générale sauf peut-être pour les points que nous verrons plus loin.

Un nombre croissant d'études démontrent que la communication par ordinateur est de plus en plus utilisée et acceptée dans le domaine de la recherche scientifique (Rosenbaum et Snyder, 1991). Plusieurs de ces études proviennent maintenant de recherches faites dans le domaine des sciences sociales et se concentrent sur les aspects psychologique, social et culturel de la communication par ordinateur, plutôt que sur les aspects techniques comme ce fut le cas jusqu'à maintenant (McClure et al., 1991). Voici quelques résultats.

Une expérience menée par Hiltz (1984, p.157) sur l'utilisation du groupe de discussion comme support de communication entre les participants de dix groupes de recherche spécialisés, démontre que la majorité des utilisateurs de réseaux électroniques est d'accord pour dire que la communication accrue avec les pairs a changé leur compréhension face aux intérêts et aux activités des autres chercheurs dans leur domaine. En général, ces usagers rapportent passer plus de temps à communiquer avec d'autres membres de leur communauté scientifique depuis qu'ils utilisent de tels systèmes. Ce type de communication semble même avoir pour effet de mieux intégrer les nouveaux venus.

Une étude de Rosenbaum et Snyder (1991), portant sur l'analyse de 36 caractéristiques des messages de cinq groupes de discussion, constate qu'il y a quelques normes importantes qui se dégagent des groupes de discussion: une forte tendance à n'inclure qu'un sujet par message; une forte tendance à personnaliser le message en incluant d'autres éléments que le nom; une tendance à ne pas faire de référence explicite à un message lorsqu'on en reprend le sujet. Les auteurs concluent à ce sujet que cette tendance se rapproche plus de la communication verbale (en personne) que de la communication écrite.

Les autres caractéristiques de cette même étude sont moins uniformes, comme par exemple la fidélité de la ligne sujet par rapport au contenu. Quant au contenu des messages, une très forte tendance se dégage quant à la fidélité des sujets par rapport au domaine couvert par le groupe de discussion. Malgré le fait qu'il n'y ait aucun contrôle sur la nature des messages expédiés, les participants restent dans les limites du sujet couvert par la liste. C'est à cette étude que le présent projet s'apparente le plus.

Une autre étude (Doty et al., 1990), basée sur des groupes témoin et des entrevues auprès de différents groupes dans des organisations de recherche, essayait entre autres de voir (1) s'il est vrai que les réseaux changent le comportement des chercheurs et (2) quelles preuves existent que les réseaux changent la communication scientifique. En réponse à la première question, les chercheurs indiquent que les réseaux ont un certain impact sur la façon d'organiser leur travail, mais peu sur ce qu'ils font comme recherche ou considèrent comme étant une bonne recherche. Les réseaux sont considérés comme de bons outils de recherche, mais doivent être utilisés à l'intérieur de la structure normative de la recherche de l'entreprise. En réponse à la deuxième question, il apparaît que les éléments de base de la communication scientifique n'ont pas changé significativement à cause de l'apparition des réseaux. Mais il faut reconnaître que les réseaux facilitent, voir même améliorent grandement la communication (distance, transfert d'information plus facile, etc.) (Cooley et Wendt, 1991).

Les réseaux peuvent faciliter l'échange d'expertise et de conseils tout en améliorant les habiletés à se conformer à certaines normes techniques (Doty et al., 1991). Il en va autrement cependant de l'utilisation des réseaux pour l'échange de documents formels de recherche ou de toutes autres sortes de publications plus officielles. Cela apparaît pour les chercheurs comme une entrave au système de reconnaissance et de récompense établi. Avec les publications imprimées, il y a un contrôle sur qui peut y accéder et à quelles fins. Ce contrôle n'existe pas dans le monde électronique (Doty, 1992). Donc, même si les chercheurs semblent adhérer sans trop de problèmes aux normes plus techniques des réseaux, il en va autrement pour les normes sociales.

Quant à eux, Kovacs et Kovacs (1991) relèvent des points intéressants face à l'importance des groupes de discussion: transfert rapide de l'information; élimination des coûts de la distribution traditionnelle; utilisation peu coûteuse; maintien des contacts avec d'autres scientifiques; possibilité de collaboration dans les recherches en cours (quoique sur ce point les auteurs ne sont pas tous prêts à l'affirmer); substitut efficace des moyens traditionnels de communication comme le téléphone et la poste.

Snyder et al. (1992) ont pour leur part analysé les normes de référence de trois groupes de discussion sur Bitnet (un réseau à l'origine de l'Internet). Il en ressort que la seule norme standard de référence est l'utilisation de la ligne sujet. Les autres caractéristiques étudiées étaient l'inclusion du message, la façon dont on réfère à une personne (nom, adresse électronique, etc.), à un message ou aux personnes de l'extérieur. Les auteurs expriment cependant certaines réserves quant aux résultats obtenus. Plusieurs facteurs peuvent influencer les normes d'un groupe de discussion. Par exemple, les membres d'un groupe avec peu de participants et qui se connaissent bien entre eux vont être portés à utiliser le prénom de quelqu'un pour y référer, plutôt que son nom complet.

La communication scientifique est faite de normes sociales assez rigoureuses qui en font un monde un peu à part. Ces normes ont en effet tendance à créer des groupes d'élite difficiles à percer. La communication par ordinateur semble cependant avoir cette capacité de briser un peu le mur de cette communauté. Par ses possibilités inestimables de communication et de transfert d'information, la communication par ordinateur permet au milieu scientifique d'élargir facilement ces horizons. Cependant, bien que les chercheurs soient prêts à se conformer à certaines normes techniques quant à l'utilisation des réseaux, il en va autrement de leur adhérence aux normes sociales. Il ressort des études

que la structure normative de la communauté scientifique n'est pas prête de disparaître à cause de l'arrivée des réseaux.

4. Méthodologie

Différentes approches ont été utilisées pour l'étude de la communication par ordinateur (Doty et al., 1991), allant de groupes témoin (" focus groups ") à l'enquête par questionnaire, en passant même par des expériences en laboratoire. La présente étude examine les messages d'un groupe de discussion en biomécanique par l'entremise de l'analyse de contenu.

L'analyse de contenu est défini comme une technique de recherche qui permet de tirer des conclusions valides et conformes à partir de certaines données dans un contexte défini (Krippendorff, 1989). L'analyse de contenu est privilégiée ici puisqu'elle est fréquemment utilisée, dans la communication, pour analyser l'échange de messages entre interlocuteurs.

L'analyse est effectuée à l'aide d'une grille développée et validée dans le cadre de cette recherche. La grille d'analyse fut construite à l'aide de certaines caractéristiques, mentionnées dans la littérature, concernant le contenu des messages (ligne sujet, demande, objet du message) ainsi que leur structure (destinataire, information sur l'expéditeur, nombre de sujet par message). Pour compléter cette grille, 15% des 279 messages des mois d'août et de novembre 1994 furent choisis au hasard et analysés. Cet échantillon a permis de relever d'autres caractéristiques permettant de compléter la grille d'analyse.

Voici la grille obtenue:

- Le nombre de sujets contenus dans le message (un, deux trois ou plus).
- Le destinataire du message (liste ou personnel).
- La fidélité de la ligne sujet du message.
- L'origine géographique du message.
- Les éléments contenu dans la signature (nom, affiliation, adresse électronique, adresse physique, téléphone, télécopieur, personnalisation).
- Les citations faites (personne, organisation, groupe de discussion, publication).
- Les références à un message de la liste (en incluant le message, par le nom de l'expéditeur, par le sujet).
- Le contenu des messages: demande de référence, sur des produits, concernant une méthodologie ou diverses; " flame "; sommaire; annonce de poste ou de conférence ou autres; commercial; commentaires sur la liste Biomch-L; Internet; et autres.

Pour vérifier la fidélité de la grille d'analyse, un deuxième échantillon, constitué de 56 messages (20% des mêmes 279 messages), a été prélevé. Deux méthodes ont été utilisées, la fidélité inter-codeur (sur l'ensemble des 56 messages) et la fidélité intra-codeur (sur 5% de ces mêmes messages). Ces messages furent codés séparément, à l'aide de la grille développée, par deux personnes différentes. Le pourcentage d'accord (nombre de jugements en accord/nombre de jugements total) fut de 95% pour la méthode inter-

codeur et de 98% pour la méthode intra-codeur, soit au-delà des 85% qui représentent la norme acceptable pour la recherche en science sociale (Rosenbaum et Snyder, 1991). La grille fut donc appliquée à l'ensemble des messages du mois d'août 1994 et l'analyse qui suit en est issue.

Une étude comme la nôtre peut cependant difficilement mesurer l'impact d'un groupe de discussion sur l'avancement des idées en science, conséquence importante de la communication scientifique. La grille d'analyse permet de coder les messages selon leur contenu général seulement. Il est impossible de savoir, par exemple, quels thèmes de recherche sont discutés. Pour aller plus profondément dans le contenu des messages et des discussions, un expert du domaine serait nécessaire, il serait plus apte à juger ce contenu spécialisé.

5. Analyse

Les résultats analysés seront séparés en deux catégories. Une première catégorie concerne la nature technique du message, souvent reliée à des normes plus générales que scientifiques. La deuxième catégorie concerne le contenu, caractéristiques qui se rapprochent plus des normes scientifiques, donc de la communication scientifique à proprement parler.

Figure 1 Résultats de l'analyse des messages du mois d'août 1994

NOEUDS	NOMBRE	%
1 sujet/message	88	98%
2 sujets/message	2	2%
adressé liste	87	97%
adressé ind.	3	3%
Fidélité sujet	86	96%
ORIGINE GEOGRAPHIQUE		
Amérique centrale	1	1%
Amérique nord	56	62%
Canada	7	8%
Etats-Unis	49	54%
Europe	18	20%
Asie	4	4%
Océanie	6	7%
Afrique	3	3%
Incertaine	2	2%

SIGNATURE

Nom et/ou prénom	73	81%
Nom seulement	0	0%
Prénom seulement	3	3%
Nom et prénom	70	78%
Affiliation	56	62%
Adresse		
électronique	41	46%
Adresse physique	33	37%
avec pays	18	20%
sans pays	15	17%
Téléphone	31	34%
Télécopieur	28	31%
Personnalisation	8	9%
Aucune signature	15	17%

CITATION

Personne	2	2%
avec affiliation	1	1%
sans affiliation	1	1%
Organisation	2	2%
Groupe de		
discussion	0	0%
Publication	2	2%
Auto-citation	1	1%

RÉFÉRENCE À UN MESSAGE

Message inclus	9	10%
Expéditeur	2	2%
Adresse		
électronique	0	0%
Sujet	11	12%
Autres	1	1%

CONTENU

Référence	13	14%
Produit	14	16%
Méthodes	22	24%
"Flame"	0	0%
Sommaire	12	13%
avec synthèse	2	2%
sans synthèse	10	11%
Annonce	22	24%

poste	8	9%
conférence	6	7%
autres	8	9%
Commercial	0	0%
Biomch-L	4	4%
normes	1	1%
autres	3	3%
Internet	3	3%
Demandes diverses	8	9%
Autres	9	0%

Nous répondrons plus précisément à la question posée en introduction à savoir s'il y a des normes, scientifiques ou sociales, qui se dégagent de ce groupe de discussion. Par le contenu des messages relevé, l'analyse nous révélera si oui ou non la communication par ordinateur semble influencer la communication scientifique ou si elle tend plutôt à remplacer les canaux traditionnels de communication.

Rappelons ici que cette étude est exploratoire et que pour rendre les résultats plus significatifs, un plus grand échantillon devrait être analysé.

5.1. Les caractéristiques techniques

5.1.1 Le nombre de sujet par message, le destinataire du message, la fidélité de la ligne sujet et l'origine géographique du message.

La première caractéristique étudiée concerne le nombre de sujet par message, 98% des messages respectent la consigne de n'envoyer qu'un sujet par message. Cette norme semble généralisée. Elle est d'ailleurs fortement encouragée par la netiquette, et respectée dans les groupes de discussion. Rosenbaum et Snyder (1991) remarquaient qu'il n'était pas rare de voir une personne poster plusieurs messages consécutifs, avec des sujets différents plutôt que de faire un seul envoi. Seulement 2% des messages contenaient deux sujets et aucun trois ou plus.

Le groupe se démarque plus des autres études en ce qui concerne le destinataire du message. Seulement 3% des messages (3 messages) étaient des messages personnels envoyés à la liste. De ces 3 messages, deux étaient en fait des erreurs, des demandes d'abonnement envoyés à la liste plutôt qu'à l'administrateur de la liste. L'étude de Rosenbaum et Snyder (1991) en dénotait 25%, mais couvrait un échantillon de 662 messages.

La fidélité de la ligne sujet est également respectée très souvent puisque 96% des messages ont effectivement comme contenu le sujet indiqué dans l'en-tête. Comme pour la caractéristique précédente (destinataire), on note une augmentation de cette norme par rapport aux études faites au début des années 90. On peut supposer que ce sont des normes qui se sont ancrées dans les habitudes des

participants de groupes de discussion au fur et à mesure que ces groupes devenaient plus populaires. L'expérience doit aussi entrer en ligne de compte. On peut en effet réaliser l'importance de ces deux éléments quand on les vit soi-même à travers nos expériences (recevoir des messages mal identifiés, recevoir des messages qui ne nous sont pas adressés, etc.). On est sûrement porté à faire plus attention à notre propre comportement.

Pour l'origine géographique, bien que ce groupe de discussion soit international, plus de 62% des messages proviennent d'Amérique du nord (54% proviennent des États-Unis et 8% du Canada), puis, 20% proviennent d'Europe, 4% proviennent d'Asie, 7% proviennent d'Océanie (tous d'Australie) et 3% proviennent d'Afrique. Dans 2% des cas, l'origine du message est incertaine. Il n'y a pas de signature et l'adresse de l'en-tête ne permet pas de déterminer l'origine avec certitude. Ce groupe étant originaire de l'Amérique du nord, cela peut expliquer la plus grande proportion des usagers venant de cette partie du continent.

5.1.2 Les éléments de la signature

Pour les éléments de la signature, 83% des messages comportent au moins un des éléments de la signature. C'est en dessous du 94% de l'étude faite par Rosenbaum et Snyder. On peut peut-être expliquer ce fait par le nombre importants de messages qui sont des annonces (24%) et qui souvent ne sont pas signés.

Le nom et/ou le prénom de l'expéditeur se retrouvent dans 81.11% des messages. Aucun de ces messages n'est signé avec le nom de famille seulement, tandis que 78% portent le nom et prénom. Seulement 3% des messages sont signés uniquement avec le prénom. C'est un reproche que l'on a souvent fait d'ailleurs aux membres des groupes de discussion qui ne signaient que par le prénom. Quand le groupe commence, on a l'impression qu'on se connaît tous et on est plus familier, mais il ne faut pas oublier qu'on ne sait pas qui nous lit et que tous ne nous connaissent pas par le fait même.

Dans 62% des cas, l'expéditeur indique son affiliation. Ce n'est pas une caractéristique que l'on retrouve dans beaucoup de recherches, mais elle est fortement recommandée par la netiquette.

Il est surprenant de voir que seulement 46% des messages indiquent une adresse électronique dans la signature. Pourtant, le groupe de discussion, comme son nom l'indique bien, se veut un moyen de communication efficace et rapide. Et comme la netiquette veut qu'on réponde directement aux personnes qui posent des questions sans passer par la liste, ce pourcentage paraît faible. Des 49 messages sans adresse électronique dans la signature, 20 étaient de plus des demandes d'information. Et comme un message de la liste le mentionnait, il est maintenant possible de réduire les en-têtes des messages ce qui occasionne parfois la perte de la ligne de l'expéditeur. Donc si l'adresse électronique ne se trouve pas dans le message, il devient impossible de répondre sans un certain délai, le temps de retrouver l'adresse de l'expéditeur.

La combinaison fortement recommandée par la netiquette (Carroll et Broadhead, 1994; Gibbs et Smith,

1993) est le nom, l'affiliation et l'adresse électronique. 39% des messages ont au moins ces trois éléments, 22% ont au moins le nom et l'affiliation, 4% le nom et l'adresse électronique et 16% le nom seulement. On entend ici par nom, la présence du prénom et/ou du nom de famille.

Les autres éléments de la signature se retrouvent à des pourcentages plus faibles: 34% indiquent le téléphone, 31% le numéro de télécopieur, 9% personnalisent leur message (dessin ou citation). L'adresse physique est indiquée dans 37% des signatures et 20% seulement indiquent le pays. Finalement, 17% des messages (15 messages) n'ont aucune signature et proviennent principalement des États-Unis (12 messages).

5.1.3 Les citations

Les citations ne sont pas très fréquentes sur Biomch-L. Seulement 7% des messages en contiennent: 2% en rapport avec des personnes, 2% des organisations, 2% des publications et 1% pour de l'auto-citation (ce qui ne comprend pas les messages ou l'expéditeur mentionne la recherche sur laquelle il travaille pour faire une demande d'information). Ces résultats se trouvent bien en deçà de ceux obtenus par Snyder et al. (1992). La grande proportion de demandes (49% des messages) pourrait en partie expliquer ce faible résultat. Rappelons également que ces messages datent du mois d'août, un mois ayant un achalandage plus faible que la moyenne et coïncidant avec la reprise des activités après les vacances estivales.

5.1.4 Référence à un autre message de la liste

Pour ce qui est d'une référence à un autre message de la liste, 17% des messages en contiennent: 10% incluent le message, 2% mentionnent le nom de l'expéditeur et 12% le sujet. Ces résultats sont faibles par rapport à ceux de l'étude de Snyder (1992). La moitié des messages (44 documents) étant des demandes, ils ne réfèrent pas à d'autres messages. C'est donc dire que peu d'échanges ou de débats sont faits à travers la liste sur des opinions ou des idées. Il faudrait vérifier avec les autres mois si cette situation reflète vraiment les activités principales du groupe. Encore une fois, le mois d'août étant une période de vacances et il se peut que cela se reflète dans les activités de la liste.

Il n'y a pas d'uniformité dans la façon de se référer à un autre message. Ces résultats vont dans le même sens que ceux obtenus par Rosenbaum et Snyder (1991) qui voyaient là une tendance des utilisateurs à se comporter dans un groupe de discussion comme lors d'une communication verbale. C'est-à-dire qu'on prend souvent pour acquis que les destinataires savent de quoi on parle, en omettant de faire référence explicitement au message en question, et quand on y fait référence, il ne semble pas y avoir de " façon de faire ".

5.2 Les caractéristiques du contenu

Contrairement aux autres caractéristiques, chaque message est classé dans au moins un noeud. Rappelons les caractéristiques retenues pour l'analyse des messages quant au contenu: le message

concerne une demande (référence, produit, méthodologie, autres); un " flame "; un sommaire suite à une demande (avec ou sans synthèse); une annonce (poste, conférence, autres); un contenu commercial; un commentaire sur la liste; Internet; tout autre sujet autre que ceux mentionnés ci-dessus.

5.2.1 Demande

Les demandes d'information de toutes sortes se retrouvent en grand nombre. En effet, près de la moitié des messages en contiennent (49% des messages). 14% des messages concernent une demande de référence; 16% une demande sur un produit; 24% touchent à de l'information sur une méthode; 9% des demandes diverses (comme par ex. adresse électronique, assurances pour un laboratoire en biomécanique).

Ce haut pourcentage a ouvert, au mois de mars 1995, un débat intéressant sur la qualité de la liste. Un des administrateurs a en effet envoyé à la liste un message faisant un sommaire de quelques commentaires reçus de la part de certains membres.

- Ne faites pas perdre leur temps aux autres.
- Consulter les archives de la liste.
- Éviter le " please subscribe me " envoyé à la liste.
- La liste devrait être scindée en deux (étudiants et experts).
- Créer un FAQ (Frequently Asked Questions).
- Encourager les discussions publiques (sur la liste) quand des questions sont posées.
- Modérer la liste.
- Créer un " Usenet News ".

La présence des étudiants semblent indisposer certains membres qui déplorent que la liste soit utilisée comme premier moyen pour chercher de l'information plutôt que comme dernier recours. C'est mauvais pour l'étudiant qui ne développe pas son aptitude à trouver de l'information pertinente et, surtout, qui ne semble pas apprendre à utiliser d'autres ressources.

Certains en viennent donc à proposer de scinder la liste pour exclure les étudiants des discussions des experts. On retrouve là une caractéristique des collègues invisibles qui est le contrôle des membres de la " société ". Mais tous ne partagent pas ce désir d'isoler les étudiants. Par contre, ils sont d'accord pour dire que trop de questions de référence sont posées sur la liste, ce qui est en partie confirmé par les résultats obtenus dans la présente recherche.

Il faut en effet se souvenir, comme le fait remarquer un des membres, qu'un des buts des groupes de discussion est de permettre l'interaction entre " ceux qui savent " et " ceux qui apprennent ". La meilleure solution pour éviter que ceux qui savent ne servent que de diffuseur de référence, c'est de sensibiliser la liste à ce sujet. La pression des pairs est en effet vue comme un élément important pour rétablir la qualité de la liste.

5.2.2 Sommaire

Les sommaires (un retour sur les réponses obtenues suite à une demande) constituent 13% des messages (2% sont de vraies synthèses tandis que 11% ne sont qu'une liste des messages reçus). Un des points mentionnés plus haut fait mention du désir des membres d'assister à ces échanges, d'encourager la discussion publique. Le fait de répondre directement à la personne qui fait la demande empêche les membres d'assister aux discussions sur un sujet qui doivent s'avérer très intéressantes pour l'avancement de la science.

Il est donc intéressant de voir que malgré le fait que la norme encourageant à répondre directement au demandeur soit très présente sur les réseaux, les chercheurs voudraient plutôt la contourner tout en étant conscients que cela risque d'engendrer beaucoup plus de messages dans leur courrier. Ce désir rejoint une des normes scientifiques touchant la mise en commun du savoir (Merton, 1973). De plus, ces échanges sur une méthode ou un sujet précis ne peuvent que faire avancer la science en encourageant les nouvelles idées.

5.2.3 Les annonces

Les annonces constituent pour leur part 25% des messages (9% pour un poste, 7% pour une conférence et 9% autres [changement d'adresse électronique (5), cours, sortie d'un numéro de périodique, parution d'un article]). La littérature couverte ne mentionnait rien sur ce fait, mais les membres de Biomch-L ne voudraient pas que la liste devienne un médium d'annonces publiques. Les annonces restent cependant liés étroitement au domaine couvert.

5.2.4 Biomch-L, Internet et autres

4% des messages concernent Biomch-L et 3% la navigation sur Internet. Quant au 10% des messages qui ne se classaient dans aucune de ces catégories, on retrouve: une tentative de débat sur la mauvaise conduite en science (2 messages), la recherche d'un correspondant, une réponse à une demande envoyée à la liste (2), des tests (2), et deux messages incompréhensibles dont un qui semble être une demande envoyée par erreur à la liste, plutôt qu'au serveur, pour avoir la liste des abonnées de la liste.

Pour finir, un point intéressant à noter, aucun des 90 messages du mois d'août n'avait un contenu commercial ou n'était un " flame ". Ce qui semble, par expérience personnelle, refléter la situation des groupes de discussion scientifique ou professionnel.

Les comportements observés à travers la liste de discussion Biomch-L ne sont pas uniformes. À part le nombre de sujets par message, la fidélité de la ligne sujet et le respect des autres membres (en évitant par exemple l'envoi de message personnel à la liste et l'envoi de message à des fins commerciales), les autres caractéristiques ne sont pas généralisées à travers les messages. Il est plus réaliste de parler de tendances que de normes.

Cette étude nous a permis d'en apprendre un peu plus sur une nouvelle forme de communication utilisée en science: les groupes de discussions par courrier électronique. Les normes émergentes sont cependant plus reliées à la structure des messages qu'à leur contenu. Les résultats de Doty et al. (1991) allaient d'ailleurs dans ce sens. Ils concluaient en effet que les méthodes de travail des chercheurs étaient grandement améliorées grâce aux possibilités de communication des réseaux, mais que pour le contenu même de ces échanges, les normes scientifiques prévalaient encore. Échanges plus concrets, par exemple sur des résultats, en milieu plus fermé.

Nous n'avons donc pu tirer de grandes conclusions sur le contenu des messages car on retrouvait surtout des demandes d'informations dont les réponses ne circulaient pas directement sur la liste. D'ailleurs, des commentaires récents (mars 1995) sur la liste soulignent un certain intérêt pour les abonnés de voir les réponses circuler. Il semble que cela pourrait susciter des débats et des échanges plus enrichissants. Ces débats et échanges nous permettraient de faire une meilleure analyse du contenu scientifique des messages.

6. Conclusion

On a vu que les chercheurs prennent place dans une communauté où la communication est régie par un ensemble de normes très strictes rattachées étroitement à un système de récompenses et un processus de publication très rigoureux. L'arrivée des réseaux vient ébranler ces habitudes depuis longtemps établies, habitudes qui avaient malheureusement rendu l'accès aux recherches de plus en plus difficile, d'une part à cause des délais de publication, et d'autre part à cause de la difficulté pour les nouveaux venus de s'intégrer aux groupes informels (collèges invisibles) et avoir ainsi le coup de pouce nécessaire pour leur assurer plus de visibilité.

Les normes scientifiques de communication sont donc confrontées aux normes de communication présentes sur les réseaux. Dans l'ensemble, les groupes de discussion scientifique semblent se conformer assez bien à ces normes quant au niveau de la structure des messages. Mais pour ce qui est du contenu, la rigidité de la communication scientifique ne semble pas être ébranlée, tout au plus facilitée. Rien ne laisse voir que les informations échangées à travers les messages pourraient éventuellement avoir autant de poids, par exemple, qu'un article publié. La reconnaissance des pairs semblent vouloir passer par les mêmes normes établies depuis si longtemps. Par contre, ce groupe permet une intégration plus facile des nouveaux venus dans les discussions et échanges prenant place sur la liste.

À la lumière des résultats obtenus quant à la nature des messages (contenu), le groupe de discussion nous apparaît ressembler aux discussions de " couloir " qui peuvent se dérouler lors de conférences, de congrès ou autres activités réunissant les membres d'une profession. Les groupes de discussion sont en effet une façon de se parler qui devient accessible pendant toute l'année. Mais pour l'instant, ces échanges sont plutôt informels.

Ces remarques nous amènent à définir de nouvelles pistes de recherche pour en savoir plus sur cet aspect des groupes de discussion. Ainsi, il faudrait d'abord vérifier à quel point le groupe de discussion favorise

les communications privées entre les membres, par courrier électronique ou à l'aide d'un autre moyen. Par exemple, si les demandes d'information sont répondues par des messages privés, nous ne connaissons pas l'ampleur des discussions qui en découlent. Par conséquent, il semble difficile de mesurer l'impact réel d'un groupe de discussion scientifique uniquement à l'aide des messages qui y circulent. Par contre, étudier les communications personnelles qui en découlent pose deux problèmes importants: l'identification des objets à étudier (qu'est-ce qui découle d'un groupe de discussion?) et le caractère privé d'une telle étude.

7. Bibliographie

- Carley, Kathleen, and Kira Wendt. "Electronic Mail and Scientific Communication." *Knowledge: Creation, Diffusion, Utilization* 12, no. 4 (1991): 406-440.
- Carroll, Jim, and Rick Broadhead. *Canadian Internet Handbook*. Scarborough: Prentice Hall, 1994.
- Cole, S., and J. R. Cole. *Social Stratification in Science*. Chicago: University of Chicago Press, 1973.
- Crane, Diane. *Invisible Colleges. Diffusion of Knowledge in Scientific Communities*. Chicago: The University of Chicago Press, 1972.
- Doty, Philip. "Electronic Networks and Social Change in Science." *Proceedings of the ASIS Annual Meeting* 29 (1992): 185-192.
- Doty, Philip, Ann Peterson Bishop, and Charles R. McClure. "The National Research and Educational Network (NREN): an Empirical Study of Social and Behavioral Issues." *Proceedings of the ASIS annual meeting* 27 (1990): 284-299.
- _____. "Scientific Norms and the Use of Electronic Research Networks." *Proceedings of the ASIS Annual Meeting* 28 (1991): 24-38.
- Fraase, Michael. *The Mac Internet Tour Guide. Cruising the Internet the Easy Way*. Chapel Hill: Ventana Press, 1993.
- Garvey, W. D., and B. C. Griffith. "Structure, Objectives, and Findings of a Study of Scientific Information Exchange in Psychology." *American Documentation* 15, no. 4 (1971): 258-267.
- Garvey, William D. *Communication: the Essence of Science*. Toronto: Pergamon Press, 1979.
- Gibbs, Mark, and Richard Smith. *Navigating the Internet*. Indianapolis: Sams Publishing, 1993.
- Ginsparg, Paul. "First Steps Towards Electronic Research Communication." *Computers in Physics* 8, no.

4 (1994): 390-396.

Hagstrom, Warren O. *The Scientific Community*. Carbondale: Southern Illinois University Press, 1965.

Hiltz, Star Roxanne. *Online Communities, A Case study of the Office of the Future*. Norwood: Ablex Publishing Corporation, 1984.

Kovacs, Michael J., and Diane K. Kovacs. "The State of Scholarly Electronic Conferencing." *Electronic Networking* 1, no. 2 (Winter 1991): 29-36.

Krippendorff, Klaus. *Content Analysis: An Introduction to its Methodology*. Newbury Park: Sage Publications, 1989.

McClure, Charles R., Ann P. Bishop, Philip Doty, and Howard Rosenbaum. *The National Research and Education Network (NREN): Research and Policy Perspectives*. Norwood: Ablex Publishing Corporation, 1991.

Merton, Robert K. *The Sociology of Science*. Chicago: University of Chicago Press, 1973.

Price, Derek J. de Solla. *Science Since Babylon*. New Haven: Yale University Press, 1961.

_____. *Little Science, Big Science*. New York: Columbia University Press, 1963.

Rice, Ronald E., and Patricia Tarin. "Staying Informed: Scientific Communication and Use of Information Sources Within Disciplines." *Proceedings of the ASIS Annual Meeting* 30 (1993): 160-164.

Rosenbaum, Howard, and Herbert Snyder. "An Investigation of Emerging Norms in Computer Mediated Communication: an Empirical Study of Computer Conferencing." *Proceedings of the ASIS Annual Meeting* 28 (1991): 15-23.

Snyder, Herbert, Howard Rosenbaum, and Carol Hert. "Referencing Norms in Computer Mediated Communication: an Exploratory Study of Computer Conferencing." *Proceedings of the ASIS Annual Meeting* 29 (1992): 93-97.

Vallée, Jacques, and al. "The Computer Conference: An Abstract State of Communication?" *Futurist* 9 (1975): 116-121.

La Text Encoding Initiative : les moyens pour ajouter de la valeur à un texte numérisé

par

Guylaine Beaudry

Cursus vol.1 no 2 (printemps 1996)

Cursus est le périodique électronique étudiant de l'École de bibliothéconomie et des sciences de l'information (EBSI) de l'Université de Montréal. Ce nouveau périodique diffuse des textes produits dans le cadre des cours de l'EBSI.

ISSN 1201-7302

cursus@ere.umontreal.ca

URL:<http://www.fas.umontreal.ca/ebsi/cursus/>

L'auteure

Après des études en musicologie, Guylaine Beaudry obtiendra une maîtrise en bibliothéconomie et sciences de l'information en 1996. Ses intérêts portent, entre autres, sur la norme SGML et les publications savantes. Elle a été éditrice de *Cursus* de janvier 1995 à janvier 1996. Dans le cadre de sa maîtrise à l'EBSI, elle réalise la conversion vers SGML de deux revues savantes: *Surfaces* et *Géographie physique et Quaternaire*.

La Text Encoding Initiative : les moyens pour ajouter de la valeur à un texte numérisé a été écrit à l'automne 1996 dans le cadre du cours *Structure et fonctions des logiciels documentaire (BLT 6351)* donné par le professeur Yves Marcoux.

Pour joindre l'auteure:beaudryg@ere.umontreal.ca

Droits d'auteur

Tout texte demeure la propriété de son auteur. La reproduction de ce texte est permise pour une utilisation individuelle. Tout usage commercial nécessite une permission écrite de l'auteur.

Introduction

Qu'est-ce qui distingue *Les Fleurs du Mal* de Charles Baudelaire sur support électronique de la même oeuvre sur support papier? Le support papier permet la lecture en diagonale, la lecture attentive, la consultation de la table alphabétique des titres des poèmes et bien sûr procure le plaisir au lecteur de la sensation et de l'odeur du papier, de l'objet *livre*. L'avènement de l'édition électronique, même s'il ne signifie pas, comme certains peuvent le penser, la disparition du support papier, amènera inévitablement une remise en question des conditions et des types d'utilisation de ce support. Dans un environnement numérisé, la matérialisation du texte peut se faire de différentes façons, le support papier étant une des possibilités. Le support électronique permet la lecture à l'écran, la lecture sur papier et éventuellement l'écoute du texte grâce aux outils de synthèse vocale. Le support électronique offre donc le choix au lecteur du contact, de la relation qu'il souhaite établir avec le texte. Il permet une distinction entre l'accès à l'information et la façon de se l'approprier. Dès qu'un texte est numérisé il devient possible d'y faire la recherche de simples chaînes de caractères, ce qui ajoute aux capacités de repérage offertes par la table des matières ou l'index. En utilisant un langage décrivant la structure logique du document de façon compréhensible pour l'ordinateur, la recherche pourra alors s'effectuer sur une chaîne de caractères en tenant compte de sa localisation dans la structure du document. Ces capacités de recherche permettent donc un accès supplémentaire à l'information que ne permet pas le texte imprimé. De plus, l'édition électronique permet d'ajouter une valeur à l'information en utilisant l'hypertexte. Il est alors possible de relier un poème à son analyse, à sa traduction, à des notes ou à d'autres versions présentées dans d'autres éditions de la même oeuvre.

Cet article présente d'abord les origines de la *Text Encoding Initiative (TEI)* et les raisons pour lesquelles SGML a été choisi pour l'encodage des textes de ce projet. Suit une introduction à la définition du type de document (DTD) élaborée par la TEI. Pour cette partie, bien que le texte comporte plusieurs définitions, une connaissance de base de SGML peut être nécessaire à une bonne compréhension. Le lecteur sera peut-être intéressé à consulter le texte suivant: [A Gentle Introduction to SGML](#). Dans la dernière partie, une liste de projets utilisant la DTD est donnée ainsi que les objectifs de la *Text Encoding Initiative* pour les prochaines années.

1. Origine de la *Text Encoding Initiative*

1.1 La valeur d'un texte sur support électronique

L'importance des possibilités du texte sur support électronique est devenue évidente. Depuis quelques années, de nombreux projets de recherche sont consacrés à l'analyse de textes par ordinateur dans des domaines comme la littérature, la philologie, la linguistique et les sciences de l'information. Dans ce courant, plusieurs institutions se sont constituées des banques de textes numérisés; ces textes numérisés deviennent de plus en plus des ressources courantes. Un sondage mené en 1989 et repris en 1991 démontre que, pour l'intervalle compris entre ces deux années, le nombre de bibliothèques faisant l'acquisition de textes électroniques a plus que doublé (Prince-Wilkin 1991, 13). Mais qu'est-ce qui fait la valeur et la qualité d'un texte électronique? La réponse est certainement liée à l'utilisation que veulent en faire les clientèles. Sans considérer le support physique, la valeur d'un texte est basée sur la qualité et la rigueur du travail d'édition. Un deuxième critère déterminant pour les textes sur support électronique est le format. Un format d'encodage doit permettre de rendre compte et de maximiser un travail d'édition de qualité en permettant aux utilisateurs de bien l'exploiter. Le format est également un facteur important pour l'utilisation du texte par différentes applications qui sont par exemple: analyse de concordance, repérage d'information, analyse de style, génération d'index selon la fréquence des mots ou production de nouvelles éditions critiques sur support papier ou électronique. Pour réaliser ces applications, le format d'encodage des textes doit permettre un accès aux mots et au mieux, à la structure logique des documents. Un format image comme format définitif des textes est donc à rejeter car seul un format qui permet l'accès aux textes et à leur structure logique permettra les applications souhaitées.

Le repérage et les autres applications peuvent être rendus très complexes en raison de la nature même des textes. Un texte peut comporter plusieurs passages écrits dans une autre langue, des annotations ou des notes, des citations, des directives

scéniques ou des tableaux. Selon le genre littéraire, l'articulation du texte se fait par chapitre, section, scène, acte, verset ou autres. Tous ces éléments, souvent implicites, sont autant d'informations susceptibles d'être utiles, voire nécessaires à l'utilisateur. Le format d'encodage du texte doit donc rendre compte, selon les besoins, de la structure logique du document.

1.2 Balisage

L'accès au texte est possible en utilisant l'ASCII (*American Standard Code for Information Interchange*), ce que plusieurs initiatives de numérisation ont choisi d'utiliser.¹ Cependant, l'utilisation de ces textes demeure limitée à la lecture ou à la simple recherche de chaînes de caractères. Pour avoir accès à un texte de façon dynamique, le balisage représente une solution intéressante.

Le balisage était le travail que faisait à l'origine le typographe. Il consistait à inscrire à la main sur un document des codes qui servaient ensuite à préparer la représentation physique du texte. La ponctuation et les espaces que nous utilisons pour articuler et délimiter les paragraphes, les phrases et les mots sont aussi une forme de balisage. Appliqué au texte sur support électronique, le balisage permet d'aller plus loin en rendant explicite pour l'ordinateur ce qui est implicite pour le lecteur. Le balisage permet d'identifier la structure logique d'un texte. Comme le mentionne Susan Hockey (1994, 684):

A text without markup can only be used for very simple applications. One analogy is trying to perform functions such as sorting and searching on a bibliographic record which does not have field divisions.

Pour répondre aux besoins des chercheurs, plusieurs systèmes de balisage ont été créés ainsi que le rapporte Susan Hockey (1994, 684-5). Au début des années soixante à Édimbourg, le système *COCOA* a été développé pour des textes écossais anciens. *COCOA* permet l'identification de la structure logique d'un texte et est utilisé par la majorité des programmes d'analyse de textes, notamment le Oxford Concordance Program. Le *Thesaurus Linguae Graecae* a aussi développé son propre système de balisage appelé *beta code*. Par la suite, sont apparus des systèmes de balisage performants, notamment TEX (Tau Epsilon Xi) qui est un système puissant pour la description des formules mathématiques. TEX est très utilisé dans le monde scientifique et est disponible dans le domaine public. Les systèmes de balisage spécifique (*procedural markup*) tels qu'utilisés par les traitements de textes comme Word ou WordPerfect sont maintenant très répandus. Il est à noter que ces systèmes de balisage sont des formats propriétaires, c'est-à-dire qu'ils sont développés par des organisations dont l'objectif est d'acquérir une part du marché.

La cohabitation de tous ces formats a nécessairement menée vers un véritable chaos (Burnard 1988). À cette étape, le balisage apparaissait déjà comme étant essentiel pour obtenir des textes de qualité, cependant chacun de ces langages a été développé pour des besoins et des applications spécifiques, ce qui contribue au manque de flexibilité. De plus, il y a peu de documentation disponible et les possibilités d'adaptabilité à de nouvelles situations sont difficiles, parfois impossibles: beaucoup de temps et d'efforts doivent être investis pour convertir les textes d'un format à un autre. À cause de ces caractéristiques, aucun de ces formats ne pouvait donc espérer être adopté comme norme officielle.

1.3 SGML

En 1986, le *Standard Generalized Markup Language* (SGML) est sanctionné ISO 8879:1986 par l'*Organisation internationale pour la normalisation* qui est l'instance de normalisation la plus élevée à l'échelle mondiale. SGML n'est pas un langage mais plutôt un métalangage qui permet de fournir un cadre syntaxique servant à la conception de grammaires pour la description de l'information sur support électronique. SGML est un langage de balisage *descriptif*, par opposition au balisage *procédural* qui lui, est orienté vers l'aspect visuel du document. SGML permet la description de la structure logique d'un document en y apposant des *balises* qui permettent d'identifier les différentes composantes d'un texte, appelées *éléments*. Par exemple, <chapitre> est une balise qui pourrait servir à la description d'un segment d'un document répondant aux caractéristiques décrivant l'élément "chapitre". Les relations entre les éléments et la définition de ces éléments sont déclarées par le concepteur dans la *Définition du type de document* (DTD) qui est une sorte de grammaire pour un genre particulier de document. L'*instance de document* qui comprend les données (le contenu), le balisage descriptif et la

référence à une DTD demeure complètement indépendante du traitement, de la plateforme et du logiciel utilisés. Cette particularité rend l'information balisée avec SGML très flexible. L'échange d'un même document entre un mini-ordinateur et un ordinateur Macintosh, tous deux utilisant des logiciels différents, est monnaie courante avec SGML.

1.4 Qu'est-ce que la *Text Encoding Initiative* (TEI)?

Le besoin de normalisation dans le balisage des textes en littérature et sciences humaines a conduit vers la *Text Encoding Initiative*. Les fondateurs de ce projet sont principalement des chercheurs en littérature et en sciences humaines qui souhaitaient un système de balisage et un format facilitant le traitement par ordinateur, l'échange et le partage des textes numérisés. La TEI est gérée et subventionnée principalement par trois associations professionnelles ayant des activités de recherche dans le domaine du traitement de textes par ordinateur: l'*Association for Computational Linguistics* (ACL), l'*Association for Literary and Linguistic Computing* (ALLC) et l'*Association for Computing and the Humanities* (ACH).²

L'objectif de la TEI est de produire un modèle général et des directives pour l'encodage de tous les genres de textes littéraires, peu importe la langue, en répondant aux exigences des chercheurs de chacun des domaines concernés. Aussi, devenait-il crucial pour les chercheurs et les universitaires que les textes créés puissent être utilisés, et plus encore, être réutilisés dans un très grand nombre d'applications, celles existantes comme celles qui ne sont pas encore imaginées ou développées.

1.5 La conférence de Poughkeepsie

La genèse de la TEI remonte à une conférence tenue au Collège Vassar à Poughkeepsie (New York) en novembre 1987. Environ trente représentants venant des domaines de l'archivistique, des sciences humaines, de la littérature et de la recherche universitaire ont reconnu non seulement l'intérêt, mais la nécessité d'une normalisation pour l'encodage et l'échange des textes. Leurs discussions ont mené à des recommandations sur la portée, la structure et le contenu que devrait avoir le système d'encodage. La conférence a résulté en une série de principes³ pour l'élaboration d'un système d'encodage des textes. Le succès de cette conférence est dû à l'implication et au regroupement d'organismes clés dans le traitement des textes par ordinateur et à l'apparition alors récente du SGML qui permettait le développement d'un système simple, flexible et ayant la capacité de satisfaire tous les besoins.

1.6 Pourquoi SGML?

Les caractéristiques du format retenu devaient permettre de remplir les besoins qui avaient déjà été identifiés, à savoir, la capacité d'exprimer des distinctions et parfois des chevauchements hiérarchiques complexes, la capacité d'assurer l'indépendance des données vis-à-vis des différents logiciels et des plateformes informatiques et la compatibilité avec les normes existantes. SGML était donc le candidat le plus apte à répondre aux besoins.

Suite à cette conférence, le coup d'envoi de la *Text Encoding Initiative* fut donné en juin 1988 grâce à l'implication financière du *National Endowment for the Humanities*. Quatre comités ont alors été mis sur pied par des chercheurs de l'Europe et de l'Amérique du Nord pour étudier les questions se rapportant à la documentation des textes électroniques, à leur représentation, à leur analyse et leur interprétation et à la syntaxe du métalangage. Ce travail a mené à l'élaboration d'une DTD et des *Guidelines for Electronic Text Encoding and Interchange*.⁴ Les grandes lignes de cette DTD de même qu'une brève description des *Lignes directrices* font l'objet de la section suivante.

2. La DTD de la TEI et les *Lignes directrices*

2.1 *Guidelines for Electronic Text Encoding and Interchange*

Les *Lignes directrices* sont des conventions pour la description physique et logique de plusieurs genres de textes. Elles proposent des solutions pour les références croisées à l'intérieur et à l'extérieur d'un document, pour l'identification arbitraire de segments de textes, pour l'alignement d'éléments parallèles occupant la même place dans la structure hiérarchique et pour le chevauchement de hiérarchies. Elles proposent également des conventions pour les liens à partir du texte vers des données textuelles, sonores ou visuelles grâce à la norme HyTime.[5](#)

En raison d'un grand souci de préservation de l'autonomie intellectuelle des chercheurs qui balisent les textes, la TEI a opté pour la définition de "lignes directrices" plutôt que de chercher à l'établir des règles immuables. Être trop contraignant aurait mené à l'impossibilité d'appliquer les principes à une situation donnée. Un schéma trop permissif, par contre, risquait de nuire à l'échange des ressources. Il fallait donc trouver un juste milieu.

Plutôt que de simplement suggérer l'utilisation du SGML, la TEI a développé une DTD spécifique. L'utilisation de celle-ci est soumise à des pratiques qui se divisent en trois niveaux: obligatoires, recommandées et optionnelles. De plus, des mécanismes explicites sont définis pour modifier ou faire des ajouts à la DTD tout en s'assurant que les documents soient conformes à la DTD de la TEI. Les *Lignes directrices* de la TEI proposent à leurs utilisateurs une démarche visant à identifier ce qui est à baliser et, ensuite, à réaliser le balisage en SGML de sorte que les textes résultants puissent être échangés électroniquement.

2.2 Modèle de la DTD de la TEI

Les *Lignes directrices* comptent 1290 pages pour décrire le modèle de la TEI. La présente description se limite à la définition de base des principales composantes de la DTD. La DTD de la TEI n'est ni monolithique ni une solution à tous les problèmes ni une collection non structurée de balises. Ce qui est proposé est plutôt un cadre de base avec une variété d'additions optionnelles répondant à des applications spécifiques. On réfère à ce modèle comme étant celui de la pizza de Chicago (le client choisit la pâte de son choix, une croûte mince par exemple, et fait ajouter les ingrédients qu'il désire) en le comparant aux menus de cafétéria ou de tables d'hôte où tous les choix sont faits à l'avance. Dans le langage des déclarations ceci peut s'illustrer de la façon suivante: [6](#)

```
<!ENTITY % croûte "(pan | mince | farcie)" >
<!ENTITY % garniture "(pepperoni | champignon | poivron | anchois ...)">
<!ELEMENT pizza - - (%croûte; fromage & tomate, (%garniture;)* )>
```

Le modèle de la DTD de la TEI est constitué de modules ou fragments de DTD qui peuvent être combinés de différentes façons selon certains principes. La DTD de la TEI n'est donc pas définie comme un seul bloc. Son architecture est modulaire, ce qui donne la possibilité de construire plusieurs DTD qui seront toutefois conformes à la TEI. Une DTD conforme à la TEI est potentiellement composée de quatre fragments de DTD:

1- Un ensemble de balises obligatoires (*core tag sets*)

Cet ensemble a deux composantes:

- a) L'ensemble des éléments et des attributs requis pour tous les genres de documents. Par conséquent cet ensemble est obligatoire.
- b) Un en-tête qui peut être comparé à une page de titre électronique (*TEI header*).

2- Un ensemble de balises de base (*base tag set*)

L'utilisateur doit ici choisir parmi les six ensembles définis qui représentent autant de catégories de textes: prose, poésie,

théâtre, transcription du discours (*transcribed speech*), dictionnaire et informations terminologique. Les ensembles de balises de base définissent les types de documents. Par exemple, TEI.dictionaries est l'ensemble qui contient la déclaration des éléments nécessaires au balisage d'un dictionnaire. Généralement, un seul ensemble de balises de base est nécessaire pour l'encodage d'un genre spécifique.

3- Ensembles de balises additionnelles (*additional tag sets*)

Ces balises permettent de répondre à des besoins particuliers. L'utilisation de ces balises est compatible avec tous les ensembles de base.

4- Ensembles de balises auxiliaires (*auxiliary tag sets*)

Ces ensembles sont composés de balises avec des rôles très précis et très spécifiques, typiquement pour la description d'éléments appartenant à une DTD indépendante de la DTD principale. Ce mécanisme est possible grâce à l'option **CONCUR** qui permet le balisage d'un document avec différentes structures hiérarchiques.

Lorsque **CONCUR** est utilisé, l'identificateur générique⁷ est précédé du nom de la deuxième DTD. Par exemple, si les citations sont balisées en utilisant une DTD représentée par le nom "QD", l'élément <q> ayant comme valeur de l'attribut "who", Jean prendra la forme suivante: <(QD) q who=Jean>. Il y a quelques limitations à cette option, mentionnons premièrement la nécessité de définir plus d'une DTD. Deuxièmement, tous les parseurs SGML ne supportent pas l'option **CONCUR**, en fait la majorité de ceux qui sont disponibles dans le domaine public ne la supporte pas.

Au plus haut niveau, tous les documents conformes à la TEI partagent un modèle commun composé d'un en-tête bibliographique et de l'élément <text>. Voici un exemple d'un document minimal conforme à la TEI :

```
<!DOCTYPE tei [ <!ENTITY TEI.prose "INCLUDE">
]>

<tei>
  <teiHeader>
    <fileDesc>
      <titleStmt>
        <title>Le plus petit document conforme à la TEI</title>
      </titleStmt>
      <publicationStmt>
        <p>Ce document n'est pas publié.</p>
      </publicationStmt>
      <sourceDesc>
        <p> Ce document est original.</p>
      </sourceDesc>
    </fileDesc>
  </teiHeader>
  <text>
    <body>
      <p>Voici le document conforme à la TEI le plus court qu'on puisse imaginer.
      </p>
    </body>
  </text>
```

</tei>

Deux modifications peuvent s'opérer sur ce modèle minimal: plusieurs éléments <tei> peuvent se suivre, cette première option permet de considérer l'oeuvre d'un même auteur comme un seul texte. Une deuxième option serait de considérer l'oeuvre d'un auteur comme une suite de textes indépendants. Le contenu de la balise <text> peut donc être une seule oeuvre ou un ensemble d'oeuvres. Dans ce deuxième cas, un élément <group> apparaît à la place de l'élément <body> de <text> et peut être composé d'une séquence d'autres <group> ou <text>. L'ensemble de balises de base (*base tag sets*) détermine la structure d'un document, plus précisément, la composition de l'élément <text> selon l'interprétation qu'on veut donner au texte.

2.2.1 Divisions

Les différentes parties d'un texte sont définies par la nature même du document. Le choix de la structure de base détermine le vocabulaire des parties ou sections d'un texte. L'articulation d'une lettre n'est pas celle d'un roman ou d'un dictionnaire. Cependant, tous ces objets tels que "chapitre", "section", "acte" ont en commun d'être incomplets en eux-mêmes et d'avoir une place définie dans la structure logique du document. Puisque pour certaines raisons culturelles ou d'application, l'appellation de ces objets peut varier, l'architecture de la TEI les considère comme le même type d'élément, soit une division représentée par l'élément <div>. Une distinction est cependant possible entre les différentes positions hiérarchiques (<div1>, <div2>, etc. jusqu'à <div7>).

Pour les éléments dont la position importe peu, un élément <div> peut être employé. Cependant, les divisions numérotées et celles qui ne le sont pas ne peuvent pas apparaître dans un même élément.

Pour préciser la nature de l'élément <div>, l'attribut "type" est utilisé; par exemple, un chapitre pourrait être identifié par la balise <div2 type='chapitre'>. Pour conjuguer avec cette variété, la définition de toutes les divisions de l'élément <text> est donnée par une "entité paramètre" appelée "%component.seq". Le contenu de cette entité paramètre variera donc d'un ensemble de balises de base à un autre.

2.2.2 Exemple d'un document comportant des divisions

```
<!-- ... -->
<body>
  <div1 type=recueil>
    <div2 type=poème>
      <titre>L'invitation au voyage</titre>
      <strophe>
        <vers>
          <segment>Mon enfant, ma soeur, </segment>
          <segment>Songe à la douceur</segment>
          <segment>D'aller là-bas vivre ensemble!</segment></vers>
        <vers>
          <segment>Aimer à loisir,</segment>
          <segment>Aimer et mourir</segment>
          <segment>Au pays qui te ressemble!</segment></vers>
        <vers>
          <segment>Les soleils mouillés</segment>
          <segment>De ces ciels brouillés</segment>
          <segment>Pour mon esprit ont les charmes</segment>
          <segment>Si mystérieux</segment>
          <segment>De tes traîtres yeux</segment>
```

```

        <segment>Brillant à travers leurs larmes.</segment></vers>
</strophe>
<refrain>
    <segment>Là, tout n'est qu'ordre et beauté,</segment>
    <segment>Luxe, calme et volupté.</segment>
</refrain>
</strophe>
<!-- ... -->
</div2>
<!-- ...-->
</div1>
</body>
<!-- ... -->

```

2.3 Classes

Les quatre cents éléments déclarés dans la DTD de la TEI peuvent être regroupés de bien des façons. La TEI identifie deux grands types de regroupement:

1. Par classes *d'attributs* (*attribute classes*)

Les éléments d'une classe d'attributs partagent le même ensemble d'attributs. Par exemple, tous les éléments avec lesquels il est possible de faire une référence croisée ou une association avec un autre élément utilisent le même ensemble d'attributs, appelé "pointer".

2. Par classes *de modèles* (*model classes*)

Les éléments d'une classe de modèles partagent les mêmes propriétés logiques dans la structure du document, ils apparaissent à la même position dans la structure SGML du document. Par exemple, tous les éléments utilisés pour décrire des corrections éditoriales ou des omissions sont membres de la classe "edit". Les éléments nécessaires pour les citations bibliographiques font partie de la classe "bibl". Un élément peut faire partie de plusieurs classes. Les classes sont définies dans la DTD de la TEI par des entités paramètres et sont utilisées pour les mises à jour, la documentation et les modifications.

2.4 TEI *Lite*

La DTD TEI *Lite* est une sélection des éléments, attributs et paramètres de la DTD complète de la TEI. L'objectif de ce sous-ensemble est de donner un ensemble de départ comprenant les éléments que chaque utilisateur de la TEI devrait connaître. L'utilisation de la DTD TEI *Lite* ne peut cependant pas rendre compte du système modulaire de la DTD complète qui est souvent nécessaire pour le balisage de genres de textes particuliers.

Malgré cela, plusieurs applications peuvent être effectuées en utilisant TEI *Lite*. Les *Oxford Text Archive* l'utilisent pour convertir vers SGML les textes déjà numérisés dans un autre format. Tous comme les *Electronic Text Centers* de l'Université de Virginie et de l'Université du Michigan l'utilisent pour le balisage de tous leurs documents. La *Text Encoding Initiative* utilise elle-même TEI *Lite* pour sa documentation technique.

2.5 L'en-tête de la TEI (*TEI header*)

La description catalographique des documents numérisés est un aspect qui a été étudié en profondeur par un comité de la

TEI. L'intérêt que suscite l'en-tête de la TEI d'un point de vue bibliothéconomique est certain. Tout en innovant, les solutions proposées s'harmonisent avec les processus déjà en place dans les bibliothèques, comme le démontre l'exemple de l'Université de Virginie qui suit cette brève description des réponses apportées par la TEI.

L'en-tête de la TEI, qui fait partie de l'ensemble de balises obligatoires, sert à décrire un document balisé pour permettre aux utilisateurs d'avoir de l'information sur le texte lui-même: la ou les source(s), les principes utilisés pour le balisage et l'historique des révisions et modifications apportées au texte. Ces informations sont nécessaires autant pour les chercheurs qui utilisent les textes que pour les catalogueurs. Aucun document n'est conforme à la TEI s'il ne comporte pas la balise **<teiHeader>**.

Les quatre parties de cet en-tête sont:

1- **<fileDesc>** peut être vu comme l'équivalent de la page titre d'un document papier. Il est difficile d'imaginer un document sans page titre, de la même façon l'élément **<fileDesc>** est le seul qui soit obligatoire pour la **<teiHeader>**. La flexibilité offerte par l'architecture de la TEI permet la description d'un texte en respectant la norme bibliothéconomique RCAA2.

2- L'élément **<encodingDesc>** décrit la relation entre le texte encodé et sa ou ses source(s). Il peut contenir, par exemple, de l'information sur le projet dans lequel s'inscrit l'encodage de ce texte ou des détails sur les décisions éditoriales qui ont été prises.

3- L'élément optionnel **<profileDesc>** permet de donner une description détaillée de ce qui caractérise les aspects non-bibliographiques du texte, tels la langue d'usage, la situation dans laquelle le texte a été produit, le nom des participants et leur rôle. La classification et les descripteurs assignés au texte font également partie de cet élément.

4- **<revisionDesc>** permet la description de l'historique des changements apportés au texte.

L'en-tête peut être très simple ou très complexe et exhaustif. Tout dépend de l'application et des besoins des usagers. Les décisions quant au niveau de précision à apporter à l'élément **<teiHeader>** sont importantes. Il ne faut pas négliger l'importance d'une très bonne description pour l'acquisition de textes électroniques.

Une application intéressante est d'utiliser l'élément **<teiHeader>** pour créer un enregistrement en format MARC. La bibliothèque de l'Université de Virginie a réalisé un tel projet (Gaynord 1994). Depuis 1994, le catalogage des textes sur support électronique est réalisé avec l'en-tête de la TEI. Ces descriptions sont ensuite téléchargées par FTP sur l'ordinateur central. Un programme en langage PERL fait alors la conversion de la description catalographique de SGML vers le format MARC. Les notices MARC sont par la suite chargées dans le catalogue d'accès public de la bibliothèque.

Quelques limitations ont été identifiées par les utilisateurs. L'élément **<revisionDesc>** offre la possibilité de documenter les changements apportés au texte électronique. Cependant, bien que ce problème touche tous les genres de textes électroniques, les *Lignes directrices* ne définissent pas quand et comment identifier une nouvelle version suite à des modifications apportées. Des discussions entre les bibliothécaires, les chercheurs et les créateurs de textes devraient avoir lieu pour arriver à des ententes définissant de façon univoque ce que constitue la "version" d'un texte.

Avec l'élément **<teiHeader>** un document transporte sa description catalographique et même davantage. Un utilisateur peut chercher dans le contenu des différents éléments pour identifier l'information ou le texte dont il a besoin. Cependant, rien dans l'en-tête ne permet de naviguer à travers le texte ou retrouver des portions de celui-ci qui répondent à certains critères.

3. Présent et futur de la *Text Encoding Initiative*

3.1 Qui utilise la DTD de la TEI?

Voici une liste de quelques projets qui utilisent la DTD de la TEI.

Women Writers Project

Ce projet a débuté en 1989 à l'Université Brown. L'objectif est de constituer une base de données avec accès au plein texte de la littérature écrite par des femmes en anglais pour la période de 1330 à 1830.

[URL: http://www.stg.brown.edu/projects/wwp/wwp_home.html](http://www.stg.brown.edu/projects/wwp/wwp_home.html)

Center for Electronic Texts in the Humanities (CETH)

Mis sur pied conjointement par l'Université de Princeton et l'Université Rutgers en 1991, le CETH a pour objectif de promouvoir le développement, la diffusion et l'utilisation des textes électroniques en sciences humaines.

[URL: http://www.ceth.rutgers.edu/](http://www.ceth.rutgers.edu/)

The Oxford Text Archive (OTA)

Géré par les Oxford University Computing Services, l'OTA rend disponible plus de 1500 titres. Son site comprend des textes électroniques de plusieurs auteurs importants en grec, en latin, en anglais et en une douzaine d'autres langues.

[URL: http://info.ox.ac.uk:80/~archive/](http://info.ox.ac.uk:80/~archive/)

American Verse Project

Il s'agit d'une nouvelle source de textes conformes à la TEI annoncée le 18 décembre 1995. Cette nouvelle initiative vient de *Humanities Text Initiative* de l'Université du Michigan et constituera une collection de textes de la poésie américaine.

[URL: http://www.hti.umich.edu/english/amverse/](http://www.hti.umich.edu/english/amverse/)

Electronic Text Center - University of Virginia Library (ETC)

Le ETC numérise et collecte depuis septembre 1992 des textes dans le but de les rendre disponibles par son service de textes en-ligne. Le Centre met également à la disposition de la communauté de l'Université de Virginie l'équipement informatique et les logiciels permettant l'analyse des textes tout en fournissant la formation nécessaire aux chercheurs pour l'utilisation de ces nouveaux outils.

[URL: http://www.lib.virginia.edu/etext/ETC.html](http://www.lib.virginia.edu/etext/ETC.html)

3.2 Le futur de la TEI

Avec la DTD et les *Lignes directrices*, la TEI a déjà accompli beaucoup dans l'établissement de fondations intellectuelles pour le balisage de textes. Naturellement, beaucoup reste à faire. Dans un article portant entre autres sur le futur de la TEI, Ide et Sperberg-McQueen (1995) énoncent les objectifs suivants:

- Production de documentation destinée à la formation des utilisateurs;

- Évaluation suite à une utilisation à grande échelle et implantation de mécanismes pour poursuivre l'extension et la modification des *Lignes directrices*;
- Mise en place de processus plus efficaces pour répondre aux demandes des utilisateurs;
- Démarches pour l'implantation de la DTD et des *Lignes directrices* dans plusieurs projets de numérisation, d'archivage et de production de textes électroniques;
- Activités de formation pour les utilisateurs;
- Extension des *Lignes directrices* comprenant des modifications, des révisions et des extensions en tenant compte de nouvelles normes telles que HyTime et DSSSL (*Document Style Semantics and Specification Language*);
- Description d'autres genres littéraires;
- Poursuite des recherches concernant la description physique d'autres types de textes comme, par exemple, des documents manuscrits, papyrus, etc.

Des collaborations s'établissent aussi avec d'autres acteurs des industries de la langue pour le développement de norme, par exemple, avec le *Expert Advisory Group on Language Engineering Standards* (EAGLES). Des membres de l'organisation de la TEI sont également impliqués dans le projet européen MULTEXT (*Multilingual Tools and Corpora*), qui s'appelait au départ "the Text Software Initiative". Les participants au projet MULTEXT développent des logiciels d'analyse de textes. Ces logiciels seront compatibles avec tout document SGML ainsi qu'avec des documents non balisés (Ide; Véronis 1994).

Conclusion

Plusieurs personnes s'accordent pour reconnaître la très grande réussite de la TEI. Les résultats sont remarquables pour son produit, son adoption et son utilisation par le milieu ainsi que pour la gestion et la coordination du projet. La TEI est aujourd'hui reconnue par la communauté universitaire, par l'industrie et par les agences gouvernementales de financement en Amérique du Nord, en Europe et au Japon. En effet, un comité TEI a été formé au Japon pour y encourager l'utilisation de TEI et coordonner les recommandations venant d'expériences japonaises.

Charles F. Goldfarb, l'inventeur du SGML, à la fin de sa préface élogieuse aux quatre numéros spéciaux de *Computers and the Humanities* sur la TEI écrit:

The vaunted "information superhighway" would hardly be worth traveling if the landscape were dominated by industrial parks, office buildings, and shopping malls. Thanks to the Text Encoding Initiative, there will be museums, libraries, theaters, and universities as well.

Les Fleurs du Mal et toute la production littéraire auront une place de choix sur les inforoutes. Grâce à SGML et à la TEI, l'utilisateur ou le lecteur pourra de plus en plus choisir la relation qu'il souhaite avoir avec le texte.

NOTES

1. Voir à ce sujet le projet Gutenberg [<URL:http://jg.cso.uiuc.edu/pg_home.html>](http://jg.cso.uiuc.edu/pg_home.html)[Retour](#)
2. TEI est subventionnée pour environ 1 000 000 US\$ par le US National Endowment for the Humanities, la Fondation Andrew W. Mellon, le Conseil national de recherche en sciences humaines du Canada et la Direction générale XIII de la Commission de l'Union européenne.[Retour](#)
3. Voir annexe 1.[Retour](#)
4. Qui seront appelées Lignes directrices dans la suite de ce texte.[Retour](#)
5. HyTime (ISO 10744) Langage de structuration hypermédia/événementiel (Hypermedia/Time-based Structuring Language).[Retour](#)
6. Exemple inspiré d'un texte de Lou Burnard [<URL:http://info.ox.ac.uk:80/~archive/teij31/ORGN.htm>](http://info.ox.ac.uk:80/~archive/teij31/ORGN.htm)[Retour](#)
7. Le mot employé pour désigner l'élément dans une balise, e.g. le mot text est l'identificateur générique de la balise (Generic identifier) . Un parseur SGML est un programme qui analyse une DTD et un document balisé avec celle-ci pour ensuite contrôler si le document comporte des erreurs de balisage. L'élément <text> est présent dans le haut de la hiérarchie de tous les ensembles de balises obligatoires. Une entité paramètre est définie pour utilisation dans une DTD et non à l'intérieur d'un texte balisé. Le contenu d'une entité paramètre est un ensemble de déclarations que le concepteur d'une DTD peut utiliser plusieurs fois. L'utilisation d'entités paramètres évite la redondance dans l'écriture et la lecture de la DTD ce qui la rend plus claire. Les entités paramètres facilitent également la gestion d'une DTD. Des modifications faites à l'intérieur d'entités paramètres évitent de modifier les déclarations d'une DTD. Les changements apportés s'appliquent automatiquement à tous les éléments touchés en une seule opération. Le signe pourcent (%) est le premier caractère d'une entité paramètre.[Retour](#)

Quelques ressources World Wide Web sur la TEI

La page d'accueil de la TEI

[<URL:http://www.uic.edu/orgs/tei/>](http://www.uic.edu/orgs/tei/)

Guidelines for Electronic Text Encoding and Interchange

[<URL:http://etext.virginia.edu/TEI.html>](http://etext.virginia.edu/TEI.html)

Où trouver la DTD de la TEI?

[<URL:ftp://ftp.ifi.uio.no/TEI>](ftp://ftp.ifi.uio.no/TEI)

[<URL:ftp://info.ox.ac.uk/pub/ota/TEI/dtd/>](ftp://info.ox.ac.uk/pub/ota/TEI/dtd/)

Texte d'introduction à la TEI

[<URL:http://info.ox.ac.uk:80/~archive/barebones>](http://info.ox.ac.uk:80/~archive/barebones)

L'architecture de la DTD de la TEI

<http://info.ox.ac.uk:80/~archive/teij31>

TEI *Lite* et une introduction à la TEI

<http://info.ox.ac.uk/~archive/teilite/>

Bibliographie

Barnard, David T.; Burnard, Lou; Gaspart, Jean-Pierre; Price, Lynne A.; Sperberg-McQuenn, C.M.; Varile, Giovanni Battista. "Hierarchical Encoding of Text: Technical Problems and SGML Solutions." *Computers and the Humanities* 29, no 3 (1995): 211-231.

Burnard, Lou. "Report of Workshop on Text Encoding Guidelines." *Literary and Linguistic Computing* 24, no 4 (1988): 131-133.

Burnard, Lou. "Rolling your own with the TEI." *Information Services & Use* 13, (1993): 141-154.

Burnard, Lou. "What is SGML and How Does It Help?" *Computers and the Humanities* 29, no 1 (1995): 41-50.

Coombs, James H.; et al. "Markup Systems and the Future of Scholarly Text Processing." *Communications of the ACM* 30, no 11 (1987): 933-947.

Dunlop, Dominic. "Practical Considerations in the Use of TEI Headers in a Large Corpus." *Computers and the Humanities* 29, no 1 (1995): 85-98.

Gaynor, Edward. "Cataloging Electronic Texts: The University of Virginia Library Experience." *Library Resources and Technical Services* 38, no 4 (October 1994): 403-413.

Gaunt, Marianne I. "Center for Electronic Texts in the Humanities." *Information Technology and Libraries* 13, (mars 1994): 8-13.

Giordano, Richard. "The TEI Header and the Documentation of Electronic Texts." *Computers and the Humanities* 29, no 1 (1995): 75-84.

Goldfarb, Charles F. "Preface." *Computers and the Humanities* 29, no 1(1995): 1.

Guédon, Jean-Claude. *Why are Electronic Publications Difficult to Classify? The Orthogonality of Print and Digital Media*. 1994.

<http://www.nlc-bnc.ca/documents/libraries/cataloging/guej1.txt>

Herwijnen, Eric Van. *SGML Pratique*. Paris: International Thomson Publishing France, 1995.

Hockey, Susan. "Developing Acces to Electronic Texts in the Humanities." *Computers in Libraries* 13, no 2 (February 1993): 41-43.

Hockey, Susan. "Evaluating Electronic Texts in the Humanities." *Library Trends* 42, no 4 (Spring 1994): 676-693.

Ide, Nancy M.; Sperberg-McQueen, C.M. "The TEI: History, Goals, and Future." *Computers and the Humanities* 29, no 1 (1995): 5-15.

Ide, Nancy M.; Véronis, Jean. "MULTTEXT (Multilingual Tools and Corpora)." *Proceedings of the 14th International Conference on Computational Linguistics, COLING'94*, Kyoto (Japan), 1994, 90-96.

Lynch, Clifford A. "Text Encoding Initiative." *Bulletin of the American Society for Information Science* 15, (June/July 1989): 15-16.

Price-Wilkin, John. "Text Files in Libraries : Present Foundations and Future Directions." *Library hi tech* 9, no 3 (1991): 7-44.

Seaman, David. *Campus Publishing in Standardized Electronic Formats - HTML and TEI*. Electronic Text Center - University of Virginia Library, November 1994.

Sperberg-McQueen, C.M. "The Design of the TEI Encoding Scheme." *Computers and the Humanities* 29, no 1(1995):17-39.

Sperberg-McQueen, C.M.; Burnard, Lou, eds. *Guidelines for Electronic Text Encoding and Interchange (TEI P3)*. Chicago/Oxford: Text Encoding Initiative, 1994.

Annexe 1

Principes émanant de la Conférence de Vassar qui a eu lieu en novembre 1987.

1. The Guidelines are intended to provide a standard format for data interchange in humanities research.
2. The Guidelines are also intended to suggest principles for the encoding of texts in the same format.
3. The Guidelines should
 - a. define a recommended syntax for the format,
 - b. define a metalanguage for the description of text-encoding schemes,
 - c. describe the new format and representative existing schemes both in that metalanguage and in prose.
4. The Guidelines should propose sets of coding conventions suited for various applications.
5. The Guidelines should include a minimal set of conventions for encoding new texts in the format.
6. The Guidelines are to be drafted by committees on
 - a. text documentation
 - b. text representation

c. text interpretation and analysis

d. metalanguage definition and description of existing and proposed schemes,

coordinated by a steering committee of representatives of the principal sponsoring organizations.

7. Compatibility with existing standards will be maintained as far as possible.

8. A number of large text archives have agreed in principle to support the Guidelines in their function as an interchange format. We encourage funding agencies to support development of tools to facilitate this interchange.

9. Conversion of existing machine-readable texts to the new format involves the translation of their conventions into the syntax of the new format. No requirements will be made for the addition of information not already coded in the texts.

Vers une définition de la lecture professionnelle

par

Carole Brouillette

Cursus vol.1 no 2 (printemps 1996)

Cursus est le périodique électronique étudiant de l'École de bibliothéconomie et des sciences de l'information (EBSI) de l'Université de Montréal. Ce nouveau périodique diffuse des textes produits dans le cadre des cours de l'EBSI.

ISSN 1201-7302

cursus@ere.umontreal.ca

URL:<http://www.fas.umontreal.ca/ebsi/cursus/>

L'auteure

Carole Brouillette a complété des études en histoire de l'art et en bibliothéconomie. L'analyse de texte et l'indexation de l'image, de même que les théories qui s'y rattachent sont les champs des sciences de l'information qui l'intéressent principalement. Ayant complété un stage de formation à l'Office national du film, elle travaille actuellement chez Provigo pour développer un système de classification intégré à des documents administratifs, photographies et vidéos.

Vers une définition de la lecture professionnelle a été écrit à l'automne 1994 dans le cadre du cours *Traitement et représentation des textes (BLT 6251)* donné par le professeur Suzanne Bertran d-Gastaldy.

Pour joindre l'auteure : brouilc@war.wyeth.com

Droits d'auteur

Tout texte demeure la propriété de son auteur. La reproduction de ce texte est permise pour une utilisation individuelle. Tout usage commercial nécessite une permission écrite de l'auteur.

Introduction

Notre étude se veut un état de la question des recherches pouvant mener à une définition de la lecture professionnelle. De par notre formation en bibliothéconomie et sciences de l'information, de par l'étude des processus d'analyse de contenu que nous avons été appelée à développer et ayant pu nous familiariser avec une gamme d'outils spécifiquement conçus pour faciliter la description des contenus textuels (thésaurus, listes de vedettes-matières, plans de classification, etc.), nous en sommes arrivée à décrire le métier d'indexeur pour une large part comme celui de *lecteur professionnel*.

Nous nous sommes donc intéressée à creuser cette notion, afin de mieux comprendre sa spécificité, car nous avons pressenti que la lecture effectuée dans le cadre d'un travail diffère de celle faite dans une autre situation, dont par exemple celle de loisir qu'on lui oppose fréquemment. De plus, la situation de "lecture-travail" subissant des changements radicaux à cause des technologies de l'information, comme plusieurs autres situations de travail, il est devenu important de comprendre la nature de cette lecture en mutation et les interstices par lesquelles se manifestent les transformations qu'elle subit. Michel Foucault illustre bien l'un des changements de trajectoire du travail du texte:

(...) l'histoire a changé sa position à l'égard du document: elle se donne pour tâche première, non point de l'interpréter, non point de déterminer s'il est vrai et quelle est sa valeur expressive, mais de le travailler de l'intérieur et de l'élaborer: elle l'organise, le découpe, le distribue, l'ordonne, le répartit en niveaux, établit des séries, distingue ce qui est pertinent de ce qui ne l'est pas, repère des éléments, définit des unités, décrit des relations. (Foucault, 1969, p. 14).

C'est dire que les lecteurs atteints par ce changement n'agissent plus en fonction d'un seul décryptage du texte, d'une transcription décodée de la matière textuelle, mais voient leurs tâches se concentrer sur la greffe de textes en territoires discursifs fertiles. Ainsi, ce type de lecteur, moins occupé à chercher un chaînon manquant dans les archives documentaires de l'humanité, s'affaire maintenant à la mise en relation des segments textuels exprimant des idées, sans donner préséance à leur apparition chronologique. On pourrait dire qu'il pratique l'hypertexte, et donc qu'il se doit de gérer les liens qu'il tisse entre les paliers textuels et les idées qui y circulent, les dénominateurs communs comme la chronologie n'étant plus régulateurs de sa recherche.

"Lire c'est donc constituer et non pas reconstituer un sens." (Goulemot, 1985, p. 91)

"(...) la lecture analytique d'un texte est une opération de commentaire, c'est-à-dire d'ajout descriptif et explicatif." (Meunier et al., 1994, p. 24).

Pour citer à nouveau Foucault, disons que l'histoire connaît un "déplacement du discontinu: son passage de l'obstacle à la pratique; son intégration dans le discours de l'historien où il ne joue plus le rôle d'une fatalité extérieure qu'il faut réduire, mais d'un concept opératoire qu'on utilise." (p. 17).

Cette façon d'articuler le discours, ou de fabriquer le texte, a des conséquences sur la recherche documentaire et sur la lecture conviée par ce travail et c'est sur ces aspects de la "génétique discursive" que nous nous pencherons plus à fond dans les pages qui suivent. Nous avons choisi une démarche de recherche qui relève davantage de la comparaison des perceptions de la lecture exprimées dans divers domaines, dont les études littéraires et la psychologie, que de la comparaison des modes d'appréhension du texte par domaine du savoir (environnement, médecine, etc.). À cette étape-ci, nous croyons important de comparer les approches de la lecture plutôt qu'il ne soit les situations de lecture comme telles.

D'autre part, il est bon de noter que les textes retenus pour faire état d'une certaine lecture professionnelle sont américains, français et québécois et datent d'au plus vingt-cinq ans. Nous avons choisi de ne pas restreindre notre corpus aux toutes dernières années, la fin des années 1960 et le début des années 1970 ayant vu naître plusieurs textes théoriques sur la lecture qui supportent en plusieurs points l'évolution de la lecture professionnelle, l'informatisation de la matière textuelle et leurs conséquences en milieu de travail.

1. Détour méthodologique via les définitions de la lecture-jouissance

Une des façons d'aborder le problème de la définition de la lecture-travail est de la comparer à d'autres types de lectures. À cet effet, on rencontre dans la littérature, sous diverses formes, une distinction assez fréquente entre la lecture du plaisir et les autres lectures.

Par exemple, Hochon et Évrard (1994) résument la distinction faite par Bellenger (1978) entre la lecture-jouissance (où l'objectif est de goûter le texte en tant que tel), la lecture-appropriation (où l'objectif est de comprendre en profondeur la teneur d'un texte), la lecture-exploration (visant la recherche d'une information supposée sive ou présente dans le texte), la lecture-assimilation (visant une connaissance du texte en tant que tel permettant par exemple d'en faire un compte-rendu), la lecture d'inspiration (en vue de rebonds vers d'autres lectures ou des réalisations externes au texte).

Malgré qu'ils indiquent que la plupart de ces lectures comportent à divers degrés tous ces objectifs, leur définition de la lecture professionnelle les inclut tous sauf le premier. Pour Hochon et Évrard, en effet, la

la lecture professionnelle se veut une lecture pouvant s'inscrire dans le long terme, un parcours individualisé, une lecture attentive, d'inspection et d'observation *scrutinative* principalement des contenus, une lecture d'exploration, et une lecture pouvant engendrer une importante activité de rédaction (Hochon et Évrard, 1994, pp. 11-12).

Aucune mention de plaisir ou de jouissance dans cette analyse de la lecture professionnelle. À tout le moins, il est question de la structure du texte, mais en fonction du rôle actif qu'elle opère dans la représentation du contenu. L'objectif de "goûter le texte en tant que tel" aurait, de ce point de vue, peu à voir avec la lecture-travail. Mais encore, s'en tenir à la seule dimension de plaisir ou plutôt par la négative de l'absence de plaisir pour esquisser une première caractéristique de la lecture-travail est discutable. On peut en effet imaginer l'impression, sans doute très proche du plaisir, ressentie par un chercheur à la lecture d'un texte ajoutant beaucoup de matière à des recherches menées depuis longtemps, permettant le partage de résultats qui donnent de la crédibilité à des phénomènes jusque-là hypothétiques, ou facilitant la mise en relation de phénomènes isolés. Le professionnel du milieu médical qui trouvera dans un texte un indice essentiel au perfectionnement du vaccin contre le sida éprouvera sans doute une satisfaction apparentée au plaisir.

Le strict plaisir est donc, de ce point de vue, insuffisant pour distinguer les deux types de lecture. Il faudra donc qualifier ce plaisir. Barthes laisse entendre que le plaisir de la lecture implique un **espace** de jouissance:

"Ce n'est pas la 'personne' de l'autre [l'auteur pour le lecteur ou le lecteur pour l'auteur] qui m'est nécessaire, c'est l'espace: la possibilité d'une dialectique du désir, d'une *imprévision* de la jouissance: que les jeux ne soient pas faits, qu'il y ait un jeu." (Barthes, 1973, p. 11).

Voilà des indices intéressants qu'il est difficile d'associer à la lecture-travail: imprévision et lien de causalité entre plaisir et révélation graduelle, accidentée du sens, ou en un mot "suspens" du sens. Nous reparlerons des prédispositions du lecteur professionnel plus loin, mais juste en reprenant l'exemple cité précédemment, on imagine mal que le plaisir du chercheur soit davantage dû à la retenue de la révélation de l'information plutôt qu'à l'information elle-même.

"Dans la perversion (qui est le régime du plaisir textuel) il n'y a pas de 'zones érogènes' (expression au reste assez casse-pieds); c'est l'intermittence, comme l'a bien dit la psychanalyse, qui est érotique: celle de la peau qui scintille entre deux pièces (le pantalon et le tricot), entre deux bords (la chemise entrouverte, le gant et la manche); c'est le scintillement même qui séduit, ou encore: la mise en scène d'une apparition-disparition." (Barthes, 1973, p. 19)

Ainsi, contrairement à la lecture-travail, la lecture du (strict) plaisir serait alimentée par la subversion de tous les modes d'accès, voire de possession.

À un autre niveau, la lecture-travail serait faite à dessein, contrairement à celle du plaisir:

"Si j'accepte de juger un texte selon le plaisir, je ne puis me laisser aller à dire: celui-ci est bon, celui-là est mauvais. Pas de palmarès, pas de critique, car celle-ci implique toujours une visée tactique, un usage social et bien souvent une couverture imaginaire. Je ne puis doser, imaginer que le texte soit perfectible, prêt à entrer dans un jeu de prédicats normatifs: c'est trop ceci, ce n'est pas assez cela; le texte (il en est de même pour la voix qui chante) ne peut m'arracher que ce jugement, nullement adjectif: c'est ça! Et encore plus: c'est cela pour moi!" (Barthes, 1973, p. 24).

"Pleasure reading is playful: it is free activity standing outside ordinary life; it absorbs the player completely, is unproductive, and takes place within circumscribed limits of space and time." (Nell, 1988, p. 2).

"Une seule condition à cette réconciliation avec la lecture: ne rien demander en échange. Absolument rien. N'élever aucun rempart de connaissance préliminaire autour du livre. Ne pas poser la moindre question. Ne pas donner le plus petit devoir. Ne pas ajouter un seul mot à ceux des pages lues. Pas de jugement de valeur, pas d'explication de vocabulaire, pas d'analyse de texte, pas d'indication biographique... S'interdire absolument de 'parler autour'. Lecture-cadeau. Lire et attendre." (Pennac, 1992, p. 127.).

Attendre sans attente ni obligation, dirions-nous. D'où la condition privée du plaisir de la lecture, assujettie à aucune prescription extérieure: "the easy sinking through the pages into soundless dreams." (Nell, 1988, p. 9).

Ainsi, la lecture-travail, si on l'oppose à cette étape-ci à celle du plaisir "pur", ne favoriserait pas l'expansion d'un espace de jouissance où clignote la révélation de l'information. Elle ferait plutôt révaloir une certaine *nudité* de l'information soit son accès immédiat. De plus, la lecture-travail serait motivée par un projet formulé *avant* le texte et re-formulé *après* le texte: ce qui l'insère dans un **processus qui s'inscrit dans le temps, qui est orienté et qui se veut productif.**

2. Lecture-travail et objet de lecture

La notion de *texte en tant que tel* retenue par Hochon et Évrard pour expliciter la notion de lecture-jouissance mérite quelques éclaircissements. Il nous semble important de ne pas faire d'équivalence systématique entre objet de lecture et type de lecture. On associe parfois, par exemple, lecture-jouissance et littérature ou lecture-jouissance et poésie, ces catégories littéraires s'étant entre autres donné pour objet le travail du texte, dirions-nous *en tant que tel*. Toutefois, il est tout aussi envisageable de faire une lecture "jouissance" que professionnelle de ces objets textuels. Donc, la lecture-travail ne dépendrait pas du texte lui-même, mais bien de certaines dispositions du lecteur. Eco (1979) dirait que la coopération du lecteur est nécessaire à l'actualisation du texte.

"Le plaisir, (...) n'est pas un *élément* du texte, ce n'est pas un résidu naïf; il ne dépend pas

d'une logique de l'entendement et de la sensation; c'est une dérive, quelque chose qui est à la fois révolutionnaire et asocial et ne peut être pris en charge par aucune collectivité, aucune mentalité, aucun idiolecte." (Barthes, 1973, p. 39).

Le professionnalisme relèverait donc davantage de l'attitude du lecteur que de la matière textuelle sur laquelle il exerce une action.

3. Dispositions du lecteur

Les deux notions de lecture professionnelle (Hochon et Évrard, 1994) et de lecture-jouissance (Bellenger, 1978), contiennent le mot "lecture" et leur complément se distingue davantage par rapport à un contexte de lecture qu'en fonction d'une pré-supposition théorique (ex.: lecture psychanalytique, etc.):

PROFESSIONNEL-LE:

relatif à la profession, qui est tel par profession (Dictionnaire Petit Robert 1);

(SITUATION DE) JOUISSANCE:

(situation de) plaisir que l'on goûte pleinement (Dictionnaire Petit Robert 1).

3.1. La compréhension

Bertrand Gervais (1991) s'est intéressé à l'acte de lecture en associant, dans une perspective extrême de disposition du lecteur, la jouissance à la progression et le travail à la compréhension.

"Il semble bien que les définitions de l'acte de lecture soient le lieu d'un conflit, d'une opposition entre deux pratiques. L'une, habituellement jugée favorablement, où la lecture d'une oeuvre est l'occasion d'un travail, de l'approfondissement d'un savoir et d'une érudition; et l'autre, jugée plutôt négativement, où elle est le lieu d'un repos, d'un divertissement. Ces deux pratiques de lecture sont souvent décrites comme le résultat de deux procès de nature différente. La lecture comme travail est une activité dite de compréhension, tandis que la lecture comme repos est une activité de progression. L'une va en profondeur: elle est lente; l'autre reste en surface: elle est rapide." (p. 20.)

Cette schématisation bi-polaire de l'acte de lecture est, nous en convenons avec Gervais, trop exclusive, trop radicale. Nous préférons, comme lui, entrevoir les deux activités compréhension-progression amalgamées et alternativement à l'oeuvre dans toute lecture. Mais l'emphase mise sur l'une des variables a certainement des effets sur les stratégies de lecture. On imagine par exemple que pour la lecture-travail, le poids de la compréhension, bien que déstabilisée par l'activité de progression, sera dominant.

Ainsi, associant dans cette perspective lecture professionnelle et compréhension, nous avons cru bon relever dans la littérature (Baker et Brown, 1984; Gibson and Levin, 1975; Stauffer, 1969) les principaux mécanismes d'auto-régularisation que l'on dit justement favoriser la compréhension, afin d'identifier les décisions éventuellement impliquées par l'acte d'une lecture dite professionnelle.

3.1.1. Établissement des objectifs de lecture

Cette disposition figure en tête de liste car toutes les autres en dépendent. Contrairement à ce qu'insinuait Pennac en parlant du plaisir de lire, la lecture-travail aurait pour principale caractéristique la détermination ou la formulation de la raison qui pousse à la lecture. L'attitude de totale réceptivité serait l'affaire du plaisir.

Gibson et Levin (1975) donnent à cet effet un exemple très simple (p. 476): le fait qu'un lecteur ait déterminé que l'objet de sa quête soit une valeur chiffrée dans un texte où la majorité des caractères sont des lettres affectera sa vitesse de lecture et son parcours visuel du texte. Il est facile d'imaginer que sa lecture sera tout autre s'il cherche plutôt à y déceler l'orientation politique de l'auteur. Et pour citer Barthes, il est clair que l'érotisme du texte ne sera pas au rendez-vous lors de cette fouille; c'est plutôt la logique du strip-tease (1973, p. 20) de l'information qui sera alors tout à l'honneur.

"Ce n'est pas là (là où il y a érotisme) le plaisir du strip-tease corporel ou du suspense narratif. Dans l'un et l'autre cas, pas de déchirure, pas de bords: un dévoilement progressif: toute l'excitation se recute; fugie dans l'espoir de voir le sexe (...) ou de connaître la fin de l'histoire (...)." (Barthes, 1973, p. 20).

Farrow (1991) a également démontré que les lecteurs professionnels que sont les indexeurs lisent les textes d'une manière particulière en fonction de l'objectif de travail auquel ils doivent répondre: indexer un texte. Considération intéressante, Bellenger va même jusqu'à lier la détermination d'un objectif de lecture et la réflexion:

"On arrive à penser plus vite quand on maîtrise les moyens (domestication de l'oeil), quand on sait ce que l'on cherche (objectif) et dans quelle intention (projet de lecture) on investit un effort pour une lecture désirée." (Bellenger, 1978, p. 109).

Enfin, Meunier et al. vont dans ce sens en affirmant que:

"Les théories classiques sur le processus cognitif de l'accès au contenu d'un texte ont toujours soutenu que cette activité dépend toujours des objectifs et des projets que se donne le lecteur." (Meunier et al., 1994, p. 22).

La lecture-travail serait donc fondée sur l'identification d'une information convoitée, à partir de laquelle organiser sa lecture. Le lecteur professionnel serait donc enclin à suivre des indications afin d'atteindre le lieu de l'information qu'il convoite, alors que le lecteur euphorique disposerait plutôt de toute carte

routière pour s'adonner avec plus de liberté à la déambulation surprenante.

3.1.2. Modification de vitesse et des stratégies de lecture en fonction des différents objectifs de lecture

À différents objectifs correspondraient différentes approches du texte. La vitesse de lecture est un des éléments pouvant varier et avec elle, le parcours du texte. Le choix d'un objectif, d'une orientation de lecture est traduit chez Bellenger par l'expression "lecture active" et cette *activité* de recherche aurait pour effet de libérer le lecteur du mot à mot:

"Le parcours flexible est en fait une alternance de parcours horizontaux et de traversées obliques sur une page. La variété de ce parcours est en relation directe avec le projet de lecture. Un parcours flexible est la preuve d'une réaction au texte, d'une bonne attention et d'une lecture active." (Bellenger, 1978, p. 110-111).

"Ainsi, face au texte (et à son auteur), le lecteur manifeste sa liberté dans son interprétation; mais il peut aller beaucoup plus loin dans ce comportement: varier sa vitesse de lecture... et même ne pas lire l'intégralité de son texte; bref, procéder à des lectures partielles: lecture d'écrémage, de recherche, de survol. Face à un nouveau texte, le lecteur définit sa stratégie, puis entame sa lecture avec flexibilité." (Richaudeau, 1994, p. 29)

Baker et Brown (1984) font référence à des expériences de Smith (1967) pour affirmer que de bons lecteurs du niveau secondaire utilisaient des stratégies de lecture différentes si on leur demandait de remarquer certains détails d'un texte ou plutôt d'en dégager une impression générale. Par contre, les lecteurs éprouvant des difficultés abordaient le texte de la même manière pour répondre aux deux demandes. De plus, ils pouvaient difficilement décrire leurs stratégies de lecture.

"Reading is as varied and adaptive an activity as perceiving, remembering, or thinking, since in fact it includes all these activities." (Gibson et Levin, 1975, p. 454)

À cet effet, Gibson et Levin (1975, p. 454-455) rapportent les propos d'un chercheur du M.I.T. décrivant sa façon de lire divers documents. Une des différences marquée entre la lecture de ses livres préférés et celle de tous les autres documents est la suivante: les premiers sont lus en totalité et d'un bout à l'autre, même lors de relecture, alors que les autres sont lus en survol d'abord, avec arrêts sur certains graphiques, titres et sous-titres. Ce n'est que s'il ne comprend pas l'article ou s'il lui semble intéressant (en fonction de ses critères à lui) qu'il décide de le lire plus attentivement.

Cette stratégie qu'adopte le lecteur professionnel pour prendre connaissance du texte, en plus d'être ajustable en fonction de l'objectif qui sous-tend la lecture, est dirions-nous "organique" et évolue au fil même de la lecture. Organique parce que transformée par la prise de connaissance elle-même.

"Dès qu'il se dessine un premier sens dans le texte, l'interprète anticipe un sens pour le tout. À son tour, ce premier sens ne se dessine que parce qu'on lit déjà le texte, guidé par l'attente d'un sens dé ;terminé. C'est dans l'élaboration d'un tel projet anticipant, constamment révisé, il est vrai, sur la base de ce qui ressort de la pénétration ultérieure dans le sens du texte, que consiste la compréhension de ce qui s'offre à lire." (Gadamer, 1976, cité par Meunier et al., 1994, p. 22.)

"They [strategies] shift with feedback (rate of knowledge) as the reader progresses." (Gibson et Levin, 1975, p. 471).

Ainsi la stratégie de lecture est ajustée avant, puis durant la lecture. Mais malgré cette souplesse de la stratégie de lecture, il est difficile pour un lecteur d'appréhender un texte en conciliant plusieurs objectifs qui n'appartiennent pas à un même registre.

"That attending to physical features of a message, such as acoustic features of speech or graphic features of writing, is not compatible with optimal reception of the semantic contents of the message is in accord with common sense, and much more objective evidence could be produced (see for instance Martin and Strange, 1968)." (Gibson et Levin, 1975, p. 470).

Cette lecture ou plutôt cette compréhension devrait donc être monovalente, pour être performante chez l'humain. Sans s'étendre sur le sujet, on imagine l'utilité de la superposition de "lectures professionnelles" effectuée à l'aide de logiciels informatiques comme SATO (Voir à ce sujet plusieurs textes de Suzanne Bertrand Gastaldy et al.) ou ALCESTE (BLOT et al. 1994).

3.1.3. Identification des éléments importants de certaines portions de texte

Si l'on veut que le parcours autre que linéaire du texte soit possible et utile, il doit être accidenté, être parsemé de pauses là où des éléments de signification s'érigent en relief. Sinon, cette lecture déboule en surface du début du texte jusqu'à la fin.

"(...) Ce parcours flexible est possible quand l'oeil cherche des 'appuis' dans un texte et que le simple jalonnement des pages permet d'accéder à ce que l'on cherche (idées principales, détails, chiffres...)." (Bellenger, 1978, p. 111).

Ainsi, l'une des tâches importantes du lecteur professionnel consisterait à trier l'information, c'est-à-dire la localiser, puis la qualifier (cette information x est pertinente en fonction de y). Évidemment, la lecture successive de textes structurés de manière identique fournit des indices facilitant le repérage de certaines informations (ex.: titres, en-têtes de sections, etc).

3.1.4. Prise en considération de la structure inhérente du texte

Il y a deux façons d'envisager la structure d'un texte: de par ses qualités graphiques et son découpage physique et par la succession des informations greffées au squelette graphique. Bellenger (1978, p. 112) parle alors de " l'architecture du texte". Le caractère, la typographie seraient des espèces d'indicateurs de changement de ton, de mots-clés, de connotations, etc. Ainsi lorsque le lecteur se trouve contraint de décider, devant un mot incompris par exemple, s'il ne fait qu'anticiper une compréhension ultérieure du mot au fil du texte ou si une stratégie immédiate pour remédier à cette incompréhension doit être mise de l'avant, il prendra dans une certaine mesure les caractéristiques plastiques de l'écriture. De la même façon, il pourrait faire de ces clignotements de l'apparence, des balises qui guident la sélection de l'information.

"[Barker, Foxon, Barber, Laufer] ont démontré d'une façon ou d'une autre l'existence d'une fonction symbolique, et non pas seulement iconique ou indicielle, des signes typographiques qui constituent un véritable code interprétatif. (...) La réalisation récurrente d'éditions de luxe et le renouveau d'intérêt pour les manuscrits calligraphiés d'une part, les nombreuses études récentes sur les modes complexes de présentation des textes et l'enluminure des manuscrits médiévaux d'autre part, appuient tous l'hypothèse fondamentale selon laquelle les formes ont un effet sur le sens." (McKenzie, 1991, pp. 37-38).

C'est dire que ces caractéristiques formelles peuvent être intégrées dans une méthodologie de la lecture. À ce propos, Bertrand-Gastaldy et Pagola (1994), dans leur procédures pour l'utilisation de SATO, en décrivant plusieurs angles d'analyse pour l'élaboration d'un thésaurus, font voir la pertinence de la prise en compte de données typographiques, que ce soit pour évaluer le degré de figement des termes d'un domaine, ou encore pour organiser la segmentation d'un texte.

Le lecteur professionnel serait donc chargé d'opérer une certaine discrimination des informations, non seulement relatives au contenu, mais présentes à la surface du texte et surtout d'établir un lien entre ces deux niveaux d'information.

3.1.5. Usage de connaissances antérieures en vue d'interpréter des informations nouvelles

Le repérage d'éléments nouveaux d'information semble se faire, selon certains (Rubenstein et al. cités par Gibson et Levin, 1975, p. 441), comme si le lecteur essayait de classer les informations dont il prend connaissance dans des compartiments créés antérieurement, au fil des lectures passées, avant de s'en ouvrir de nouveaux. C'est dire que dans un premier temps la nouvelle information serait vue comme celle n'appartenant pas à l'ancienne ou nécessitant un ajustement de l'ancienne.

"Toute rétention suppose un effort de prélèvement, de classement et de réactivation des informations visées." (Bellenger, 1978, p. 121).

"Les connaissances antérieures du lecteur déterminent la quantité et la qualité des

informations qu'il peut significativement traiter dans un texte et qui, par conséquent, affectent directement sa reconstruction de sens. Les connaissances antérieures du lecteur se partagent en deux grandes catégories: les connaissances générales sur le monde et les connaissances langagières." (Tardif, 1989, p. 11.)

Outre l'espèce de catalogue de connaissances générales activé par le lecteur, la couleur de son catalogue, la teinte que lui font subir ses intérêts entre en ligne de compte. Et ici, non seulement il est question des intérêts du lecteur, mais bien de ses intérêts actuels (ou actualisés au fil du temps).

"(...) la lecture est nécessairement cette activité liée à la totalité de l'individu, à ce qu'il est, à ce qu'il vit, à son projet actuel." (Foucambert, 1976, p. 38).

"The scientists attended to information that was especially relevant to their interests and needs." (Wyatt, 1993, p. 51)

Ainsi, le lecteur professionnel, non seulement devrait-il avoir un projet de lecture, mais ce projet devrait s'inscrire, pour être compréhensible par d'autres, dans un cadre plus large, qui lui inclut une communauté avec laquelle il est en contact depuis un certain temps (nous y reviendrons section 3.3): une communauté de chercheurs, une entreprise, etc. Bref, le lecteur professionnel devrait être en mesure de mettre son projet de lecture en perspective, c'est-à-dire de reconnaître les motivations qui le poussent à lire un texte, celui-là, qui porte sur tel aspect de tel sujet, écrit par tel auteur, etc.

D'ailleurs, Afflerbach (1990, cité par Wyatt, 1993, p. 50) a constaté lors d'une étude effectuée auprès d'étudiants complétant un doctorat en chimie et en anthropologie, que le recours aux connaissances antérieures, en vue de l'identification des idées principales d'un texte, était beaucoup plus important lors de la lecture de textes liés à leur champ d'expertise. De manière similaire, Pritchard (1990, cité par Wyatt, 1993, p. 50) a observé que les stratégies du lecteur qui s'appuient sur son savoir antérieur étaient généralement plus nombreuses et plus sophistiquées à la lecture de textes divulguant des informations appartenant à sa propre culture plutôt qu'à une culture étrangère.

3.1.6. Évaluation du texte concernant sa clarté, sa complétude (completeness) et sa cohérence

Wyatt (1993) a effectué une étude visant à évaluer les stratégies mises à l'oeuvre par certains lecteurs professionnels pour faciliter leur compréhension. On a choisi pour ce faire une population d'enseignants universitaires, ayant une activité de recherche importante. On leur a demandé de choisir des articles qu'ils considéraient important de lire dans le cadre de leur fonction. Enfin, on a convenu que la forme des articles devait être de type scientifique. Une série d'attitudes décrivant la lecture de chacun fut notée. La seule qui fut commune chez tous les lecteurs fut l'évaluation du texte lu en termes de qualité de la revue de littérature et des citations, de perspective théorique, de méthodologie, d'analyse, de résultats (surtout pour les apports nouveaux de la recherche), de conclusion générale.

Le jugement serait donc l'une des facultés les plus actives, à divers niveaux, du lecteur professionnel. En plus de devoir juger des limites de son projet de lecture, il serait amené à décider avec discernement des stratégies à adopter pour prendre connaissance du texte, pour ensuite produire une appréciation critique de son contenu et du lien qu'entretient ce texte avec ses intérêts, qui, dans un mouvement circulaire, sont à la base de son projet de lecture.

3.1.7. Capacité de réaction face à certaines incompréhensions

Le jugement du lecteur en situation de travail, nous l'avons déjà mentionné, intervient en plusieurs endroits. Ce processus qui revient plus ou moins à insérer le texte dans un contexte qui le dépasse (il est mauvais en regard de tel intérêt, etc.; tel mot est incompréhensible pour tel lecteur ayant tel bagage de connaissance, etc.). Mais le lecteur professionnel est appelé non seulement à qualifier ainsi le texte dont il prend connaissance, mais aussi à réagir devant les achoppements de la signification.

"Realizing that one has failed to understand is only a part of comprehension monitoring; one must also know what to do when comprehension failures occur. (...) If readers decide to take strategic action, a number of options are available. They may store the confusion in memory as a pending question (Anderson 1980), in hopes the author will soon provide clarification. Or readers may decide to take action immediately and so reread, look ahead in the text, or consult outside sources." (Baker and Brown, 1984, p. 41)

Ainsi, ces lecteurs non seulement jugent la matière textuelle dans plusieurs de ses strates significatives, mais décident des limites du déploiement de leurs stratégies de compréhension lorsque confrontés à certaines incompréhensions.

3.1.8. Sélection de critères appropriés pour l'évaluation de la compréhension

On a peine à imaginer le jugement comme étant exercé sur le seul texte sans y voir un déséquilibre dans la transmission de l'information. Le lecteur, pour être cohérent, devrait pousser l'évaluation jusque sur l'évaluateur, jusque sur lui-même: il devrait pouvoir rendre compte de sa compréhension en fonction de critères tout aussi déterminés, puisque tout aussi étroitement liés au projet de lecture que ceux ayant guidé son appréciation du texte. C'est à cette étape que la récapitulation des stratégies s'opère: "J'ai pu sauter ce bout de texte car j'ai déjà pris connaissance de ce dont il y était question; ce néologisme est plus ou moins explicite, il m'oblige à une recherche, cette citation est erronée car j'ai déjà lu le texte auquel on réfère, etc."

3.2 Principe d'économie

Gibson et Levine (1975) associent au principe de souplesse et d'adaptabilité de la lecture un principe d'économie. En effet, la décision qui donne faveur à l'un des mécanismes d'appréhension du texte de la part du lecteur semble, dans plusieurs cas, pouvoir se placer dans une perspective économique. Pourquoi le lecteur professionnel escamoterait-il des passages de texte? Pourquoi lirait-il d'abord les

graphiques? Pourquoi émettrait-il un jugement sur un texte sinon pour que sa démarche fasse preuve d'une certaine *rentabilité*?

"A second major principle [the first one is that reading is an adaptive process] is the trend toward economy in the adult reader. This trend breaks down into two important subprinciples:

1. The reader will direct his attention to processing textual material in the most economical way he can:
 - a. by selecting relevant information;
 - b. by ignoring irrelevant information;
 - c. by processing the largest units that are appropriate for the task;
 - d. by processing the least amount of information compatible with the task.
2. (...) adaptive reading is characterized by continual reduction of information. The reduction is accomplished in three ways:
 - a. processing is reduced in proportion to the number of alternatives that could succeed;
 - b. alternatives are reduced by the application of rules and constraints which are structural variables in the text;
 - c. alternatives are reduced by using old information, both conceptual and inferred from cooccurrence requirements in the text, to comprehend new information."

(Gibson et Levin, 1975, p. 482)

Énoncé en ces termes, l'indice de rentabilité serait une valeur distinctive de la lecture-travail, celui-ci faisant office de mesure du travail dans la société actuelle. Toutefois, malgré que la dimension économique de la lecture professionnelle nous semble intéressante, elle mérite quelques nuances quant aux affirmations voulant qu'on ne l'envisage qu'en termes de *réduction*. Certains voient plutôt dans l'analyse textuelle un ajout d'information, une plus-value du texte:

"Dans certains cas, on pourra ajouter des informations sur le locuteur, l'allocutaire, la situation, le contexte, le temps, le statut illocutoire, etc. Bref, le texte doit être soumis à des analyseurs qui en décrivent les constituants sémiotiques. Il est important de noter que le texte qui est alors l'objet de l'analyse n'est plus le texte de départ mais le texte amplifié de ces multiples niveaux de description." (Meunier et al., 1994, p. 25).

Évidemment ces deux approches ont presque vingt ans de différence: alors que la première concentrait l'expertise sur l'extraction synthétique d'un type d'information moyennant un certain rendement (rapport qualité/quantité), la seconde semble plutôt miser sur l'aspect combinatoire des informations textuelles, (rapport qualité/étendue).

L'économie au sens d' "organisation des divers éléments d'un ensemble; manière dont sont distribuées les parties" (Dictionnaire Petit Robert 1, 1986, p. 600), malgré qu'elle puisse être exprimée différemment selon les approches, serait donc une variable à considérer dans la définition d'une lecture professionnelle.

3.3 Appartenance du lecteur professionnel

L'une des caractéristiques distinctives du "mature reader" selon Gray et Rogers (cités par Stauffer, 1969, p. 28) est la suivante: "awareness of himself as a responsible group member". Celle-ci figure parmi quatre autres qui se comparent & agrave; celles relevées précédemment pour esquisser un profil du lecteur professionnel: prédisposition pour la lecture, intérêt défini, intérêt croissant de façon exponentielle ("expanding spiral of interest"), compétences impliquant le jugement, la compréhension, etc. Aussi, nous nous proposons de considérer cette appartenance à un groupe comme un trait intéressant pour le cas qui nous occupe.

On a dit que la jouissance de la lecture était un acte privé:

"Il [l'homme] habite en bande parce qu'il est grégaire, mais il lit parce qu'il se sait seul. Cette lecture lui est une compagnie qui ne prend la place d'aucune autre, mais qu'aucune autre compagnie ne saurait remplacer. Elle ne lui offre aucune explication définitive sur son destin mais tisse un réseau serré de connivences qui disent le paradoxal bonheur de vivre alors même qu'elles éclairent l'absurdité tragique de la vie. En sorte que nos raisons de lire lui sont aussi étranges que nos raisons de vivre. Et nul n'est mandaté pour nous réclamer de comptes sur cette intimité-là." (Pennac, 1992, p. 175).

"Le plaisir (...) c'est une dérive, quelque chose qui est à la fois révolutionnaire et asocial et ne peut être pris en charge par aucune collectivité (...)." (Barthes, 1973, p. 39).

Le professionnalisme impliquerait, lui, une appartenance minimale à un groupe: un cadre universitaire, un employeur, etc., puisqu'intégré dans un processus de production discursive. La lecture-travail, sans être communication elle-même, serait subordonnée à une communication *éventuelle*, qui puisse se dérouler hors de l'espace textuel au moment de rendre compte, de divulguer, de publier, etc. Ici, donc, il ne faudrait pas associer l'appartenance sociale à la contrainte improductive, mais à un espace où la lecture peut être délibérément récupérée.

À ce propos, Gervais critique la perception qu'ont certains auteurs (Hatt, 1976; Eco, 1985) du texte comme objet de communication pour proposer à la place d'envisager le texte comme objet de travail:

"Le texte n'est donc pas éliminé mais son rôle n'est plus celui de régie unique; il n'est plus la seule donnée, la seule source d'information de ce qu'est l'acte de lecture. Il n'est qu'une des trois variables identifiées [1. mandat que cherche à atteindre le lecteur; 2. degré d'activation et de développement des processus mis en jeu à la lecture (affectif,

argumentatif, cognitif, perceptif, symbolique); 3. texte] qui forment cet horizon conceptuel permettant de penser la lecture comme une activité aux formes multiples. (...) Pour conserver à la lecture ses raisons d'être, un équilibre est visé entre ce qui vient du texte et ce qu'apporte le lecteur en aptitudes et en mandats." (Gervais, 1991, pp 18-19).

C'est de ce mandat que relève l'appartenance à un groupe évoquée plus haut. Le lecteur *mandaté* serait professionnel. Vu sous cet angle, une telle lecture se veut davantage *matière* à communication qu'*objet* de communication.

Synthèse

Le mot-clé "lecture-travail" n'existe pas encore dans les bases de données bibliographiques ou les index, pas plus que "lecture professionnelle". Et bien que ce sujet soit traité par plusieurs, au moins par bribes, il n'en demeure pas moins volatile, éparé. C'est donc en empruntant plusieurs voies sinueuses que nous nous sommes approché de la notion de lecture professionnelle.

Il aura fallu la définir par opposition à la lecture-plaisir, à cet effet lieu réservé de l'imprévisible, du jeu, de l'érotisation (au sens où l'entend Barthes), de la gratuité. Par un jeu de miroir inversé, on a pu déduire que la lecture-travail serait inscrite dans un processus productif.

Puis, en se référant à une certaine sémiologie de la réception, on a déterminé que la lecture professionnelle relevait davantage de l'attitude du lecteur que de l'objet de lecture. En situation de travail, le lecteur serait prédisposé au déploiement de multiples antennes cognitives orientables. D'abord, l'établissement clair des objectifs de lecture en vue d'atteindre efficacement l'information identifiée comme étant requise, ce qui implique que l'on aborde plus ou moins le texte "un angle à la fois". Ces objectifs amènent la détermination d'un angle d'attaque puis un parcours du texte, une certaine flexibilité dans la vitesse de lecture, voire un abandon du mot à mot, une certaine segmentation du texte ou encore une discrimination de certains éléments du texte, sémantiques comme graphiques. Puis, le lecteur tisserait des liens entre cette lecture et une part de connaissances précédemment acquises, générerait les réactions qui en découlent (incompréhension, par exemple) et évaluerait le texte, puis sa compréhension du texte.

D'autre part, la lecture professionnelle serait régie par certains principes d'économie qui se traduisent tantôt par la réduction de l'information à une quantité moindre d'éléments textuels (certains mots-clés, un résumé, etc.), tantôt par l'ajout (plus-value) de propriétés identifiées au cours de la lecture et utiles pour l'analyse.

Finalement, le lecteur professionnel étant impliqué dans un processus productif et sa lecture étant éventuellement *matière* à communication, celui-ci, de par son action de lecture, traduirait indubitablement son appartenance à un groupe (social, intellectuel, idéologique, industriel).

Conclusion

Notre projet visait à définir un tant soit peu la lecture faite dans le cadre du travail. On a pu suivre la piste de certaines intuitions ou réaffirmer ce qui parfois peut paraître évident. Mais ce que cette recherche nous a aussi permis de constater est que la lecture professionnelle, tout comme les autres procédés productifs, pour voir son rendement augmenter, appelle le développement de certains outils, de certains "facilitateurs". À cet effet, la lecture professionnelle se trouve aux confins de deux tranches historiques, dans une espèce de clairière théorique entre les habitudes de lecture acquises à partir du support papier et les autres, encore embryonnaires, qu'encouragent les développements de l'informatique.

Comme nous l'avons laissé entendre ci-haut, l'humain ne peut en un court laps de temps et sans parcourir de nombreuses fois le texte, enregistrer systématiquement l'ensemble des informations inscrites à l'intérieur de plusieurs registres de lecture (ex.: niveau orthographique, lexical, typographique, sémantique, etc.). L'ordinateur peut lui faciliter cette tâche, et surtout, de par la réduction des délais de traitement qu'il occasionne, lui permettre de l'élaborer et de lui donner une portée autrefois pratiquement inaccessible. De plus, l'information et les sources d'information prolifèrent et commandent une plus grande efficacité de l'analyse de l'information. De même, l'information, pour chacune des instances en cause dans notre société, a acquis une valeur. Et cette valeur progresse dans le temps et fluctue en fonction des milieux qui définissent cette valeur. Des outils sont devenus essentiels pour répondre à la demande d'analyse de ces données dont la complexité et les réseaux dans lesquels elles circulent n'ont cessé de croître. Toutes ces données laissent croire que la lecture-travail connaîtra sous peu des développements qui la rendront indissociable des outils informatiques. Ce qui est, à notre avis l'un des principaux traits qui la distingue de la lecture-plaisir.

Alors que cette dernière se développe en vase clos dans chacun des univers privés des individus, la lecture-travail exige, de par sa fonction intégrée dans un réseau d'intérêts, une certaine concertation quant à son développement. Par exemple, les dirigeants d'entreprises devront accepter de voir la notion de lecture professionnelle évoluer et s'attendre à fournir des instruments aux professionnels chargés d'analyser l'information, sans quoi leur compétitivité s'en trouvera atteinte. Le milieu de l'éducation devra s'ajuster à cette réalité: l'apprentissage de LA lecture comme d'un processus unique est dépassé; il faudra songer à l'enseignement de stratégies diverses de lectures, en fonction de diverses exigences et cadres de lecture. Si le lecteur ne peut faire autrement qu'être autodidacte en situation de plaisir, il serait souhaitable que ce ne soit pas le cas en situation de travail et que la lecture-travail soit intégrée à l'ensemble des curriculums de formation.

Si cette transformation de l'analyse est entamée, elle est loin d'être enracinée. Beaucoup de recherches devraient à cet effet être entreprises concernant les interfaces "lecteur professionnel/machine", les limites vers lesquelles doivent tendre les concepteurs de logiciels d'aide à l'analyse pour que ces outils soient à la fois très souples dans leurs fonctions et relativement simples à utiliser, la pédagogie de l'intégration des interfaces de recherche et d'analyse de données informatiques à tous les niveaux d'enseignement.

Voilà plusieurs avenues pour la profession d'indexeur que l'on dit souvent menacée par l'informatisation. Il faudra sous peu remédier à la panique qui atteint plusieurs lecteurs de notre époque face à l'infinie prolifération de l'information et à la paranoïa qui ne cesse de les accabler par crainte qu'un document extraordinaire ne demeure inaperçu dans la masse documentaire. Déjà Foucault laissait entendre en 1969 que l'ère du repérage était révolue et qu'une brèche était taillée dans celle du travail constructif de liens entre les fragments d'une constellation d'informations, i.e. de l'analyse. Les artefacts ne sont donc plus les seuls objets de découverte historiques; leur permutation, leur orientation le sont également. La fouille est maintenant vectorielle.

Plusieurs indexeurs sont plus qu'au courant de ces développements, ils s'appliquent à augmenter la propension et à démontrer la crédibilité de ces méthodes. Souhaitons-leur une intégration sociale utile qui justifie quelques moments de lecture-plaisir bien méritée...

Bibliographie

Baker, L. et Brown, A.L. "Cognitive Monitoring in Reading." In: Flood, J. (éd.). *Understanding Reading Comprehension: Cognition, Language, and the Structure of Prose*. Newark (Del.): International Reading Association, 1984 : 21-44.

Barthes, R. *Le plaisir du texte*. Paris : Seuil, 1982 (1973(c)).

Bellenger, L. *Les méthodes de lecture*. Paris : Presses Universitaires de France, 1985 (1978(c)).

Bertrand-Gastaldy, S. et équipe. "Un prototype de système-expert pour l'aide à l'analyse des jugements." In: *Informatique et droit: Congrès international: La pointe des réalisations et les débats actuels, Montréal 1992, 30 septembre au 3 octobre*. Montréal : AQDIJ, 1994, section C1.3.

Bertrand-Gastaldy, S. et Pagola, G. *Le contrôle du vocabulaire et l'indexation assistés par ordinateur: une approche méthodologique et un procédurier pour l'utilisation de SATO*. Montréal : Université de Montréal : E.B.S.I., 1994. [Sera publié par l'ASTED].

Bertrand-Gastaldy, S. et Pagola, G. "L'analyse de contenu textuel en vue de la construction de thésaurus et de l'indexation assistées par ordinateur; applications possibles avec SATO." *Documentation et bibliothèques*, vol. 38, no 2 (avril-juin 1992) : 75-89.

Bertrand-Gastaldy, S. "L'indexation assistée par ordinateur: un moyen de satisfaire les besoins collectifs et individuels des utilisateurs de bases de données textuelles dans les organisations." *ICO Québec*, vol. 2, no 3 (1990) : 71-91.

Blot, I.; Hammer, B. et Le Roux, D. "Traitement des questions 'ouvertes': utilisation d'ALCESTE, outil d'assistance à l'analyse." *ICO Québec*, (printemps 1994) : 101-105.

- Chartier, R. (éd.). *Pratiques de la lecture*. Marseille : Rivages, 1985.
- Eco, U. *Lector in fabula: le rôle du lecteur*. Paris : Grasset (Livre de poche), 1985 (1979(c)).
- Farrow, J.F. "A Cognitive Process Model of Document Indexing." *Journal of Documentation*, vol. 47, no 2 (Juin 1991) : 149-166.
- Foucambert, J. "Qu'est-ce que lire (chapitre 3)." In: *La manière d'être lecteur*. Paris : Sermap, 1976 : 36-53.
- Foucault, M. *L'Archéologie du savoir*. Paris : Gallimard, 1976 (1969(c)).
- Gervais, B. *Tensions de lecture*. Montréal : UQAM: GREL (Série Recherches et documents), 1991.
- Gibson, E.J. et Levin, H. *The Psychology of Reading*. Cambridge : MIT Press, 1978 (1975(c)).
- Hatt, F. *The Reading Process: a framework for analysis and description*. London : Clive Bingley and Hamden (Conn.) : Linnet Books, 1976.
- Hochon, J.-C. et Evrard, F. "Lecture professionnelle et gestion personnalisée de documents textuels." *ICO Québec*, (printemps 1994) : 9-18.
- McKenzie, D.F. *La bibliographie et la sociologie des textes*. Paris : Éditions du Cercle de la Librairie, 1991.
- Meunier, J.G.; Bertrand-Gastaldy, S. et Paquin, L.C. "La gestion et l'analyse des textes par ordinateur: leur spécificité dans le traitement de l'information." *ICO Québec*, (printemps 1994) : 19-28.
- Nell, V. *Lost in a Book: The Psychology of Reading for Pleasure*. New Haven and London : Yale University Press, 1988.
- Paquin, L.-C. "La lecture experte." *Technologie, idéologie et pratique*, vol. 10, no 2-4 (1994) : 209-222.
- Paquin, L.C. "Une lecture experte des textes juridiques." In: Thomasset, C.; Côté, R. et Bourcier, D. (éds). *Le droit saisi par ordinateur*. Cowansville : Les Éditions Yvons Blais, 1993 : 413-440.
- Pennac, D. *Comme un roman*. Paris : Gallimard, 1992.
- Richaudeau, F. "Théories de la lecture et perspectives: de l'oralité des signes à la perception mosaïque." *Revue de Bibliologie*, no 40 (1994) : 26-32.

Rochon, Y. "Utilisation de SATO dans le domaine de l'évaluation environnementale." *ICO Québec*, (printemps 1994) : 97-100.

Stauffer, R.G. "Reading, a thinking process." In: *Directing Reading Maturity as a Cognitive Process*. New York : Harper & Row, 1969 : 3-31.

Sticht, T. G. "Comprehending Reading at Work." In: Just, M.A. et Carpenter, P.A. (éds.). *Cognitive Processes in Comprehension*. Hillsdale (NJ) : Lawrence Erlbaum Associates, 1977.

Tardif, J. "Les retombées de la psychologie cognitive dans les recherches en lecture." In: Boyer, J.-Y. et Lebrun, M. (éds). *L'actualité de la recherche en lecture*. Montréal : ACFAS (Les Cahiers scientifiques 71), 1990 ((c)1989) : 9-31.

Wyatt, D.; Pressley, M. et al. "Comprehension Strategies, Worth and Credibility Monitoring, and Evaluation: Cold and Hot Cognition when Experts Read Professional Articles that are Important to Them." *Learning and Individual Differences*, vol. 5, no 1 (1993) : 49-72.

Des mini-centres d'information de réalité virtuelle

par

Hélène Guertin

Cursus vol.1 no 2 (printemps 1996)

Cursus est le périodique électronique étudiant de l'École de bibliothéconomie et des sciences de l'information (EBSI) de l'Université de Montréal. Ce nouveau périodique diffuse des textes produits dans le cadre des cours de l'EBSI.

ISSN 1201-7302

cursus@ere.umontreal.ca

URL:<http://www.fas.umontreal.ca/ebsi/cursus/>

L'auteure

Hélène Guertin est professeur en bureautique au Collège de Maisonneuve, à Montréal. Elle poursuit également une maîtrise en sciences de l'information. Elle s'intéresse actuellement à l'ensemble des aspects des autoroutes de l'information.

Des mini-centres d'information en réalité virtuelle a été écrit à l'hiver 1995 dans le cadre du cours *Technologie de l'information (BLT 6029)* donné par le professeur James Turner.

Pour joindre l'auteure: guertinh@ere.umontreal.ca

Droits d'auteur

Tout texte demeure la propriété de son auteur. La reproduction de ce texte est permise pour une utilisation individuelle. Tout usage commercial nécessite une permission écrite de l'auteur.

INTRODUCTION

L'ordinateur est pour certaines personnes une machine puissante mais ennuyeuse et bornée, surtout liée aux études ou au travail. Depuis quelques années cependant, des génies de l'informatique travaillent à changer cette machine en un créateur d'univers virtuels qui risquent de bouleverser notre culture et notre société.

La réalité virtuelle (RV) est cette nouvelle technologie qui crée des environnements artificiels et permet à l'être humain de vivre une expérience sans qu'aucun objet réel ne soit présent. Il s'agit d'une expérience en trois dimensions, entièrement calculée par l'ordinateur. L'utilisateur peut interagir sur l'environnement virtuel comme s'il y était physiquement présent, effectuer des gestes «naturels», manipuler des objets virtuels, par exemple, sans jamais avoir à se déplacer. L'expérience peut être vécue individuellement ou à plusieurs.

La RV est un champ jeune : les images de synthèse en trois dimensions sont nées il y a une vingtaine d'années. C'est le jeune informaticien californien, Jaron Lanier qui inventa l'expression «réalité virtuelle» et qui fit connaître cette technologie au grand public. La RV, pendant longtemps le domaine réservé de l'aviation, est maintenant accessible à de nombreux autres secteurs comme la robotique, l'architecture, le design, la médecine, l'éducation, les arts, etc. Des centres de recherche universitaires, des entreprises privées et la NASA travaillent aujourd'hui à créer des produits et services en réalité virtuelle. Mais c'est par l'industrie ludique que la technologie de la RV rejoint la majorité de ses adeptes au Japon, dans plusieurs pays d'Europe et aux États-Unis.

Nous allons aborder dans le présent texte deux niveaux de la réalité virtuelle : le plus commun permet de regarder dans le monde virtuel par une «fenêtre sur le monde» (WoW ou Window on World). Cette fenêtre peut être aussi grande qu'un écran de cinéma ou aussi petite qu'un écran d'ordinateur. Nous analyserons également un niveau dit «supérieur», l'immersion, qui exige le port d'un équipement spécial.

En quoi cette nouvelle technologie peut-elle intéresser les spécialistes en sciences de l'information? Pour répondre à cette question, nous jetterons un coup d'oeil sur des expériences de premier niveau, et tenterons par la suite d'imaginer d'autres scénarios futuristes d'immersion, susceptibles d'intéresser les spécialistes en sciences de l'information les plus curieux.

Il apparaît essentiel à ce stade de distinguer la «bibliothèque virtuelle» de la «bibliothèque en réalité virtuelle». Les deux termes font appel à l'utilisation d'un ordinateur branché à un réseau externe (Internet par exemple) pour rejoindre une bibliothèque ou un centre de documentation. La «bibliothèque

virtuelle» permet de consulter, à distance, le catalogue en ligne de bibliothèques et de centres de documentation dispersés dans le monde sans devoir se déplacer. La «bibliothèque en réalité virtuelle» fait appel à la technologie de la RV pour construire un lieu composé d'images, illustrant à la fois un contenu informationnel et un décor qui pourrait ressembler ou non à une bibliothèque traditionnelle. Grâce à un équipement spécial, l'utilisateur a l'impression qu'il butine de l'information, en temps réel, dans le décor apparaissant à l'écran.

1. DESIGN ET PROGRAMMATION D'UN MONDE VIRTUEL

Créer des mondes virtuels implique deux tâches : la première consiste à construire les formes physiques du paysage et les objets qui le composent, l'autre à programmer le comportement de ces objets. Il existe aujourd'hui des environnements dont la programmation et le design graphique ont été conçus pour des environnements virtuels spécifiques.

Le design du monde virtuel et de ses objets doit d'abord être défini dans une base de données. Chaque objet a une localisation et une orientation dans le monde virtuel et ces données sont en interaction les unes avec les autres.

Il existe des programmes qui fournissent des objets virtuels primaires comme des cubes, des cônes, des sphères, mais la méthode commune est de construire des objets à partir de polygones. On pourrait, à la rigueur, construire des objets virtuels à partir de formes rectangulaires, donc relativement faciles à dessiner, mais il faut admettre que, dans pareil cas, on serait bien loin de la réalité.

Le créateur d'un environnement virtuel peut attribuer des couleurs et une texture aux objets qui le composent. Plus il ajoute de textures et de couleurs foncées, plus les demandes faites au logiciel et à l'équipement seront grandes; la rapidité à rendre l'image à l'écran ira en décroissant. Il en va de même pour l'éclairage qu'on peut vouloir simple ou sophistiqué. Il en revient aussi au créateur d'embellir son lieu virtuel en important des images en 3D de banques de dessins. Il peut ajouter, par exemple une plante ou une peinture, détails qui faciliteront l'orientation de l'utilisateur dans l'environnement virtuel. La RV permet donc à l'être humain de faire preuve de créativité et d'originalité.

Un ensemble de programmes est nécessaire pour travailler en réalité virtuelle. Il est possible d'importer gratuitement des logiciels (REND386 et AVRIL) à partir d'Internet.¹ Ces programmes n'exigent pas de super ordinateurs, mais il faut savoir qu'un appareil désuet augmente le délai de réponse et dégrade la perception qu'a l'utilisateur de se retrouver dans un monde virtuel.

2. NIVEAUX DE RÉALITÉ VIRTUELLE

2.1 Une «fenêtre sur le monde» : un modèle calqué sur le monde réel

2.1.1 Expérience de la RV en bibliothéconomie

Alan Poulter, de l'University de Loughborough en Angleterre, spécialiste en création de bibliothèques en réalité virtuelle², prône l'utilisation de la RV lorsque les collections sont disponibles en ligne ou sur CD-ROM. L'usage de la RV est également intéressant si la bibliothèque est géographiquement éloignée ou lorsqu'une institution est fermée. Ce spécialiste ne voit aucun intérêt pour le moment à fréquenter une bibliothèque virtuelle dont la structure physique et intellectuelle serait inconnue de l'utilisateur. Lorsque l'objectif vise exclusivement le butinage, il est préférable, selon cet auteur, de calquer l'environnement virtuel sur l'environnement réel. Point n'est besoin de puissantes stations de travail ni d'équipement futuriste : un simple micro-ordinateur 386 ou 486PC ainsi qu'un logiciel en RV suffisent.

2.1.2 Butinage dans une bibliothèque en RV

Pour se déplacer dans la bibliothèque en RV de niveau «fenêtre sur le monde», l'utilisateur peut utiliser une souris en 3D. Il clique sur une porte qui l'amène dans une grande salle où sont classés les livres, comme dans la bibliothèque qu'il fréquente. Cette salle mène à d'autres portes virtuelles donnant accès aux auteurs, aux titres, aux sujets.

Les sections ne présentant aucun intérêt pour l'utilisateur sont éliminées et celles contenant des documents intéressants, placées dans un endroit bien en évidence. En somme, chacun peut organiser la bibliothèque en RV comme il l'entend, un peu comme on trie et filtre la réalité visuelle et sonore autour de soi.

Dans un modèle suggéré par Poulter³, les livres sont rangés sur les tablettes, clairement identifiés par des codes apparaissant sur leur tranche. Les livres trop grands sont coupés pour entrer sur les étagères. L'utilisateur peut prendre un objet (ici un livre virtuel) et le déplacer. Puisque les objets semblables sont rangés ensemble, le butinage permet aussi de regarder les autres objets placés sur les tablettes voisines. Si l'utilisateur choisit de voir les détails bibliographiques d'un document en particulier, il clique une fois sur la souris : la fiche pour chacune des sources pertinentes sur les étagères apparaît alors en surbrillance (avec certains logiciels). Si l'utilisateur clique deux fois, la fiche bibliographique sera imprimée en fin de session, à la sortie, sur l'imprimante virtuelle.

La programmation pourrait être plus sophistiquée, permettant même le butinage assisté par un guide offrant de l'aide auditive. Ce discours digitalisé serait activé par l'utilisateur et lui donnerait de l'information sur le contenu d'une section ou sur la manière de lire les objets. Pour éviter de se perdre, l'utilisateur laisserait tomber de petites roches virtuelles sur son chemin, comme le Petit Poucet.

2.2 L'immersion

Les tenants de l'immersion estiment que le niveau «fenêtre sur le monde» n'est pas vraiment de la RV puisqu'il est facile de «tomber» de la fenêtre. En effet, lorsque l'utilisateur tourne la tête, c'est la salle où il travaille qu'il voit. Il en va tout autrement au niveau immersion : lorsque le participant tourne la tête, il ne voit pas la pièce réelle où il travaille mais le monde autour, celui de la réalité virtuelle apparaissant à l'écran. Ceci est possible grâce à une nouvelle génération d'interfaces informatiques homme-machine

plus ou moins sophistiquées, créée pour les utilisateurs de l'immersion. La qualité des expériences dans cette technologie dépend largement de l'équipement et des logiciels dont dispose l'utilisateur. Le logiciel AVRIL (*A Virtual Reality Interface Library*), disponible via l'Internet, se contente d'un 386PC ou davantage et d'outils d'interface peu coûteux.

2.2.1 L'équipement sensoriel

Il existe sur le marché une panoplie de casques et de gants sensoriels. Ces outils sont équipés de capteurs reliés à l'ordinateur, et l'informent des moindres mouvements de l'utilisateur. En clair, les capteurs prennent les mouvements de la tête et de la main de l'utilisateur dans le monde réel et grâce aux calculs effectués par l'ordinateur, les transfèrent dans le monde virtuel affiché à l'écran.

Le *casque sensoriel*, à la fois visuel et auditif comprend des lunettes contenant deux écrans vidéos miniatures, un pour chaque œil, pour une vision stéréoscopique, comme dans la vie réelle. Le casque est généralement couplé à des écouteurs. Une lentille de grossissement spécialement adaptée s'interpose entre l'œil et l'écran, si bien que l'image occupe la quasi-totalité du champ visuel.⁴ L'ordinateur est informé de la position et de l'orientation de l'utilisateur par les capteurs magnétiques placés sur le casque. Ce sont les mouvements de la tête de l'opérateur qui permettent de changer l'axe de la vision. L'ordinateur, informé en permanence de ces mouvements, présente instantanément à l'utilisateur la vue qui correspond à la direction de son regard.⁵ Si l'utilisateur est désorienté ou perdu, le système va enregistrer automatiquement sa position et le ramener au bon endroit.

Le *gant sensoriel* ou gant de données reconnaît les gestes de la main et permet la navigation dans l'environnement virtuel et l'interaction avec les objets de cet environnement. Par des capteurs fixés sur le gant, l'ordinateur enregistre en permanence les moindres mouvements et changements d'orientation, et recalcule, en temps réel, les images qu'il envoie à l'utilisateur.⁶ C'est en pressant des boutons spécifiques sur le gant sensoriel que les données sont transférées sous une forme reconnaissable par l'ordinateur, permettant ainsi au participant d'interagir avec les objets. Quand il avance la main, le programme affiche à l'écran l'image virtuelle de sa main en 3D. S'il bouge la main, l'image de cette main s'anime en même temps à l'écran. Le gant permet non seulement de désigner un objet virtuel, mais également de le saisir et de le déplacer.

Le prix de l'équipement est fonction de la performance et de la sophistication des mondes virtuels que l'on construit. Cependant, ces prix ont énormément chuté. Aujourd'hui, on peut obtenir un ensemble complet pour moins de 1000\$ US comprenant logiciel, casque et gant de données.⁷

2.2.2 Problèmes d'immersion

Les casques de données et les capteurs magnétiques sont étroitement surveillés relativement aux impacts sur la santé de champs électromagnétiques placés près du corps. Les effets à long terme ne sont pas encore connus et les spécialistes estiment que l'usage de systèmes de RV doit par conséquent être limité à de courtes périodes.

Cette mise en garde ne devrait cependant pas servir de prétexte pour se croiser les bras aux spécialistes en sciences de l'information. Ils ne doivent pas espérer que la réalité virtuelle soit une mode passagère.

VERS DES MINI-CENTRES D'INFORMATION EN RV

3.1 Inventer son propre centre d'information en RV

Bien sûr, certains mordus du multimédia voudront créer eux-mêmes leur propre mini-centre d'information en réalité virtuelle; ce centre n'aura de limites que celles de l'imagination de son propriétaire. Non seulement ces utilisateurs le voudront esthétique, mais également pratique. En effet, pourquoi recréer un espace d'abord conçu pour les documents imprimés si l'on est obligé par la suite d'élaguer l'information moins pertinente. Par ailleurs, un environnement pré-dessiné, avec des formes un peu trop «carrées», risque de déplaire à ceux qui, ayant une connaissance de l'infographie, désireront utiliser leur talent et leur imagination pour créer un lieu virtuel vraiment à leur image.

Peut-on concevoir une bibliothèque en RV sans avoir de connaissances en sciences de l'information (SI)? Peut-on remplir un centre d'information de documents pertinents sans être spécialisé en SI? Sûrement, mais la participation d'un spécialiste est hautement souhaitable; l'idéal serait une équipe composée d'un expert en SI, d'un informaticien et d'un infographiste. Il reviendrait au premier d'acquérir l'information, de veiller à ce qu'elle soit traduite en objets virtuels, modifiables selon les besoins et facilement repérable par les clients éventuels.

Un mini-centre d'information spécialisé pourrait être garni de renseignements provenant de plusieurs bases de données. L'information y serait bien organisée, et accessible 24 heures par jour sur Internet. Ainsi, un usager désireux d'obtenir des renseignements sur un sujet précis pourra s'adresser à l'un de ces mini-centres, que ce soit gratuitement ou moyennant rétribution. Reste à savoir ce qu'il adviendra de la question des droits d'auteurs, mais il s'agit là d'un autre débat.

3.2 Le monde virtuel en réseau

Grâce aux télécommunications, des rencontres entre usagers de la RV ont maintenant lieu à distance, bien que cette pratique soit encore loin d'être généralisée. Ce n'est que lorsque les réseaux de communications seront plus performants et la technologie de la RV plus avancée que des mini-centres d'information autonomes, programmés en RV, seront accessibles via le réseau Internet. Des usagers pourront alors s'y rencontrer virtuellement, via une ligne téléphonique. Coiffés du casque sensoriel, les correspondants auront l'impression d'une rencontre face à face. Le gant leur permettra de butiner et de repêcher l'information désirée dans l'environnement virtuel de leur choix et de télécharger cette information dans leur micro-ordinateur.

Compte tenu de l'ampleur du travail que nécessite la programmation en RV, il y a un intérêt, du moins

pour le moment, à ce que les mini-centres en RV se regroupent autour d'un grand thème, par exemple les «autoroutes de l'information». Chaque mini-centre se spécialiserait sur un volet lié aux inforoutes et offrirait toute l'information disponible et d'actualité s'y rattachant. Situé sur leur propre serveur, chacun proposerait un lien aux autres centres du réseau. En fait, chacun des participants au réseau serait responsable, à la fois, du contenu et du décor relié à son sujet, ainsi que de la mise à jour de l'information dans son propre mini-centre.

Serait-il possible de créer un méga-centre d'information en RV? Sans doute, mais reproduire le décor des bibliothèques d'une grande université par exemple nécessiterait beaucoup de temps. L'avantage des mini-centres privés par rapport à un méga-centre public réside principalement dans l'efficacité et la rapidité avec laquelle l'information y serait ajoutée, retranchée et avec laquelle le décor serait modifié suivant les besoins des clientèles et le développement de l'infographie et du multimédia. Cependant, un méga-centre d'information en RV, à accès payant, supporté par des investisseurs privés, ne semble pas improbable.

3.3 Le spécialiste en SI face à la réalité virtuelle

Être spécialiste en SI n'est pas chose facile à l'heure actuelle. Bien souvent ces spécialistes doivent travailler dans un environnement traditionnel tout en se formant dans leurs temps libres aux nouvelles technologies. Par ailleurs, un grand nombre d'entre eux attendent que les «modes» passent et prennent du retard face à une clientèle jeune et de plus en plus familière avec les nouvelles technologies. Il est vrai que la publicité faite autour de l'équipement de la RV - casque, gant et combinaison sensoriels, qui ressemblent souvent à des gadgets - a de quoi effrayer les gens sérieux.

Quelles seront les conséquences de l'utilisation de cette technologie pour les spécialistes en SI? Actuellement, les bibliothèques, les centres de documentation et les centres d'archives sont déjà affectées par un accès de plus en plus généralisé à Internet et aux nouvelles technologies de l'information. Les gouvernements, le monde des affaires, le monde de l'éducation et de la recherche sont de plus en plus «branchés» ou, du moins, sont intéressés à l'être. Déjà, le secteur privé s'est approprié d'une masse d'information importante dans les secteurs de l'économie et des affaires. Il est permis de penser qu'éventuellement, des mini-centres d'information en RV iront chercher de l'information un peu partout sur le réseau et tenteront, eux-aussi, de se substituer aux bibliothèques traditionnelles. Il apparaît donc urgent que les bibliothèques actuelles réfléchissent sur le rôle de client-serveur qu'elles pourraient jouer et ce, avant que le secteur privé et une multitude d'utilisateurs ne le fassent à sa place.

Il se pourrait bien que dans un futur rapproché, les spécialistes en SI n'oeuvrent pas du côté des bibliothèques, mais plutôt dans le secteur privé ou parmi les internautes. La situation est alarmante pour les bibliothèques mais pas pour les spécialistes en sciences de l'information en général. Il suffira que ceux-ci prennent le virage, ce qu'ils peuvent faire indépendamment des bibliothèques. Il serait dommage cependant pour la société que les bibliothèques attendent ou agissent trop lentement, retardant ainsi l'accessibilité générale à des trésors d'information.

Un grand nombre de chercheurs, de professeurs et d'étudiants parcourent régulièrement les bases de

données des bibliothèques du monde entier sans avoir à se déplacer. La présence et la signification des bibliothèques traditionnelles vont donc s'estomper face au développement des nouvelles technologies, qui fournissent de l'information instantanément à des milliers d'utilisateurs. Les spécialistes en SI doivent donc être en mesure d'offrir le plus rapidement possible une gamme de services basés sur les besoins d'une clientèle qui pense en termes d'ordinateur et de réseaux et peut être, dans un avenir rapproché, de réalité virtuelle.

Les spécialistes de l'information doivent, dès maintenant, non seulement comprendre la technologie de la RV et s'informer, mais également s'impliquer.

Comme la majorité des travailleurs du secteur quaternaire, les spécialistes de l'information devront retourner à l'école de l'image pour apprendre à visualiser l'information et s'ils en ont le talent, participer au design des bibliothèques en RV en collaboration avec un spécialiste de l'infographie. Ils devront également chercher à comprendre la «culture» de cette technologie et à rejoindre une clientèle composée à la fois d'amateurs de Nintendo et de chercheurs provenant d'un grand nombre de disciplines.

3.4 L'avenir de la RV

L'usage généralisé de la réalité virtuelle dans les bibliothèques publiques et scolaires semble improbable à l'heure actuelle mais on sait à quel point les nouvelles technologies de l'information évoluent rapidement. Qui aurait pu prévoir, il y a cinq ans, que des millions de personnes à travers le monde communiqueraient entre eux par l'intermédiaire d'Internet?

Bien sûr, la technologie n'est pas encore mature et l'illusion, loin d'être parfaite, mais la RV avance à grand pas. D'innombrables applications seront la source de très gros profits si cette technologie continue sur cette envolée. Les outils d'interface grand public sont de plus en plus perfectionnés, moins chers et plus légers. Certains experts prédisaient que ce marché atteindrait 190 millions \$ en 1995.⁸

CONCLUSION

On ne peut contrôler le développement de la technologie: les ordinateurs doublent leur performance très rapidement et de nouveaux types de logiciels apparaissent régulièrement. Par ailleurs, les livres et les bibliothèques ne vont pas disparaître du jour au lendemain; une grande partie de la population ne possède aucune connaissance de l'informatique et adore le contact avec un livre ou une revue sur support papier.

Puisque, le monde de l'éducation s'implique de plus en plus dans les nouvelles technologies de l'information, les institutions traditionnelles risquent de prendre du retard face à une clientèle jeune et informatisée qui les désertera graduellement, non seulement par goût mais aussi pour des raisons d'efficacité en recherche d'information. À moins de changements majeurs et rapides, ces institutions se verront donc majoritairement fréquentées par une population plus âgée et souvent peu encline à s'adapter

aux nouvelles technologies.

Les effets de la technologie joints aux besoins changeants des sociétés en mouvement ont mené vers l'ère de l'information. Ces changements sont façonnés par un ensemble de forces : les intérêts économiques, les intérêts politiques et les technologies en compétition. Les effets combinés de ces forces sont difficiles, voire impossibles à prévoir. Les nouvelles technologies modifient nos comportements, nos modes de pensée et, malheureusement pour certains, nous font oublier les leçons du passé. C'est dans cet univers orienté "technologie" que le spécialiste en SI doit situer son plan de carrière. On peut penser que les personnes en mesure de bâtir des centres d'information en réalité virtuelle, de les remplir d'objets pertinents et qui, en plus, seront capables de trouver rapidement et efficacement de l'information précise dans une multitude de sites répartis dans le monde, ont un avenir prometteur.

En résumé, il nous semble peu probable que les projets en réalité virtuelle voient le jour dans les bibliothèques traditionnelles. Les spécialistes en SI intéressés par les nouvelles technologies qui incluent maintenant la RV, ont donc intérêt à penser que leur avenir ne se trouve pas dans le fonctionnariat, mais dans le secteur privé emplois plus fragiles, il est vrai, mais combien plus dynamiques..

NOTES

1. Logiciel REND386 <URL:<http://sune.uwaterloo.ca/pub/rend386/deve.5.zip>> Logiciel AVRIL <URL:<http://sune.uwaterloo.ca/pub/avril/avril.zip>> [Retour](#)
2. Alan Poulter. "Building a browsable virtual reality library." Aslib Proceedings 46, no 6 (June 1994): 151-155. Alan Poulter. "Towards a virtual reality library." Aslib Proceedings 45, no 1 (Jan. 1993): 11-17. [Retour](#)
3. Alan Poulter. "Towards a virtual reality library." Aslib Proceedings 45, no 1 (Jan. 1993): 14-15. [Retour](#)
4. Jean Segura. "Des mondes virtuels." Sciences et Avenir no 535 (sept. 1991): 31. [Retour](#)
5. Édouard Launet. "Les folies de la réalité virtuelle." Sciences & Vie no 909 (juin 1993): 82. [Retour](#)
6. Christophe Agnus. "Le monde en trompe l'oeil." L'Express no 2091 (8 août 1991): 36. [Retour](#)
7. Robin Hollands. "Essential garage peripherals." Virtual Reality World 3, no 1 (Jan.-Feb. 1994): 56. [Retour](#)
8. Michel Truchon. "Échos de Silicon Valley." Le Soleil (dim. 28 août 1994): C2. [Retour](#)

BIBLIOGRAPHIE

Articles de périodiques en anglais

- Blau, Brian et Kevin Yurica. "Designing VR applications for the desktop." *Virtual Reality World* (sept.-oct. 1994): 17-21.
- Gay, Eben. "Building virtual worlds on a school budget." *Virtual Reality World 2*, no 6 (nov.-déc. 1994): 45-47.
- Hollands, Robin. "Essential garage peripherals." *Virtual Reality World 3*, no 1 (janv.-févr. 1995): 56-57.
- Jacobson, Bob. "The many virtues of the virtual..." *Virtual Reality World 3*, no 1 (janv.-févr. 1995): 60.
- McLellan, Hilary. "Sharing the vision: interview with Dr. Thomas Furness." *Virtual Reality World 1*, no 2 (été 1993): p-t.
- Miller, Carmen. "Dr. Thomas A. Furness III, virtual reality pioneer." *Online 16*, no 6 (nov. 1992): 14-27.
- Stone, Bob et Andy Connell. "The UK's virtual reality and simulation initiative: one year later." *Virtual Reality World no 5* (sept.-oct. 1994): 50-58.

Articles de périodiques en français

- Agnus, Christophe. "Le monde en trompe l'oeil." *L'Express no 2091* (8 août 1991): 36-38.
- Dumais, Nelson. "La réalité virtuelle se nomme VAPS." *Informatique et Bureautique 14*, no 3 (mars 1993): 35-38.
- Launet, Édouard. "Les folies de la réalité virtuelle." *Science & Vie no 909* (juin 1993): 80-89.
- Schifres, Alain. "Une époque virtuellement formidable." *L'Express no 2176* (25 mars 1993): 18-21.
- Segura, Jean et Dominique Chouchan. "«Réalité virtuelle», un plongeon dans l'image?" *La Recherche 22*, no 229 (févr. 1991): 232-239.
- Segura, Jean. "Des mondes virtuels." *Sciences et Avenir no 535* (sept. 1991): 28-32.
- Simard, Anne-Marie. "Des mirages plein les yeux." *L'Actualité 18*, no 6 (15 avr. 1993): 77-78.

Stone, Robert J. "De la «réalité virtuelle»... à la réalité industrielle." *La Recherche* 24, no 251 (févr. 1993): 216-218.

Articles de quotidiens en français

Baritault, Alain. "La réalité virtuelle se met à la portée du grand public." *La Presse* (2 août 1994): C6.

Dutrisac, Robert. "Le Centre d'expérimentation et de développement en informatique multimédia." *Le Devoir* (27 août 1994): B1.

Truchon, Michel. "Échos de Silicon Valley." *Le Soleil* (28 août 1994): C2.

Articles de périodiques en bibliothéconomie

Chartier, Philippe. "La réalité virtuelle : voyage au-delà du gadget." *Documentation et bibliothèques* 41, no 1 (janv.-mars 1995): 39-46.

Emerson, Toni. "Mastering the art of VR : on becoming the HIT Lab cybrarian." *The Electronic Library* 11, no 6 (déc. 1993): 385-390.

Shaw, Debora. "Libraries of the future : glimpses of a networked, distributed, collaborative, hyper, virtual world." *Libri* 44, no 3 (1994): 206-223.

Poulter, Alan. "Building a browsable virtual reality library." *Aslib Proceedings* 46, no 6 (juin 1994): 151-155.

Poulter, Alan. "Towards a virtual reality library." *Aslib Proceedings* 45, no 1 (janv. 1993): 11-17.

Surprenant, Thomas. "Libraries, Information, and Virtual Reality." *Wilson Library Bulletin* (oct. 1991): 95-97, 136.

De l'accès sujet au principe de respect des fonds: pour une diffusion plus large de nos archives

par

Jean-François Hamel

Cursus vol.1 no 2 (printemps 1996)

Cursus est le périodique électronique étudiant de l'École de bibliothéconomie et des sciences de l'information (EBSI) de l'Université de Montréal. Ce nouveau périodique diffuse des textes produits dans le cadre des cours de l'EBSI.

ISSN 1201-7302

cursus@ere.umontreal.ca

URL:<http://www.fas.umontreal.ca/ebsi/cursus/>

De l'accès sujet au principe de respect des fonds: pour une diffusion plus large de nos archives a été écrit à l'hiver 1994 dans le cadre du cours *Recherche en archivistique (BLT 6621)* donné par le professeur Carol Couture.

Pour joindre l'auteur:hamelje@ere.umontreal.ca

Droits d'auteur

Tout texte demeure la propriété de son auteur. La reproduction de ce texte est permise pour une utilisation individuelle. Tout usage commercial nécessite une permission écrite de l'auteur.

Introduction

Nous convenons tous que la conservation et la diffusion des archives en vue de leur exploitation sont les deux finalités du travail de l'archiviste. Beaucoup d'efforts ont été mis dans le traitement et l'élaboration des instruments de recherche depuis plusieurs années. La normalisation de la description vient couronner ce travail. Cependant, jusqu'à ce jour, les archivistes se sont peu impliqués dans cet aspect de la diffusion qui consiste à jouer un rôle actif auprès des administrateurs et du grand public pour les sensibiliser à l'importance de conserver leur patrimoine archivistique.

Par ailleurs, nous retrouvons depuis la guerre une clientèle élargie et diversifiée de chercheurs (Duchemin, 1983, p.5), qui restent encore peu nombreux et le plus souvent restreints aux champs traditionnels de la recherche historique. Dans le contexte actuel de nos sociétés en crise où l'utilitarisme et les restrictions budgétaires prennent une place prédominante, nous devons nous demander si les archives justifient les dépenses qu'elles nécessitent quand elles servent si peu. L'archiviste d'aujourd'hui a-t-il encore un rôle à jouer, une mission à remplir?

Une plus grande sensibilisation de nos concitoyens au monde des archives aiderait, certes, le financement et le développement des services d'archives, nous évitant de déplorer d'autres fermetures comme celle du Petit séminaire de Québec.¹ Mais n'est-ce pas là la manifestation d'un problème de culture, d'un malaise au sein de notre société? Quoiqu'il en soit, il est urgent de trouver tous les moyens pour faire servir les archives davantage.

L'accès sujet pourrait être l'un de ceux-là qui rendrait à plus de chercheurs potentiels un accès plus facile au contenu des documents d'archives. Mais ce type d'accès n'est pas sans venir remettre en cause dans notre pratique l'opportunité d'appliquer le principe de respect des fonds, de revenir sans le vouloir, pour des raisons de commodité et d'efficacité, au classement par sujet tel qu'il se pratiquait avant Natalis de Wally.

L'informatique pourrait peut-être résoudre notre problème de l'accès aux archives tant au niveau des principes que de celui de l'efficacité pratique. Un fait semble devenir évident: l'informatique a autant besoin de l'archivistique pour authentifier les documents qu'elle crée, que nous avons besoin d'elle pour un traitement efficace et une diffusion plus large des archives. Les possibilités de la "recherche en ligne" sont à notre disposition pour donner accès au plus grand nombre. Mais peu de choses ont encore été faites de ce côté.

Il fallait sans doute passer par l'étape obligée de la normalisation² avant d'espérer mettre un jour en commun les informations contenues dans les archives que nous avons eu la tâche de conserver depuis

des siècles. Par contre, cet effort de normalisation, pour être cohérent, aurait dû se faire aussi en fonction d'une intervention à l'actif, en amont de la chaîne documentaire, parce que c'est dans les bureaux des producteurs de fonds que l'informatique impose déjà ses normes aux archives.

Ainsi, nous croyons que dans un contexte d'informatisation des archives, l'accès sujet devient un atout important qui peut ouvrir les portes des services d'archives et susciter l'intérêt de nouveaux chercheurs à travers une clientèle d'utilisateurs potentiels jusqu'ici difficiles à identifier. En prenant connaissance des besoins en information qu'exprimeront ces nouveaux chercheurs dans leurs requêtes, nous serions capables de trouver de nouvelles utilités aux archives et de les faire servir davantage.

Notre "réflexion archivistique" portera d'abord sur la façon dont l'accès sujet peut élargir notre clientèle malgré les obstacles et les difficultés à indexer les documents d'archives. Nous démontrerons ensuite que l'informatique devient une nécessité pour réaliser de nouveaux instruments de recherche à condition de pouvoir intervenir en amont, soit au stade actif de la création des documents. Enfin, nous ferons voir que l'accès sujet ne remet pas en cause le principe de respect des fonds et que celui-ci doit être le principe directeur qu'actualise la classification.

1 Une diffusion plus large par l'accès-sujet

1.1 La clientèle des archives

Les études de clientèles actuelles ne sauraient donner qu'un portrait du chercheur traditionnel, que l'archiviste connaît bien, sans pour autant définir toute la potentialité du produit lui-même, qui pourrait servir à d'autres types de chercheurs et répondre à la question: quel est le type d'archives qui intéresse les chercheurs? Une meilleure connaissance de notre produit pourrait sans doute donner de meilleurs résultats et préparer de nouvelles stratégies en vue d'attirer de nouveaux chercheurs.

Comme le rapportent plusieurs enquêtes soulignées par le *Groupe de travail canadien sur l'indexation par sujet*, les différents types d'archives attirent des clientèles différentes et "les utilisateurs d'archives forment un groupe de recherche bien diversifié" (BCA, 1992, p. 21). La plus forte proportion d'utilisateurs se recruterait du côté de l'organisme parrain, suivie du groupe des universitaires. Les archivistes eux-mêmes pourraient être considérés comme le groupe le plus important, tout comme les généalogistes à certains endroits. Les autres chercheurs proviendraient principalement des médias ou seraient des historiens amateurs.

La diversité des groupes de chercheurs entraîne également des besoins diversifiés en information ainsi que des questions différentes auxquelles doit répondre l'archiviste. De plus, seulement une minorité d'utilisateurs d'archives ont tout le temps voulu pour effectuer leurs recherches et étudier la pertinence de nombreux documents d'archives. Plusieurs chercheurs négligent de consulter des sources d'archives parce qu'ils ne peuvent pas compter sur un repérage efficace et rapide.

"Il apparaît manifeste, à la lumière des résultats obtenus par les diverses recherches... que l'approche traditionnelle du repérage ainsi que le fait que l'archiviste soit l'intermédiaire obligé entre l'utilisateur et les documents ne sont plus possibles. Dans l'ensemble, la consultation des archives augmente de façon considérable, surtout par des usagers des secteurs non traditionnels. Les archivistes doivent plus que jamais servir une clientèle hétérogène ayant des besoins et des attentes différents. Le temps que les utilisateurs et les archivistes doivent consacrer au repérage et à l'étude des documents se réduit. Les ressources en personnel sont trop limitées et le volume de documents trop grand pour que les archivistes continuent à jouer le rôle d'intermédiaire privilégié dans le repérage des documents" (BCA, 1992, p 26).

Il nous apparaît donc évident que de meilleurs instruments de recherche permettraient de récupérer des chercheurs traditionnels tout en rendant plus accessibles les archives à des clientèles potentielles non traditionnelles. Mais pour cela, il faut des systèmes de repérage qui pourront s'adapter à des utilisateurs différents en répondant à la variété de leurs besoins. Les instruments de recherche actuels ne peuvent vraiment satisfaire que les habitués des services d'archives. Dans leur forme actuelle, nous irions jusqu'à dire que ces instruments rebutent et qu'ils sont un frein à l'accès et à la diffusion.

1.2 Les avantages de l'accès sujet

L'accès par sujet, lié à l'utilisation de la "recherche informatique en ligne", pourrait permettre à un grand nombre d'individus qui ne fréquentent pas les services d'archives de se manifester et ainsi, par leurs questions, de nous faire connaître les nouvelles utilités qu'ils entendent trouver dans l'exploitation des archives. Nous aurions ainsi à la fois une meilleure connaissance des utilisateurs possibles de nos services et un regard neuf sur ce à quoi nos archives peuvent servir.

L'accès par sujet, sans remettre en question le classement par provenance des documents selon les activités et les actions spécifiques qui les ont fait naître, permettrait de tenir davantage compte des sujets abordés par les documents et de revaloriser ainsi l'accès à la valeur d'information. Il permettrait à un éventail plus grand de chercheurs de retrouver plus efficacement l'information qu'ils sont venus chercher tout en ayant la possibilité de situer cette information dans son contexte organique d'origine pour mieux en comprendre la genèse. Dans cette optique, l'accès sujet deviendrait un complément à l'accès par provenance, qui consiste à chercher de l'information sur un sujet donné en navigant au hasard dans les fonds et les séries susceptibles de contenir de l'information pertinente sur ce sujet.

1.3 Les obstacles à la diffusion et à l'accès

La difficulté d'une approche par sujet tient au fait que les instruments de recherche actuels sont bâtis en suivant le cadre de classification qui organise et structure les archives. C'est en effet par le cadre de classification spécifique au fonds, que l'archiviste rend effectif le principe de respect des fonds dans son volet interne. La description des archives dans les instruments de recherche procède selon le même principe allant du général au particulier, soit du fonds à la série - rarement jusqu'à la pièce -, tout en respectant la structure du plan de classification. Cependant, cette description se limite généralement à énumérer les titres et les dates des documents sans parler des sujets traités dans les archives. "Le

classement et la description des archives ne permettent donc pas le repérage des documents selon le sujet de recherche" (Hiller, 1987, p.13). C'est à l'archiviste de référence que revient la tâche de combler cette lacune par sa propre connaissance du fonds. Il devient en quelque sorte le "fichier sujet" du service.

Or cela n'est pas sans poser quelques problèmes qui sont reliés entre autres à la mobilité des archivistes. Cette situation de fait rend le chercheur dépendant de l'archiviste pour sa recherche et l'oblige presque à tout coup à venir le consulter en personne (Hiller, 1987, p.14-15).

Ainsi, il faut bien reconnaître que les instruments de recherche actuels ne sont peut-être pas des plus adéquats pour la recherche et qu'en dernier ressort, c'est l'archiviste qui devient "l'instrument ultime". Cette situation, qui accapare l'archiviste de référence et rebute le chercheur lui-même dont l'autonomie est restreinte, est, selon nous, un obstacle majeur à une plus grande diffusion des archives et à leur exploitation.

On aura beau argumenter, justifier l'importance et le rôle des archives auprès de l'administrateur, sensibiliser le grand public, il demeurera toujours que l'utilisateur des archives sera un individu spécifique ayant des besoins particuliers. Cette personne demeure le plus souvent issue de l'institution même ou du milieu académique. Les archives s'adressent avant tout aux chercheurs et ne sont qu'accidentellement des objets de curiosité, des objets de musée.

Heureusement, l'application des règles de description de documents d'archives (RDDA) permettra d'envisager des changements majeurs pour faciliter l'identification de sujets dans les zones de "portée" et "contenu". Mais à conditions de faire convenablement la description et l'indexation de ces zones. Il y a donc encore place à l'amélioration.

1.4 Les difficultés inhérentes à l'indexation

Si le repérage de l'information en archivistique s'est surtout développé sans le recours à l'accès sujet, cela tient à la nature même du document d'archives qui n'a pas été créé à l'origine comme un outil de documentation (Caya, 1982, p.32). Les archives sont créées pour des raisons administratives et non culturelles. Contrairement au livre, "le sujet d'un document d'archives ne peut être défini à partir de certaines caractéristiques externes comme le titre" (Langlois 1987, p. 39). Plusieurs documents n'ont d'ailleurs pas de titre et leur contenu peut également comprendre plusieurs sujets. De plus, leur structure interne varie d'un type de document à l'autre.

Ensuite, comme nous l'avons signalé, la description, le classement et les instruments de recherche en archivistique reposent sur le principe de respect des fonds, qui rend compte de la nature organique des archives, du lien qui les unit entre elles. Il n'est donc pas étonnant de constater qu'en matière d'indexation les archivistes manquent d'expertise puisque, toujours selon les mêmes principes, le repérage préconisé est conçu en fonction de la provenance, du créateur.

Pour qu'un système de repérage par sujet puisse être valable, il doit pouvoir prendre en considération ces

principes archivistiques. On ne peut pas considérer uniquement la valeur d'information des documents. Il faut tenir compte aussi de la valeur de témoignage des archives ainsi que de leur caractère historique qui évolue dans le temps avec l'organisme producteur lui-même (BCA, 1992, pp. 31-33).

Ainsi, le *Groupe de travail sur l'indexation par sujet* en arrive à la conclusion que l'indexation par sujet doit respecter l'aspect archivistique des documents, mais propose d'utiliser certaines méthodes propres à la bibliothéconomie tels les *fichiers d'autorité* pour les personnes physiques ou morales et les lieux. Il suggère de faire une distinction entre les entrées sujets et les entrées créateurs. De plus, "les sujets doivent être indexés en fonction des niveaux de description et renvoyés directement à ces derniers"³ (BCA, 1992, p. 50). Enfin, pour respecter le caractère des documents, le vocabulaire des termes historiques doit être contrôlé.

Les difficultés qui surgissent lors de l'indexation des archives et des instruments d'archives, militent en faveur d'une réforme en profondeur des instruments de recherche traditionnels. Leur normalisation, qui a été faite jusqu'ici, est un compromis dans la bonne direction.⁴ Cependant, les efforts ne touchent pas à ce qui se fait actuellement au stade actif. Ils ne portent que sur les "archives historiques".

Or l'indexation est un processus long et coûteux qui oblige l'archiviste à devenir un indexeur compétent. Dans ce contexte, l'indexation assistée par ordinateur ne permettrait-elle pas de trouver des solutions qui puissent faciliter la tâche de l'archiviste et s'appliquer dès la création des documents afin de mieux préparer les futurs versements? N'y aurait-il pas moyen d'envisager, avec les nouvelles technologies, un outil de repérage capable d'aller au-delà du contenant pour donner directement accès au contenu, à l'information du document?

2 L'Informatisation: une nécessité

Certains logiciels de repérage permettent déjà l'accès direct à la valeur de l'information des documents en autant que ceux-ci soient numérisés. La recherche plein texte à l'aide de procédures de recherche booléennes ou autres, en langage naturel ou contrôlé, fait de l'accès sujet aux documents une réalité qui peut donner l'impression à plusieurs qu'il n'est dès lors plus nécessaire d'établir une classification ou d'élaborer des descriptions. La machine à repérer se moque, en effet, du contexte de création.

Par ailleurs, s'il se crée des outils de repérage pour les documents au stade actif, ne conviendrait-il pas de les récupérer et de les utiliser lorsque les archives sont au stade de la conservation permanente? Dans l'éventualité où tous les documents ont été numérisés, qu'advient-il des instruments de recherche traditionnels?

2.1 Les instruments de recherche traditionnels

Jusqu'ici, la tendance s'exprime davantage par l'indexation du contenu des instruments de recherche plutôt qu'à l'indexation du contenu des documents comme le préconise aussi le Bureau canadien des

archives. D'ailleurs, il y a peu d'instruments qui décrivent à la pièce comme c'est le cas dans l'inventaire analytique. La description qui se fait du général au particulier, à partir du niveau du fonds, s'arrête le plus souvent au niveau du dossier, comme c'est le cas pour le répertoire numérique.

On affirmera alors que la description donne au chercheur des indices où trouver ce qu'il cherche et non pas si ce qu'il cherche est contenu dans les documents réunis dans un dossier. Le chercheur devra vérifier lui-même en passant chacune des pièces en revue.

Dans cette description, nous retrouvons davantage des informations sur le type de document (rapport, procès-verbaux, correspondance, etc.), leur chronologie ou la période couverte et d'autres renseignements par rapport à la provenance, au créateur ou à l'événement qui peut-être le "sujet principal" du dossier ou de la série.

Ainsi, au lieu d'avoir une indexation qui reflète le contenu d'une véritable condensation, nous retrouvons des clés d'accès typiquement archivistiques, tirées à même les titres - qui portent eux-mêmes peu de sens - et relatives à des personnes (index onomastiques), à des lieux, à des périodes et à des faits historiques (zone portée et contenu depuis les RDDA). Ce type d'indexation permet de repérer des informations sur le créateur ou la provenance; mais on peut se demander s'il aide véritablement tous les types de recherche.

Par contre, qu'en est-il des autres chercheurs, de ceux qui, de prime abord, ont de la difficulté à identifier et à exprimer leurs besoins en information sur un sujet donné? Ce type de chercheur pensera-t-il seulement qu'il peut consulter des archives dans sa quête d'information? A part certains instruments tel que le répertoire thématique, rien ne peut aider le chercheur non traditionnel, sinon l'archiviste en personne. Nous retrouvons cette clientèle de chercheurs principalement dans les bibliothèques, mais non dans les services d'archives.[5](#)

Afin que l'indexation des sujets contenus dans les documents soit précise, c'est-à-dire, sans créer trop de bruit ou de silence, elle doit être exécutée en profondeur soit jusqu'à la pièce. Ce travail représente évidemment de gros investissements en temps et en argent.[6](#) Quand on considère le peu de ressources dont disposent les archivistes, il serait étonnant que nous puissions produire de meilleurs résultats que ceux que nous avons aujourd'hui sans compter sur l'aide informatique.

2.2 L'apport de l'informatique

L'informatisation des archives peut aider l'archiviste à produire de nouveaux outils de recherche davantage adaptés ou adaptables aux besoins en information de tout genre d'usagers.

Pour y arriver, il faut dépasser le simple stade de la gestion automatique des documents pour nous orienter vers une gestion de l'information contenue dans le document: ce qui implique une conversion du format analogique au format numérique de la représentation des données contenues dans les documents

de toutes sortes Il faut donc dépasser la "gestion des stocks". Pour cela, nous devons envisager la numérisation des archives comme une solution avantageuse à plusieurs points de vue malgré les problèmes technologiques et le manque de ressources. Nous croyons d'ailleurs, que les problèmes pratiques causés par les technologies informatiques sont de faux problèmes, parce qu'ils ne remettent pas en question les fondements de l'archivistique⁷ d'une part et, d'autre part, parce qu'il pourra toujours y avoir une solution technique à un problème technique. Mais il est évident que toute nouvelle technologie a une incidence sur notre agir et qu'elle est un facteur de mutation de nos habitudes de penser par les possibilités qu'elle engendre. Il ne s'agit donc plus ici de se contenter d'informatiser les instruments de recherche dans leur forme actuelle. Il faut en créer d'autres.

Ainsi, la possibilité d'interroger et de traiter tout un document informatisé et stocké dans une base de données laisse entrevoir la possibilité d'un accès direct via un logiciel de recherche plein texte.⁸ Avec de larges masses de documents informatiques, des systèmes experts seraient bientôt en mesure de trouver les informations pertinentes aux usagers en conciliant les approches des traitements statistico-linguistiques et l'enquête cognitive pour reconstituer l'expertise des spécialistes en analyse documentaire.⁹

De fait, bien que les services d'archives n'aient encore bien souvent à gérer que les documents sur supports traditionnels, il est urgent de considérer dès l'actif l'ampleur que prend la création des documents informatiques, et d'envisager une conversion graduelle de nos archives vers la forme numérique. Le bureau sans papier (Dumais, 1989, p. 33)¹⁰ est en voie de devenir une réalité qui va convertir les services d'archives en un nouvel "archeion" de l'information organique consignée.

La reconnaissance optique des caractères (OCR) pour le texte (Goldwaser, 1991) et l'intelligence artificielle pour la recherche automatisée sont deux atouts qui doivent nous permettre de miser sur les nouvelles façons d'accéder rapidement à l'information en utilisant l'indexation assistée par ordinateur. Cette technique alléger grandement la tâche de l'indexeur ou de l'archiviste indexeur, si celui-ci doit indexer chaque document à conservation permanente.

Or, nous ne pouvons pas encore compter sur une recherche plein texte effectuée de façon autonome par le chercheur dans un document d'archives. Nous ne pouvons pas non plus compter sur l'archiviste pour indexer à la pièce chaque document. Nous nous retrouvons donc dans la position où il faut se contenter d'indexer la description de nos documents d'archives dans les instruments de recherche traditionnels. Mais de cette façon, l'archiviste risque de se cantonner dans une attitude passive puisque de plus en plus de documents, qui arriveront des créateurs, seront produits électroniquement et sous forme numérique.

Il n'y a pas de doute que l'importance des documents numérisés ira en s'accroissant, toutefois, les archivistes sont encore dépourvus quant à leur capacité à les traiter de façon conventionnelle. En outre, ces documents ordinolingués seront déjà normalisés en fonction des besoins des usagers à l'actif pour permettre entre autres l'échange de données (EDI), leur reproduction, leur stockage et leur repérage.¹¹ Cette normalisation sera organisée bien avant que l'archiviste ait complété l'élaboration de ses propres normes de description, et ce, sans tenir compte de la normalisation technologique en cours effectuée par

des informaticiens et des gestionnaires, lesquels demeurent peu sensibles aux préoccupations archivistiques.

Il ne fait pas de doute que, dans ce contexte informatique, les instruments de recherche élaborés à l'actif devront être récupérés par les dépôts d'archives, car comment, en effet, les archivistes responsables des archives définitives pourraient-ils se payer le luxe d'ignorer les outils qui donnent déjà accès à l'information?

Il apparaît donc important que les archivistes, en toute logique, s'impliquent dès la création des documents pour mieux préparer les futurs versements en structurant d'abord l'information pour les besoins du créateur même s'il ne s'agit pas toujours uniquement d'archives.¹² Les nouvelles technologies provoquent à notre insu des mutations au sein de notre profession qui devra s'adapter pour survivre à son nouvel environnement de travail.

Cela veut également dire que les archivistes devront tenir compte des avancées technologiques en ce qui a trait au stockage et au repérage de l'information sans pour autant négliger d'appliquer les principes archivistiques dans un univers documentaire informatisé.

Ainsi, pour pouvoir appliquer ces principes, il est important que le plan de classification puisse être élaboré dès l'actif tout comme le calendrier de conservation et que les deux outils puissent faire l'objet d'un programme informatique intégré à un système capable de gérer et stocker tous les documents produits par l'organisme dès l'instant de leur création, jusqu'au moment de leur disposition finale en vue de leur destruction ou de leur conservation permanente.¹³

3 L'intervention de l'archiviste en amont

3.1 L'application du principe de respect des fonds à l'actif

Devant la facilité de l'accès à l'information offert par un système automatisé, il serait tentant de faire l'économie du principe de respect des fonds pour se concentrer uniquement sur le repérage. De la sorte, l'accès par sujet viendrait remettre en question la pertinence du principe.

"Désormais, la nécessité de structurer méthodiquement de vastes ensembles d'archives échappe à l'archiviste. L'ordinateur ayant saisi le vocabulaire d'indexation porté sur les bordereaux des versements successifs effectuera lui-même la besogne. Il rendra possible les investigations méthodiques et systématiques demandées autrefois aux inventaires écrits (Collin, 1990, p. 7).

De plus, pour nombre de chercheurs non traditionnels, la mise en contexte n'est pas une nécessité absolue. Les photos, par exemple, n'ont pas besoin de mise en contexte quand elles ne servent qu'à illustrer. L'image parle d'elle-même, et les archives photographiques en sont réduites à une matière à l'état brut.

Cependant, nous sommes enclin à croire que les principes archivistiques ont encore leur place et qu'ils sont une nécessité dont nous ne pouvons pas faire l'économie sans en subir les inconvénients, bien que cela n'apparaisse pas évident de prime abord pour le producteur du fonds.

En effet, si le principe de respect des fonds découle de la nature même des archives, s'il s'enracine véritablement dans la réalité, il doit trouver quelque part sa justification. Et cela, non seulement en regard des archives définitives, mais aussi des archives courantes. Une meilleure connaissance du principe devrait entraîner des conséquences positives sur l'intervention archivistique, en particulier en amont du cycle de vie des documents.

L'application du principe de respect des fonds devrait permettre en premier lieu de garantir l'intégrité des documents pour leur valeur de témoignage en ce qui concerne l'utilisation quotidienne de l'information qui sert d'abord l'administrateur. Lors du Congrès international des archives tenu à Montréal en 1992, Charles Dollar soulignait que:

"New information technologies have not changed and nor are they likely to change the fundamental nature of records as evidence of action and transactions with a specific context of creation and use. It is the record, not information in general, that is the domain of archives. It was refreshing to hear Liv Mykland's reaffirmation of the centrality of the concept of provenance which - as several speakers also noted - *is the basis for assessing the trustworthiness and reliability of electronic records*. As the volume of material in electronic form grows, users increasingly will require that the information handling service community provide them with the robust indicators of trustworthiness, reliability, and validity that only archival theory offers" (Dollar, 1992, p.2).

La valeur du principe de respect des fonds tient en ce qu'il permet d'organiser et de traiter, de façon cohérente et conséquente, tous les documents à l'actif. L'accès par sujet n'enlève rien à la nécessité de classer les documents pour mieux les retrouver en particulier sur un "disque dur", et surtout, pour mieux en identifier la provenance et le créateur lui-même.

Il importe de rappeler que les archives sont créées pour servir en premier lieu le créateur. L'information organique et consignée qu'il vient chercher doit aussi demeurer dans son contexte de création pour des raisons évidentes de probation, de valeur juridique et légale. A priori, il ne peut pas, lui non plus, se contenter d'une information dissociée de son contexte. Pour avoir un sens et s'intégrer à la prise de décision, l'information est nécessairement liée à un dossier, à une affaire (Laroche, 1971, p. 26).

Mais le danger existe bien de penser que parce que nous pouvons avoir accès de plus en plus facilement, grâce à l'informatique, au contenu de chaque document pris individuellement, il n'est plus nécessaire d'organiser cette information et en particulier de l'organiser selon les principes archivistiques reconnus.

3.2 La mise en oeuvre du principe de respect des fonds par la

classification

La classification des documents, si elle veut vraiment servir l'administration ou le créateur tout comme le chercheur, doit se faire selon les principes archivistiques.

Si le document d'archives est un produit organique issu des activités d'un organisme, cela implique qu'il n'est pas isolé mais intégré dans un contexte documentaire (ou informationnel). Celui-ci comprend la circulation du flot continu de plusieurs informations interreliées dans le but d'informer et de prendre des décisions, de leur donner suite, tout ceci dans le cadre d'une action spécifique elle-même comprise dans une activité plus globale issue d'un mandat relatif à une situation donnée (Laroche, 1971, p. 26). L'action ou l'affaire est le principe qui sous-tend la création des archives. Il doit en être l'élément unificateur duquel découle a posteriori le principe de respect des fonds, et cela, à tous les niveaux à partir du document lui-même jusqu'au fonds.

Le classement à l'actif doit donc refléter cette nécessité archivistique parce que pour les besoins de l'administration, il doit tenir compte du fait que les documents sont conservés ensemble dans le contexte qui les a vu naître. Ceci afin de pouvoir produire la preuve et pour ajouter à la valeur de témoignage qu'ils possèdent non pas isolément, mais en relation avec les autres documents de la chaîne décisionnelle. Il s'agit de conserver intact le processus décisionnel. C'est pourquoi les activités ou les fonctions d'un organisme doivent être considérées comme la base de la division de la classification parce qu'elles découlent de l'action qui est le principe structurant qui se situe à la genèse de tout fonds d'archives (Laroche, 1971, p.15).

Donc on peut dire que, dès l'actif, pour des raisons administratives, légales et financières, les archives ne peuvent pas être réduites à leur seule valeur "primaire" d'information et qu'elles doivent constituer un tout organique interrelié qui forme une structure, une forme.

Si l'accès sujet peut aider à retrouver l'information, il doit dépasser la simple information primaire. Il ne doit pas empêcher l'accès par la provenance à l'information contextuelle.

Évidemment, l'utilisateur ou le créateur du fonds n'a pas à y référer à ce contexte parce qu'il le connaît et le vit. Ce n'est pas le cas du chercheur dans un service d'archives.

Ainsi, pour faciliter le repérage de l'information à l'administrateur, on élabore des guides de classification qui permettent de produire des index dont les mots-clés sont dégagés à même le titre des séries, des dossiers et de d'autres descripteurs qui sont puisés dans le contenu de ces séries et de ces dossiers qui portent sur le sujet dont il est question en fonction du degré de profondeur que l'on veut atteindre. Il convient donc d'uniformiser les méthodes de repérage à l'actif et en particulier les règles d'indexation qui doivent avant tout tenir compte de la clientèle institutionnelle.[14](#)

Cependant, si cette "indexation archivistique" pouvait se faire à l'actif pour répondre aux besoins du créateur, serait-elle encore capable de répondre aux besoins des chercheurs de tout horizon par la suite?

De plus, même si le plan de classification est la pierre angulaire où s'articule le principe de respect des fonds dans son volet interne, qu'advient-il de la mise en contexte du fonds telle que produite par l'archiviste aux différents niveaux de la description dans les instruments de recherche traditionnels conçus pour les archives définitives?

Cette information sur le créateur et le fonds reste toujours primordiale pour le chercheur historien et pour les autres car la mise en contexte permet de recréer l'action ou l'événement qui a donné naissance à l'information et lui donne son sens.

Cela veut dire qu'il faudrait que l'archiviste vienne compléter l'instrument de recherche élaboré à l'actif par des descriptions contextuelles pour satisfaire ainsi les autres utilisateurs des archives quand celles-ci sont rendues au stade de leur conservation permanente.

Ainsi, même si l'informatique nous permet d'envisager une approche à la fois par sujet et par provenance pour que cette information ait un sens et que les documents aient une valeur ajoutée de preuve et de témoignage, il faut faire en sorte que l'ordinateur ne vienne pas briser le lien organique qui doit unir les informations entre elles.

"Archivist must ensure that archives and archival functions are not submerged in a sea of electronic information and thereby lose their distinctive identity. We can do this by demonstrating in our interaction with other participants in the information handling-service arena that archival science is the only discipline that attaches significance to the contextual relations of records and provides tools for an understanding of them. We must continue to emphasize how vital the maintenance of provenance is for all users of electronic information " (Dollar, 1992, p. 3).

3.3 Pour un système informatique intégré

En fait, tout tient dans la structure que l'on doit donner à l'information et aux bases de données.

"Mais ce qui ne souffre plus de doute, c'est la nécessité de maintenir un principe régulateur, qui soit de nature proprement archivistique. Quant tout paraîtra se volatiliser, quand le lien de ficelles qui enserre nos papiers enliassés se défera, il n'en sera que plus nécessaire de renforcer idéalement le lien structurel, même s'il faut pour cela demander à l'ordinateur de le mettre en évidence. Pourquoi non? La recherche d'une structure est un programme comme un autre" (Laroche, 1971, p. 48).

Le plan de classification est donc le premier instrument de recherche que doit récupérer ou créer l'archiviste à partir de la masse documentaire. Mais sa tâche doit aller plus loin. Il doit décrire les liens qui existent entre les documents et les séries, établir le contexte de la création, situer le fonds lui-même par rapport aux autres et expliquer la place qu'il occupe dans l'histoire d'une société. Ces descriptions apparaissant dans des fichiers séparés doivent être ajoutées à l'instrument de recherche de base qu'est le plan de classification qui sera complété par l'index sujet qui doit l'accompagner.

Dans un contexte informatique, cela pourrait donner naissance à une base de données relationnelle de type hypertexte ou hypermédia. Le niveau de description devrait à ce moment-là s'arrêter à la série et à la sous-série. Pour le dossier, l'indexation du titre et du contenu des pièces devrait pouvoir se faire avec l'assistance de l'ordinateur. Quant au chercheur, il pourrait toujours se servir de l'index ou du thésaurus ainsi généré ou faire appel à un logiciel d'investigation plein texte. Les liens électroniques entre les différents fichiers permettraient de préserver alors le lien organique entre les documents.

Tout en garantissant le contexte de création du document par l'application du principe de respect des fonds actualisé dans le plan de classification à l'actif, l'ordinateur deviendrait l'outil puissant par lequel pourraient s'établir encore plus sûrement les liens organiques entre les documents; mais surtout, il deviendrait un instrument de recherche qui permettrait enfin de renouer avec les avantages de l'accès par sujet.

Conclusion

L'accès sujet permet d'enjamber la paroi du fonds ou du contenant tout comme celle des séries. Pour le chercheur qui veut avoir de l'information sur un sujet donné, le contexte n'a pas toujours d'importance et le respect du fonds lui importe peu. Il semblerait donc que l'accès sujet remette en cause l'utilité du respect des fonds. Mais comme nous l'avons vu plus haut, le principe de respect des fonds est une nécessité pour le classement parce qu'il est seul à prendre en compte la nature même des documents d'archives et leur structuration. Dans un contexte documentaire informatisé, il doit être le principe régulateur qui permet de classer et de gérer toutes les archives peu importe leur support, afin d'en assurer l'authenticité et la provenance.

L'informatique, en retour, offre des possibilités d'exploitation des archives encore insoupçonnées qui peuvent libérer l'archiviste de tâches ingrates et ouvrir les fonds à de nouvelles clientèles de chercheurs. L'ordinateur, couplé à la "recherche en ligne" devient un moyen puissant de diffusion des archives par les archives que l'archiviste ne peut plus se permettre de négliger. L'informatique redonne un sens au rôle que doit jouer l'archiviste dans l'organisation de l'information organique consignée en autant que celui-ci prenne conscience de l'importance de son expertise dans le cadre du virage technologique.

Ainsi, l'archiviste ne doit plus se contenter d'attendre s'il veut encore servir à quelque chose demain. Il doit, s'il veut toujours agir comme conservateur du patrimoine archivistique et gardien de la mémoire collective, s'impliquer dès l'actif afin de normaliser la forme que prendront les archives de demain.

D'imposants défis se dressent sur son chemin.

Notes

1 Cette fermeture est survenue à l'automne 1992 et a suscité de vives réactions de la part des milieux

culturels, de la recherche et des archives au Québec.

2 L'opération de normalisation contribuera certainement à clarifier, à conforter ou encore à développer les notions servant d'assises à la discipline archivistique. Les retombées concrètes seront nombreuses et ce, peu importe que l'on utilise les ressources informatiques ou non. Pourtant c'est l'utilisation de l'informatique à des fins de gestion des archives ou encore la gestion des archives informatiques qui milite le plus en ce sens" (Rousseau, 1992, p. 146).

3 Il faut souligner ici que souvent, dans la pratique, les archivistes ne respectent pas ce principe qui oblige à faire la description sans sauter de niveau, en passant du général au particulier. Nous l'avons nous-même expérimenté. A ce sujet, il est intéressant de noter que Hugo Stibbe, dans la méthode qu'il préconise pour décrire les notices d'autorité, semble également aller à l'encontre de ce principe. Il faut dire que rarement un fonds qui n'est pas clos et complet peut être décrit d'abord dans sa globalité."La procédure suggérée peut sembler contredire le principe qui veut que l'on décrive en allant du général au spécifique. Dans les faits, pourtant, la procédure respecte le principe, en ce sens que nous décrivons l'ensemble tel que nous le percevons d'après les documents dont nous disposons au moment de la préparation de la description. Ce principe doit être appliqué et vu sous cet angle." (Stibbe, 1993, p.79)

4 Selon Langlois, "l'accès sujet en archivistique ne pourra être possible sans une réforme en profondeur des instruments de recherche. Le manque de ressources humaines et matérielles interdit aux archivistes de recourir à des solutions extrêmes, si nécessaires soient-elles." (Langlois, 1987, p.47). Il propose d'opter pour des réformes progressives. Ces réformes devraient porter, selon lui, sur l'établissement de consensus sur les méthodes de traitement. Nous croyons que l'effort de normalisation actuel s'inscrit dans cette optique, mais qu'elle devrait englober également le traitement à l'actif.

5 L'interrogation de bases de données archivistiques à partir de terminaux situés dans les bibliothèques permettrait à ces chercheurs d'avoir accès aux ressources des services d'archives qu'ils ne fréquentent jamais. Mais pour cela, il faut que les archives puissent être interrogées de la même façon que pour toute autre documentation: par sujet.

6 Une telle indexation, si elle n'est pas conduite par un expert, risque de produire beaucoup de bruit, sans pour autant éliminer tous les silences, et ainsi mettre les chercheurs sur de fausses pistes.

7 "The question... is whether modern archival really requires extensive revision before it can be applied fully to computer records. If archivists take a close look at what the theoretical principles behind archival practices actually entail, the answer must be "no". The appraisal process as it applies to electronic records shows that, despite the different media, there is no fundamental difference in evaluating paper or computerized information; both are assessed according to their administrative, legal, evidential, and informational values within the context of the record-keeping system in which they were created. In terms of the arrangement of electronic records, the theoretical principles merely need to be reformulated, not redefined, to reflect a more conceptual, less physical orientation... a good system of archival description should focus on the context of the information itself, not its physical form" (Bailey,

1990, p. 194).

8 À titre d'exemple de logiciels de recherche plein texte, mentionnons entre autres le logiciel "Naturel" relativement facile d'utilisation, et le logiciel SATO, qui est un système d'analyse et de gestion de données textuelles. "Il permet l'annotation de documents multilingues, le repérage "sur mesure" des éléments du texte et l'analyse qualitative ou quantitative du document ou de ses parties." (Centre d'ATO, 1992, p.4)

9 "Une équipe du centre ATO.CI travaille à la conception d'un prototype de système expert pour l'"Aide à l'analyse documentaire (Tri, classification, indexation) des jugements, de façon à faciliter la tâche des conseillers juridiques de la Société québécoise d'information juridique (SOQUIJ)" (Bertrand-Gastaldy, 1992, p.11).

10 Au sujet du bureau sans papier, voir aussi les auteurs suivants: Beaver (1993, p.215), Symonds (1992, p.88) et Tinkel (1993, p. 32).

11 Concernant l'échange de données informatiques, voir Neumeister (1992, pp. 29-32).

12 En effet, l'archiviste qui s'implique dès la création des documents et qui travaille à la gestion des documents informatiques, ne sera pas étonné si on lui confie la tâche de gérer aussi les autres types de documentation non organique. Pour l'administrateur, seule compte l'information dont il a besoin pour sa prise de décision. Le rôle de l'archiviste doit donc être polyvalent s'il veut bien s'intégrer à l'entreprise. Il doit agir avant tout comme un spécialiste de l'information et répondre aux besoins de tous les usagers bien que ce travail puisse être partagé entre différentes spécialisations.

13 Sans vouloir entrer dans la science-fiction, un tel système - comprenant essentiellement une base de données - formerait en quelque sorte le fonds lui-même de l'institution. Celle-ci pourrait gérer de la même façon les fonds privés qu'elle aurait acquis en cohabitation permanente avec ses propres archives courantes, intermédiaires et définitives en autant qu'un tel système ait la capacité suffisante de tout conserver ensemble et de permettre un accès rapide. L'utilisation de "juke boxes" et de disques optiques numériques pourrait être associée à ce système de base de données.

14 Évidemment, l'un des problèmes ou obstacles est que le personnel affecté à cette tâche n'a pas toujours la formation requise. D'où, encore une fois, la nécessité de l'intervention de l'archiviste à l'actif.

Bibliographie

Bailey, Catherine. "Archival theory and Electronic Records". *Archivaria* 29 (Winter 1989-90):180-196.

Beaver, David. "Pushing beyond paper." *MacUser* 9, no 1 (Jan. 1993): 215(6).

Bertrand-Gastaldy, Daoust, François et al. *Les traitements statistico-linguistiques et l'enquête cognitive comme moyens de reconstituer l'expertise des spécialistes en analyse documentaire: le cas de la jurisprudence*. (Cahier de recherche no 2). Montréal: Centre de Recherche en Cognition et Information (ATO-CI), Université du Québec à Montréal, 1992.

Bureau canadien des archivistes (BCA). Groupe de travail sur l'indexation par sujet. *L'indexation par sujet en archivistique: rapport du Groupe de travail sur l'indexation par sujet*. Ottawa: BCA, Comité de planification sur les normes de description, (publ. no. 4), 1992.

Caya, M. "La diffusion par les instruments de recherche en archivistique", *Archives* 14 , no 1 (juin 1982): 31-50.

Centre d'analyse de textes par ordinateur (Centre d'ATO). *Fiche descriptive du logiciel SATO: version 3.6.* Montréal: Université du Québec à Montréal, 1992.

Collin, Hubert. "La mise en oeuvre du principe de respect des fonds. Réflexion sur l'état actuel." *Archives* 22 , no 2 (automne 1990): 3-8.

Dollar, Charles M. *Seizing the opportunity: archivists in the information age.* Comments Delivered at the Fourth Plenary Session XII International Congress of Archives. Montreal:1992.

Dollar, Charles. "Archivists and Records Managers in the Information Age." *Archivaria* 36 (Autumn 1993): 37-52.

Duchemin, Michel. *Les obstacles à l'accès, à l'utilisation et au transfert de l'information contenue dans les archives: une étude RAMP*. Paris: UNESCO, 1983.

Dumais, Nelson. "Remplacer la paperasse par l'archivage électronique, c'est rentable". *Informatique & bureautique*. Montréal: (décembre 1989): 32-37.

Goldwasser, Daniel. "Lecture automatique: la lecture optique sur le chemin de l'automatisme". *Documentaliste-Sciences de l'information*, 28, no 3 (1991)

Hiller, Marc. "L'archiviste de référence: instrument de recherche ultime?". In *Réflexions archivistiques*. Sous la direction de Carol Couture. Montréal: Université de Montréal, École de bibliothéconomie et des sciences de l'information (1987):11-23.

Langlois, Égide. "L'accès-sujet en archivistique: vers une réforme en profondeur des instruments de recherche?". In *Réflexions archivistiques*,. Sous la direction de Carol Couture. Montréal: Université de Montréal, École de Bibliothéconomie et des sciences de l'Information (1987): 37-52.

Laroche, Carlo. "Que signifie le respect des fonds? Esquisse d'une archivistique structurale."

Supplément au no 73 de la *Gazette des archives* . Paris, (1971): 57.

Neumeister, Bernard. "L'EDI et la GED: La synergie." *Archimag* (1992): 56, 29-32.

Rousseau, J.-Y. "La normalisation et l'informatisation des archives". In *La normalisation en archivistique: un pas de plus dans l'évolution d'une discipline: mélanges Jacques Ducharme*. Sous la direction de Carol Couture, avec la collaboration de Frédérick Brochu. Québec, Documentor; Association des archivistes du Québec, 1992.

Stibbe, Hugo. "L'application de la notion de fonds." *Archives* 25, no1 (été 1993): 51-83.

Symonds, W.C. "Getting rid of paper is just the beginning; going from forms to electronic data means reinventing business." *Business Week* 21, no 298 (Dec. 1992): 88(2).

Définition et application de critères d'évaluation d'outils de recherche dans Internet

par

Jean-Noël Plourde

Cursus est le périodique électronique étudiant de l'École de bibliothéconomie et des sciences de l'information (EBSI) de l'Université de Montréal. Ce nouveau périodique diffuse des textes produits dans le cadre des cours de l'EBSI.

ISSN 1201-7302

cursus@ere.umontreal.ca

URL: <http://www.fas.umontreal.ca/ebsi/cursus/>

L'auteur

En cette seconde et dernière année de sa maîtrise, M. Plourde est inscrit dans le profil de recherche. Ses intérêts de recherche portent, d'une part, sur les documents structurés et, d'autre part, sur la diffusion et le repérage de l'information dans Internet. Il a un baccalauréat en mathématiques. Il a également suivi des cours en informatique (gestion et intelligence artificielle) et en histoire de l'art.

Son expérience de travail se situe principalement dans le domaine de l'informatique comme analyste. Depuis son retour aux études, il donne des cours en micro-informatique à temps partiel. Auparavant, il a travaillé deux années en France à titre de consultant en base de données dans le milieu bancaire. Il a commencé sa carrière en informatique à la Société Radio-Canada où il a occupé différentes fonctions pendant une dizaine d'années et oeuvré principalement dans les systèmes d'information.

Critère et évaluation d'outils de recherche des ressources dans Internet a été écrit à l'automne 1995 dans le cadre du cours *Structures et fonctions des logiciels documentaires (BLT 6351)* donné par le professeur Yves Marcoux.

Pour joindre l'auteur: jn.plourde@sympatico.ca

Droits d'auteur

Tout texte demeure la propriété de son auteur. La reproduction de ce texte est permise pour utilisation individuelle. Tout usage commercial nécessite une permission écrite de l'auteur.

INTRODUCTION

Il est d'ores et déjà reconnu que le repérage d'information dans Internet constitue une tâche difficile. Plusieurs outils, spécialisés dans le repérage d'information, tels Lycos ou Yahoo, ont été développés afin de répondre à ce besoin pressant. Est-ce que tous ces sites offrent les mêmes services? Vous l'avez deviné, bien sûr, la réponse est non: chacun possède ses particularités.

Dans la première partie de cette étude, nous établissons des critères d'évaluation qui nous permettront de mieux comprendre les forces, les faiblesses, les similitudes et les différences entre ces outils de recherche et d'orienter le choix d'un outil de recherche. Nous suggérons que ce choix soit fonction, d'une part, des services offerts et, d'autre part, des besoins de l'utilisateur. Cette première partie se divise en quatre sections: la première aborde la typologie des outils de recherche, la seconde porte sur les services de mise à jour, la troisième traite des services de repérage et la dernière étudie les services de présentation des résultats.

Dans la seconde partie, nous évaluons treize outils de recherche en nous basant sur les critères suggérés.

Fondamentalement, les critères d'évaluation d'outils de recherche sont similaires à ceux utilisés lors de l'évaluation d'un ouvrage de référence électronique. Ces critères sont: l'objectif du site, la crédibilité de l'organisation responsable, la couverture, l'objectivité, l'auditoire visé, l'organisation, la fréquence de mise à jour, le coût, l'originalité, le format de présentation, la convivialité, l'historique des recherches, les menus d'aide, l'utilisation d'opérateurs booléens, la troncature, et c. Nous ne passerons pas en revue l'ensemble de ces critères mais nous nous attarderons spécifiquement à ceux qui sont liés aux services de mise à jour, de repérage et de présentation des résultats que nous expliciterons plus loin.

Avant de passer au vif du sujet, il est nécessaire d'apporter quelques précisions. Dans la littérature, les sites spécialisés en repérage d'information sont désignés sous plusieurs dénominations. Parmi celles-ci on retrouve: robots (*robots*, *webcrawlers*), service de repérage (*search service*) et moteurs de recherche (*search engines*). De notre point de vue, cette terminologie porte à confusion. En effet, les robots constituent un service en amont dont le but est de recueillir les données afin de mettre à jour une base de données W3. Le service de repérage est, quant à lui, un service en aval qui correspond naturellement à l'interface utilisateur pour effectuer la recherche. Enfin, les moteurs de recherche offrent des services

pour tirer profit des données et établir le pont entre la base de données W3 et l'interface utilisateur.

Dans cette étude, nous utiliserons l'expression «outils de recherche» telle que définie dans le Vocabulaire d'Internet préparé par Marcel Bergeron et Corinne Kempa, terminologies à la Direction des services linguistiques de l'Office de la langue française: programme qui indexe le contenu de différentes ressources d'Internet, et plus particulièrement de sites W3, qui permet à l'internaute de rechercher de l'information à partir d'un navigateur W3, selon différents paramètres, en se servant de mots clés, et qui permet d'avoir accès à l'information ainsi trouvée.

De plus, dans le cadre de cette étude, nous ferons une distinction sémantique entre "site Web" et "base de données W3". Un "site Web" sera un site qui diffuse de l'information dite primaire (telles les pages d'accueil personnelles) et une "base de données W3" sera un site qui diffuse de l'information dite secondaire (tels Lycos ou Yahoo).

Vous trouverez une bibliographie à la fin de ce travail. Celle-ci est également disponible en ligne à l'URL suivant:

<http://tornado.ere.umontreal.ca/~plourdej/moteurs.html>

Nous y avons noté les références. De plus, à cet URL, d'autres références sur l'évaluation d'outils de recherche sont disponibles.

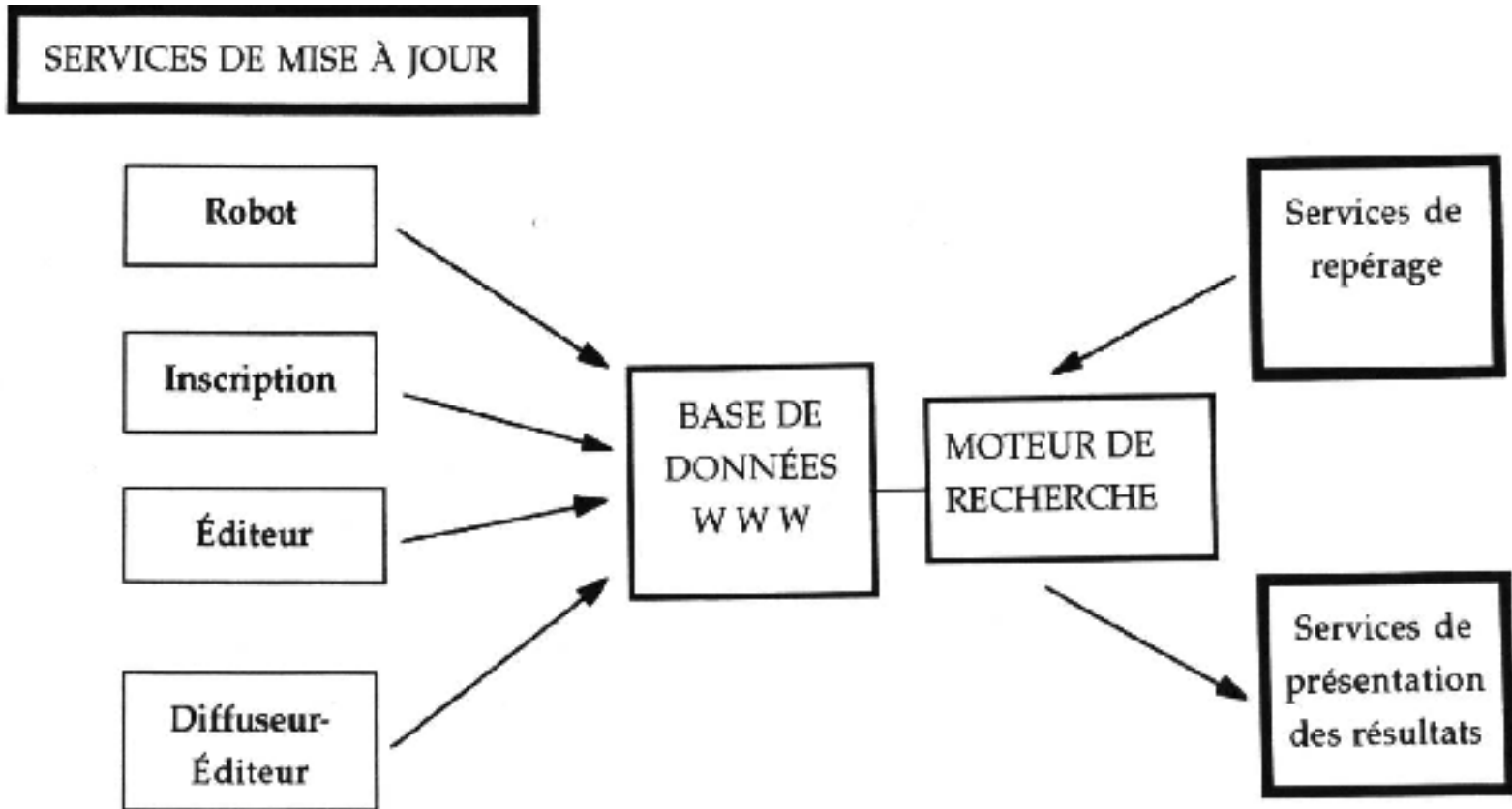
Cette étude a eu lieu pendant les mois de septembre à décembre 1995. Une mise en garde s'impose: compte tenu de l'évolution rapide des outils de recherche, il est fort probable que les fonctionnalités offertes par ces outils auront évolué et que certaines de nos remarques ne s'appliqueront plus dans les mois à venir.

1.TYPOLOGIE DES OUTILS DE RECHERCHE

Un outil de recherche se compose d'une partie technique et d'une partie fonctionnelle. La partie technique touche plus spécifiquement les éléments suivants: le type de serveur (PC, Mac, Sun), le système d'exploitation (UNIX, Microsoft Windows), le nombre de serveurs (sites miroirs), le type de base de données (fichiers plats, bases de données relationnelles ou orientées objets), le langage de programmation (C++, Pascal), le langage d'interprétation (Perl, Tcl), le moteur de recherche (WAIS, GLIMPSE, PURSUIT, etc.) et sûrement d'autres éléments. Bien que très intéressante, cette composante ne fera pas l'objet de notre étude.

La partie fonctionnelle porte plus spécifiquement sur les services associés à l'exploitation d'une base de données W3. Nous porterons une attention particulière aux services de mise à jour des données, à ceux de repérage et à ceux de présentation des résultats. Les services de mise à jour ont pour mission de colliger les données du Web afin de constituer une base de données W3. Les services de repérage ont, quant à eux, la mission d'offrir une interface utilisateur doublée d'un moteur de recherche pour l'exploitation de ces

données. Finalement, les services de présentation des résultats correspondent tout simplement aux fonctions offertes pour l'exploitation des données repêchées. La figure suivante illustre ces différents services.



Les outils de recherche peuvent être classifiés selon plusieurs types. Pour n'en citer que quelques-uns, il y a ceux qui produisent des index (Lycos), des répertoires (Yahoo), des métarépertoires (URL Square, métarépertoire au Japon) et des métaindex (All-in-One). Dans le cadre de cette étude, nous nous limiterons aux outils de recherche de type index et répertoire.

1.1. Type index

Tel que son nom l'indique, ce type est associé à tous les outils de recherche dont les fonctions de repérage reposent principalement sur l'utilisation d'index. Nous y retrouvons, entre autres, des sites tels: ALIWEB, CUI World Wide Web Catalog, InfoSeek, Lycos, Open Text Web Index, Web Crawler, WWW Home Pages Harvest Broker et WWW Worm.

Un des avantages de ce type réside dans le fait que l'utilisateur n'a pas à connaître la catégorie (et la structure hiérarchique) dans laquelle pourrait exister l'information recherchée. L'information n'est pas compartimentée en catégories. La recherche s'opère principalement par concordance avec un modèle (*pattern matching*). Cette approche peut entraîner de bons taux de rappel mais aussi beaucoup de bruit, c'est-à-dire une baisse du taux de précision. Fondamentalement, ce type se rapproche de celui utilisé pour la recherche par sujet dans les OPACs traditionnels. Cependant, contrairement à ces derniers, il n'y a pas de vocabulaire contrôlé. Notons également que, généralement, le type index est basé sur des techniques automatisées pour répertorier l'information produite dans le Web.

1.2. Type répertoire

Ce type est associé à tous les outils de recherche dont les fonctions de repérage reposent principalement sur une classification afin d'organiser l'information selon une propriété (thématique, chronologique, etc.). En général, c'est la classification de type thématique qui est privilégiée. Dans ce type, nous retrouvons, entre autres, les sites tels: Einet Galaxy, GNN's Whole Internet Catalog, The World-Wide Web Virtual Library, Yahoo, The Clearinghouse for Subject-Oriented Internet Resource Guides [U n. de Mich.].

Lorsque l'utilisateur connaît le domaine de sa recherche et la structure hiérarchique liée à un outil de recherche, le type répertoire peut s'avérer très intéressant. En catégorisant l'information, il est facile de butiner d'un document à l'autre portant sur un même sujet. Cette approche permet de réduire le taux de bruit mais aussi d'augmenter le taux du silence (en supposant qu'un document est classifié dans une seule catégorie). Fondamentalement, ce type est analogue au plan de classification que l'on retrouve dans les bibliothèques traditionnelles (mais en beaucoup moins élaboré). Notons également que, généralement, le type répertoire est basé sur des techniques manuelles pour répertorier l'information produite dans le Web. En général, il y a une valeur ajoutée. Bien que certains de ces outils offrent la recherche à l'aide d'index, celle-ci est beaucoup plus limitée que dans les outils de type index.

2. LES SERVICES DE MISE A JOUR

La pierre d'assise de tout système d'information est sans contredit la base de données sur laquelle repose celui-ci. Les éléments utilisés pour constituer les bases de données W3 varient énormément d'un outil de recherche à l'autre. Nous allons examiner les services de mise à jour selon les critères suivants: le mode opératoire de mise à jour des données diffusées dans le Web, les types d'objets répertoriés, l'information retenue des objets répertoriés, la taille et la fréquence de mise à jour de la base.

2.1. Le mode opératoire de mise à jour de la base de données W3

Les méthodes utilisées pour la mise à jour de la base de données W3 sont principalement le robot, l'inscription, l'éditeur et le diffuseur-éditeur. Notons que ces méthodes ne sont pas mutuellement exclusives. Un outil de recherche peut utiliser l'une ou l'autre, ou toutes ces méthodes à la fois. Par exemple, la base de données Yahoo utilise les méthodes de robot, de l'inscription et de l'éditeur.

2.1.1. Le robot

Le *robot* est sans aucun doute l'une des méthodes les plus répandues pour l'exploration du Web. Les sites tels Infoseek, Lycos, Open Text Web Index, WebCrawler, Whole Internet Catalog, Harvest Broker, WWW Worm et Yahoo utilisent cette méthode. Parfois, dans la littérature, on emploie des noms encore plus évocateurs tels *spiders* (araignée), *worms* (ver de terre, se faufiler) ou *Web Wanderer* (vagabond du Web). Malgré l'exotisme des images évoquées par ces noms, la réalité est toute simple: un robot (*spiders*, *worms* ou *Web wanderer*) est tout simplement un programme qui s'exécute sur un ordinateur relié au Net

et qui explore systématiquement celui-ci de manière à colliger l'information présente. C'est l'automatisation et la systématisation de ce que l'on fait de chez soi en se baladant dans le Web (et pourtant on est bien assis devant son ordinateur...).

Avec le mécanisme du robot, le diffuseur d'information (vous et moi) joue un rôle complètement passif car le robot se charge de rechercher l'information diffusée. Celui-ci ne demande pas son reste (ni votre permission) et visite lien par lien la toile du Web selon un algorithme de recherche. Cet algorithme suit généralement un de ces deux principes: par la nature hiérarchique des liens dans une page Web, l'exploration faite par les robots suit une recherche en profondeur (l'algorithme est basé sur l'exploration récursive du premier lien hypertextuel rencontré) ou en largeur (l'algorithme repose sur l'exploration de tous les noeuds d'une page avant de descendre dans un des liens hypertextuels). Pour obtenir plus d'information sur les robots, sur les techniques utilisées et sur les controverses à leur sujet, je vous invite à consulter les excellents articles de Martijn Koster (voir bibliographie).

Une dernière remarque: la tâche des robots, soit d'explorer le Web dans sa totalité, constitue un problème de taille et, dans les faits, pratiquement impossible à résoudre. C'est un peu comme si vous vouliez parcourir tous les chemins de la planète, de la manière la plus efficace, en un temps minimal.

2.1.2. L'inscription

Plusieurs outils de recherche offrent aux utilisateurs d'inscrire leurs publications. Parmi celles-ci, nous retrouvons InfoSeek, Lycos, Open Text, WebCrawler, Harvest Broker, WWW Worm et Yahoo. Notons toutefois que le rôle du diffuseur de l'information primaire est, en général, très minime et se limite à donner uniquement une adresse URL. Par la suite, l'outil de recherche *visitera* (fort probablement à l'aide d'un robot) l'URL indiqué pour en extraire l'information.

Vous avez produit un article sur le Web et désirez le faire connaître? Inscrivez-vous (ou plutôt inscrivez votre URL préféré) aux bases citées précédemment. Par ailleurs, pour exploiter efficacement ces bases de données en vue de vous publiciser, nous vous conseillons de lire la section suivante intitulée *L'information retenue des objets répertoriés*. En effet, l'information répertoriée et retenue varie énormément d'un outil de recherche à l'autre.

2.1.3. L'éditeur (guest editors)

La méthode de l'éditeur est utilisée uniquement dans les outils de recherche de type répertoire. Ces sites profitent de personnes qui se spécialisent dans un domaine précis et qui s'occupent d'en maintenir la cohésion.

Les sites Galaxy et Whole Internet Catalog font partie de cette catégorie. Les diffuseurs sont également encouragés à proposer des sujets. Ceux-ci seront inclus dans la base de données uniquement sous l'approbation d'un responsable du domaine (guest editors). Évidemment, contrairement aux techniques automatisées, cette technique est plus coûteuse en ressources humaines mais permet généralement une valeur ajoutée à l'information répertoriée.

2.1.4. Le diffuseur-éditeur

Bien que très peu utilisée, nous en faisons quand même mention. Cette méthode, assez originale, est toute à l'opposé de celle des robots vue précédemment. Le diffuseur d'information primaire joue un rôle beaucoup plus actif. En effet, l'outil de recherche est tributaire du bon vouloir des diffuseurs d'information primaire. Ce sont eux qui décident des objets à diffuser (c'est la revanche des écrivains sur les éditeurs).

Néanmoins, le diffuseur doit se conformer à une norme pour décrire l'objet de sa diffusion. Le seul site qui utilise cette méthode, à notre connaissance, est celui d'Aliweb. Pour tout objet de diffusion, le diffuseur prépare un fichier, selon des normes prescrites, puis le signale à l'outil de recherche qui se chargera de récupérer la description contenue dans le fichier pour l'inclure dans sa base. De plus, l'un des services de l'outil de recherche visitera régulièrement le site Web pour prendre en compte toute modification éventuelle à la description.

Cette méthode comporte plusieurs avantages. La plus importante concerne la description de l'objet à diffuser. Contrairement aux robots qui ne font qu'extraire les données du Web, cette méthode permet une valeur ajoutée par le diffuseur. Celui-ci peut, en effet, décrire l'objet diffusé, lui associer des mots clés et le classer selon la norme émise. En contrepartie, actuellement, Aliweb comporte peu d'éléments (environ 6,000 documents).

2.2. Les types d'objets répertoriés

Le type d'information retenue varie d'un outil de recherche à un autre. Avant de pouvoir interroger une base de données, il faut nécessairement alimenter celle-ci. Plusieurs questions se posent et parmi celles-ci la plus importante consiste à déterminer les types d'objets à répertorier (HTML, Gopher, Ftp, etc.). Cette question est fondamentale à la définition des services de repérage qui seront offerts par la suite.

Par exemple, la base de données de LYCOS est constituée uniquement de documents de type HTML, GOPHER et FTP. Les fichiers de type Usenet (groupe de discussion) ou Telnet sont tout simplement ignorés par ses robots. Pour obtenir de l'information d'une page Web, l'outil est certes intéressant. Par contre, si vous cherchez de l'information produite dans un groupe de discussion, à la fine pointe de l'actualité, nous vous recommandons InfoSeek ou The Whole Internet Catalog. En opposition à Lycos qui essaie d'indexer le Net, les responsables de Whole Internet Catalog s'emploient à trouver les liens les plus prolifiques (*hotlist*) en consultant les groupes de discussion ou les listes *What's New*.

Un second exemple est la base de données InfoSeek. C'est une base à but lucratif mais qui permet l'accès gratuit à un sous-ensemble des produits offerts. Cette base est, en fait, constituée de plusieurs bases de données. D'une part, elle offre des références à des pages Web et à des messages récents de groupes de discussion. D'autre part, InfoSeek offre, contre rémunération, un accès à des données autres que celles du Net: plus de 80 périodiques informatiques, certaines agences de presse, la base de donnée Medline et plusieurs autres.

Un dernier exemple est la base de données de Clearinghouse for Subject-Oriented Internet Resource Guides. Cette base contient environ 400 guides spécialisés dans plusieurs sujets que l'on retrouve dans Internet. Ces guides sont écrits en format HTML ou en format texte. Ils peuvent contenir des références à des documents de type W3, Gopher ou FTP mais celles-ci sont parfois décrites en texte sans lien hypertextuel.

2.3. L'information retenue des objets répertoriés

A chaque objet répertorié correspond un enregistrement dans la base de données. Selon l'outil de recherche, la nature de l'enregistrement peut varier énormément. Minimale, celui-ci contient l'adresse URL et le titre de l'objet répertorié. Il peut également, par exemple, contenir les en-têtes, des mots clés ou, mieux encore, l'intégralité des pages Web. Évidemment, la qualité des services de repérage sera directement proportionnelle à la richesse de cet enregistrement.

Par exemple, à chaque document W3 répertorié par LYCOS correspond un enregistrement contenant le titre, une partie de l'en-tête, les cent mots les plus significatifs, les vingt premières lignes du texte, la taille en octets, la taille en nombre de mots et le nombre de liens. Morale de l'histoire, si vous voulez inscrire votre page W3 chez Lycos, prenez garde à vos vingt premières lignes!

La base de données de Open Text, qui contient l'intégralité de chaque page Web répertoriée, associe à chaque document un extrait, un titre, le premier en-tête, l'URL et les liens hypertextuels. C'est de loin le site qui exploite le mieux la structure logique des documents HTML. Par ailleurs, WebCrawler indexe tout le contenu des documents répertoriés mais n'exploite pas la structure logique de ceux-ci. Pis encore, CUI World Wide Web Catalog indexe uniquement les titres des documents répertoriés!

2.4. La taille de la base de données

Bien qu'elle dépende directement de l'information retenue des objets répertoriés et du mode opératoire du repérage des données, la taille de la base de données est un critère souvent utilisé dans les études comparatives d'outils de recherche. C'est pour quoi nous avons décidé de l'inclure dans notre étude.

La taille des bases est, comme vous pouvez l'imaginer, très variable. D'ailleurs les différentes études sont parfois contradictoires à cet effet. Les chiffres varient énormément selon la définition donnée aux documents d'une base de données (par exemples, considère-t-on un seul URL par page ou l'ensemble des URLs contenus dans une page, est-ce un document intégral ou uniquement son URL, etc.) et la date à laquelle ont été prises les statistiques (l'évolution très rapide du Web rend vite ces chiffres obsolètes). Par conséquent, nous devons utiliser ces chiffres avec précaution, à titre indicatif uniquement et éviter toute conclusion hâtive.

Par exemple, la base de données GNN Whole Internet contient 2,000 liens en comparaison de 10,75 millions pour Lycos et pourtant le GNN's Whole Internet n'est pas *moins bon* que Lycos. Comme on a vu

précédemment, GNN Whole Internet est un outil de recherche de type répertoire qui utilise la méthode éditeur alors que Lycos est un outil de recherche de type index qui utilise la méthode des robots. Tous deux ont des visées bien différentes. La première fonctionne manuellement, ce qui lui confère une valeur ajoutée, alors que la seconde est entièrement automatisée. Si vous êtes intéressés aux sujets de dernière heure de l'Internet, il peut être intéressant de visiter GNN Whole Internet.

De même, la base de données de Clearinghouse for Subject-Oriented Internet Resource Guides est constituée uniquement de 400 guides sur les ressources de l'Internet. Ceci n'enlève rien à la qualité et à l'intérêt de celle-ci.

Ceci étant dit, il est quand même intéressant de pouvoir comparer la taille des bases de même nature. Dans les bases de type index, d'une part, nous avons Lycos avec 10,75 millions d'URL avec des références au document primaire et, d'autre part, nous retrouvons WWW Worm, Open Text et InfoSeek qui ont respectivement 3 millions, 1 million et 400,000 documents intégraux. Dans les bases de type répertoire, on retrouve Galaxy avec des index de 100,000 documents intégraux et Yahoo avec 66,000 références.

2.5. La fréquence de mise à jour de la base de données

Les données du Net sont très volatiles: à chaque jour s'ajoutent de nouveaux documents, d'autres sont modifiés ou détruits (dead link). Particulièrement pour les outils de recherche qui utilisent les robots, il est important que ces bases soient constamment remises à jour. Par exemple, la base de données Lycos est entièrement reconstruite chaque semaine. Environ 10,000 documents s'y ajoutent hebdomadairement. Open Text suit la même fréquence. WebCrawler, quant à lui, remet à jour sa base à toutes les six semaines.

Les bases de données Yahoo et Harvest Broker opèrent une reconstruction journalière de leur base respective. Pour d'autres sites gérés manuellement, tels Whole Internet Catalog et WWW Virtual Library, la responsabilité incombe aux éditeurs (difficilement quantifiable).

A cause de la méthode diffuseur-éditeur utilisée par Aliweb, l'ajout de nouvelles références est laissé à l'entière responsabilité des diffuseurs d'information primaire. Similairement, la base de données Clearinghouse for Subject-Oriented Internet Resource Guide est enrichie de nouveaux guides, préparés et maintenus par des étudiants en bibliothéconomie et science de l'information (SILS, University Library at University of Michigan), à intervalle de 5 à 8 mois.

3. LES SERVICES DE REPÉRAGE

L'interface utilisateur, plus ou moins complexe selon les sites, comporte l'entrée des requêtes de recherche (termes recherchés et paramètres de la recherche). Nous porterons notre attention uniquement sur les interfaces utilisant des formulaires HTML. Évidemment, en fonction des services offerts, l'interface utilisateur varie énormément entre les outils de recherche. De plus, il n'y a pas de normalisation et chaque

outil possède sa propre interface. Nous allons examiner les services liés à la documentation offerte en ligne et aux fonctions pour le repérage.

3.1. La documentation

La documentation pour les services de repérage aide les utilisateurs à atteindre deux objectifs. Le premier est d'évaluer la pertinence de la base, c'est-à-dire sa nature (objets répertoriés), ses objectifs, son autorité, etc. Le second est la maîtrise et l'utilisation efficace des services de repérage et la vérification du comportement de ces services (obtient-on les résultats escomptés?).

Par exemple, l'outil de recherche Galaxy permet, à l'aide de liens hypertextuels, de trouver de l'information à propos de l'organisation responsable du site. De plus, une introduction complète et détaillée avec des exemples de stratégie de recherche et une page portant sur les nouvelles entrées (*What's New*) sont disponibles. De loin, il s'agit de l'un des meilleurs avec Harvest Broker et WWW Worm.

En contrepartie, nous retrouvons CUI World Wide Web Catalog qui n'offre aucun service d'aide. WebCrawler, quand à lui, offre une documentation correcte mais sans plus. Notons toutefois, à la défense de WebCrawler, que celui-ci offre peu de fonctions et que l'utilisation de son interface est très simple. En conclusion, plus les services de repérage sont sophistiqués, plus claire et précise doit être la documentation.

3.2. Les fonctions de repérage

Nous avons vu, dans la section portant sur les services de mise à jour, que les types d'objets et l'information retenue varient d'un outil de recherche à l'autre. L'organisation structurelle des données a un impact direct sur les outils de repérage et sur la qualité des fonctions offertes. Restreindre l'indexation à certains champs améliorera la précision de nos recherches. De même, la notion de pertinence est très importante en recherche de l'information. C'est pourquoi plusieurs outils de recherche utilisent le principe de pondération sur les mots recherchés.

De plus, une panoplie d'opérateurs sont utilisés pour raffiner le repérage. Nommons, entre autres, les opérateurs booléens, les opérateurs de voisinage et les opérateurs de chaînes de caractères. Notons que ces opérateurs reposent essentiellement sur le principe de concordance de modèle (pattern matching). Par ailleurs, certaines bases offrent des langages de recherche qui permettent l'utilisation d'expressions régulières.

Néanmoins, contrairement aux recherches automatisées dans les bases de données traditionnelles, plusieurs fonctions ne sont toujours pas offertes par les outils de recherche. Pour n'en citer que quelques-unes, prenons les opérateurs numériques et l'historique de recherche. Actuellement, il est impossible d'effectuer une requête comportant une comparaison de dates. De même, aucun outil ne permet d'établir une session usager de manière à conserver un historique des recherches. Il est à noter qu'une telle

fonctionnalité pourrait théoriquement être assurée par l'amélioration de la gestion de l'historique de navigation dans les logiciels clients W3 pour les pages comportant des formulaires.

3.2.1. Les champs indexés

L'indexation du texte intégral constitue un moyen, critiquable certes pour le bruit résultant, qui peut s'avérer intéressant. Par ailleurs, l'indexation selon la structure logique d'un document est un moyen efficace pour cibler de l'information. Par exemples, le titre, les en-têtes, le résumé ou les premières lignes d'un texte forment à coup sûr des entités pour lesquelles les mots s'y rattachant ont une plus grande pertinence. Mais ce ne sont pas les seules. En plus de la structure logique du document, il y a les noms d'URL et les noms de fichiers qui forment également des entités très pertinentes pour certains types de repérage.

Prenons l'exemple de Open Text Web Index qui offre quatre outils de repérage. Parmi ceux-ci, nous retrouvons le Power Search qui, selon plusieurs analystes, est l'interface la plus flexible du Web. Cette base contient les documents intégraux et la structure logique de ceux-ci, du moins en partie. Il est possible de faire des recherches sur le titre, le premier en-tête, «_l'extrait_», les liens hypertextuels ou tout simplement le texte complet.

L'outil de recherche WWW Worm offre, quant à lui, quatre champs indexés pour la recherche: le titre, le nom du fichier HTML, l'URL et les références à un URL (références croisées). Ce dernier type de recherche permet de déterminer, par exemple, quels sont les documents de la base de données WWW Worm qui contiennent un lien hypertextuel à votre page d'accueil. La recherche dans un URL offre la possibilité, par exemple, de rechercher des films en format MPEG (l'URL se termine par «.mpeg») provenant de la nasa (l'URL contient le mot «nasa» dans le nom de l'ordinateur hôte). Cet outil peut s'avérer très intéressant lorsqu'on connaît, du moins en partie, la nature de l'objet à repêcher.

3.2.2. La pondération des mots

Une autre technique utilisée pour améliorer le repérage est d'associer un «_poids_» à chacun des mots présents dans un document répertorié. Cette technique a plusieurs variantes. Un mot reçoit un poids plus ou moins grand selon le nombre de fois et/ou les positions où il apparaît dans le document. Par exemple, un mot apparaissant dans le titre d'un document recevra un poids plus grand qu'un mot n'apparaissant que dans le corps du document. De même, un mot apparaissant deux fois recevra un poids plus grand qu'un mot n'apparaissant qu'une seule fois. Ces poids associés aux mots présents dans un document sont utilisés lors d'une recherche pour calculer la «_pertinence estimée_» du document par rapport à la requête formulée. D'autres facteurs peuvent participer également au calcul, tels la proximité des termes recherchés dans le document et le fait qu'ils apparaissent dans le même ordre que dans la requête ou non. Une des méthodes les plus populaires pour calculer la pertinence estimée d'un document par rapport à une requête est celle de type WAIS.

Quelle que soit la méthode utilisée, la pertinence estimée est représentée par une «_pondération_» numérique, qui s'exprime habituellement sous la forme d'un nombre compris entre 0 et 100, 100

représentant la plus grande pertinence estimée possible. Ainsi, la pondération d'un document en réponse à une requête de recherche donne une certaine indication de sa pertinence probable par rapport à cette requête.

Prenons l'exemple de l'outil de recherche Lycos. Celui-ci classe les résultats, c'est-à-dire les références aux documents repêchés, par ordre décroissant des pondérations calculés pour chaque document en fonction des termes recherchés (*relevancy ranking of terms*). En d'autres mots, à chaque document de la base est calculé un poids (*weight terms*). Cette pondération est calculée à partir du nombre d'occurrences des termes recherchés, de leur position dans le document (le titre, les premières lignes du texte, etc.) et de leur proximité (s'il y a lieu). C'est sans aucun doute l'outil qui exploite le mieux cette technique.

L'outil de recherche WebCrawler utilise aussi l'algorithme de pondération. Néanmoins, cet algorithme est moins bien développé que celui de Lycos car, malgré l'indexation intégrale du texte, cette base n'utilise pas la structure logique du document (en-tête, citation, etc.). La pondération repose uniquement sur le nombre d'occurrences du terme recherché sans tenir compte de sa position dans le texte.

3.2.3. Les opérateurs booléens

Tous et chacun connaissent l'importance, dans la recherche automatisée, des opérateurs booléens (AND, OR, NOT). Plusieurs outils de recherche offrent ce type de fonction. Cependant, peu d'entre eux permettent une utilisation rigoureuse de ces opérateurs (nous qualifions ceux-ci de pseudo-booléens).

Les outils de recherche Open Text Web Index et Einet Galaxy, quant à eux, permettent une utilisation des opérateurs booléens (AND, OR, NOT) de manière précise.

L'outil de recherche Lycos permet de spécifier le nombre minimum de termes à être présents dans les résultats sans pouvoir déterminer lesquels parmi la liste des termes fournis (c'est un pseudo ET/OU booléen). Il est également possible d'utiliser la négation d'un terme mais cette négation n'est pas exclusive. En effet, les documents contenant le terme recevront tout simplement une pondération moins significative et apparaîtront plus loin dans la liste des résultats. Toutes les fonctions de repérage offertes par Lycos reposent sur le concept de pondération.

En contrepartie, la base de données CUI World Wide Web Catalog considère uniquement un OU implicite entre les termes de recherche.

3.2.4. Les opérateurs de voisinage

En recherche automatisée, nous utilisons souvent des opérateurs de voisinage telles l'adjacence et la proximité. Peu d'outils de recherche offrent ce type de service.

La base de données Open Text Web Index permet l'adjacence. Un point intéressant, bien que ne faisant pas partie des services offerts par Lycos, les concepts d'adjacence et de proximité sont quand même pris

en considération. En effet, la pondération calculée repose, en partie, sur la distance entre les termes recherchés.

3.2.5. Les opérateurs de chaînes de caractères

Les opérateurs de chaînes de caractères correspondent le plus souvent à ceux que l'on connaît dans les traitements de texte. Nous retrouvons principalement les caractères génériques pour la troncature ou la concordance avec un masque (*wild card*), les sous-chaînes (*searching substring*) et les phrases (*searching phrase*). De plus, certains outils de recherche permettent la recherche sur des mots apparentés (*approximative matches*). La distinction entre les majuscules et les minuscules est également supportée par quelques outils.

Plusieurs outils de recherche permettent l'utilisation de la troncature. Nous retrouvons, entre autres, Lycos et Harvest Broker.

3.2.6. Les langages de recherche

Certains outils de recherche offrent des langages pour raffiner nos recherches ou l'utilisation d'expressions régulières. Pour n'en nommer que quelques-uns, citons le WWW Worm qui utilise le programme «egrep» d'Unix et la base de données CUI World Wi de Web Catalog qui demande une connaissance du langage PERL ou encore Harvest Broker qui permet d'utiliser des expressions régulières. Mais, comme tout langage, l'investissement en temps pour en faire l'apprentissage en décourage plus d'un!

4. LES SERVICES DE PRÉSENTATION DES RÉSULTATS

Là encore, la diversité règne. La présentation des résultats est aussi vitale que l'ensemble des services vus jusqu'à maintenant. En effet, quels sont les avantages d'obtenir 1,000 documents si les résultats sont difficilement exploitables? Nous allons examiner ces services selon la documentation offerte et les fonctions de présentations.

4.1. La documentation

La documentation pour les services de présentation des résultats assiste l'utilisateur à interpréter correctement les résultats. Par exemple, comme on a vu précédemment, plusieurs outils de recherche offrent un classement des résultats obtenus selon leur pondération par rapport à la requête de recherche (cf. 3.2.2). Malheureusement peu d'entre eux expliquent le calcul de cette pondération.

4.2. Les fonctions de présentation des résultats

Pour faciliter la compréhension des résultats, plusieurs techniques sont disponibles. Nous retrouvons principalement l'affichage de la requête utilisée et des formats de présentation variables (abrégé, moyen ou complet). Minimale, les résultats contiennent l'URL et le titre du document repêché. Cette

approche minimaliste est utilisée par CUI World Wide Web Catalog. L'URL est, en général, un lien hypertextuel menant au document primaire. A ces caractéristiques, d'autres outils de recherche joignent un extrait du texte avec les termes associés à la recherche mis en évidence (mots en contexte), la pondération, la taille du document et, parfois, la possibilité de réutiliser des résultats obtenus pour effectuer une autre recherche (*see similar page*).

Dans Lycos, lors de sa requête, l'utilisateur peut préciser un nombre maximum de résultats (pour limiter le temps d'attente) et un seuil minimum de pondération (pour diminuer le bruit). De plus, il peut sélectionner un des trois types de format de présentation des résultats.

L'interface de sortie de WebCrawler est très simple. Il est possible de spécifier le nombre maximum de résultats (10, 25 ou 100). Les résultats sont ordonnés par ordre décroissant de pondération et contiennent uniquement le titre et la pondération. Ce peu d'information rend difficile l'évaluation des résultats.

L'outil de recherche Open Text Web Index permet en sortie de sélectionner une option pour faire apparaître dans leur contexte les termes de recherche. Un extrait accompagne les documents repêchés. Une option intéressante permet de rechercher des documents similaires à celui qui est sélectionné (*Find Similar Pages*). L'algorithme est simple et repose sur le nombre d'occurrences de mots présents dans le document.

Harvest Broker offre la possibilité de sélectionner certains éléments, telles la pondération et la description, pour qu'ils apparaissent dans les résultats.

5. ÉVALUATION D'OUTILS DE RECHERCHE

Plusieurs des évaluations d'outils de recherche publiées ont été produites, du moins en partie, en fonction des résultats obtenus à des tests de performance (*benchmark*). La nôtre diffère de celles-ci dans la mesure où nous portons notre attention principalement sur les critères que nous avons présentés précédemment. Nous avons évacué délibérément les notions de rappel et de précision des résultats pour nous attarder uniquement aux services offerts par un outil de recherche.

Cette évaluation n'a aucune prétention scientifique. Elle est divisée en deux sections selon le type index ou répertoire et porte sur treize outils de recherche. Tel que mentionné dans la première partie de ce travail, le type index est associé à tous les outils de recherche dont les fonctions de repérage reposent principalement sur l'utilisation d'index. Le type répertoire est, quant à lui, associé à tous les outils de recherche dont les fonctions de repérage se basent principalement sur un plan de classification.

A chaque outil de recherche est associé une évaluation globale: soit (*) pauvre, (**) passable, (***) bon et (****) très bon. Si l'information concernant un critère d'évaluation n'est pas disponible pour un outil de recherche, nous inscrivons un point d'interrogation, «_?_», à sa description. Il peut arriver également que l'on ajoute, à ce point d'interrogation, de l'information entre parenthèses dans le cas où l'information n'a pu être vérifiée.

Tel que mentionné en introduction, il est fort probable que les fonctionnalités offertes par les outils de recherche auront évolué depuis la fin de nos travaux et que certaines de nos remarques ne s'appliquent plus.

5.1. Les outils de recherche de type index

5.1.1. ALIWEB

L'originalité d'ALIWEB découle de son mode opératoire pour la mise à jour de l'information diffusée dans le Web, c'est-à-dire la méthode du diffuseur-éditeur. Le diffuseur d'information primaire est mis à contribution dans l'élaboration des éléments servant au repérage. Il doit remplir un gabarit (*template-type*), selon le type d'objet à diffuser, qui permet, entre autres, d'ajouter une description, des mots clés, une date de mise-à-jour, des adresses électroniques.

Le repérage peut s'effectuer selon le type d'objet diffusé et les zones indexées tels le titre, la description, les mots clés et l'URL. En outre, il est possible de restreindre la recherche à un domaine. Par exemple, on peut effectuer une recherche uniquement sur les documents produits au Canada (ca). Par ailleurs, la recherche peut porter sur une sous-chaîne, un mot complet ou une expression régulière.

Pour les résultats en sortie, l'utilisateur peut spécifier, en plus du titre et de la pondération, les zones telles la description, les mots clés et l'URL. De plus, l'utilisateur a mainmise sur le nombre de résultats maximum désiré.

Bref, un outil dont nous pensons beaucoup de bien. Par contre, nous avons quelques réserves. En premier lieu, la documentation est plutôt déficiente sur la constitution du gabarit et sur l'utilisation d'expressions régulières. De plus, la pondération sur les résultats en sortie n'est expliquée nulle part. Finalement, cette base contient peu de références (environ 6,000). Malgré tout, nous croyons que cette méthode comporte plusieurs avantages. Cet outil mérite un détour et nous lui attribuons trois étoiles (***)

-
- **URL:** <http://web.nexor.co.uk/public/aliweb/aliweb.html>
 - **Type:** Index

DOCUMENTATION

- Bonne (aide, exemples, organisation) mais pauvre en ce qui concerne les expressions régulières et la construction d'un fichier à diffuser

SERVICE DE MISE À JOUR

- **Mode opératoire de mise à jour:** Diffuseur-éditeur
- **Objets répertoriés:** FTP, GOPHER, pages W3

- **Information retenue:** Titre, URL, une description et selon le type de gabarit (template-type) d'autres informations supplémentaires tels le nom de l'organisation, l'adresse électronique, des mots clés, date de mise à jour, etc.
- **Taille:** 6,000 références
- **Fréquence de mise à jour:** Journalière pour les informations contenues dans la base et selon la bonne volonté des diffuseurs pour les ajouts.

SERVICE DE REPÉRAGE

- **Moteur de recherche:** ? (Perl Script)
- **Opérations:** Mots multiples, sous-chaînes, expression régulière (syntaxe Perl)
- **Zones d'index:** Titre, description, mots clés, URL
- **Divers:** Restriction à un domaine (un pays)
- **Ce qui n'est pas offert (documenté):** opérateurs booléens; opérateurs de voisinage

SERVICE DE PRÉSENTATION DES RÉSULTATS

- **Format:** Selon les éléments sélectionnés par l'utilisateur
 - **Limite du nombre de résultats:** Oui
 - **Paramètres de recherche affichés:** Oui
 - **Description:** Titre, pondération, description, mots clés et d'autres champs selon le type de gabarit.
 - **Évaluation:** Bon (***)
-

5.1.2. CUI W3 CATALOG

L'un des outils de recherche qui nous apparaît comme étant le plus décevant. Aucune documentation ou aide en ligne n'est disponible. On y indique que l'on peut utiliser les expressions régulières en langage PERL sans aucune explication. Les documents sont ajoutés à partir de listes What's New ou Hot List provenant de divers serveurs.

La présentation des résultats est correcte mais là encore sans explication sur le contenu. Une description et une date de mise à jour accompagnent ces résultats. Le seul point positif est la possibilité de refaire une recherche à partir du formulaire contenant les résultats.

En résumé, nous ne pouvons pas dire que cet outil nous a enchanté. Nous lui accordons seulement une étoile (*).

-
- **URL:** <http://cuiwww.unige.ch/w3catalog>
 - **Type:** Index

DOCUMENTATION

- Pauvre et inexistante pour l'interprétation des résultats

SERVICE DE MISE À JOUR

- **Mode opératoire de mise à jour:** Éditeur
- **Objets répertoriés:** *Hot List*, *What's New* et des références d'outils de recherche de type répertoire
- **Information retenue:** Titre, URL et une description
- **Taille:** 40,000 références
- **Fréquence de mise à jour:** Journalière

SERVICE DE REPÉRAGE

- **Moteur de recherche:** ? (Perl Script)
- **Opérations:** Phrase, sous-chaîne, expression régulière (syntaxe Perl)
- **Zones d'index:** Aucune
- **Ce qui n'est pas offert (documenté):** opérateurs booléens, distinction entre les majuscules et les minuscules

SERVICE DE PRÉSENTATION DES RÉSULTATS

- **Format:** Un seul
 - **Limite du nombre de résultats:** Non
 - **Paramètres de recherche affichés:** Oui
 - **Description:** Titre et une description. Notons toutefois qu'il n'y a pas de pondération
 - **Évaluation:** Pauvre (*)
-

5.1.3. HARVEST BROKER

Harvest est utilisé par plusieurs sites de recherche (appelés broker). Nous en parlerons de manière générique. Cet outil s'appuie, pour le repérage d'information diffusée, sur les méthodes du robot, de l'inscription et de l'éditeur. Il est un de ceux qui possèdent la documentation la plus complète. Celle-ci comprend même des explications sur les expressions régulières.

Selon l'implantation, Harvest Broker fonctionne avec GLIMPSE ou WAIS comme moteur de recherche. GLIMPSE est de loin le moteur le plus évolué. En effet, il permet l'utilisation d'un sous-ensemble d'expression régulière (dont la troncature), d'un opérateur d'adjacence, d'opérateurs booléens, de mots apparentés et plus encore. Aussi, il est possible de restreindre notre recherche à des zones tel l'auteur, le titre, l'URL, les mots clés ainsi que plusieurs autres selon le type d'implantation.

En sortie, nous obtenons l'URL, le titre et la taille du document. L'utilisateur peut indiquer le nombre de résultats maximum et sélectionner l'affichage de la pondération, de l'URL, de la description (si elle existe)

et des références croisées lors de sa requête.

Un de nos outils préférés. Un exemple à suivre pour tous ceux qui désirent développer un outil de recherche. Nous lui accordons le maximum, c'est-à-dire quatre étoiles (****).

-
- URL: <http://town.hall.org/Harvest/brokers/www-home-pages/>
 - Type: Index

DOCUMENTATION

- Excellente (aide, exemples, exemples avec expression régulière)

SERVICE DE MISE À JOUR

- **Mode opératoire de mise à jour:** Robot, éditeur et inscription
- **Objets répertoriés:** pages W3, FTP, News (selon le site)
- **Information retenue:** Titre, URL, mots clés, liens hypertextuels (URL citation), le texte intégral et d'autres attributs (varient selon le type de moteur utilisé)
- **Taille:** 45,000 documents
- **Fréquence de mise à jour:** Journalière

SERVICE DE REPÉRAGE

- **Moteur de recherche:** WAIS ou GLIMPSE (ce dernier est beaucoup plus puissant)
- **Opérations:** Selon le moteur utilisé, l'adjacence, sous-chaîne, mot exact, mots multiples, mot apparentés, opérateurs AND et OR, phrase, troncature, expression régulière (avec GLIMPSE), la distinction entre les majuscules et les minuscules
- **Zones d'index:** Titre, URL, mots clés, liens hypertextuels (URL citation), le texte intégral et d'autres attributs (varient selon le moteur utilisé)

SERVICE DE PRÉSENTATION DES RÉSULTATS

- **Format:** Un seul
- **Limite du nombre de résultats:** Oui
- **Paramètres de recherche affichés:** Oui
- **Description:** Titre, pondération, URL, citation et description
- **Évaluation:** Très bon (****)

5.1.4. INFOSEEK

InfoSeek est à but lucratif. Heureusement, un accès gratuit est disponible à une partie des données. C'est un des outils de recherche les mieux appréciés dans plusieurs études comparatives (la qualité se paie).

Bien qu'étant de type index, InfoSeek offre également une classification. On peut considérer cette base comme un hybride index-répertoire. La recherche s'opère sur les noms de fichiers et leur contenu. Il est également possible de consulter une douzaine de sujets (topics). L'utilisateur peut spécifier le type d'objet désiré soit des pages Web, des groupes de discussion ou des FAQ d'Internet. La documentation est excellente.

Plusieurs fonctions sont disponibles pour le repérage. L'utilisation de guillemets permet de spécifier des phrases pour la recherche. De plus, InfoSeek porte une attention aux noms propres (les mots en majuscule) de manière à les distinguer des noms communs. Des opérateurs de voisinage, tels l'adjacence et la proximité, sont également offerts. Par contre, il n'y a pas d'opérateurs booléens mais néanmoins l'utilisateur peut spécifier des mots obligatoires (+) ou en proscrire (-).

En sortie, nous obtenons, en plus du titre, la pondération, la taille et le type du document, une description, des références croisées et le tout par ordre décroissant de pondération. Une particularité intéressante d'InfoSeek est la fonction Find similar pages qui, tel que son nom l'indique, permet de rechercher d'autres ressources similaires à un résultat obtenu.

Un produit de grande qualité qui, comme Harvest Broker, vaut le déplacement. Nous lui attribuons quatre étoiles (****).

-
- URL: <http://www.infoseek.com>
 - **Type:** Index et répertoire

DOCUMENTATION

- :
- Excellente (organisation, aide, exemples, FAQ, etc.)

SERVICE DE MISE À JOUR

- **Mode opératoire de mise à jour:** Robot et inscription (par courrier électronique)
- **Objets répertoriés:** pages W3, Gopher, FTP, NewsGroup, FAQ
- **Information retenue:** Texte intégral et les 300 premiers caractères du document
- **Taille:** 400,000 documents W3
- **Fréquence de mise à jour:** Mensuelle

SERVICE DE REPÉRAGE

- **Moteur de recherche:** Propriétaire (Python programming language, similaire à Perl)
- **Type d'objets:** W3, Newsgroup, FAQ, Reviewed pages (ces pages sont répertoriées dans les topics) et les topics (sujets)
- **Opérations:** L'adjacence, la proximité (à l'intérieur de 100 mots), mots obligatoires, mots non-désirés, phrase, utilisation de la virgule et des majuscules pour distinguer les noms propres, recherche à l'intérieur d'un sujet
- **Zones d'index:** selon le type d'objet
- **Divers:** L'utilisation de majuscules indique que l'on recherche des noms propres (tels Microsoft, Plourde Jean-Noel, etc.)
- **Ce qui n'est pas offert (documenté):** opérateurs booléens (quoique dans les opérations on peut forcer la présence d'un mot ou proscrire sa présence); troncature
- **Remarque:** Dans la version commerciale, il est possible de préciser un intervalle de dates pour le repérage.

SERVICES DE PRÉSENTATION DES RÉSULTATS

- **Format:** Un seul
- **Limite du nombre de résultats:** 10 (démon) et 200 (ceux qui sont inscrits)
- **Paramètres de recherche affichés:** Oui
- **Description:**

Type d'objet (utilisation d'une icône pour indiquer s'il s'agit d'une page W3 ou d'un titre Gopher, etc.), titre, pondération, URL, taille, description.

- **Autres fonctions:** Similar Pages
 - **Remarque:** Contre rémunération, il est possible d'accéder à MDX Health Digest, Hooves Company Profiles, Wire (BusinessWire, PR Newswire, Newsbyte, News Network, Reuter business report), Computer Periodicals (InfoWorld, Computer World, Newsbytes), Computer Select (Byte, Business Week, PC World, Wall Street Journal) et plusieurs autres bases encore....
 - **Évaluation:** Très bon (****)
-

5.1.5. LYCOS

C'est probablement l'outil de recherche le plus populaire du Web. C'est à tout le moins celui qui est le plus fréquemment cité dans les études comparatives. Tout comme InfoSeek, cet outil est à but lucratif. Par contre, son utilisation est gratuite et ses revenus proviennent de sources publicitaires.

Lycos est constitué de deux bases: une petite d'environ 500,000 URL - où chaque URL possède une description - et une grande d'environ 10,75 millions d'URL (qui contient également les éléments de la petite base). Les objets répertoriés sont de type Gopher, FTP et page Web. La documentation est complète. En outre, cette documentation décrit les différences majeures entre Lycos et ses principaux concurrents Yahoo et InfoSeek.

Les fonctions de recherche permettent des opérateurs pseudo-booléens (ALL/ANY en spécifiant un nombre de termes) ainsi que la négation. Le moteur de recherche s'appelle PURSUIT. Tout élément de la base est pondéré selon la position de celui-ci dans les textes. Par exemple, un mot situé dans le titre ou dans le premier paragraphe d'une page aura un plus grand poids. Il est possible aussi de spécifier le degré de précision des termes recherchés (loose, fair, good, close, strong match). Nous devons admettre cependant que nous n'avons pas saisi toute la subtilité entre ces paramètres!

En sortie, on peut préciser le nombre maximum de résultats et trois formats d'affichage. La présentation est très bien faite. A chaque paramètre de recherche est associé le nombre de documents trouvés. De plus, pour chacun des résultats avec le format complet, il y a l'URL, le titre, les 200 premiers caractères dans les zones de type en-tête, les 100 mots les plus significatifs, les 20 premières lignes ou 20% du document (le plus petit des deux) et d'autres informations. On retrouve également la pondération, le nombre de termes trouvés, le degré d'adjacence et parfois un extrait. Un point fort intéressant: tous les paramètres de recherche sont mis en surbrillance dans l'extrait ce qui facilite l'évaluation des résultats retournés.

Cet outil est incontournable pour le repérage d'information dans le Web. Il constitue probablement celui dont les résultats de sortie sont les mieux présentés. Nous lui accordons le maximum, quatre étoiles (****).

-
- URL: <http://www.lycos.com>
 - Type: Index

DOCUMENTATION

: Excellente (aide, exemples, FAQ, organisation)

SERVICE DE MISE À JOUR

- **Mode opératoire de mise à jour:** Robot et inscription
- **Objets répertoriés:** pages W3, Gopher et FTP
- **Information retenue:** Titre, entêtes, URL, 100 mots les plus significatifs du document, liens hypertextuels (citation) et les vingt premières lignes.
- **Taille:** Il y a deux bases. La première (petite) est constituée d'environ 500,000 références alors que la seconde (grande) est constituée d'environ 10 millions d'entrée.
- **Fréquence de mise à jour:** Hebdomadaire

SERVICE DE REPÉRAGE

- **Moteur de recherche:** PURSUIT
- **Opérations:** pseudo-et et pseudo-ou (ALL/ANY), pseudo-négation
- **Zones d'index:** Aucune
- **Remarque:** Bien qu'a priori, Lycos semble offrir peu, en termes de fonctionnalités de repérage, il

utilise un moteur basé sur la pondération de toute l'information répertoriée. Par exemple, le poids accordé à un mot sera plus grand si celui-ci est dans le titre ou dans les premières lignes du texte. En conséquence, Lycos pallie l'absence d'opérateurs de repérage et de zones d'index par une pondération. De plus, plusieurs modalités de repérage sont offertes (loose, fair, good, close, strong match).

SERVICES DE PRÉSENTATION DES RÉSULTATS

- **Format:** Trois formats disponibles
 - **Limite du nombre de résultats:** Oui
 - **Paramètres de recherche affichés:** Oui Seuil minimum de pondération: Oui
 - **Description** complète: Titre, pondération, nombre de mots trouvés/nombre de mots recherchés, le degré d'adjacence des termes de recherche, URL, extrait.
 - **Remarque:** Les paramètres de recherche sont en surbrillance dans les résultats (mots en contexte)
 - **Évaluation:** Très bon (****)
-

5.1.6. OPEN TEXT INDEX

C'est probablement l'outil de recherche qui exploite le mieux, avec Harvest Broker, la structure logique des documents. Cette base contient l'intégralité des pages répertoriées. La documentation est très bien faite. Plusieurs fonctions de recherche sont disponibles. L'utilisateur peut spécifier une zone tels le titre, la première en-tête, l'URL ou l'extrait pour le repérage ou tout simplement explorer le texte intégral. De plus, sont présents les opérateurs booléens, l'opérateur de négation, l'adjacence et la proximité.

Bien que n'offrant qu'un format en sortie, Open Text présente les résultats adéquatement. Ceux-ci comprennent, en plus du titre, la pondération, la taille et une description. De plus, à chacun des résultats, l'utilisateur peut demander la recherche de pages similaires tout comme Harvest Broker. Il est également possible d'obtenir les lignes du document primaire dans lesquelles apparaissent les mots recherchés. C'est le seul outil qui, à notre connaissance, offre cette dernière fonctionnalité.

Cet outil de recherche nous apparaît très intéressant. Pour ceux qui se préoccupent de l'exploitation de la structure logique des documents, nous le leur conseillons vivement. Nous lui accordons quatre étoiles (****).

-
- URL: <http://www.opentext.com>
 - Type: Index

DOCUMENTATION

- Excellente (aide, exemples)

SERVICE DE MISE À JOUR

- **Mode opératoire de mise à jour:** Robot et inscription
- **Objets répertoriés:** pages W3
- **Information retenue:** Texte intégral et la structure logique du document HTML
- **Taille:** 1 million de documents W3
- **Fréquence de mise à jour:** Hebdomadaire

SERVICE DE REPÉRAGE

- **Moteur de recherche:** PAT
- **Opérations:** Opérateurs booléens AND, OR, BUT NOT, opérateurs d'adjacence (followed by) et de proximité (near), phrase, troncature
- **Zones d'index:** Titre, première entête, URL, mots clés, liens hypertextuels (URL citation), extrait ou texte intégral

SERVICE DE PRÉSENTATION DES RÉSULTATS

- **Format:** Un seul
 - **Limite du nombre de résultats:** Oui
 - **Paramètres de recherche affichés:** Oui
 - **Description:** Titre, pondération, taille, URL, extrait
 - **Autres fonctions:** *Find similar Pages*
 - **Évaluation:** Très bon (****)
-

5.1.7. WEBCRAWLER

Cet outil de recherche est également très populaire dans le Web. La documentation est correcte, les fonctions de recherche sont limitées (opérateurs pseudo-booléens) et la présentation des résultats est plus que succincte (bien que l'on puisse spécifier le nombre maximum de résultats). Malgré tout, ce site doit sa popularité à sa simplicité. Un outil à employer par les néophytes et les réfractaires d'opérateurs de recherche.

Nous lui accordons une étoile (*) .

URL: <http://webcrawler.com>

- **Type:** Index

DOCUMENTATION

Bonne (aide, exemples)

SERVICE DE MISE À JOUR

- **Mode opératoire de mise à jour:** Robot et inscription
- **Objets répertoriés:** pages W3, Gopher et FTP
- **Information retenue:** Titre et plein texte
- **Taille:** 150,000 documents W3
- **Fréquence de mise à jour:** Environ aux six semaines

SERVICE DE REPÉRAGE

- **Moteur de recherche:** ?
- **Opérations:** pseudo-ET et pseudo-OR (ALL/ANY), sous-chaîne
- **Zones d'index:** Aucune
- **Ce qui n'est pas offert (documenté):** adjacence, distinction minuscule/majuscule, phrase

SERVICE DE PRÉSENTATION DES RÉSULTATS

- **Format:** Un seul
 - **Limite du nombre de résultats:** Oui
 - **Paramètres de recherche affichés:** Oui
 - **Description:** Pondération et titre
 - **Évaluation:** pauvre (*)
-

5.1.8. THE WWW WORM

Cet outil de recherche sert principalement à répertorier des citations. La base de données contient l'intégralité des pages Web répertoriées. La documentation est bien faite et comprend des exemples de stratégies de recherche.

Une des particularités de cet outil est d'offrir la possibilité d'obtenir des références croisées. En plus du texte intégral, la base contient des index sur les URL, les liens hypertextuels et le titre. Les fonctions de repérage offrent des opérateurs pseudo-boulogiens ainsi que des expressions régulières selon la terminologie de egrep (sous Unix).

L'utilisateur peut choisir le nombre maximum de résultats mais un seul format de sortie est disponible. Ce format, assez succinct, comprend le titre et l'URL d'un document dans lequel le lien repêché apparaît (citation).

Un site intéressant à visiter pour ses références croisées mais dont la présentation des résultats est plutôt déficiente. Nous lui accordons deux étoiles (**).

WWW WORM

- URL: <http://www.cs.colorado.edu/home/mcbryan/WWW.html>
- Type: Index

DOCUMENTATION

- Excellente (aide, exemples)

SERVICE DE MISE À JOUR

- **Mode opératoire de mise à jour:** Robot et inscription
- **Objets répertoriés:** pages W3
- **Information retenue:** URL, liens hypertextuels (citation) et titre
- **Taille:** 3 millions de références
- **Fréquence de mise à jour:** Variable

SERVICE DE REPÉRAGE

- **Moteur de recherche:** ?
- **Opérations:** pseudo ET et OU (ALL/ANY) et expression régulière (egrep d'UNIX)
- **Zones d'index:** URL, liens hypertextuels (citation), titre et le nom du fichier HTML
- **Ce qui n'est pas offert (documenté):** opérateurs de proximité, d'adjacence, la troncature, la distinction minuscule/majuscule, phrase
- **Remarque:** un mot d'une requête doit comporter au moins trois caractères

SERVICE DE PRÉSENTATION DES RÉSULTATS

- **Format:** Un seul
- **Limite du nombre de résultats:** Oui
- **Paramètres de recherche affichés:** Oui
- **Description:** Titre et une référence dans laquelle le document est citée
- **Évaluation:** Passable (**)

5.2. Les outils de recherche de type répertoire

5.2.1. CLEARINGHOUSE FOR SUBJECT-ORIENTED INTERNET RESSOURCE GUIDES

Fonctionnant comme un catalogue, cet outil de recherche contient environ 400 guides, classés dans environ une douzaine de catégories, portant sur les ressources de l'Internet. La documentation (What's New) est plutôt succincte.

Les guides ont été rédigés principalement par des étudiants en bibliothéconomie et science de l'information de l'Université du Michigan. Ils sont écrits en HTML ou en format texte. Les utilisateurs sont encouragés à proposer leur propre document. Ce tte base offre également un outil de recherche rudimentaire pour les guides écrits uniquement en format texte.

Un site intéressant pour les ressources de l'Internet mais qui n'utilise pas suffisamment les possibilités d'HTML et des liens hypertextuels. Nous lui accordons une seule étoile (*).

-
- URL: <http://www.lib.umich.edu/chhome.html>
 - Type:

Répertoire

DOCUMENTATION

Faible

SERVICE DE MISE À JOUR

- **Mode opératoire de mise à jour:** Éditeur
- **Objets répertoriés:** Guides écrits en HTML ou en format texte
- **Information retenue:** Les guides
- **Taille:** 400 guides
- **Fréquence de mise à jour:** variable (2, 5 ou 8 mois)

SERVICE DE REPÉRAGE

- **Moteur de recherche:** VERONICA (pour les textes dans les menus Gopher) Répertoires: 12 catégories
- **Remarque:** assez décevant comme moteur de recherche, il est préférable de travailler directement dans les répertoires.
- **Évaluation:** Pauvre (*)

5.2.2. GALAXY (EINET)

Cette base de données W3 est divisée en une douzaine de catégories et est maintenue par des éditeurs (*guest editors*). Comme dans l'exemple d'InfoSeek, on peut qualifier cette base d'hybride répertoire-index car il est possible d'utiliser une recherche par index. La documentation est complète et excellente.

C'est un des rares outils de recherche de type répertoire qui permet des requêtes complexes avec des opérateurs booléens. Les fonctions de troncature et de mots apparentés sont également disponibles. De

plus, on peut spécifier une recherche sur des zones tels le titre et les liens hypertextuels ou encore sur tout le texte. On peut également préciser la nature de l'objet recherché tels les titres de gopher, les ressources Telnet ou les pages de Galaxy.

L'utilisateur a le choix parmi trois formats de sortie. Le format de type long est très riche en information. En plus du titre, de la pondération et de la taille, ce format contient également un extrait du document, les mots les plus fréquents (avec leur occurrence), des en-têtes, la date du document et plus encore. Certainement un des sites qui présentent le mieux, avec Lycos, les résultats. De notre point de vue, c'est un des meilleurs outils de type répertoire. La diversité des services offerts est impressionnante et nous lui accordons quatre étoiles (****).

-
- URL: <http://galaxy.einet.net/galaxy.html>
 - Type:

Répertoire et index

DOCUMENTATION

- Très bonne (aide, exemples, etc.)

SERVICE DE MISE À JOUR

- **Mode opératoire de mise à jour:** Robot et inscription
- **Objets répertoriés:** pages GALAXY W3, Newsgroups, FTP, What's New, Gopher, Telnet et plus encore
- **Information retenue:** ?
- **Taille:** 100,000 documents
- **Fréquence de mise à jour:** ?

SERVICE DE REPÉRAGE

- **Répertoire:** 11 catégories
- **Moteur de recherche:** ?
- **Opérations:** opérateurs booléens complet (AND, OR, NOT), phrase, mots apparentés, troncature.
- **Zones d'index:** URL, titre, extrait et première entête

SERVICE DE PRÉSENTATION DES RÉSULTATS

- **Format:** Trois
- **Limite du nombre de résultats:** Non
- **Paramètres de recherche affichés:** Oui
- **Description:** Titre, pondération, taille, extrait, mots fréquents (avec leur fréquence), outline, source,

type et dates de création et d'indexation.

- **Évaluation:** Très bon (****)
-

5.2.3. WHOLE INTERNET CATALOG

Une des particularités de ce catalogue est la possibilité de consulter également les bases de données Yahoo et WebCrawler à partir d'une interface unique. The Whole Internet Catalog est constitué d'environ une douzaine de sujets, chacun divisé en d'au tres collections (il y a trois niveaux hiérarchiques). La méthode de l'éditeur est, là aussi, également utilisée. Les utilisateurs sont encouragés à soumettre leurs idées. La documentation est bonne.

Les liens (plus de 2,000) sont choisis dans les listes *Hot List* ou *What's New*. Il n'y a pas de fonctions de recherche. Il faut connaître, comme la plupart des bases de type répertoire, la nomenclature choisie. A chaque lien est as socié un compteur qui calcule le nombre d'accès. Nous vous suggérons fortement cette base pour connaître les dernières nouveautés du Web et nous lui attribuons deux étoiles (**).

- URL: <http://nearnet.gnn.com/wic/>
- **Type:**

Répertoire

DOCUMENTATION

- Bonne

SERVICE DE MISE À JOUR

- **Mode opératoire de mise à jour:** Éditeur
- **Objets répertoriés:** Hot List, What's New
- **Information retenue:** ?
- **Taille:** 2,000 références
- **Fréquence de mise à jour:** ?

SERVICE DE REPÉRAGE

- **Répertoires:** 14 catégories
 - **Évaluation:** Passable (**)
-

5.2.4. THE WORLD WIDE WEB VIRTUAL LIBRARY

Cet outil de recherche épouse parfaitement la réalité de l'Internet. Il est complètement distribué sur plusieurs sites. Plus de cinquante sujets sont maintenus indépendamment par différents collaborateurs qui font office d'éditeurs. Cette base est très riche en information. Par contre, la documentation est déficiente.

Plusieurs classifications sont disponibles. Une d'entre elles ordonne les sujets tout simplement par ordre alphabétique. Il est également possible d'obtenir les sujets selon la classification LC (Library of Congress). Par contre, aucune fonction de recherche n'est offerte.

Cet outil est plus difficile à utiliser comparativement à Yahoo et Galaxy. Cet état de fait est probablement dû au manque d'homogénéité dans la présentation des pages produites par les différents collaborateurs. Néanmoins, un site fort intéressant et nous lui accordons deux étoiles (**).

-
- URL: <http://www.w3.org/hypertext/DataSources/bySubject/Overview.html>
 - Type:

Répertoire

DOCUMENTATION

- Inexistante

SERVICE DE MISE À JOUR

- **Mode opératoire de mise à jour:** Éditeur
- **Objets répertoriés:** ?
- **Information retenue:** ?
- **Taille:** ?
- **Fréquence de mise à jour:** ?

SERVICE DE REPÉRAGE

Répertoires: Plusieurs types de classification sont offertes (dont celle de la Library of Congress)

- **Remarque:** Intéressant comme site mais difficile à s'orienter.
- **Évaluation:** passable (**)

5.2.5. YAHOO

Probablement l'outil de recherche de type répertoire le plus populaire dans le Web. Plus d'une vingtaine de catégories pouvant contenir jusqu'à quatre niveaux hiérarchiques.

Des fonctions de recherche sont également disponibles de manière à spécifier le titre, l'URL ou la description d'un document. L'outil de recherche est assez rudimentaire (opérateurs pseudo-booléens, pseudo-opérateurs de proximité). La documentation est très bonne et permet à l'utilisateur de naviguer correctement dans le catalogue.

L'utilisateur peut en outre spécifier le nombre maximum de résultats. Il y a un seul format de présentation. Notons toutefois que les paramètres de recherche sont mis en surbrillance dans les résultats. Un outil de recherche de type répertoire qui est certes intéressant. Nous lui attribuons trois étoiles (***)

-
- URL: <http://www.yahoo.com>
 - Type:

Répertoire et index

DOCUMENTATION

- Très bonne (aide, exemples, etc.)

SERVICE DE MISE À JOUR

- **Mode opératoire de mise à jour:** Robot et inscription
- **Objets répertoriés:** pages W3 et Reuters Newswire
- **Information retenue:** ?
- **Taille:** 66,000 références
- **Fréquence de mise à jour:** Journalière

SERVICE DE REPÉRAGE

- **Répertoire:** 14 catégories
- **Moteur de recherche:** ?
- **Opérations:** sous-chaîne, mot exact, distinction minuscule/majuscule, pseudo ET et OU (ALL/ANY).
- **Zones d'index:** Aucune

SERVICE DE PRÉSENTATION DES RÉSULTATS

- **Format:** Un seul
 - **Limite du nombre de résultats:** Non
 - **Paramètres de recherche affichés:** Oui
 - **Description:** Catégorie, titre, extrait
 - **Remarque:** Il n'y a aucune pondération mais les mots de recherche sont mis en surbrillance.
 - **Évaluation:** Bon (***)
-

EN CONCLUSION

Nous avons vu que les outils de recherche diffèrent dans leur type, leurs services de mise à jour, de repérage et de présentation des résultats. La nature des objets répertoriés et l'information retenue dans les bases constituent les fondements du système d'information et ont un impact direct sur la qualité du repérage et les besoins de l'utilisateur. La qualité du repérage s'évalue en fonction des services offerts pour formuler nos requêtes. Enfin, il faut que les résultats obtenus contiennent suffisamment d'information pour que l'utilisateur puisse en déterminer la pertinence.

On a vu également que des services, tels les opérateurs de comparaison et l'historique des recherches, sont absents dans les services de repérage des outils de recherche contrairement à ce qui existe, en général, en informatique documentaire traditionnelle. Il sera intéressant d'observer l'évolution des services offerts par les outils de recherche à cet effet.

Plusieurs critères d'évaluation n'ont pas été abordés. Ceux-ci, croyons-nous, devraient faire l'objet de travaux futurs. Un élément fort important qui devra être étudié, à notre avis, est la gestion des caractères diacritiques dans les outils de recherche. On connaît déjà les problèmes causés par une mauvaise gestion de ceux-ci en informatique documentaire.

Suite aux résultats obtenus de notre évaluation d'outils de recherche, nous avons quelques commentaires et suggestions à émettre. Malgré sa faible évaluation, l'outil de recherche WebCrawler serait un bon point de départ pour le néophyte vu sa simplicité d'utilisation. Pour exploiter la structure logique des documents, l'outil de recherche Open Text Index Web constitue, avec ses documents intégraux, un site très intéressant. Bien entendu, les classiques tels Lycos, Yahoo et InfoSeek sont à visiter. En terminant, nous vous conseillons les outils de recherche Galaxy et Harvest Broker: de notre humble point de vue, elles se démarquent avantageusement des autres bases en termes fonctionnels.

Une dernière mise en garde: toute évaluation et comparaison d'outils de recherche, incluant l'étude présente, constitue une photo prise à un moment bien précis. Contrairement au monde des logiciels de bureautique qui est beaucoup plus stable, l'évolution constante du Web rend ces comparaisons rapidement caduques. Par contre, les critères établis au cours de cette étude sont à notre avis beaucoup moins éphémères.

BIBLIOGRAPHIE

Courtois, Martin P., William M. Baer, and Marcella Stark. "Cool Tools for searching the Web: A Performance Evaluation." *Online* (novembre 1995): 15-32.

Koster, Martijn. Robots in the Web: threat or treat? 1995. URL: <http://web.nexor.co.uk/mak/doc/robots/robots.html>.

Lanteigne, Diane. BLT-6451 Recherche documentaire automatisée 1: Présentation du travail individuel. École de bibliothéconomie et des sciences de l'information: 1995.

Leighton, H. Vernon. *Performance of Four World Wide Web (WWW) Index Services: Infoseek, Lycos, WebCrawler and WWWorm*. 1995. URL: <http://www.winona.msus.edu/services-f/library-f/webind.htm>.

Liu, Jian. *Understanding WWW Search Tools*. Septembre 1995. URL: <http://www.indiana.edu/~librcsd/search/>.

Matrix of WWW Indices: A comparison of Internet indexing tools. 1995. URL: <http://www.sils.umich.edu/~fprefect/matrix/>.

Notess, Greg R. "Searching the World-Wide-Web: Lycos, WebCrawler and More." *Online* (1995): 48-53.

Paul, Kathryn, and Kathleen Matthews. *Tools and Techniques for Searching the Web: Subject Trees and Search Engines*. 1995. URL: http://burns.library.uvic.ca/KWM_Post_CLA.html.

Randal, Neil. "Search Engines: Powering Through the Internet." *PC Computing* (septembre 1995): 165-168.

Scales, B. Jane, and Elizabeth Caulfield Felt. "Diversity on the World Wide Web: Using Robots to Search the Web." *Library Software Review* 14, no. 3 (septembre 1995): 132-136.

Stanley, Tracey (ECLTSS@lucs-01.novell.leeds.ac.uk). Searching the World Wide Web with Lycos and InfoSeek. Computing Service, University of Leeds, UK: 24 octobre 1995. URL: <http://www.leeds.ac.uk/ucs/docs/fur14/fur14.html>.

Winship, Ian R. (ian.winship@unn.ac.uk). World Wide Web searching - an evaluation. Information Services Department, University of Northumbria at Newcastle, UK, 1995. URL: <http://www.bubl.bath.ac.uk/BUBL/IWinship.html>.

Notes:

Le taux de rappel se définit comme étant le nombre de documents pertinents sélectionnés divisé par le nombre de documents pertinents dans la base.

Le taux de précision se définit comme étant le nombre de documents pertinents sélectionnés divisé par le nombre de documents sélectionnés.

Le taux de bruit se définit comme étant la différence entre 1 et le taux de précision.

Le taux de silence se définit comme étant la différence entre 1 et le taux de rappel.

En général, à quelques exceptions près tel Galaxy, ces index se limitent aux rubriques et sous-rubriques du plan de classification.

Sauf si le robot suit un protocole le soumettant à certaines règles (voir les articles de Martijn Koster à cet effet).

Cette norme est appelée Internet Anonymous FTP Archives (IAFA) for publishing information on the Internet with anonymous FTP.

La documentation n'est pas explicite sur la méthode utilisée pour calculer les mots les plus significatifs.

Quoique la documentation ne fournisse aucune explication concernant la définition d'une ligne (car il n'y a pas de notion de ligne en HTML)..

Le cas de Lycos est intéressant. Les chiffres, selon les études, varient de 3.6 à 10 millions de documents.

Les formulaires HTML permettent de contrôler la saisie des données à l'aide de cases à cocher, de boutons de radios, de zones de saisie de texte, etc. Les outils de recherche qui n'utilisent pas de formulaires sont de plus en plus rares.

InfoSeek est l'exception. Pour les membres enregistrés à InfoSeek, il est possible d'utiliser des comparaisons de dates. Malheureusement ce service n'est pas offert gratuitement.

Selon l'outil de recherche, un extrait peut correspondre soit aux premières lignes du document, soit aux entêtes du document, soit aux titres d'un menu gopher, etc.

WAIS: Wide Area Information Service.

Dans plusieurs études nous retrouvons cette caractéristique. Notons toutefois que celle-ci correspond tout simplement au principe d'adjacence.

Ce choix s'effectue dans l'interface de saisie de données (voir display option)..

Voir à cet effet la bibliographie ci-jointe.

Les types d'objets sont classifiés de la manière suivante: organisation, document, service, user, software, etc.

La proximité se définit à l'intérieur de 100 mots.

Ceci étant dit, il faut quand même prendre un certain recul sur les commentaires émis. Lycos, tout comme InfoSeek, est à but lucratif.

Les documents informatiques: pour une classification efficace

par

André Vigneau

Cursus vol.1 no 2 (avril 1996)

Cursus est le périodique électronique étudiant de l'École de bibliothéconomie et des sciences de l'information (EBSI) de l'Université de Montréal. Ce nouveau périodique diffuse des textes produits dans le cadre des cours de l'EBSI.

ISSN 1201-7302

cursus@ere.umontreal.ca

URL: <http://www.fas.umontreal.ca/ebsi/cursus/>

L'auteur

L'auteur est titulaire depuis 1994 d'une maîtrise en bibliothéconomie, option archivistique, de l'EBSI de l'Université de Montréal. Il est depuis deux ans archiviste à la Société de transport de la Communauté urbaine de Montréal (STCUM) où il participe activement à l'organisation des archives définitives. L'auteur a réalisé, sous la direction de la STCUM, deux publications relatives aux fonds d'archives de cette entreprise. Par ailleurs, l'auteur est responsable de la saisie des sommaires et des résumés de la revue Archives (AAQ) sur le site WEB de l'Association des archivistes du Québec (en construction).

André Vigneau, *Les documents informatiques: pour une classification efficace* a été écrit à l'hiver 1994 dans le cadre d'un cours donné par le professeur Carol Couture.

Pour joindre l'auteur: vigneaua@ere.umontreal.ca

Droits d'auteur

Tout texte demeure la propriété de son auteur. La reproduction de ce texte est permise pour une utilisation individuelle. Tout usage commercial nécessite une permission écrite de l'auteur.

INTRODUCTION

L'irruption massive de la micro-informatique dans les entreprises et organismes à partir du début des années 80 a radicalement et irrévocablement changé la nature du travail de production et de soutien à celle-ci. Bien sûr, l'informatique existe avant cette décennie. Elle est toutefois le fief d'entreprises de forte taille ou d'organismes gouvernementaux et académiques qui doivent manipuler des masses substantielles de données et de dossiers sur des ordinateurs centraux ou des mini-ordinateurs. Ceux qui saisissent les données le font sur des terminaux simples et ne peuvent contrôler en aucune façon le traitement des données, puisque celui-ci est exclusivement effectué par le personnel technique affecté spécifiquement à cette tâche. De plus, les données enregistrées sont pour la plupart de nature numérique, utilisées surtout à des fins de gestion et de statistiques.

En 1981, IBM lance l'ordinateur personnel PC/XT (Gookin, 1990, p. 5). Enfin la puissance des ordinateurs centraux ou des mini-ordinateurs parvient aux entreprises, aux organismes et aux particuliers à des coûts acceptables. Cette démocratisation de l'informatique amène, pour le meilleur et pour le pire, une démocratisation de la création et de l'utilisation de l'information. Pour le meilleur, parce qu'elle simplifie de beaucoup certaines tâches répétitives, qui peuvent désormais être automatisées, et qu'elle permet une accélération des activités de l'organisme. Pour le pire, parce qu'on commence à présumer du caractère d'infailibilité générale de l'ordinateur, trait attribuable à des exploits inattaquables en matière de puissance de calcul. Présomption d'infailibilité par exemple à retracer, par une fonctionnalité de recherche intégrée, un document précis sur les 5 000 que peut contenir un disque rigide de 250 mégaoctets. Pour empêcher le repérage, il suffit que le nom tapé ne soit pas identique à celui recherché pour que la recherche n'aboutisse pas et que les difficultés commencent.

Cette *faillibilité* du système de repérage des documents (par une opération de recherche logique, nous ne sommes pas capables d'accéder au document physique), nous fait apparaître les limites de ce cerveau synthétique. Si fulgurante que soit sa capacité de traitement des données qu'il reçoit ou possède déjà en mémoire, l'ordinateur est totalement dépourvu d'intelligence et d'intuition. Ces lacunes relèguent ainsi notre boîte de verre et de plastique au rang d'un simple outil à mémoire. Fabriqué par l'humain, pour l'humain, il est toujours en mesure de faire précisément ce qu'on lui demande, mais jamais ne fait-il plus.

Dans la réalité administrative quotidienne, où de plus en plus tout doit être mesuré et justifié à l'aune de la productivité et de la rentabilité, il devient difficile, sinon impossible, d'accepter que des documents créés et enregistrés sur ordinateur ne puissent être repérables là même où ils sont créés. Face à toutes les fonctionnalités techniques de l'outil, et au constat que seules l'intelligence et la créativité humaines peuvent organiser efficacement *tous* les documents consignés sur les différentes mémoires d'un ordinateur, il nous apparaît évident qu'un travail *sur* l'ordinateur doit précéder un travail *par* ordinateur. Or l'information la plus immédiatement accessible à quiconque utilise un ordinateur pour son travail est un ensemble de représentations symboliques du contenu de l'ordinateur: lettres, chiffres ou icônes. Le travail *sur* l'ordinateur consistera donc à agencer ces représentations symboliques pour qu'elles facilitent le repérage de l'information à son niveau le plus "superficiel", c'est-à-dire à celui de l'écran. Il s'agira en fait de créer une classification efficace pour les documents informatiques. Un document bien classifié pourra être repéré beaucoup plus facilement qu'un autre qui ne l'aurait pas été avec la même rigueur.

Notre présentation comporte donc deux parties. Une première où nous rappelons les grands principes qui guident la classification et une deuxième où nous tentons de concilier la rigueur de la classification avec la puissance technique de l'informatique, tout en étudiant la pertinence d'appliquer un plan de classification déjà existant pour des documents sur support papier à un ensemble de documents informatiques.

1. NATURE ET FONCTION DE LA CLASSIFICATION

De la manière la plus générale possible, la classification est le processus intellectuel par lequel des choses, des êtres animés ou des idées peuvent être répartis en catégories, pour éventuellement les différencier davantage, si la quantité l'exige. Au départ très larges et peu nombreuses, tout en étant très populeuses parce qu'elles comportent peu de critères d'exclusion, ces catégories deviennent de plus en plus étroites et nombreuses en quantité, et peu peuplées en termes de spécimens, parce que leurs critères d'exclusion ont fortement augmenté. Le résultat de ce processus fera en sorte que là où régnait le désordre et l'incompréhension, émerge une structure beaucoup plus apparente qui nous livre déjà une partie de l'information, encore une fois dans sa manière la plus large. Il est évident que ce processus ne peut *refaire* l'ordre naturel des choses, c'est-à-dire celui qui est le fruit d'une évolution qui s'étend sur des milliards d'années. Là où toutefois la classification peut changer le cours des choses, c'est dans le domaine du travail de l'homme, et plus spécifiquement sur les représentations de ce travail, c'est-à-dire sur les documents produits et reçus dans le cadre de ses activités. "Our society recognizes some documents as records, because they carry out or document transactions². Because records are accepted within this social and legal framework as evidence of an act, they are retained in record-keeping systems designed to serve the needs of the people and organizations that created or received them" (Bearman, 1993, p. 17).

Pourquoi classifier? Ce n'est pas tout de diviser en grandes ou petites unités semblables l'ensemble des documents produits par une personne physique ou morale, pour ensuite en effectuer le classement et puis le rangement dans le bon tiroir du classeur pertinent. La classification ne peut pas servir qu'à la désignation du meilleur lieu d'entreposage pour un document.

L'administration courante est une chose éminemment dynamique, où l'information créée précisément et facilement doit pouvoir être repérée, et utilisée, exactement de la même manière, c'est-à-dire avec précision et facilité. La finalité de la classification n'est donc pas de diviser pour diviser, jusqu'à l'ultime atomisation. Nous pensons plutôt que la finalité en est une de repérage, et idéalement de repérage rapide. A ce propos, Jean-Yves Rousseau mentionne justement qu'"Un système de classement bien adapté aux besoins accélère le classement des documents et *leur repérage subséquent* [je souligne]" (Rousseau, 1979, p. 4). D'autre part, si dans ses commentaires sur la classification, Theodore R. Schellenberg situe son intervention au niveau gouvernemental, il ne nous apparaît pas déplacé de les transposer sur l'ensemble des champs d'application de l'archivistique: "In maintaining its records for current use, a government agency is concerned mainly with keeping them *so that they can be found quickly when they are wanted* [je souligne]" (Schellenberg, 1956, p. 52). Enfin, on ne saurait être plus clair sur la finalité de la classification en lisant les quelques lignes suivantes, extraites du Guide de gestion des archives de l'Université de Montréal: "C'est dans le but d'assurer une utilisation maximale de l'information contenue dans les documents que les administrations se dotent de systèmes qui augmentent l'efficacité administrative en facilitant la classification et le rangement des documents et, *surtout, en rendant leur repérage plus rapide* " [je souligne]" (Université de Montréal, 1989, p. 1).

Une fois établie la pertinence de la classification, il s'agit de déterminer les principes généraux qui en balisent le processus. En reprenant le parallèle avec l'idée de la classification universelle mentionnée plus haut et en l'appliquant aux documents, nous devons établir des catégories générales dans lesquelles des caractéristiques croissantes de spécificité et d'exclusivité permettent de "caser" à *un* endroit donné, un document donné.

Quels sont alors les axes qui gouvernent l'établissement du plan de classification? Pour les déterminer, nous procéderons en utilisant un exemple particulier, celui d'un organisme. Chemin faisant, nous rejetterons graduellement certains modes de regroupement des documents, pour progressivement restreindre notre choix aux axes les plus pertinents.

En examinant une entreprise dans son fonctionnement quotidien, et en analysant la documentation qu'elle produit pour se présenter aux yeux de sa clientèle, par exemple des rapports annuels ou des documents de promotion, nous pouvons saisir les liens organisationnels qui unissent, par exemple la direction générale à la division de la comptabilité, à celle de la recherche et du développement, ou encore à celle des communications. Les documents relatifs à une année fournissent

une photographie sans trucage de ce qui *est*, maintenant (voir Figure 1).



Figure 1: les liens organisationnels

Or, la consultation de ces mêmes documents sur une période de dix ans par exemple, révèle selon toute probabilité, à moins que l'organisme soit d'une stabilité à toute épreuve, des changements substantiels quant à sa structure et à ses divisions administratives. Une même activité peut ainsi passer d'une division à une autre, et cette dernière division peut fort bien disparaître quelques années après. L'établissement d'un plan de classification sur une structure organisationnelle risque de compliquer inutilement et en permanence les choses. D'ailleurs, cette pratique ne semble pas recommandée: "[...] les structures administratives passent avec le temps et se modifient de telle sorte qu'on ne saurait fonder sur elles un cadre de classement ayant suffisamment de permanence" (Archives Nationales de France, 1970, p. 191). D'autre part, le théoricien Schellenberg ne semble pas lui aussi particulièrement enchanté par cette perspective: "Such a division into organizational classes is possible and advisable only in governments whose organization is stable, and whose functions and administrative processes are well-defined." (Schellenberg, 1956, p. 59).



Figure 2: le déplacement d'une activité d'une division à une autre.

Deuxième axe possible, celui constitué par les sujets. Dans l'organisme mentionné plus haut, ce processus de classification se traduirait dans la réalité par une division presque infinie en autant de sujets que l'organisme peut générer ou recevoir dans le courant de ses activités. Nous pourrions retrouver dans ses filières, des dossiers aussi diversifiés et disparates que "Inondation", "Conception du rapport annuel", "Entretien des ordinateurs", "Party de la compagnie" ou "Employés surnuméraires". En plus de la relative imprécision de ces "étiquettes" lors de leur création, leur utilisation n'est pas garante d'efficacité. En effet, si le roulement du personnel est assez élevé, la connotation attachée à une étiquette déterminée peut varier d'un employé à un autre. Le résultat sera que, pour un document correctement classé par une première personne, le même document, se voyant attacher une connotation différente par une deuxième personne, sera classé dans un tout autre dossier. Les conséquences à long terme sont facilement prévisibles: une confusion croissante et une impossibilité de retracer un document rapidement.

Cette approche sujet, fondamentale pour la bibliothéconomie, ne convient certes pas à l'archivistique, du moins pour les niveaux supérieurs de classification. En poursuivant cette logique, nous ne pouvons pas établir la structure des activités de l'organisme puisqu'aucun lien n'existe entre un sujet et un autre, chacun vivant en parallèle, sans jamais rien révéler de la raison pour laquelle il existe. De nouveau, Pérotin nous éclaire sur ce point: "Dans cet ordre d'idées, ce que l'on peut faire de plus mauvais et de plus "anti-archivistique", c'est d'agir comme pourrait le faire - à juste titre, lui - un bibliothécaire et d'établir un grand tableau des *sujets* (ou plutôt des *objets*) possibles des différents documents, articles ou groupes d'articles, suivant lesquels on répartirait ensuite ceux-ci" (Archives Nationales de France, 1970, pp. 191-192).

Ce bref survol des possibilités de découpage symbolique des activités d'un organisme a fait ressortir les faiblesses d'une vision trop collée au *document*, au détriment de la *mission* de l'organisme. Et c'est justement par cette mission que nous sommes davantage en mesure d'élaborer une structure de classification. Dans les documents constitutifs d'un organisme, qu'il soit privé ou public, grand ou petit, quelques lignes sont réservées à la mention des raisons pour lesquelles l'organisme a été créé. Ces raisons lui sont spécifiques et particulières. Bien sûr, tout ce que l'organisme fait ne lui est pas exclusivement spécifique. Il y a les notions relatives à la gestion de la mission, c'est-à-dire tout le support logistique requis pour sa réalisation, et ceux plus spécifiques liés à ce qui donne à l'organisme ses caractéristiques de différenciation par rapport à un autre organisme. Une entreprise fera de la coupe de bois, une autre fournira des conseils en implantation de systèmes informatiques, une autre enfin transformera une matière brute en un produit raffiné. Toutes trois sont

différentes par leurs activités spécifiques et toutes trois se ressemblent, dans une certaine mesure, par leur support à ces activités spécifiques.

La mission d'un organisme est sans doute ce qui varie le moins dans son cycle de vie. C'est pour cela qu'un système de classification qui, à un premier niveau général, reflète davantage la mission que les documents, a plus de chance de favoriser le repérage rapide de l'information. Or que présente dans ses plus grandes lignes tout organisme sinon des fonctions auxquelles se rattachent un nombre variable d'activités, desquelles éventuellement émanent des documents. Comme nous en avons déjà fait état, ces fonctions peuvent s'intituler, à titre d'exemple, "Ressources humaines", "Trésorerie et finances", ou encore "Ventes et commercialisation". En archivistique, la division selon l'axe des fonctions semble être celle qui se mérite l'approbation la plus unanime:

"La systématique des fonctions administratives peut être ce [solide] fondement. Cette notion est malheureusement assez malaisée à cerner; aussi, plutôt que d'en tenter une définition théorique (telle que: "action de la collectivité publique dans un domaine déterminé" ou "type d'action de la collectivité publique"), paraît-il préférable de donner des exemples. Ainsi, la fonction judiciaire est connue de tout le monde; il en est de même de la fonction enseignante, de la fonction fiscale, de la fonction militaire, etc. C'est sur ces concepts majeurs, rangés dans un ordre logique (à la rigueur alphabétique) que l'on pourra établir des cadres de classement." (Archives Nationales de France, 1970, p. 192).

Pour sa part, Schellenberg, toujours en concentrant son regard sur les structures gouvernementales, que nous osons étendre à tout système de classification, public ou privé, affirme que:

"Thirdly, public records, as a rule, should be classified in relation to function. They are the result of function; they are used in relation to function; they should, therefore, be classified according to function" (Schellenberg, 1956, pp. 62-63).

Ayant déterminé le premier niveau de notre système de classification, soit celui des catégories alignées sur les grandes fonctions, les niveaux plus spécifiques seront l'objet de la démarche suivante. S'il a été possible de s'en tenir facilement à une seule directive pour notre premier niveau, pour les niveaux spécifiques, la tâche se complique quelque peu. En effet, au-delà des similitudes fonctionnelles entre des organismes assez différents (ressources humaines, comptabilité, etc.), la manière avec laquelle ils réalisent les actions et les activités afférentes à ces fonctions varie considérablement. L'un privilégiera les correspondances personnalisées et très étoffées pour s'adresser à une clientèle huppée; l'autre misera davantage sur des promotions grand public et tout en faisant ce que tous deux appellent des "relations publiques".

Pour nous donner un aperçu des axes par lesquels nous pouvons diviser les sous-catégories, nous reprenons succinctement la liste des modes de classement présentés par Couture et Rousseau (1982, pp. 114-116), à l'exclusion du mode "fonctionnel", déjà présenté plus haut:

- Sujet: principal, secondaire et tertiaire.
- Type: selon le document produit (factures, état de compte, directives, etc).
- Provenance: selon la personne physique ou morale responsable de la création du document.
- Classement mixte: combine aussi bien le classement par sujet que par type, ou encore il incorpore un autre mode de classement.

Pour sa part, Schellenberg (1956, p. 59) précise que:

"The broadest or primary classes may be established on the basis of the major functions of an agency, the secondary classes on the basis of activities, and the most detailed classes should comprise individual file units, or aggregates of file units, that are established in relation to transactions pertaining to persons, corporate bodies, places, or topics."

Comme nous le voyons, il n'y a pas de consensus établi sur les niveaux inférieurs, ou plutôt sur la structure secondaire et

tertiaire du système de classification. Toutefois il ne faut jamais perdre de vue le but premier de la classification, qui est le repérage rapide des documents. Et cette opération se faisant le plus souvent par des gens qui n'ont pas conçu le système, son utilisation doit en être d'autant plus simple. "Tout individu peut comprendre le système et s'en servir rapidement; il n'est pas la propriété de spécialistes" (Couture et Rousseau, 1982, p. 111).

Ce qui importe donc pour les niveaux secondaire et tertiaire est la cohérence dans les subdivisions. Par conséquent, il ne faudra pas confondre à un *même* niveau hiérarchique des activités, comme l'embauche du personnel, à des types de documents, comme les formulaires d'offres de services, ou les états financiers.

Pour pouvoir refléter les nouvelles activités de l'organisme, le plan de classification devra faire preuve d'une grande capacité d'adaptation et permettre l'ajout de nouveaux éléments. "Le bon système de classement *doit en outre être souple* de façon à tenir compte des futurs développements de l'organisme. Si la firme s'engage dans de nouvelles activités, le système de classement doit être en mesure de pouvoir les inclure sans qu'il soit nécessaire d'y apporter des modifications majeures. " (Rousseau, 1979, p. 7).

En conclusion à cette première partie, nous proposons que la classification est le processus par lequel un document cesse de "flotter" d'une manière imprécise dans le "vague administratif" pour trouver son point d'ancrage dans la réalité documentaire de l'organisme, représentée par le plan de classification, lui-même portrait de la vie quotidienne de l'organisme. Par son passage dans la grille du plan, le document se "fixe" à un endroit précis, réalisant ainsi le lien entre la vie organique et la vie documentaire (voir Figure 3).



Figure 3: le document se fixe dans la vie de l'organisme.

2. APPLICATION DE LA CLASSIFICATION AUX DOCUMENTS INFORMATIQUES

Après avoir présenté les caractéristiques de la classification nous tenterons de déterminer si un organisme qui possède déjà un plan de classification bien structuré pour la gestion de ses documents-papier peut utiliser ce même plan suite à une informatisation de plus en plus poussée de ses activités et opérations. Puis nous proposerons quelques moyens simples qui peuvent aider à la gestion à l'écran des documents produits ou reçus par l'organisme, en faisant une meilleure exploitation du système d'exploitation de l'ordinateur.

L'informatisation croissante de l'organisme amène surtout deux conséquences: une augmentation de la quantité d'informations consignées, et la possibilité désormais d'avoir des documents qui ne sont jamais autre chose que des documents électroniques et qui peuvent ne jamais connaître d'existence sur papier.

C'est sans doute justement le fait qu'un document informatique, qui est *avant toute chose* un document, donc une information consignée, puisse ne pas exister sur papier, donc sur un support traditionnel, qui laisse les archivistes perplexes. D'un côté, ils savent que l'information contenue dans l'ordinateur est le fruit d'une activité voulue par l'organisme; qu'elle possède une valeur certaine; que par sa présentation à l'écran, elle ressemble de plus en plus à une éventuelle copie papier du document. En revanche, ils semblent composer plus difficilement avec le fait que cette information consignée puisse ne jamais franchir l'écran de verre, tout en pouvant, par le biais de la télématique, faire le tour du monde derrière cet écran, et ne jamais se matérialiser en un objet qu'un administrateur, ou un archiviste, peut avoir, littéralement, entre les mains. Toujours, toutefois, cette information consignée conserve son statut de document.

Pour mieux comprendre cette nécessaire dissociation entre le support et l'information, reprenons un passage

particulièrement éclairant d'un article de Trudy Huskamp Peterson (1988, p. 84):

"Records (the institutional form of the generic "document") have three parts: a base, an impression upon the base, and information. The crux of the definition is that the information is fixed by the impression upon the base. The definition does not specify the type of base, nor the type of impression, nor the character of the information, nor the length of time it is fixed.

Bearing that definition in mind, it is clear that format makes no difference to the fundamental nature of a document or a record. So, too, format makes no difference to archival principles, for these principles relate solely to the fact that the product of the activity is a record. Further, the principles relate primarily to the selection, maintenance and use of the information in the records; practice relates to the base and means or recording (the impression) employed. "

Peterson nous amenant sur le terrain des principes de l'archivistique, peut-être pouvons-nous jeter un bref coup d'oeil du côté de son principe moteur, soit celui du respect des fonds, et plus particulièrement à son deuxième niveau, c'est-à-dire celui du respect de l'ordre interne. C'est ce deuxième niveau qui peut faire en sorte que les documents témoignent non seulement des activités, mais aussi de leur importance administrative dans l'organisme. Peterson nous éclaire ensuite sur l'articulation de la classification à l'informatique:

"Turning to the issue of arrangement and machine-readable records, provenance remains the most significant principle for the archivist. Who created the records is the fundamental question, and the answer varies just as it does with paper records. The key is that someone somewhere had to create the records to fulfill an organizational need, and that fact is the key to provenance" (Peterson, 1988, p. 86).

Dès lors, n'est-il pas possible d'harmoniser un plan de classification traditionnel, c'est-à-dire axé principalement sur le support papier, à la réalité intangible de l'informatique. Prenons l'exemple de la comptabilité. Avant l'arrivée de l'ordinateur, les documents comptables existent déjà sous des formes papier fortement caractérisées: grand livre, états financiers, journal des ventes, journal des débours, livre de caisse, etc. Que font les tableurs électroniques? Ils reproduisent à l'écran la feuille de papier, donc le type de document, sur laquelle le responsable de ces opérations consigne ses informations financières. Comme le dit Pieyns, "[...] nombre des archives nouvelles sur supports informatiques sont les successeurs sur un support différent des documents créés sur des supports traditionnels antérieurement à l'utilisation des ordinateurs." (Pieyns, 1988, p. 203).

Dans le cas précis de la comptabilité, une des différences fondamentales, en opposition à la feuille de papier, c'est que la feuille électronique comporte plusieurs fonctionnalités qui lui donnent une valeur ajoutée. Entre autres, enfouies sous les colonnes et les rangées, des formules mathématiques (déterminées par l'utilisateur) permettent à la feuille de se "calculer elle-même", si je puis dire. Mais en bout du compte, l'information utilisée est toujours de nature comptable; elle provient toujours d'une activité ou d'un ensemble d'activités reliées *directement* à l'entreprise; et elle s'insère toujours dans une hiérarchie administrative logique. La finalité du document, c'est-à-dire dans ce cas-ci faire état d'une activité comptable, demeure la même, elle ne change pas simplement parce que les opérations sont faites sur un ordinateur plutôt qu'à l'aide d'une simple calculatrice. Si la finalité et le créateur du document demeurent les mêmes, peu importe le support, la classification du document dans le plan ne devrait pas varier puisque le plan de classification a été fait non pas selon les supports mais selon les fonctions et que le document que nous avons dans les mains, ou sous les yeux, est une conséquence directe d'une activité reliée à une seule des grandes fonctions de l'organisme. Ainsi, à l'instar de Charles M. Dollar, nous ne pensons pas qu'une distinction doive s'appliquer parce que le document est sur support informatique: "New information technologies have not changed and nor are they likely to change the fundamental nature of records as evidence of actions and transactions with a specific context of creation and use." (Dollar, 1994, p. 449).

Ne pourrions-nous pas appliquer ce raisonnement à l'ensemble des documents informatiques créés ou reçus dans un organisme? Que ce soit du traitement de texte, de la conception graphique, du courrier électronique ou de la production audio-visuelle, le document n'a pas été créé simplement à des fins d'amusement. Le document a un but bien particulier,

qui est celui de contribuer à une activité, voulue et planifiée par l'organisme, et souvent, de témoigner de cette activité. L'information qu'il transmet par informatique est certes différente par sa nature même: elle est médiatisée, par opposition à la lecture "humaine" de documents papier, par exemple. Mais ce qui demeure le point le plus important est que ce que le document transmet reste toujours de l'information, peu importe le support.

Si nous examinons l'ensemble des documents, et par extension, l'ensemble des fonctions, nous constatons que leur représentation, c'est-à-dire le plan de classification, révèle plus d'informations que la somme des documents qu'il contient. Pourquoi? Parce que le plan de classification livre le contexte de la vie de l'organisme. Que ses activités soient consignées sur papier ou sur support informatique, leur contexte ne change pas: elles s'inscrivent toujours dans la vie de l'organisme. Elles ne connaissent pas d'existence parallèle, selon qu'elles soient informatisées ou non.

Le document individuel ne pouvant plus être différencié désormais par son support, lorsqu'il s'agit de sa classification, l'ensemble des documents d'un organisme, qu'ils soient sur papier ou sur informatique, peuvent à notre avis être "fondus" dans un seul plan de classification. Et non seulement pensons-nous que ce transfert du monde réel au monde virtuel est possible, mais aussi l'estimons-nous des plus souhaitables. A cet effet, nous reprenons de nouveau la recommandation de Couture quant à la simplicité du système de classement. "Tout individu peut comprendre le système et s'en servir rapidement. Il n'est pas la propriété des spécialistes" (Couture et Rousseau, 1982, p. 111). Alors si un plan de classification existe déjà pour les documents papier dans l'organisme, pourquoi en créer un nouveau qui ne s'appliquerait qu'aux documents informatiques? Si une telle démarche était entreprise, ce classement se ferait alors seulement sur la base du support. Nous pensons qu'à l'intérieur d'un organisme, ce système de classement pourrait s'avérer aussi menaçant que le classement par sujet pour les grands ensembles de documents. Pourquoi "faire compliqué" quand on peut "faire simple", et qu'au surplus tous s'y retrouvent?

Comment donc réaliser le transfert de la structure du plan de classification sur papier à la surface de l'écran de l'ordinateur? C'est précisément par les fonctionnalités du système d'exploitation que la création d'un plan de classification à l'écran peut se réaliser. Créés par l'utilisateur, et identifiés à l'aide de caractères alphanumériques qu'il leur a attribués, les "répertoires" ou les "dossiers" peuvent constituer les intitulés de chacune des grandes catégories du plan de classification. A partir du plan de classification sur papier, nous pouvons reprendre chacun de ses niveaux et l'appliquer à une hiérarchie de présentation à l'écran, peu importe la plate-forme utilisée. Précisons qu'il s'agit bien ici d'établir une correspondance identique entre le plan de classification de l'organisme et la structure de présentation à l'écran de tous les répertoires (principaux et subordonnés) et fichiers que l'organisme a saisi dans son système informatique. Pour illustrer notre propos, nous prendrons les fonctionnalités les plus usuelles des systèmes d'exploitation des plates-formes PC-DOS, Macintosh et, pour refléter la croissance du phénomène des réseaux, nous ajouterons les fonctionnalités du système UNIX, tel qu'utilisé sur le réseau dorsal de l'Université de Montréal.

Sur les plates-formes PC-DOS et Unix, la création de répertoires se fait par la commande <mkdir>, alors que sur la plate-forme Macintosh, le répertoire étant appelé un "dossier", la commande s'appelle "Nouveau dossier". A l'aide du plan de classification, nous reconstituons toute l'arborescence du plan par la création de répertoires, de sous-répertoires et de "sous-sous répertoires". (Voir figures 4 et 5).

```
tornado[388]___ /|> * pwd
/usagers/tornado/xyzxyzxy/CLASSIFICATION_UNIX
tornado[389]___ /|> * dir
total 22
drwx-----  9 xyzxyzxy bibliol      512 Apr 16 10:46 .
drwxr-xr-x  14 xyzxyzxy bibliol     1024 Apr 16 11:26 ..
drwxr-xr-x   8 xyzxyzxy bibliol      512 Apr 15 22:55 100-000_Gestion_generale
drwxr-xr-x   7 xyzxyzxy bibliol      512 Apr 15 18:56 200-
```

```

000_Gestion_des_droits
drwxr-xr-x  7 xyzxyzxy bibliol  512 Apr 15 22:47 300-
000_Gestion_financiere
drwxr-xr-x  6 xyzxyzxy bibliol  512 Apr 16 11:37 400-
000_Gestion_ress_humaines
drwxr-xr-x  7 xyzxyzxy bibliol  512 Apr 16 10:43 500-000_Production
drwxr-xr-x  5 xyzxyzxy bibliol  512 Apr 16 10:50 600-
000_Ventes_commercialisation
drwxr-xr-x  7 xyzxyzxy bibliol  512 Apr 16 12:04 700-
000_Recherche_Developpement
tornado[3]___ /|> * cd 100-000*

```

Figure 4: le premier niveau du plan de classification sur UNIX. Par la commande <cd 100-000*>, j'accède au niveau inférieur, présenté en Figure 5. Notez qu'il n'y a pas de caractères diacritiques actuellement sur UNIX. En annexe 3, je reprends toute la structure du plan, à partir de mon entrée dans le réseau dorsal de l'Université.

```

tornado[276]___ /|> * cd 100-000*
tornado[277]___ /|> * dir
total 8
drwxr-xr-x  8 xyzxyzxy bibliol  512 Apr 15 22:55 .
drwx----- 9 xyzxyzxy bibliol  512 Apr 16 10:46 ..
drwxr-xr-x  5 xyzxyzxy bibliol  512 Apr 15 22:56 100-
100_Conseil_administration
drwxr-xr-x  2 xyzxyzxy bibliol  512 Apr 15 16:24 100-
200_Direction_generale
drwxr-xr-x  2 xyzxyzxy bibliol  512 Apr 15 16:24 100-
300_Organisation_gest_controle
drwxr-xr-x  2 xyzxyzxy bibliol  512 Apr 15 16:26 100-
400_Approvisionnement
drwxr-xr-x  2 xyzxyzxy bibliol  512 Apr 15 16:27 100-
500_Communications_Rel_publicques
drwxr-xr-x  2 xyzxyzxy bibliol  512 Apr 15 16:34 100-
600_Ress_documentaires
tornado[278]___ /|>

```

Figure 5: le deuxième niveau du plan de classification sur UNIX. Par la commande <cd 100-000*> de la Figure 4, j'ai accédé au niveau "100_***", où sont présentées les activités spécifiques reliées à la fonction "100-000_Gestion_Generale".

Là où la principale difficulté pourrait survenir, ce serait dans l'intitulé des catégories sur la plate-forme PC-DOS (Annexe 2). En effet, des contraintes d'écriture limitent d'abord à 8 caractères, suivis d'un point, suivi de trois autres caractères, la longueur des titres, et d'autre part certains symboles ne sont pas acceptés (ex.: "/", "\", "."). Pour contourner cette difficulté, une légère adaptation du plan de classification pourrait s'avérer nécessaire, mais seulement au niveau très superficiel des intitulés, la structure n'étant en rien modifiée par ces considérations techniques. A titre d'exemple, nous

pouvons nous contenter de la seule cotation, sans utiliser l'intitulé au complet, en prenant soin toutefois de substituer au point (.), le trait de soulignement (_) "100.000." ne pouvant plus être utilisé, à cause des points, il faudra alors taper: "100_000_" pour que le système PC-DOS accepte l'intitulé.

Sur les trois plates-formes, une fois ce premier travail accompli, l'insertion des documents dans les répertoires pertinents peut alors commencer. A ce moment, l'ordinateur reprend son rôle d'outil à mémoire et permet l'enregistrement ainsi que le classement, par sa représentation à l'écran, du document dans la bonne catégorie de classification, l'ordinateur ne jouant plus qu'un rôle de filière électronique. Diverses fonctionnalités de recherche, incorporées au système d'exploitation feront en sorte que si un document venait à ne pas être repéré directement par la consultation de l'arborescence des documents, il pourrait l'être indirectement par ces fonctionnalités. Ensuite, une fois repéré et consulté, un simple déplacement du document vers le répertoire approprié évitera, lors d'une recherche ultérieure, de recourir à cette méthode de compensation.

D'autre part, puisque le plan de classification à l'écran est une copie conforme du plan de classification réalisé sur papier, la sélection de la bonne catégorie de rangement peut aussi bien se faire sur papier qu'à l'écran, et ce pour tous les supports. Par exemple, on peut ainsi déterminer la catégorie d'un document informatique, à partir de la copie papier du plan³ ou encore celle d'un document-papier que nous recevons de l'extérieur de l'organisme, à partir du "plan-écran".

Non seulement l'organisme possède-t-il, avec le micro-ordinateur, un outil remarquable de création, de traitement et de contrôle de l'information, mais plus encore, il est déjà en mesure de mieux l'exploiter, simplement en utilisant les fonctionnalités de base contenues dans le système d'exploitation. Ainsi facilité par la classification, le repérage de l'information créée sur ordinateur est-il nettement plus efficace.

CONCLUSION

Après avoir mentionné les changements apportés au monde du travail suite à l'introduction de la micro-informatique et fait ressortir les problèmes de classification et de repérage de l'information, nous nous sommes penchés sur les principes de la classification. A ce propos, il apparaît clairement qu'une classification fondée sur les fonctions de l'entreprise se présente comme étant celle qui est la plus susceptible d'en suivre en toute cohérence les changements organisationnels. Puis, pour réaliser l'application de la classification aux documents informatiques, nous nous sommes servis des fonctionnalités de base de trois systèmes informatiques, dont un qui témoigne plus particulièrement du phénomène croissant des réseaux. Facilité par l'outil même qui sert à créer l'information, son repérage s'avère d'autant plus efficace que sa classification, par le biais d'un plan de classification qui est le même sur papier et sur écran, témoignera des fonctions et activités de l'organisme.

Ainsi, en utilisant les principes fondamentaux de l'archivistique et les capacités intrinsèques de l'informatique, une classification efficace des documents informatiques semble moins complexe à réaliser qu'on aurait pu l'envisager. La mission de l'organisme, ses fonctions et ses activités peuvent ainsi se refléter autant par ses documents papier, informatiques ou mixtes. Ce qui compte c'est de ne jamais se laisser obnubiler par l'outil au détriment de la démarche intellectuelle de classification.

NOTES

(1) Terminal simple: qui ne permet aucun traitement autonome des données saisies, par opposition à terminal intelligent, dans O'LEARY, WILLIAMS, et al. *Éléments d'informatique*È.

(2) Bearman précise plus haut dans le même texte que "transactions" veut dire "by which is meant actions taken in the course of conducting their business, rather than "commercial" transactions". [je souligne], p. 17.

(3) Surtout dans le cas où le système d'opération ne permet pas d'alterner facilement entre le logiciel d'application et le système (comme peut le faire par exemple la plate-forme Macintosh, avec son Système 7).

BIBLIOGRAPHIE

ARCHIVES NATIONALES DE FRANCE. *Manuel d'archivistique: Théorie et pratique des Archives publiques de France*. Chapitre IV: Le classement et la cotation, 1970: 187-227.

ARCHIVES NATIONALES DU CANADA. *Comment gérer les répertoires et les fichiers dans votre ordinateur*. Ottawa: Ministère des Approvisionnements et Services Canada, 1993, 25 p.

BAILEY, Catherine. "Archival theory and electronic records". *Archivaria* 29 (1989-1990): 180-196.

BEARMAN, David. "Record-keeping systems". *Archivaria* 36 (Autumn 1993): 16-36.

BERGERON, Pierrette. "La gestion des archives électroniques". *Archives* 23, no 3 (hiver 1992): 51-70.

BLACK, Kelly; MOTZ, Arlene. "Standardizing PC file names using DOS directories". *Records Management Quarterly* 23, no 1 (January 1989): 14-17.

CHOUINARD, Denys; ROUSSEAU, Jean-Yves; ROY, Madeleine. "La gestion des archives informatiques: l'expérience de l'Université de Montréal". *Archives* 21, no 4 (printemps 1990): 41-58.

COUTURE, Carol; ROUSSEAU, Jean-Yves. *Les archives au XXe siècle. Une réponse aux besoins de l'administration et de la recherche*. Montréal: Université de Montréal, Secrétariat général, Service des archives, 1982, 491 p.

CUNLIFFE, William; MILLER, Michael. "Writing a general records schedule for electronic records". *American Archivist* 52, no 3 (1989): 350-356.

DOLLAR, Charles M. *Electronic records management and archives in international organizations: a RAMP study with guidelines*. Paris: UNESCO, RAMP programme, 1986, PGI -86/WS/12, 165 p.

DOLLAR, Charles M. "Seizing the opportunity: archivists in the information age". *Archivum*, Actes du XIIe congrès du CIA XXXIX (1994): 449-455.

FLETCHER, Patricia T. "Electronic records management in state government: planning for the information age". *Records Management Quarterly* 24, no 4 (1990): 26-32.

GOOKIN, Dan. *Staying with DOS. How to get the most from your computer without changing your operating system*. Ventana Press, 1990, 243 p.

HJORTH, Karen; TRNING, Dan. "Files and files". *Janus* 1 (1992): 5-15.

HOLLAND, Michael E. "Adding electronic records to the archival menagerie: appraisal concerns and cautions". *Provenance (Georgia Archivist)* 8, no 1 (1990): 27-44.

L'HUILLIER, Hervé; M'BAIREH, Lisette. "Archives papiers et archives électroniques: étude prospective de la gestion des archives dans un milieu fortement équipé en bureautique". *La Gazette des archives* 127 (1984): 299-341.

- MARCOUX, Yves. "Les formats de documents électroniques en archivistique". *Archives* 26, nos 1-2 (été-automne 1994): 85-100.
- MCAFFEE, John. "Electronic filing in the 80s". *Records Management Quarterly* 17, no 4 (October 1983): 12-14.
- NATIONAL ARCHIVES AND RECORDS ADMINISTRATION. *National Archives strategy for the development and implementation of standards for the creation, transfer, access and long-term storage of electronic records of the Federal Government*. Report No. NARA/TIP-90/08. Washington: NARA Archival Research and Evaluation Staff, 1990, 28 p.
- O'LEARY, Timothy; WILLIAMS, Brian, et al. *Éléments d'informatique*. Montréal: McGraw-Hill, 1991, 256 p.
- PETERSON, Trudy Huskamp. "Machine-readable records as archival material". *Archivum*, Actes du XIe congrès du CIA XXXV, (1988): 83-88.
- PIEYNS, Jean . "Méthodes de classement des nouvelles archives". *Archivum*, Actes du XIe congrès du CIA XXXV, (1988): 201-204.
- PLAYOUST, Arlette. "La place de l'informatique aux archives". *La Gazette des Archives* 141, (Juillet 1988): 97-114.
- ROBERGE, Michel. *La classification universelle des documents administratifs*. La Pocatière: Documentor, 1985, 247 p.
- ROBERGE, Michel. *Comment gérer vos documents administratifs sur disquette*. La Pocatière: Documentor, 1985, 66 p.
- ROBERGE, Michel. *La gestion de l'information administrative. Application globale, systémique et systématique*. Québec: Documentor, 1992, 298 p.
- ROBERTS, David. "The disposal of electronic records in office automation systems of the Australian Public Service". *Archives and Manuscripts* 17, no 2 (1989).
- ROUSSEAU, Jean-Yves. "Le choix d'un système de classement et son application". *Archives* 11, no 1 (1979): 3-19.
- ROUSSEAU, Jean-Yves. "La normalisation et l'informatisation des archives". in *La normalisation en archivistique. Mélanges Jacques-Ducharme*, sous la direction de Carol Couture. Québec: Documentor, AAQ, 1992, 141-147.
- SAFFADY, William. *Managing electronic records*. ARMA International, 1992, 184 p.
- SCHELLENBERG, T.R. *Modern archives: principles and techniques*. Chapter VII: classification principles. pp. 52-64. Chicago: University of Chicago Press, 1956, 247 p.
- SCOTT, P.J. et al. "Archives and administrative change: some methods and approaches (part 3)". *Archives and manuscripts* 8, no 1 (June 1980): 41-54.
- UNIVERSITÉ DE MONTRÉAL. *Guide de gestion des archives*. Montréal: Secrétariat général, Service des archives, 1989, 1-18.
- WALCH, Victoria Irons. "The role of standards in the archival management of electronic records". *American Archivist* 53, no 1 (1990): 30-43.



École de bibliothéconomie et des sciences de l'information
de l'Université de Montréal

Numéro en cours

[Vol. 8 no 1 \(printemps 2005\)](#)

Numéros précédents

[Vol. 7 no 1 \(printemps 2004\)](#)

[Vol. 6 no 1 \(automne 2001\)](#)

[Vol. 5 no 1 \(automne 1999\)](#)

[Vol. 4 no 1 \(automne 1998\)](#)

[Vol. 3 no 2 \(printemps 1998\): Non publié](#)

[Vol. 3 no 1 \(automne 1997\)](#)

[Vol. 2 no 2 \(printemps 1997\)](#)

Vol. 2 no 1 (automne 1996)

Vol. 1 no 2 (printemps 1996)

Vol. 1 no 1 (octobre 1995)

Né d'une initiative étudiante, **Cursus** est le périodique électronique étudiant de l'École de bibliothéconomie et des sciences de l'information (EBSI) de l'Université de Montréal et diffuse des textes produits dans le cadre des cours de l'École. **Cursus** est un périodique spécialisé en sciences de l'information publié deux fois par année (automne et printemps).

Les principaux objectifs de **Cursus** sont de servir de complément aux activités pédagogiques, de permettre aux étudiantes et étudiants de l'EBSI de diffuser leurs travaux et réflexions tout en participant à un processus d'édition en soumettant leur texte et finalement d'apporter un contenu original en français sur Internet.

En se voulant une vitrine de l'EBSI, CURSUS est enfin le moyen d'offrir aux étudiants un moyen de se faire connaître par leur nouvel environnement professionnel.

Remerciements

Cursus peut paraître grâce à la participation des professeurs et à l'utilisation du matériel informatique de l'École de bibliothéconomie et des sciences de l'information. Le périodique bénéficie de subventions provenant de la Corporation des bibliothécaires professionnels du Québec, de l'École de bibliothéconomie et des sciences de l'information, du Service aux étudiants de l'Université de Montréal et de l'Association des étudiantes et des étudiants de l'EBSI.

Après un long silence de près de trois années, **Cursus** vous revient avec la présentation du premier numéro de son septième volume. Qu'il nous soit permis, pour l'année 2004, de remercier chaleureusement la disponibilité du comité de lecture et les auteurs qui ont soumis un article. **Cursus** ne peut se développer qu'avec la complicité des étudiants: nous attendons vos textes

pour le prochain numéro.

Informations générales

Accès et diffusion

À chaque parution de **Cursus**, la table des matières du numéro accompagnée des résumés des textes sera diffusée sur des listes de discussions avec les adresses des sites où les lecteurs pourront consulter le périodique en version intégrale.

Droits d'auteur

Tout texte demeure la propriété de son auteur. Cependant, l'auteur permet à **Cursus** de publier, distribuer et archiver son texte électroniquement. Après sa parution dans **Cursus**, l'auteur peut republier son texte dans un autre périodique à la condition que la mention de la parution dans **Cursus** soit clairement indiquée. La reproduction des textes parus dans **Cursus** est permise pour une utilisation individuelle. Tout usage commercial nécessite une permission écrite de l'auteur.

Protocole de rédaction
