

# Une analyse de l'utilisation d'Internet à la bibliothèque publique

Par

Andrea Görgeç

Shirley Lew,

et Ann Curry

*School of Library, Archival and Information Studies*

(Faculté des études en bibliothéconomie, en gestion des archives et de  
l'information)

Université de la Colombie-Britannique

---

## Table des matières

---

<b>I.</b>	<b>Introduction</b>	<b>2</b>
<b>II.</b>	<b>Examen de la documentation</b>	<b>4</b>
<b>III.</b>	<b>Méthode</b>	<b>7</b>
<b>IV.</b>	<b>Résultats par catégorie globale</b>	<b>19</b>
<b>V.</b>	<b>Pages Web des administrations publiques</b>	<b>46</b>
<b>VI.</b>	<b>Sites dans des langues autres que l'anglais</b>	<b>48</b>
<b>VII.</b>	<b>Périodiques en ligne</b>	<b>52</b>
<b>VIII.</b>	<b>Utilisation d'Internet sur semaine</b>	<b>53</b>
<b>IX.</b>	<b>Utilisation d'Internet dans les succursales</b>	<b>55</b>
<b>X.</b>	<b>Problèmes liés à la recherche</b>	<b>57</b>
<b>XI.</b>	<b>Recommandations pour les études à venir</b>	<b>62</b>

### Liste des tableaux

<b>Tableau I: Catégorie globales par ordre d'importance</b>	<b>19</b>
<b>Tableau II: Résultats par format ou par source</b>	<b>20</b>
<b>Tableau III: Sites en langues autres que l'anglais (en ordre d'importance)</b>	<b>50</b>
<b>Tableau IV: Catégories les plus fréquemment utilisées dans les succursales</b>	<b>55</b>
<b>Annexe 1: Liste des domaines</b>	<b>64</b>
<b>Références consultées</b>	<b>57</b>

## I. Introduction

---

Depuis les dernières années, Internet est de plus en plus présent dans bon nombre de bibliothèques publiques canadiennes. Les ressources auxquelles Internet permet d'accéder sont jugées par de nombreux bibliothécaires d'une importance équivalente aux documents que l'on trouve traditionnellement en bibliothèque et constituent ainsi une composante essentielle, et non facultative, de la collection (McClure, Bertot et Zweizig 1993, 3). En outre, de nombreux clients des bibliothèques s'attendent maintenant à ce qu'Internet soit offert parmi les services réguliers des bibliothèques publiques. Les longues files de clients devant les terminaux Internet dans les bibliothèques du Canada attestent la popularité de cette technologie. Toutefois, certains croient que les avantages d'offrir un accès Internet dans les bibliothèques publiques ne compensent pas les dépenses et difficultés liées à ce service. Certaines des opinions négatives liées à l'utilisation d'Internet reposent sur une perception selon laquelle Internet sert le plus souvent à des fins inappropriées, par exemple, pour lire le courrier électronique ou consulter des sites sexuellement explicites.

Mais que savons-nous vraiment de la façon dont Internet est utilisé dans les bibliothèques publiques? Quels sites consultent les clients? À ce jour, la recherche sur Internet n'a porté que sur deux aspects : les modèles d'utilisation d'Internet par les bibliothécaires et d'autres groupes professionnels et les niveaux de connectivité à Internet dans les bibliothèques publiques (Tillotson, Cherry et Clinton 1995, 109). Jusqu'à présent, aucune étude n'a été menée pour étudier la façon dont les clients des

bibliothèques publiques utilisent Internet dans ces établissements. Nous ne disposons d'aucune donnée pour appuyer les hypothèses ou énoncés proposés au sujet d'Internet et de son utilisation. Dans un article de 1997 sur l'*American Internet Decency Act*, l'auteur mentionne que, même si aucune étude n'a été menée sur ce que consultent les Américains sur Internet, il est fort probable que l'utilisation qui en est faite soit en grande partie sans conséquence (« The Statues of Libertines? » 1997, 296). De telles déclarations montrent bien que les conclusions tirées quant à l'utilisation d'Internet ne reposent sur aucun fondement scientifique. Le but de la présente étude est de commencer à combler les lacunes de la documentation en étudiant plus particulièrement l'utilisation d'Internet par les clients de la bibliothèque publique de Burnaby (BPB), en Colombie-Britannique.

Ce projet de recherche, qui est en fait une étude pilote, repose sur des données recueillies à partir des relevés de transaction des serveurs Web de la bibliothèque publique de Burnaby et vise à répondre aux questions suivantes :

1. Quels types de sites les clients consultent-ils?
2. Quels sont les sujets recherchés les plus fréquemment?
3. Existe-t-il des différences entre les succursales quant aux types de sites consultés?
4. Existe-t-il des différences quant aux types de sites visités selon les jours de la semaine?
5. À quelle fréquence les sites du gouvernement sont-ils consultés?
6. À quelle fréquence les sites de magazines ou de cybermagazines sont-ils

consultés?

7. À quelle fréquence les sites dans une autre langue que l'anglais sont-ils consultés?
8. Quelles sont ces autres langues?

La faisabilité de recueillir et d'analyser des données pertinentes sur Internet constitue une difficulté sous-jacente de la réalisation des objectifs de notre projet. Dans le cadre de cette étude, nous avons utilisé l'analyse des relevés de transaction (ART) pour étudier les modèles d'utilisation d'Internet. L'ART, en tant que méthode, est utilisée fréquemment dans l'étude des catalogues de bibliothèques en ligne, mais n'a jamais été employée comme outil d'analyse de l'utilisation d'Internet. Les chercheurs se sont également fixés comme objectif de déterminer l'efficacité de cette méthode pour analyser l'utilisation d'Internet.

## **II. Examen de la documentation**

---

De nombreuses études ont été effectuées sur l'utilisation d'Internet dans les bibliothèques. Comme on l'a mentionné plus tôt, bon nombre de ces études visaient l'incidence et la présence d'Internet, plutôt que son utilisation en tant que source d'information. D'autres études examinent l'utilisation d'Internet par les bibliothécaires ou d'autres groupes de professionnels. La plupart de ces études sont fondées sur les réponses fournies dans le cadre de questionnaires. Comme certaines restrictions s'appliquent quant au nombre de réponses que l'on peut donner à seule question, les

données découlant des questionnaires sont assez limitées. Un questionnaire bien rédigé se révèle plus efficace qu'une analyse des relevés de transaction pour déterminer pourquoi les gens ont recours à Internet, mais il est moins utile pour dégager avec précision ce que les gens consultent sur Internet.

L'étude de Perry (1995) est liée à la nôtre : elle examine l'utilisation d'Internet par de « véritables utilisateurs d'Internet ». Cette étude vise à déterminer l'utilisation d'Internet par la communauté internationale des internautes en leur posant certaines questions, notamment : « Qui utilise Internet ? » et « Quelle information recherche-t-on ? ». Perry généralise les modèles d'utilisation d'Internet en fonction des résultats d'un sondage qui n'a été transmis qu'aux serveurs de listes et pour lequel il n'a reçu que 84 réponses. Les participants à son étude font une utilisation très spécialisée d'Internet, ce qui ne permet pas à Perry de généraliser comme il l'a fait. Même si ses questions sont utiles et pertinentes aux objectifs de notre étude actuelle, les possibilités de réponse pour chacune de ces questions nous fournissent peu de données. Par exemple, lorsque l'on demande aux utilisateurs quel type d'information ils recherchent sur Internet, ils n'ont le choix qu'entre les catégories suivantes : acoustique, graphique, statistique, ou texte.

Une étude menée par la bibliothèque publique de Seattle (BPS) (Harvey et Horne, 1995) tente d'évaluer le niveau de satisfaction des utilisateurs quant à l'accès à Internet et aux services offerts grâce à cette technologie par le réseau des bibliothèques publiques de Seattle. Les questions posées et le but de l'étude n'ont pas été définis avec précision, mais l'étude fait référence aux réponses données par les

clients concernant l'information à laquelle ils accèdent par Internet. Toutefois, le rapport de l'étude manque de rigueur scientifique et ne précise pas l'utilisation que font les clients d'Internet. Il ne fait que souligner certains commentaires positifs des répondants et mentionner certaines utilisations du Web. D'autres études sur l'utilisation d'Internet visent certains groupes d'utilisateurs particuliers, notamment l'étude menée par Lazinger, Bar-Ilan et Peritz (1997) à l'université hébraïque de Jérusalem. Par le biais d'un sondage, les auteurs de l'étude ont tenté de déterminer quelle utilisation font les membres de la faculté des ressources sur Internet.

Des chercheurs de la bibliothèque de l'Université de Toronto ont mené une étude plus exhaustive sur l'utilisation d'Internet (Tillotson, Cherry et Clinton, 1995). Ils ont eu recours à une combinaison de plusieurs outils : un sondage en ligne, des groupes de discussion et une analyse des relevés des connections Telnet pour contrôler l'utilisation des ressources Internet particulières accessibles à partir d'UTLink, le système d'information électronique et le catalogue en ligne sur format texte de la bibliothèque de l'Université de Toronto. Cette étude se distingue des autres par son analyse des sites visités par le biais d'UTLink et le nombre d'appels de fichier. L'analyse des relevés, les groupes de discussion et les questions du sondage ont été utilisés en combinaison afin de répondre, notamment, à des questions qui nous intéressent dans le cadre de notre étude : Que visitent les clients de la bibliothèque sur Internet? À quels sites accèdent-ils? (Tillotson, Cherry et Clinton, 1995, 190). L'Université de Toronto a réussi à atteindre les objectifs de son étude sur l'utilisation des ressources Internet, par ailleurs limitées, offertes par le biais d'UTLink, mais

n'essaie pas d'appliquer ses résultats au Web dans son ensemble, comme nous tentons de le faire dans la présente étude. La portée de l'étude de l'Université de Toronto se limite aux 1 325 destinations Telnet auxquelles les utilisateurs de la bibliothèque ont accès par UTLINK. De plus, les éléments du questionnaire qui portent sur le but de l'utilisation d'Internet compromettent l'intégrité des données recueillies. Par exemple, les catégories de réponses à la question portant sur la principale raison de recourir aux ressources sur Internet ne s'excluent pas mutuellement. En effet, il y a possibilité de confusion quant à la différence entre « chercher de l'information » et presque toutes les autres réponses offertes.

### **III. Méthode**

---

Dans le cadre de la présente étude, les chercheurs ont employé une variante de la méthode d'analyse des relevés de transaction (ART). L'ART est habituellement utilisée pour étudier l'utilisation des catalogues en ligne dans les bibliothèques publiques et universitaires. C'est une méthode qui permet habituellement d'examiner les transactions entre une personne et un catalogue en ligne. L'ART a surtout été utilisée pour comprendre la façon dont les utilisateurs font leurs recherches dans les catalogues en ligne, en enregistrant et analysant les chaînes de recherche (Peter, 1993). La principale force de cette méthode réside dans sa capacité d'enregistrer les données d'une façon discrète et objective. Elle permet d'établir une image de la recherche en ligne fondée sur les chaînes de recherche réelles de l'utilisateur, plutôt que sur le souvenir qui lui en reste (Connaway, Budd et Kochtanek,



1995, 143). Les chercheurs dans ce domaine proposent d'utiliser l'ART dans l'avenir comme outil de développement des collections afin d'étudier non seulement comment les utilisateurs font leurs recherches dans le catalogue en ligne, mais ce qu'ils cherchent (Peters 1993, 57; Simpson, 1989, 61). Les chercheurs qui ont travaillé sur la présente étude croient qu'il est logique d'appliquer cette méthode aux relevés dans lesquels sont consignés les sites Internet visités par les clients des bibliothèques afin d'analyser les façons dont cette « base de données » est utilisée. La présente étude a pour but de mettre cette hypothèse à l'épreuve, mais également d'étudier l'utilisation d'Internet à la bibliothèque publique de Burnaby.

Pour la présente étude, les chercheurs ont utilisé des relevés tirés du serveur Web de la ville de Burnaby pour les terminaux Internet de la bibliothèque publique. Ces relevés consignent les adresses Web des pages que les clients ont visitées ou tenté de visiter à partir des terminaux de la BPB au cours de la période de l'étude. En revisitant les sites consignés dans le relevé et en les classant en fonction de certaines catégories prédéfinies, les chercheurs ont pu dégager les types de sites et les sujets qui intéressent les clients de la bibliothèque.

On a remis aux chercheurs des relevés individuels pour les sept jours qui se sont écoulés entre le 2 décembre et le 8 décembre 1998. Les relevés contiennent les données liées à tous les terminaux Internet du réseau de la bibliothèque de Burnaby : il y a quatre terminaux à la bibliothèque centrale Metrotown et un terminal dans chacune des autres succursales, soit Cameron, Kingsway et McGill. Le relevé du dimanche 6 décembre ne reflète que les transactions à Metrotown parce que les autres

succursales étaient fermées. Les heures d'ouverture varient légèrement pendant la semaine et la fin de semaine. Toutes les succursales, sauf celle de Cameron, sont ouvertes de 9 h à 21 h du lundi au vendredi, de 9 h à 18 h le samedi. La succursale de Cameron est ouverte le lundi de 13 h à 21 h, du mardi au jeudi de 10 h à 21 h et du vendredi au samedi de 10 h à 18 h. Seule la succursale Metrotown est ouverte le dimanche de 12 h à 16 h.

Tous les terminaux étaient équipés d'écrans spéciaux : seul l'utilisateur du terminal peut voir le contenu de l'écran. Aucun des terminaux n'était doté d'un logiciel de filtrage pour bloquer l'accès à des documents sexuellement explicites, et les utilisateurs étaient autorisés à accéder à des adresses de courrier électronique, du type de celles fournies par Hotmail. Les utilisateurs des terminaux à la BPB doivent entrer dans le système à l'aide du code barres de leur carte de bibliothèque et de leur numéro de téléphone. Cette information est utilisée par l'ordinateur pour mettre fin à la session après 30 minutes. Toutefois, aucune information permettant d'identifier les utilisateurs des terminaux n'apparaît dans les relevés.

## **Le relevé**

Il s'agit d'un relevé de transaction ordinaire consigné dans tous les serveurs

Web. Le relevé de transaction fournit les données suivantes :

- le numéro Internet (ou le code d'identification unique) du terminal utilisé à la BPB
- la date et l'heure de la transaction
- le protocole et l'adresse Web du fichier demandé ou téléchargé vers le terminal
- le statut de la transaction (il s'agit d'un code comme 200 ou 404 qui indique si la transaction a été réussie ou non)

—

Chaque ligne du relevé représente une seule transaction effectuée par un utilisateur d'un terminal Internet de la BPB. Le relevé de transaction enregistre les activités effectuées par le biais du serveur Web. Dans la plupart des cas, ces relevés sont analysés par les propriétaires des serveurs Web afin d'évaluer la demande à l'égard de certaines pages Web ou d'autres ressources Internet accessibles par le biais de leur serveur. Ils sont généralement intéressés à la provenance et à la nature des demandes (Stout, 1997). Dans le cas qui nous concerne, les terminaux de la BPB sont reliés à Internet par le serveur Web de la ville de Burnaby. Ainsi, les employés chargés des systèmes à la ville de Burnaby ont pu extraire les lignes des relevés de transactions qui représentent les demandes faites à partir des terminaux de la BPB et les fichiers transférés par le biais du serveur Web de la ville. Pour ce faire, ils ont isolé les lignes qui contenaient les numéros Internet de ces terminaux.

Chaque fois qu'un utilisateur consulte une page Web en cliquant sur un lien ou en inscrivant une adresse Web dans la fenêtre prévue à cette fin, une ligne est enregistrée. Le relevé contient une ligne pour chaque page demandée et chaque fichier transféré à partir d'un terminal de la BPB. Cela comprend tous les fichiers .gif et .jpg qui apparaissent sur la page, ainsi que le fichier html représentant la page dans son ensemble. Toutes les autres fenêtres d'exploration ouvertes par le biais de programmes d'application Java sont consignées dans une seule ligne; il en va de même de toutes les demandes réorientées vers une nouvelle adresse, suite à l'annulation d'une adresse ou à une erreur dans le libellé de l'adresse. Bref, toute

transaction liée à tout genre de fichier est consignée comme une seule ligne dans le relevé. Par conséquent, une page consultée par un client de la BPB pourrait être représentée par plusieurs lignes dans le relevé. Les pages qui contiennent de nombreux fichiers graphiques seront représentées par un plus grand nombre de lignes que celles qui contiennent principalement du texte.

### **Échantillon**

Les transactions quotidiennes donnent lieu à des milliers de lignes. En tout, 194 644 lignes ont été relevées au cours des sept jours de l'étude. Il était donc nécessaire d'établir un échantillon à étudier. Les chercheurs ont donc choisi un échantillon aléatoire composé de chaque 10<sup>e</sup> ligne des relevés. Toutefois, après avoir commencé l'analyse, ils ont décidé qu'il serait plus approprié d'opter plutôt pour chaque 20<sup>e</sup> ligne afin d'éviter la surreprésentation d'une seule page Web. Par exemple, une page saturée de fichiers graphiques peut totaliser 14 lignes; ainsi, il est possible que deux lignes représentant la même page se trouvent dans l'échantillon. Lorsque cette situation se présente, les chercheurs visitent la page deux fois (à la première ligne et à la onzième, par exemple) et la consignent comme si elle avait fait l'objet de deux visites distinctes. Les chercheurs croient qu'un échantillon composé de chaque 20<sup>e</sup> ligne, plutôt que de chaque 10<sup>e</sup> ligne, est plus susceptible de fournir des données provenant de plus d'une page.

Les données de la succursale Metrotown recueillies entre le 3 et le 5 décembre représentent toutefois un échantillon composé de chaque 10<sup>e</sup> ligne, car il s'agissait du

premier échantillon analysé. Les chercheurs ont choisi de conserver les données déjà recueillies pour cette succursale en raison de la nature exploratoire de l'étude.

Avant de composer l'échantillon, les chercheurs ont éliminé des relevés toutes les lignes qui contenaient des données inclassables et par conséquent inutilisables (voir l'explication ci-dessous concernant les fichiers éliminés). On a ainsi gagné du temps, car il n'a pas été nécessaire de rejeter ces données plus tard au cours du processus. En outre, on augmentait les probabilités que l'échantillon contienne des lignes utiles.

Les fichiers ou types de fichiers retirés étaient les suivants :

1) fichiers .gif

Ces derniers ont été retirés parce qu'ils apparaissent généralement comme des boutons ou des icônes sur des pages Web, et non pas en tant que page Web en tant que telle; par conséquent, leur page « centrale » demeure dans le relevé et a la même chance que toute autre page d'être intégrée à l'échantillon.

2) Page d'accueil de la BPB

Il s'agit de la page par défaut qui apparaît à l'écran lorsque l'utilisateur entre dans le système. Par conséquent, elle ne constitue pas un choix conscient du visiteur. On a laissé d'autres pages du site Web de la BPB dans le relevé.

3) Bandeaux publicitaires

Les chercheurs ont retiré le plus de fichiers de bandeaux publicitaires possible, fichiers qu'ils sont parvenus à identifier à partir des modèles d'adresse Web. Ces fichiers ont été retirés parce qu'il était difficile de déterminer sur quel site ces bandeaux étaient apparus. Il est impossible de classer ces fichiers comme tels parce qu'ils n'apparaissent jamais sur la même page. Voici un exemple de fichier de bandeau publicitaire retiré de l'échantillon : <http://ad.preferences.com/>.

#### 4) Fichiers .js, .class et .cs

Ces fichiers ont été retirés parce que les chercheurs ne pouvaient plus y accéder et qu'ils étaient par conséquent inclassables.

#### 5) Lignes vierges

On a retiré les lignes qui ne contenaient aucune information dans le champ réservé à l'adresse Web.

Ces fichiers ont été retirés des relevés à l'aide d'une application d'Unix appelée Egrep. Egrep peut reconnaître des chaînes de caractères et a été utilisée pour extraire tous les fichiers des relevés, sauf ceux ayant une chaîne de caractères particuliers. Par exemple, Egrep peut localiser toutes les lignes ayant des chaînes de caractères se terminant par .gif et produire un nouveau fichier à partir des lignes restantes. Au total, on a retiré 115 264 lignes des relevés : il en est donc resté 79 380. Ces lignes ont ensuite été numérotées et chaque 20<sup>e</sup> ligne a été copiée dans des échantillons de

Dans la présente étude, on a analysé en tout 4 712 lignes.

tant que page Web comme telle, plus particulièrement dans le cas de documents sexuellement explicites. Les chercheurs n'ont pas voulu perdre de données sur les

Par ailleurs, le fait de retirer les fichiers .jpg ne réduisait pas la taille des relevés de façon marquée, et les chercheurs ont donc conclu que ces fichiers ne les

possible de pages Web individuelles.

### **Limites**

qu'elles ne dévoilent rien au sujet des utilisateurs et de leurs besoins en information au moment de leur consultation (Kurth, 1993, 98; Simpson, 1989, 36). Même si les relevés

base de données, grâce au contexte dans lequel les adresses Web apparaissent dans le relevé, il n'est pas possible de tirer des conclusions solides sur les intentions des

est amplifié par la nature même d'Internet. L'utilisation d'hypertexte et de liens

graphiques favorise un type de consultation tout à fait différent de celui des catalogues

qui présente un problème, mais plutôt le pourcentage de visites accidentelles qui ne reflètent pas véritablement les intentions des clients. Par exemple, il est facile de cliquer sur un lien fourni par un moteur de recherche ou un autre document et de visiter (visite qui est d'ailleurs consignée) un site à caractère sexuel sans avoir vraiment eu l'intention de le faire.

L'autre facteur qui complique l'analyse des relevés de transaction est la nature instable de la base de données (Kurth 1993, 99), une instabilité encore plus marquée sur Internet. Il peut être difficile ou impossible de refaire des recherches déjà effectuées par un utilisateur en raison de la nature dynamique des ressources Internet. Les sites peuvent disparaître ou être révisés du tout au tout, du jour au lendemain. À maintes reprises au cours de la présente étude, les chercheurs n'ont pas été en mesure d'accéder à un site que l'utilisateur avait visité avec succès; par ailleurs, ils ont réussi à consulter un site pour lequel l'utilisateur avait reçu un message d'erreur 404. Ces sites ont tout de même été répertoriés, car le fait de cliquer sur un lien ou d'accéder à une adresse Web suppose au moins une intention de consulter le site.



## Processus de collecte des données

données dans une base de données Microsoft Access. On a recueilli les renseignements suivants :

- le jour de la semaine
- la succursale de la BPB
- le code de classification assigné
- les sous-titres des sites de magazines, de cybermagazines, de sites gouvernementaux ou de sites dans une autre langue que l'anglais
- la langue des sites qui n'étaient pas en anglais
- l'adresse Web

Les recherchistes se sont intéressés non seulement aux tendances générales quant à l'utilisation d'Internet, mais également à l'identification de tout écart dans les modèles d'utilisation entre les différentes journées de la semaine et les différentes succursales. Consultez les sections VII et IX pour une description des résultats.

On a beaucoup réfléchi à la conception d'une classification appropriée pour répartir les sites Web entre diverses catégories. En suivant l'exemple du OCLC qui utilise la classification décimale de Dewey pour classer les ressources Internet dans la base de données Net First, les recherchistes ont expérimenté avec les valeurs Dewey dans un relevé d'essai fourni par la BPB. Le processus s'est révélé laborieux et il a fallu déployer énormément d'efforts pour trouver une place à chaque site Web dans un modèle de classification très rigide. Même si la classification Dewey est efficace pour fournir l'emplacement d'un ouvrage sur les étagères de la bibliothèque, elle l'est beaucoup moins pour classer les sites Internet visités dans le cadre de la présente étude. La plupart de la documentation sur Internet est unique à ce médium et n'a pas

de corrélation directe avec les ouvrages ou autres documents que l'on trouve dans les ressources sur Internet, mais de dégager des modèles d'utilisation. Les chercheurs ont donc jugé plus approprié d'établir une liste de catégories.

dans les répertoires Internet populaires comme Yahoo! et Magellan, et reflète également les propres expériences des chercheurs avec Internet et les problèmes soulevés dans l'analyse du relevé d'essai. Les thèmes liés aux annonces personnelles, bienfaisance et sans but lucratif ont été ajoutés pendant le déroulement de l'étude pour tenir compte des choix des clients à l'égard de ces catégories. Les chercheurs ont sites dans une autre langue que l'anglais, les sites de magazines et les sites gouvernementaux, après avoir commencé le travail d'analyse. Dans tous les cas, les en fonction des nouveaux thèmes établis. La section X présente une discussion sur les thèmes et des suggestions d'améliorations pour toute étude ultérieure.

subjectif. Il a exigé des consultations constantes entre les deux chercheurs responsables de la classification, avant et pendant le processus, afin de garantir un

---



#### IV. Résultats par catégorie globale

---

l'échantillon, l'analyse statistique qui s'appuie sur le test du chi carré n'est pas appropriée. Dans ce contexte de recherche, on lui préfère le calcul de totaux et de pourcentages bruts, et l'utilisation de termes comme « significatif », « peu élevé », « élevé », etc., de façon descriptive et non dans un sens statistique. Par souci de clarté, certaines des catégories initiales ont été regroupées sous un titre plus global. Ces catégories sont suivies d'un astérisque\*.

**Tableau I : Catégories globales par ordre d'importance**

	<b>Catégorie</b>	<b>Nbre de lignes</b>	<b>% du nbre total de lignes</b>
1.	Courrier électronique	1445	30,7 %
2.	Inclassable	600	12,7 %
3.	Divertissement	386	8,2 %
4.	Affaires	381	8,1 %
5.	Moteurs de recherche	353	7,5 %
6.	Sexe	256	5,4 %
7.	Ordinateurs	208	4,4 %
8.	Sports	174	3,7 %
9.	Sciences humaines/sociales	131	2,8 %
10.	Éducation	112	2,4 %
11.	Passe-temps	115	2,4 %
12.	Lieux d'intérêt et voyages	107	2,3 %
13.	Journaux	86	1,8 %
14.	Autres médias d'information	73	1,5 %
15.	Petites annonces	66	1,4 %
16.	Sciences	25	1,0 %
17.	Répertoires	20	0,4 %
18.	Santé	19	0,4 %
19.	Organismes de bienfaisance	15	0,3 %
20.	Droit	3	0,1 %

**Tableau II : Résultats par format ou par source**

<b>Format ou source</b>	<b>Nbre de lignes</b>	<b>% du nbre total</b>
Périodiques en ligne (magazines & publications électroniques)	116	2,5 %
Langue autre que l'anglais	437	9,3 %
Sites Web des administrations publiques	71	1,5 %

### **1. Courrier électronique\***

Cette catégorie inclut le courrier électronique (« courriel »), les groupes de conversation et les cartes de souhaits électroniques. Elle représente le plus fort pourcentage de connexions au réseau, avec 30,7 p. 100, ou un total de 1 445 lignes de l'échantillon avec une marge assez large. Par exemple, la deuxième catégorie en importance est le divertissement, avec 8,2 p. 100. On ne peut pas interpréter ces statistiques au premier degré, étant donné que les chiffres dépendent largement de la nature et des modèles d'utilisation du courriel et des sites de groupes de conversation.

Le principe de la messagerie électronique est relativement unique, dans la mesure où il s'accompagne d'un certain nombre d'activités que l'on ne trouve pas couramment sur les autres sites Internet. Chaque fois qu'une fonction ou une activité liée au courriel est exécutée, elle est enregistrée comme une ligne dans le journal des transactions. Par exemple, les clients vont accéder à la page d'accueil de la messagerie électronique, ouvrir leur compte, accéder à leurs messages, lire les messages un par un, y répondre et peut-être faire de la gestion de fichiers. Chaque

étape et chaque activité est enregistrée et consignée. Chaque client peut avoir de nombreux profils de courriel. Dans un court délai, le processus d'envoi de courriel peut générer un nombre disproportionné de lignes de relevé par rapport aux autres types d'activités exécutées sur Internet.

De la même façon, l'utilisation des groupes de conversation par un seul client peut représenter un grand nombre de lignes de relevé, ce qui va naturellement faire augmenter les statistiques et l'utilisation apparente de ces services.

Pour illustrer la façon dont les sites de courriel et les groupes de conversation peuvent être surreprésentés par rapport aux autres types de sites, on compare une session de conversation du journal du 8 décembre à une autre session durant laquelle divers sites sont visités. Les deux sessions ont eu lieu le même jour, pendant la même durée sur différents terminaux de la même succursale de la bibliothèque. Il semble assez évident que la session de conversation est imputable à un seul client, étant donné que les sites visités proviennent d'un groupe de conversation dans une langue autre que l'anglais et qu'il n'y a eu aucune interruption durant cette session pour visiter d'autres sites. Aucune des deux sessions n'a enregistré de pause durant laquelle Internet n'était pas utilisé. Le tableau ci-dessous compare le nombre de lignes de relevé générées pour chaque session :

Session de conversation	de la ligne n° 5400 ∩ 5820 =	420 lignes =	21 lignes dans l'échantillon
Autre type de session	de la ligne n° 1620 ∩ 1860 =	240 lignes =	12 lignes dans l'échantillon

Ces deux sessions ont eu lieu à la même heure du jour sur une période d'une heure et quarante-cinq minutes. D'après les chiffres ci-dessus, les groupes de conversation ont généré presque deux fois plus de lignes dans l'échantillon de relevé que les autres types de sessions. Si l'on multiplie ces chiffres par les centaines de visites à des groupes de conversation effectuées pendant sept jours, on obtient un chiffre très élevé.

## **20 Inclassable\***

Cette catégorie représente 12,7 p. 100, ou 600 lignes du relevé. Elle inclut les données collectées pour les lignes qui ont produit des messages d'erreur, contenaient des bandeaux publicitaires dont la page d'origine n'a pu être déterminée, qui ne contenaient aucune donnée ou contenaient des données qui étaient inclassables et dont la source ne pouvait pas être vérifiée.

Plusieurs raisons justifient la classification d'un site dans cette catégorie. Un grand nombre de ces raisons seront abordées dans la section qui traite des problèmes. À plusieurs reprises, les chercheurs ont rencontré des messages d'erreur lorsque le relevé indiquait qu'un client avait réussi à visiter un site. Certains sites étaient protégés par des mots de passe et étaient donc inaccessibles aux chercheurs. Très souvent, le fichier mentionné dans le relevé les conduisait à une partie de page ou à un fichier .jpg. Ils ont souvent pu vérifier la page d'où provenait cet élément. Les chercheurs ont pu remonter dans un site en supprimant des portions de l'URL de droite à gauche, et déterminer ainsi l'emplacement du fichier .jpg ou d'un autre élément d'une page donnée. Lorsqu'il leur était impossible de procéder ainsi, ils plaçaient la

ligne dans la catégorie « inclassable ». Certaines pages de cette catégorie ne contenaient tout simplement aucune donnée.

Un grand nombre de fichiers de bandeaux publicitaires se trouvaient encore dans le relevé, même après le retrait de bon nombre de ces fichiers des relevés bruts avant que l'on n'utilise les échantillons pour chaque journée, parce qu'un grand nombre de ces fichiers sont difficiles à identifier.

### **30 Divertissement\***

Cette catégorie représente 8,2 p. 100, soit 386 lignes de l'échantillon. Ces chiffres reflètent les totaux correspondant aux catégories suivantes : musique, télévision/films, radio, jeux et divertissements divers. Globalement, on pourrait dire que cette catégorie contient les sites offrant une culture de masse.

Pour tous les domaines, les pages Web liées à l'industrie du divertissement et celles qui donnent des renseignements généraux sur le divertissement font partie de cette catégorie. Par exemple, la musique peut contenir aussi bien des sites Web consacrés à un artiste en particulier que des sites sur l'aspect commercial de l'industrie musicale. Les sites qui traitent des arts en tant que discipline ou domaine d'étude n'ont pas été pris en compte. Par exemple, les chercheurs ont classé un site présentant différents styles et marques de guitares dans la catégorie « arts » plutôt que dans « divertissement musical », parce que ce site n'avait pas une vocation commerciale. Ils admettent faire preuve d'une relative subjectivité lorsqu'ils décident si un site offre une culture de masse ou un contenu artistique. Selon eux, il existe une différence



essentielle qui tient à l'aspect commercial de cette catégorie.

Cette catégorie incluait également les jeux informatiques. Parmi toutes les catégories de divertissement, les sites de jeux enregistrent le nombre le plus élevé de visites, avec 28,8 p. 100 des sites de divertissement. La popularité de ce thème par rapport aux autres a surpris les chercheurs. Il semble qu'il existe une communauté de joueurs à la fois nombreux et connaisseurs, qui utilisent Internet pour échanger des stratégies, et pour critiquer et noter les jeux. Par exemple, plusieurs sites de jeux indiquent en détail comment tricher dans certains jeux informatiques. Il a souvent fallu du temps pour classer ces sites. Le langage utilisé n'était pas courant et renvoyait à une terminologie propre à la culture des jeux informatiques.

#### **40 Catégories « Affaires »**

Les catégories « Affaires » combinées représentent 8,1 p. 100, ou 381 lignes de l'échantillon. Ce thème va fort probablement gagner en popularité, étant donné qu'un nombre croissant d'entreprises découvrent les avantages de la présence sur Internet, et que de plus en plus de gens connaissent la variété des renseignements commerciaux figurant sur Internet. Chaque catégorie est décrite individuellement ci-après.

##### **Affaires - Immobilier**

Cette catégorie représente 10,0 p. 100 (38 lignes) du nombre total de pages consacrées aux affaires. Chaque page qui traite d'immobilier, qu'il s'agisse de publicités ou de services, a été classée ici. Les sites classés sont les listes de maisons

et de propriétés, et les pages Web des agences immobilières.

Globalement, les chiffres sont assez bas pour cette catégorie, et un tiers des visites enregistrées ont été générées par une seule session Internet. Les chercheurs prévoyaient que les utilisateurs accéderaient à des renseignements locaux sur l'immobilier, mais ont constaté que certains clients des bibliothèques cherchaient également des renseignements sur l'immobilier dans d'autres pays, en particulier en Asie. Les pages asiatiques consacrées à l'immobilier représentent un fort pourcentage de l'ensemble des pages Web en chinois visitées.

### **Affaires – Finances & investissements**

Cette catégorie représente 18,1 p. 100 (69 lignes) du nombre total de pages consacrées aux affaires. Les pages financières portent sur les banques et les services bancaires, les taxes, les cartes de crédit et les actions et autres placements.

Les 69 pages classées dans la catégorie « Finances et investissements » sont généralement consacrées au marché boursier et contiennent d'autres données sur les placements, ainsi que sur les services bancaires. Une page est apparue à plusieurs reprises au cours de l'étude : elle portait sur les placements dans l'or et s'intitulait Gold-Eagle.

**Affaires – Industrie**

Cette catégorie représente 24,1 p. 100 (92 lignes) de l'ensemble des pages consacrées aux affaires. Les sites Web classés dans cette catégorie portent sur la foresterie, la pêche, l'agriculture, le transport lourd, l'exploitation minière, le pétrole et les autres industries primaires. Afin de distinguer cette catégorie de la catégorie scientifique, on a décidé que les pages dont l'objectif premier est la vente, la promotion ou la description d'un secteur particulier seraient placées dans cette catégorie. Les pages qui traitent de techniques, procédures ou théories scientifiques précises appartiennent à la catégorie scientifique. Ces critères ont été appliqués à toutes les pages, y compris à celles que gère une corporation ou une compagnie.

Les recherchistes prévoient que, dans cette catégorie, les sites les plus visités seraient les sites gouvernementaux, dont les pages contiennent beaucoup de renseignements de cette nature. Ils pensaient également trouver un grand nombre de pages Web consacrées à la foresterie et à la pêche (qui sont des industries associées à la Colombie-Britannique), mais ce ne fut pas le cas. En fait, ce sont les sites consacrés au transport lourd et à l'industrie minière qui ont été les plus visités.

La variété de sites visités dans cette catégorie est intéressante. Un site Web plutôt singulier considère la prospection pétrolière comme une quête spirituelle, guidée par la main de Dieu. Les pages plus représentatives de cette catégorie sont celles de Boeing, d'Airbus, des constructeurs d'automobiles comme Mercedes et d'une grande variété de sites consacrés à l'exploitation aurifère.

**Affaires – Produits & services**

Cette catégorie représente 25,7 p. 100 (98 lignes) de l'ensemble des pages consacrées aux affaires. Elle comprend les pages Web qui présentent un produit commercial ou des services professionnels comme la consultation, la comptabilité, la conception graphique, les télécommunications, les services de livraison, etc. Il ne faut pas la confondre avec celle des achats, qui est composée, selon les recherchistes, des pages Web visant exclusivement à faire des transactions commerciales en ligne à propos de divers produits.

Après avoir analysé la grande variété de pages Web intégrées à cette catégorie, les recherchistes pensent que les études futures devraient envisager une subdivision plus détaillées des catégories dans ce domaine. L'information sur les produits et services, par exemple sur les ordinateurs, fait partie de la catégorie générale « Ordinateurs ». Dans toute étude future, l'ensemble des catégories devraient être définies plus en détail, ce qui permettrait de distinguer les aspects des produits et services d'autres types d'information plus générale.

**Emploi**

Cette catégorie représente 22,0 p. 100 (84 lignes) de l'ensemble des pages consacrées aux affaires. Elle comprend les pages Web qui offrent des emplois, donnent des conseils sur les carrières ou les curriculum vitae et des trucs à propos des entrevues. Il faut noter que 32 p. 100 des 84 pages Web consacrées à l'emploi qui ont été visitées provenaient de sites des gouvernements fédéral ou provinciaux, ou des administrations municipales du Canada. Sur tous les sites des administrations

publiques visités (tous sujets confondus), 69 p. 100 étaient liés à l'emploi. Les recherchistes pensent que les services d'offre d'emplois parrainés par le gouvernement fédéral seraient plus efficaces s'ils proposaient un lien entre ces pages et le site Web de la BPB. Les pages Web des compagnies qui affichent leurs propres offres d'emploi ou décrivent les possibilités de carrière qu'elles offrent ont également été intégrées à la catégorie « Emploi ».

### **50 Moteurs de recherche**

Cette catégorie représente 7,5 p. 100 (ou 353 lignes) de l'échantillon. Elle inclut toute page Web qui tente de proposer un type d'organisation général ou l'accès à des ressources sur Internet. Cela comprend aussi bien les sites des moteurs de recherche/des répertoires de sujets, comme Yahoo! ou AltaVista, que la liste de sujets de la BPB, qui contient des liens avec certaines ressources Internet.

Les chiffres relatifs à cette catégorie sont élevés, en raison de la définition large qui a été utilisée. Elle inclut la page My Yahoo!, qui permet à un utilisateur de personnaliser l'information obtenue grâce au répertoire de sujets Yahoo!. Certains de ces sites ne sont pas en anglais ou ont été visités par des personnes qui cherchaient des sites dans une autre langue que l'anglais.

Les recherchistes ont décidé de classer dans des sous-catégories tous les moteurs de recherche ou répertoires de sujets qui affichent la chaîne de recherche utilisée et/ou les occurrences renvoyées par le système. Par exemple, une page trouvée par AltaVista qui contient des occurrences provenant de l'équipe de hockey

des Flames de Calgary et une chaîne de recherche « Calgary Flames » serait classée sous « Sports ».

## **60 Sexe**

### **Érotisme**

Cette catégorie représente 3,3 p. 100 (ou 154 lignes) de l'échantillon. Selon les chercheurs, les sites érotiques montrent des pages composées d'images de corps nus ou de situations érotiques sans aucune intention littéraire ou artistique, et dont l'objectif fondamental consiste à provoquer un désir sexuel chez l'utilisateur.

Là encore, il faut examiner les chiffres en tenant compte de la nature et de la structure de la plupart des sites à vocation sexuelle. Ce type de site contient de nombreuses images explicites (p.ex. des fichiers .jpg), qui entraîne généralement une augmentation supérieure à la normale du nombre de lignes de relevé à chaque visite. Pour chaque image sur un tel site, une ligne est enregistrée.

Les chiffres finaux sont eux aussi exagérés, en raison de la façon dont les sites à caractère sexuel sont visionnés. Comme nous l'avons mentionné, ces sites contiennent essentiellement des images accompagnées de très peu de texte, et sont donc destinés à être regardés, et non lus. L'analyse des relevés permet de constater que les visites de ces sites sont brèves mais très actives. Les utilisateurs ont tendance à parcourir les pages du site rapidement et à naviguer vers d'autres sites périphériques tout aussi rapidement. Les chercheurs ont également observé que la majorité des sites à caractère sexuel encouragent ce type de navigation. Sur une page Web traditionnelle,

il existe des liens avec d'autres pages qui sont placés de façon stratégique, et qui promettent toujours des images explicites « plus impressionnantes et de meilleure qualité ».

Une situation particulière illustre très bien ce point. Sur une période de quatre heures, le relevé du 4 décembre a enregistré une séquence presque ininterrompue d'utilisation des sites à caractère sexuel. Notre échantillon a totalisé 120 lignes (2 400 lignes dans les relevés bruts). Les chercheurs ont comparé ce chiffre à trois autres sessions enregistrées le même jour, durant la même période de quatre heures et à partir de relevés générés par différents terminaux Internet de la même succursale. Pour les trois autres sessions, les sujets et le type de sites visités variaient. Ces sessions ont généré entre 28 et 46 lignes dans l'échantillon. Tout comme le courriel et les groupes de conversation, les sites à caractère explicitement sexuel peuvent être surreprésentés dans les données du relevé. Le chiffre 46, qui est élevé, comprend un grand nombre de visites à des groupes de conversation, ce qui démontre le bien-fondé de l'analyse de la section précédente.

Cette catégorie peut également subir l'influence de certaines techniques utilisées pour attirer davantage de visiteurs sur les sites à caractère sexuel. Des expressions et termes de recherche courants parsèment ces sites, de manière à ce qu'ils soient trouvés chaque fois qu'un utilisateur les inclut dans sa recherche. D'autres sites ont manipulé leur URL de sorte qu'il soit extrêmement similaire à des sites très connus comme Amazon.com. Malheureusement, les chercheurs n'avaient pas accès à l'information décrivant la façon dont un client a découvert un site, et n'ont donc pas pu

quantifier l'impact qu'a eu sur les résultats l'installation de termes clés sur les sites ou la manipulation des URL. Par contre, certains éléments indiquent que ces méthodes ont eu une certaine incidence sur les relevés. Par exemple, la majorité des sites à caractère sexuel apparaissent dans les relevés comme des blocs consécutifs. Ce type de situation caractérise également les sites à caractère non sexuel. À au moins deux reprises, les sites à caractère sexuel sont apparus une seule fois sur le relevé, et apparemment hors contexte par rapport aux pages qui les précédaient et les suivaient. Même si de telles anomalies ont peut-être d'autres origines, les chercheurs considèrent qu'il s'agit d'un problème important qu'il faut examiner dans le cadre de l'analyse finale et d'études futures.

### **Pornographie**

Cette catégorie représente 1,9 p. 100 (ou 90 lignes) de l'ensemble de l'échantillon. Selon les chercheurs, les sites pornographiques contiennent des pages qui montrent des images explicites d'actes sexuels, de sexe assorti de violence, de pédophilie ou de zoophilie, sans aucune intention littéraire ou artistique, et dont l'objectif fondamental consiste à provoquer un désir sexuel chez l'utilisateur.

La séparation entre érotisme et pornographie permet de faire la distinction entre les sites qui présentent des actes sexuels extrêmement douteux, voire même illégaux, et les sites qui se contentent d'afficher des photos de personnes nues qui ne pratiquent pas d'acte sexuel flagrant.

Lors de leur analyse finale, les chercheurs n'ont trouvé aucune page montrant



des actes de violence, des enfants ou des animaux, même si certains sites suggéraient ces thèmes par des hyperliens contenant des graphiques. La majorité des sites classés dans cette catégorie présentent plutôt des actes sexuels entre deux ou plusieurs personnes.

Comme nous l'avons vu plus haut, les statistiques relatives à cette catégorie sont elles aussi biaisées par la façon dont les sites à caractère sexuel sont visionnés, et par le nombre disproportionné d'images graphiques qui y sont associées.

La distinction entre les sites érotiques et pornographiques a parfois été difficile à établir. Il arrive qu'une page Web soit érotique selon la définition établie par le chercheur, mais contienne des liens graphiques dont les images pourraient être qualifiées de pornographiques. Les chercheurs ont décidé de classer uniquement la page qu'affichait la ligne du relevé, sans tenir compte de ce que l'intégralité de ce site pouvait contenir. Ils partent du principe selon lequel souvent les liens ne décrivent pas de façon exacte ce que l'utilisateur va vraiment voir s'il choisit cette option.

Par ailleurs, dans certains cas, les chercheurs n'ont pas pu visionner les pages sur les sites dont il fallait être membre ou abonné. Ils ont alors basé leur classification sur la page d'accueil de ces sites.

### **et produits de nature sexuelle**

Cette catégorie représente 0,3 p. 100 (ou 12 lignes) de l'échantillon. Les chercheurs y ont classé les sites contenant des instructions, des discussions ou des conseils liés au sexe et aux questions de nature sexuelle, par exemple les MTS, les

moyens de contraception, etc. Cela inclut l'affichage ou la description de produits sexuels.

Les chercheurs ont choisi de créer une catégorie distincte pour ce thème en raison de la controverse qui entoure l'information de nature sexuelle. Parce que les terminaux Internet de la BPB ne sont pas filtrés, il était également dans l'intérêt des chercheurs de vérifier comment les clients accédaient à un grand nombre de sites de cette nature. Cela leur a également permis de traiter les sites à caractère explicitement sexuel différemment de ceux qui se veulent informatifs et éducatifs, mais portent également sur la sexualité dans une certaine mesure.

Compte tenu de ces critères, certains sites Web ne pouvaient pas être clairement assimilés à une de ces catégories de sites à caractère sexuel. Par exemple, le site baptisé « Men on the Net » combine de l'information sur les questions de santé et d'autres questions touchant les homosexuels, et l'accès à des histoires érotiques et à des photos explicites.

Un très petit nombre de pages de cette catégorie traitent de questions de santé, ce qui est logique, compte tenu du nombre relativement peu élevé de pages à vocation médicale consultées par les clients des bibliothèques en général, comme nous le verrons plus loin. La plupart des pages de cette catégorie traitent du sexe en tant que mode de vie ou sont des sites commerciaux qui proposent des produits sexuels comme des préservatifs.

## **70 Ordinateurs**

Cette catégorie représente 4,4 p. 100 (208 lignes) de l'échantillon. Elle inclut les pages Web portant sur les logiciels, le matériel, les produits et services, l'édition Web, les didacticiels, les services offerts sur les sites Web, etc.

Un grand nombre de pages du site Microsoft étaient présentes dans cette catégorie. Les clients de la bibliothèque ont accédé aussi bien à des pages donnant des renseignements sur les produits Microsoft qu'à des pages contenant des didacticiels sur l'édition de sites Web. Les autres pages placées dans cette catégorie sont très diverses. Les chercheurs pensent qu'il faudrait définir cette catégorie plus en détail en vue d'études futures.

Le nombre relativement élevé d'ordinateurs explique sans doute également pourquoi Internet peut être une source précieuse d'information que la bibliothèque ne pourrait pas offrir dans ses collections. Certaines données anecdotiques révèlent que la Bibliothèque publique de Burnaby stocke un grand nombre d'ouvrages sur l'informatique, mais a encore de la difficulté à répondre aux demandes de ses clients. D'après l'expérience des chercheurs qui ont consulté Internet pour des questions de nature technique liées à la présente étude, Internet contient énormément de renseignements à jour sur l'informatique, auxquels on peut accéder rapidement et selon différents niveaux de complexité. Si l'on se base sur les statistiques relatives à cette catégorie, Internet peut être un complément précieux aux collections d'ouvrages sur l'informatique que possèdent les bibliothèques publiques.

## **80 Sports**

Cette catégorie représente 3,7 p. 100 (174 lignes) de l'échantillon. Elle inclut non seulement des pages Web consacrées aux nouvelles sportives et aux athlètes, mais également des pages qui traitent du sport en tant qu'industrie. Par exemple, les sites qui font la promotion d'une marque d'articles de sport ont été placés dans cette catégorie, de même que les pages qui donnent les résultats sportifs les plus récents. Les chercheurs y ont également intégré les pages Web extraites du cahier des sports des journaux en ligne ou de la chronique sportive des services de nouvelles.

L'un des sports qui dominent cette catégorie est le soccer. Il faut mentionner la présence de sites consacrés à certaines équipes dans des langues autres que l'anglais. Les chercheurs ont eu l'impression que les clients de la bibliothèque trouvaient là un autre moyen de rester en contact avec leur pays d'origine. Par exemple, les relevés contiennent fréquemment des pages Web dans une autre langue que l'anglais extraites du cahier des sports des journaux en ligne ou de la chronique sportive des services de nouvelles.

L'autre sport que l'on rencontre beaucoup dans cette catégorie est la lutte, avec le site de la Fédération internationale de lutte (WWF). Rétrospectivement, les chercheurs envisagent de classer ces pages Web dans le divertissement télévisé, qui s'applique davantage à cette activité. Contrairement au soccer, les sites consacrés à la lutte présentent des chiffres sur ce sport et des produits commerciaux connexes, plutôt que des nouvelles récentes.

## **90 Sciences humaines et sociales\***

Cette catégorie représente 2,8 p. 100 (131 lignes) de l'échantillon. Un certain nombre de catégories individuelles ont été éliminées en raison des chiffres peu élevés que chacune représentait. Il s'agit des catégories suivantes : histoire, arts, sciences sociales, et langue et littérature.

Cette catégorie contient un grand nombre de pages représentant ce que les chercheurs, à défaut d'un autre terme, ont appelé la « contre-culture ». Ils y ont inclus les pages Web concernant ce que l'on considère habituellement comme des styles de vie et des centres d'intérêts marginaux, par exemple la culture « gothique », les vampires, etc., sous la dénomination « Sciences sociales ». Internet est l'endroit idéal pour l'expression de ce type de sous-culture, parce qu'il permet l'échange d'idées entre personnes de tous les pays. Une majeure partie de ces documents n'est d'ailleurs disponible que sur Internet, dans la mesure où elle correspond aux intérêts et aux choix de vie personnel des internautes. Les chercheurs ne prévoyaient pas classer de tels sites dans la catégorie « Sciences sociales », mais ils ont dû s'y résoudre lorsque ce type de documents est apparu sur les relevés. Toute étude future de cette nature devrait créer une sous-catégorie à part pour ce thème.

Les autres pages Web appartenant à cette catégorie traitent généralement de sujets plus généraux comme l'archéologie, la psychologie, les musées, les beaux-arts et la musique. Les pages Web classées dans la catégorie « Sciences sociales » sont souvent moins bien définies que celles des autres domaines. Cela indique qu'il faut

réexaminer ce thème et le subdiviser en un plus grand nombre de sous-catégories plus pertinentes.

## **100 Éducation**

Cette catégorie représente 2,4 p. 100 (ou 112 lignes) de l'échantillon. Elle contient les pages Web qui décrivent les cours offerts et les bourses, et donnent des renseignements généraux sur les établissements d'enseignement ou leurs départements.

Internet fournit des renseignements à la fois complets et à jour sur le thème de l'éducation, et les chercheurs n'ont pas été surpris de constater que ces sites étaient très visités. Ils constituent une bonne source d'information, et les relevés contenaient des pages Web provenant à la fois d'établissements locaux et étrangers. L'information relative aux établissements locaux est généralement facile à trouver dans la bibliothèque publique sous la forme de catalogues de cours, mais l'information sur les établissements nationaux et étrangers peut être plus difficile à trouver.

Les sites classés dans cette catégorie sont généralement très clairement définis. Les chercheurs ont décidé d'intégrer l'information relative à certains départements à la catégorie « Éducation » plutôt qu'à un thème plus précis (p. ex., la Faculté des sciences de l'éducation de l'Université de la Colombie-Britannique, plutôt que les « Sciences »), en raison de l'importance accordée à la place du département au sein de l'établissement, plutôt qu'à la discipline qu'il enseigne.

## **110 Passe-temps / Intérêts personnels\***

Cette catégorie représente 2,4 p. 100 (115 lignes), de l'échantillon. Elle comprend des passe-temps tels que le jardinage, la cuisine, la menuiserie et l'artisanat, entre autres, de même que des pages personnelles et des sites d'achats.

La catégorie « achats » a été créée pour le classement des sites Web qui offrent des produits pour la vente en ligne. Cette catégorie n'a été utilisée que 49 fois. De manière générale, on suppose que les clients de la bibliothèque utilisent Internet pour acheter des biens, ce qui n'est pas toujours le cas, et les recherchistes n'ont pas voulu tirer de conclusions erronées à cet égard. Par exemple, la librairie virtuelle Amazon.com vend des livres en ligne, mais elle présente également des critiques de livres et fournit de l'information sur les auteurs. Ainsi, on a classé les pages Web d'Amazon.com contenant de l'information sous « Langues et littérature ». Par ailleurs, les pages Web de ce site consacrées à l'achat de livres ont été classées dans la catégorie « Achats ». Ce critère s'appliquait aussi aux sites Web commerciaux similaires. On a considéré certains sites, par exemple, les enchères et les catalogues en ligne, comme des sites d'achats.

Les recherchistes ont constaté avec étonnement que les sites associés aux passe-temps étaient très peu visités. On ne sait pas vraiment pourquoi il en est ainsi, mais le fait que la bibliothèque offre d'autres ressources, par exemple, des livres, sur divers passe-temps pourrait expliquer ce résultat. Par exemple, les utilisateurs ont visité très peu de sites contenant des recettes ou ayant trait au jardinage, ce qui pourrait être attribuable au fait que la bibliothèque offre de nombreux ouvrages sur ces sujets. Par ailleurs, les sites visités avaient trait à des passe-temps « éclectiques » par

exemple, la simulation de vols, qui font appel aux capacités multimédias d'Internet.

Les recherchistes admettent que la catégorie des passe-temps est problématique. Elle suppose une certaine interprétation subjective de ce qu'est un passe-temps. Les recherchistes ont examiné des sites tels que Yahoo et Magellan pour définir cette catégorie.

L'échantillon renfermait peu de pages Web personnelles. Un site personnel revenait constamment dans les échantillons, mais il ne portait pas sur un thème en particulier, ce qui laisse supposer que le client de la bibliothèque connaissait le site.

## **120 Lieux d'intérêt et voyages**

Cette catégorie représente 2,3 p. 100 (107 lignes), de l'échantillon. Elle regroupe des pages Web qui contiennent de l'information précise sur un endroit, par exemple, la météo, ou sur l'industrie touristique en général.

Bon nombre des sites classés dans cette catégorie fournissent de l'information détaillée sur des villes ou des pays. Ces sites contiennent de l'information à l'intention des voyageurs éventuels ou des recherchistes, mais ils offrent aussi souvent des liens avec d'autres sources de renseignements sur l'actualité, des journaux et d'autres renseignements pouvant intéresser les gens qui vivent à l'extérieur de leur pays d'origine. De nombreux lieux d'intérêt sont représentés dans cette catégorie, mais il semble que quelques utilisateurs ont effectué plusieurs visites pendant la semaine pour obtenir de l'information sur des pays précis, surtout des pays d'Afrique comme le Soudan, l'Algérie et l'Érythrée. Comme on ne sait pas qui étaient les utilisateurs des



terminaux Internet, il s'agit seulement d'une impression compte tenu du contenu des sites et du contexte du relevé dans lequel ils se trouvaient.

Une des pages figurant dans cette catégorie est intéressante pour les chercheurs, car elle donne une bonne idée du profil des utilisateurs qu'ont dressé les chercheurs pendant l'analyse des relevés. Dans l'échantillon du 7 décembre, on trouvait au début d'un bloc de lignes un site d'information météorologique pour Reno, au Nevada. On a classé ce site dans la catégorie « Lieux d'intérêt et voyages ». La ligne suivante représentait une page contenant des renseignements sur les activités offertes à Las Vegas. Les lignes suivantes correspondaient à des sites portant sur les robes de mariées et les chapelles de Las Vegas. Il semble que l'utilisateur ait opté pour un mariage à Las Vegas sur un thème hawaï en!

### **130 Journaux\***

Cette catégorie représente 1,8 p. 100 (86 lignes), de l'échantillon. Elle regroupe des journaux en ligne du Canada (en anglais et dans d'autres langues), des États-Unis et d'autres pays. On a également noté la langue utilisée lorsque cette langue n'était pas l'anglais.

L'importance relativement faible de cette catégorie s'explique par l'importance démesurée accordée aux sites de courrier électronique et à caractère sexuel pour les raisons que nous avons déjà exposées. D'après les chercheurs, on doit accorder une attention particulière aux sites de journaux, et ce, pour plusieurs raisons. Ces sites contiennent beaucoup de texte et peu de graphiques ce qui fait qu'on les trouve moins

fréquemment dans les relevés. Par ailleurs, comme les journaux contiennent essentiellement du texte et de nombreuses sections, la lecture de ces sites demande plus de temps que la consultation des pages à caractère sexuel, par exemple.

Les résultats de l'analyse de la catégorie « Journaux », ont amené les chercheurs à conclure que l'accès à Internet dans les bibliothèques publiques est important et même nécessaire, et qu'Internet constitue un moyen unique pour les utilisateurs de conserver un lien avec leur pays d'origine ou s'informer sur d'autres endroits d'intérêt. Cette impression a été ressentie par les chercheurs tout au long de l'étude. Soixante-dix-neuf pour cent de tous les journaux consultés (68 lignes), étaient publiés à l'extérieur du Canada et des États-Unis. En outre, 91 p. 100 de ces journaux étaient dans une langue autre que l'anglais, les principales langues étant le russe, le turc, l'espagnol et les langues parlées dans l'ex-Yougoslavie. Compte tenu du grand nombre d'Asiatiques qui vivent à Burnaby, ce pourcentage aurait probablement été encore plus élevé si la bibliothèque avait été dotée du logiciel de reconnaissance de la langue nécessaire au moment de l'étude. Par ailleurs, le fait que les journaux asiatiques sont vendus dans de nombreux endroits peut également expliquer le résultat. L'impression selon laquelle Internet constitue un lien essentiel avec l'étranger est renforcée par le fait que bon nombre des sites de journaux publiés dans une langue autre que l'anglais ont été visités plusieurs fois pendant la semaine.

#### **140 Autres médias d'information\***

Cette catégorie représente 1,5 p. 100 (73 lignes), de l'échantillon. Elle comprend

les pages Web de stations de télévision et de radio, des services d'information en ligne, de même que des pages Web traitant de politique et d'actualité qui ne sont toutefois pas associées à une agence de presse officielle.

Les résultats de l'analyse des sites de cette catégorie ne sont pas aussi uniformes que ceux de la catégorie « Journaux ». De nombreux sites d'agences d'information ont été visités et la majorité de ces sites étaient en anglais. Par exemple, les sites de CBC Newsworld, de CNN et de la BBC ont tous été fréquemment visités pendant la semaine. Il convient de signaler la popularité des pages d'actualité qu'on trouve sur les moteurs de recherche tels que Yahoo et AltaVista. Bon nombre des informations accessibles par le truchement des moteurs de recherche et des répertoires sont toutefois fournies par des réseaux de télévisions, par exemple, ABC News.

Les médias d'information qui diffusent dans une langue autre que l'anglais sont relativement présents dans cette catégorie, quoique dans une moindre mesure que dans la catégorie « Journaux ». Les pages Web créées dans une langue autre que l'anglais représentaient 40 % de toutes les pages visitées dans la catégorie « Autres médias d'information ». Comme c'était le cas dans la catégorie « Journaux », les langues parlées dans l'ex-Yougoslavie, le russe, le turc et l'espagnol comptaient parmi les plus fréquentes. Le fait qu'on retrouve souvent ces langues dans l'échantillon donne à entendre qu'un groupe de clients réguliers de la bibliothèque utilisent fréquemment les terminaux Internet.

**150 Petites annonces**

Cette catégorie représente 1,4 p. 100 (66 lignes), de l'échantillon. Elle regroupe des pages Web qui contiennent des annonces similaires à celles qu'on trouve dans le cahier des annonces classées des journaux, par exemple, des annonces personnelles et autres petites annonces.

On a ajouté cette catégorie en cours d'analyse des relevés, car cette utilisation d'Internet n'avait pas été prévue et n'a été notée qu'après le début de l'étude (les relevés qui avaient déjà été classés ont fait l'objet d'une nouvelle analyse). Les annonces personnelles dominent dans cette catégorie. Très peu de sites contenaient de l'information caractéristique de celle qu'on trouve dans les annonces classées d'un journal.

Les chercheurs ont constaté qu'il existe une diversité de sites d'annonces personnelles sur Internet destinés à pratiquement tous les groupes de population. Par exemple, l'échantillon contenait des sites de rencontre pour aider les musulmans, hommes et femmes, à trouver un conjoint, des sites de rencontre et de mariage asiatiques et un site contenant la description de femmes d'Europe de l'Est. Bon nombre de ces sites contenaient la photographie de la personne qui avait placé l'annonce, ce qu'on ne trouve pas dans les petites annonces des journaux.

**160 Sciences et santé**

La catégorie « Sciences » représente 1,0 p. 100 (45 lignes), de l'échantillon et la catégorie « Santé », 0,4 % (19 lignes). Les pages Web traitant de sujets scientifiques comme la biologie, les mathématiques, la chimie, la physique ou la géologie ont été

regroupées dans la catégorie « Sciences » tandis que les pages ayant trait à la santé, y compris le domaine pharmaceutique, ont été classées dans la catégorie « Santé ».

Les chercheurs avaient prévu une plus grande utilisation des sites inclus dans ces deux catégories. Notamment, le nombre de visites de pages Web ayant trait à la santé est beaucoup plus faible que ce qu'on avait prévu.

La période au cours de laquelle l'étude a été menée peut expliquer en partie cette situation. En effet, l'étude a été effectuée pendant la première semaine de décembre, soit juste avant la période des examens scolaires. Comme les étudiants n'avaient pas de travaux scolaires à faire, ils avaient moins besoin du genre d'information fournie sur ces sites.

## **170 Répertoires, organisations**

La catégorie « Répertoires » représente 0,4 % (20 lignes) de l'échantillon et la catégorie « Organisations », 0,3 % (15 lignes). On entend par répertoire une liste de noms ou d'adresses d'employés, de fonctionnaires du gouvernement, ou autres, qu'on trouve habituellement dans les répertoires de la bibliothèque. La catégorie « Organisations » comprend les organismes de bienfaisance qu'on n'a pas pu classer sous un thème précis. Les catégories « Répertoires » et « Organisations » ont été ajoutées après le début de l'analyse du relevé pour répondre à un besoin perçu.

Même si les sites de répertoires n'ont pas été visités souvent, on les a placés dans une catégorie distincte, car il était difficile de les inclure dans une des catégories existantes et compte tenu de l'importance d'Internet pour obtenir les coordonnées des

personnes-ressources. On ne pouvait malheureusement pas savoir à quelle fréquence Internet était utilisé pour obtenir ce genre d'information lors de l'examen de sites qui n'étaient pas précisément des répertoires. Par exemple, une personne qui visite la page d'accueil d'une société d'avocats peut simplement chercher à obtenir l'adresse de la société ou de l'information au sujet du droit. Selon l'expérience qu'ont les chercheurs de la navigation sur Internet, on peut penser que cet outil constitue une bonne source d'information sur les personnes-ressources, mais comme on ne sait pas à quelles fins les utilisateurs naviguent sur Internet, il n'a pas été possible de déterminer à quelles occasions un site a été visité dans le but d'obtenir ce genre d'information.

Le nombre de pages Web d'organisations qui ont été visitées est plus élevé que ce que laissent entendre les statistiques susmentionnées. En effet, les chercheurs ont choisi de classer les organisations par domaine, dans la mesure du possible. Par exemple, la page Web d'une association scientifique est classée dans la catégorie « Sciences ». Les organismes de bienfaisance ne peuvent généralement pas être associés à un domaine; les chercheurs ont donc jugé qu'il serait préférable de créer une catégorie distincte pour ces organismes. La catégorie a été établie pendant les travaux de classification et elle pourrait donner un aperçu plus clair de ce qu'elle regroupe. Les sites qu'on a classés dans cette catégorie comprennent ceux d'organismes de défense des droits de la personne et le site d'OXFAM.

**180 Droit**

Cette catégorie représente 0,1 % (3 lignes) de l'échantillon. La catégorie

« Droit » a été incluse pour permettre la classification des pages Web sur les sociétés d'avocats, les lois et la législation ainsi que les institutions juridiques.

Cette catégorie regroupe des sites comme un outil de recherche HotBot pour le Justice Institute of B.C., un guide sur l'*Employment Standards Act* produit par le gouvernement de la Colombie-Britannique et la page Web d'une société d'avocats spécialisée en droit des brevets et en droit de la propriété intellectuelle. Internet est un bon outil pour la diffusion de ce genre d'information. En augmentant la connaissance des utilisateurs au sujet des catégories d'information qu'on peut trouver sur Internet et en améliorant les capacités de recherche de ces derniers, on pourra peut-être favoriser une plus grande utilisation des documents juridiques. Il pourrait être intéressant d'effectuer le suivi de l'utilisation de l'information juridique sur une plus longue période pour déterminer si la faible utilisation constatée constitue la norme ou si elle correspond à la période de l'étude.

## **V. Pages Web des administrations publiques**

---

La détermination des sites du gouvernement s'est faite de deux façons. D'abord, on a considéré les sites qui ne portaient sur aucun thème particulier comme des sites « généraux » des administrations fédérale, provinciales, municipales ou de pays étrangers. Ensuite, on a classé les sites traitant d'un sujet particulier (p. ex., les sciences ou l'emploi), dans la catégorie pertinente, dans un premier temps, puis on a créé une sous-catégorie correspondant au palier de gouvernement.

### **Gouvernement fédéral**

Les pages du gouvernement fédéral représentaient 0,7 % (32 lignes), de l'échantillon, et ces pages traitaient de divers sujets. Vingt-deux pour cent (7 lignes), des sites ont été classés dans la catégorie des sites généraux, soixante-six pour cent (21 lignes), contenaient de l'information relative à l'emploi et douze pour cent des sites ont été classés dans les catégories « Affaires-Industrie » ou « Sciences ». Les chercheurs ont noté un grand nombre de visites sur la page Web « Guichets emploi » de Développement des ressources humaines du Canada.

### **Gouvernement provincial**

Cette catégorie représente 0,3 % (13 lignes), de l'échantillon. Les pages incluses dans cette catégorie se rapportent aux divers gouvernements provinciaux du Canada.

Comme c'est le cas pour les pages du gouvernement fédéral, les pages relatives à l'emploi sont les plus fréquentes, soit 6 pages sur 13. On a classé quatre pages dans la catégorie des pages générales, soit deux de la Colombie-Britannique et deux de la Nouvelle-Écosse. Une page entrait dans la catégorie « Droit » et deux, dans la catégorie « Lieux d'intérêt et voyages ».

Les pages gouvernementales ont été classées dans quelques catégories seulement, car il est difficile de savoir à quelles fins on utilise ces sites à la bibliothèque. Les statistiques sur les sites Web obtenues à partir des serveurs des gouvernements



pourraient fournir de meilleures données à cet égard.

### **Administration municipale**

Cette sous-catégorie ne représentait que 0,04 % (2 lignes), de l'échantillon. Le site du service des parcs, des loisirs et de la culture de la ville de Burnaby est l'un des sites qui ont été visités. Le client qui a visité ce site y a peut-être eu accès à partir de la page d'accueil de la bibliothèque publique de Burnaby. En effet, le site Web de la BPB offre quelques liens avec des sites gouvernementaux.

L'autre site visité dans cette sous-catégorie contenait un rapport stratégique de la ville de Vancouver produit par le comité permanent de la planification et de l'environnement. Ce genre de document peut être difficile à trouver sur support papier si le client de la bibliothèque ne connaît pas les documents gouvernementaux, mais on peut facilement le trouver sur Internet en utilisant un moteur de recherche.

### **Gouvernement étranger**

Cette catégorie représente 0,5 % (24 lignes), de l'échantillon. Elle comprend tous les sites créés par des gouvernements étrangers, y compris les sites des ambassades.

Les données recueillies pendant une seule journée contribuent à gonfler les statistiques dans cette catégorie. Au cours d'une session, un client a navigué longuement sur un site du gouvernement iranien. La consultation des sites de gouvernements étrangers variait d'une journée à l'autre. Le site de la législature du

Texas entre dans cette catégorie, tout comme le site d'une ambassade à partir desquels on peut télécharger une demande de passeport.

## **VI. Sites dans des langues autres que l'anglais**

---

La catégorie des sites dans des langues autres que l'anglais présentait un intérêt particulier pour la BPB et pour les chercheurs. Cette catégorie a donc fait l'objet d'une analyse plus détaillée. On a répertorié les pages Web en langues autres que l'anglais en utilisant le titre « Langue autre que l'anglais » et en précisant la langue utilisée. (p. ex., Sports - langue autre que l'anglais - espagnol). Dans les cas où on ne pouvait pas classer une page Web dans une catégorie, on utilisait la catégorie « général » des sites en langue autre que l'anglais. Seulement 39 des 437 sites ont été classés ainsi.

Les pages Web en langue autre que l'anglais représentaient 9,3 % (437 lignes), de l'échantillon. Cette façon de regrouper les sites fait ressortir l'importance des pages Web en langues autres que l'anglais qui ont été consultées. Les sites en langues autres que l'anglais arrivent au troisième rang, toutes catégories confondues, derrière les catégories « Courrier électronique » et « Divertissement » (à l'exclusion des sites impossibles à classer).

On a répertorié des sites dans seize langues différentes dans le cadre de l'étude. Les cinq langues retrouvées le plus fréquemment dans l'échantillon sont les suivantes (en ordre de fréquence) : le russe, les langues de l'ex-Yougoslavie, l'espagnol, l'allemand et le chinois. La fréquence élevée des pages Web en russe peut

être attribuable en partie aux nombreuses visites sur un groupe de conversation au cours des sept jours de l'étude. Il est intéressant de signaler que toutes les visites de pages Web russes ont été effectuées à la succursale Metrotown de la BPB. Ces constatations ont été corroborées par le personnel de cette succursale qui fait état de la présence de clients russes qui utilisent fréquemment les terminaux Internet de la bibliothèque.

**Tableau III : Sites en langues autres que l'anglais (en ordre d'importance)**

Langue	N <sup>bre</sup> de lignes	% de sites en langues autres que l'anglais
1. Russe	127	29,1 %
2. Langues yougoslaves	99	22,7 %
3. Espagnol	85	19,5 %
4. Allemand	37	8,5 %
5. Chinois	21	4,8 %
6. Français	20	4,6 %
7. Turc	18	4,1 %
8. Langues asiatiques	8	1,8 %
9. Coréen	8	1,8 %
10. Japonais	3	0,7 %
11. Slovaque	3	0,7 %
12. Polonais	2	0,5 %
13. Tchèque	2	0,5 %
14. Hongrois	2	0,5 %
15. Suédois	1	0,2 %
16. Italien	1	0,2 %

Les pages Web rédigées dans les langues de l'ex-Yougoslavie portent sur des sujets plus variés que les sites en russe. Les journaux en ligne et les autres services d'information étaient les sites les plus fréquemment visités par les membres de ces groupes linguistiques. Les guides portant sur l'utilisation d'Internet rédigés dans les

langues parlées dans l'ex-Yougoslavie, par exemple, le Serbian Café, ont été consultés fréquemment, comme on peut le constater en examinant l'échantillon. On y trouve autant les actualités que des liens avec des sites Internet.

Les pages Web en espagnol se distinguent des autres, et ce, pour plusieurs raisons. Par exemple, les sites en espagnol sont les seuls qu'on retrouvait dans toutes les succursales. En outre, la nature des pages espagnoles visitées était très diversifiée, ce qui fait qu'il était difficile de classer ces pages dans une catégorie. Les sites les plus souvent visités ont été classés dans les catégories « Journaux », « Sciences humaines et sociales », « Affaires » et « Magazines ».

Les pages en allemand contenaient surtout de l'information sportive, y compris un groupe de conversation sur le soccer allemand. Toutes les occurrences des pages Web allemandes ont été observées à la succursale Metrotown. Ces constatations peuvent donner à entendre qu'il y a un seul utilisateur ou un petit groupe d'utilisateurs. Comme pour les autres langues, les sites de journaux allemands ont été visités à de nombreuses occasions.

Compte tenu de l'importance de la communauté chinoise à Burnaby, le nombre de pages en chinois qui ont été visitées est probablement moindre que ce qu'il aurait été si la BPB avait disposé du logiciel nécessaire pour afficher les caractères chinois au moment de l'étude. Les pages relatives à l'immobilier ont été les plus populaires de toutes les pages visitées. Cette constatation est toutefois quelque peu trompeuse, car toutes les pages d'immobilier et, par conséquent, la plupart des pages chinoises ont été visitées au cours d'une seule session. L'utilisation des sites en français, qui

venaient après le chinois au chapitre de la fréquence des visites, était répartie plus également sur la période de l'étude, ce qui indique une utilisation plus régulière. Contrairement aux autres langues, on a noté une lacune au niveau des sites de nouvelles en chinois, ce qui est probablement attribuable au fait que la bibliothèque ne possède pas le logiciel nécessaire. Cela peut aussi être dû au fait que les journaux en chinois sont accessibles à la BPB et dans les kiosques à journaux de la ville.

Les recherchistes ont porté un intérêt particulier à l'utilisation très précise des sites en français. En effet, les sites visités étaient presque exclusivement du Québec et d'Algérie. Les pages Web du Québec portaient sur les villes de Montréal et de Québec, de même que sur des magazines et des journaux en français comme *Le Devoir* et *Planète Québec*. Les sites algériens étaient essentiellement à caractère politique et comprenaient deux journaux en ligne, à savoir *Liberté* et *Le Watan*.

Les recherchistes estiment que la présence de pages en langues autres que l'anglais dans les relevés met en lumière les avantages de l'accès à Internet dans les bibliothèques publiques. Les constatations décrites dans le présent document et relatives à chaque catégorie illustrent le caractère unique d'Internet comme source d'information, information qu'on ne pourrait pas trouver ailleurs à la bibliothèque ou ailleurs. L'utilisation prépondérante et régulière des journaux et autres médias d'information dans presque toutes les langues autres que l'anglais donne à entendre qu'Internet permet aux clients de la bibliothèque de rester en contact avec leur province ou leur pays d'origine.

## **VII. Périodiques en ligne (magazines et publications électroniques)\***

---

Aux fins de l'étude, les pages Web pages des magazines en ligne et des publications électroniques ont été consignées de deux façons. On a attribué à chacune d'elles une catégorie séparée ainsi qu'une sous-catégorie. Les magazines et les publications électroniques ont été classés dans leur propre catégorie uniquement lorsque le site était trop vaste pour correspondre à un thème en particulier. On a attribué les sous-catégories magazines ou publications électroniques à toutes les pages Web, qu'elles soient classées ou non dans un autre domaine ou dans leur propre catégorie. À titre d'exemple, chaque site de magazines a été autant que possible classé dans son domaine, puis une sous-catégorie a été précisée afin de souligner les deux aspects du site. Cette classification plus précise permet aux bibliothécaires de comparer l'utilisation des périodiques en ligne à celle des magazines sur papier.

On a parfois eu de la difficulté à déterminer, cependant, si certains sites constituaient en fait une version en ligne des magazines sur papier, des publications électroniques ou simplement d'une publication en série mise à jour irrégulièrement. À cause de cette incertitude, les chercheurs ont décidé d'intégrer les magazines et les publications électroniques dans une seule catégorie intitulée périodiques en ligne.

Le pourcentage total de page Weg consacrées aux périodiques en ligne a été calculé en fonction du nombre total de lignes de relevé associées aux sous-catégories des magazines et des publications électroniques. Par conséquent, les périodiques correspondent à 2,5 % (116 lignes) de l'échantillon. De ce total, les trois domaines les

plus souvent visionnés étaient les sciences sociales, les ordinateurs et les sports. Ces domaines se prêtent aux publications régulières et aux mises à jour. Ainsi, un magazine en ligne allemand sur les sports est apparu un nombre de fois. L'absence des magazines bien connus ou que l'on retrouve dans les kiosques à journaux, tels que Time, Maclean's ou Sports Illustrated a lieu d'étonner.

---

### **VIII. Utilisation d'Internet sur semaine**

---

Dans cette étude, l'analyse des habitudes d'utilisation d'Internet en fonction du jour de la semaine est très limitée et ce, pour plusieurs raisons. Parce que les relevés de transactions ont été effectués pendant une semaine seulement, il est difficile de comparer les jours étant donné qu'il n'existe qu'un ensemble de données par jour. En outre, les heures d'ouverture de la BPB sont différentes la semaine et la fin de semaine, ce qui a une incidence sur la taille et la diversité des sites Web enregistrés chaque jour. Ainsi, la succursale Metrotown est ouverte 12 heures par jour sur semaine, 9 heures le samedi et 4 heures le dimanche. En outre, le nombre d'utilisateurs était inférieur dans les relevés compilés pendant les heures d'ouverture plus courtes. Pour effectuer une comparaison de l'utilisation d'Internet selon le jour de la semaine, il aurait fallu recueillir un échantillon plus vaste portant sur plusieurs semaines.

Compte tenu de ces limites, on peut cependant faire quelques observations. L'échantillon du 6 décembre (un dimanche) semble se démarquer des échantillons des

autres jours de la semaine. Ce jour-là 51 % des utilisateurs ont eu recours à l'Internet pour l'envoi de messages par courrier électronique, et cette proportion est beaucoup plus élevée que pour les autres jours. En outre, c'est le seul jour de la semaine où personne n'a visité de sites à caractère sexuel. Il est difficile de déterminer pourquoi l'usage d'Internet est si différent le 6 décembre. L'utilisation élevée du courrier électronique et l'absence de sites à caractère sexuel ne semble pas typique de l'usage d'Internet en général. Pour déceler des tendances réelles, il faudrait recueillir un échantillon plus vaste pour plusieurs dimanches. Il faut également tenir compte du fait que la succursale Metrotown de la BPB est ouverte seulement quatre heures le dimanche.

## **IX. Utilisation d'Internet dans les succursales**

---

En comparant l'utilisation d'Internet dans les succursales, on a constaté que les domaines consultés par les utilisateurs sont essentiellement les mêmes. On doit cependant se rappeler que les heures d'ouverture de la succursale Cameron sont moins longues que celles des autres succursales. Du lundi au samedi, elle est ouverte dix heures de moins que les autres succursales. Metrotown est également la seule succursale qui soit ouverte le dimanche. Pour analyser l'utilisation d'Internet dans les diverses succursales, on a fait une liste des catégories selon la fréquence d'utilisation dans chaque succursale, puis on a fait une comparaison. Les résultats sont énumérés dans le tableau suivant (la catégorie inclassifiable est exclue) :

### **Tableau IV : Catégories les plus fréquemment consultées dans les succursales**



<b>Metrotown</b>	<b>Cameron</b>	<b>Kingsway</b>	<b>McGill</b>
1. Courrier électronique	Courrier électronique	Courrier électronique	Courrier électronique
2. Affaires	Divertissement	Moteurs de recherche	Divertissement
3. Moteurs de recherche	Moteurs de recherche	Affaires	Moteurs de recherche
4. Sexe	Petites annonces	Divertissement	Ordinateurs
5. Divertissement	Sports	Sports	Affaires
6. Ordinateurs	Affaires/Ordinateurs	Ordinateurs	Passe-temps

Comme on peut le constater, certaines catégories prédominent dans toutes les succursales. L'ordre dans lequel elles apparaissent diffèrent légèrement, mais en général les résultats sont similaires. Le courrier électronique, les moteurs de recherche, les affaires, le divertissement et les ordinateurs figurent dans chaque liste. Le nombre de visites aux sites de petites annonces est très élevé à Cameron, ce qui constitue une anomalie étant donné que cette catégorie figure au 15<sup>e</sup> rang lorsque les résultats des succursales sont combinés. Les passe-temps figurent uniquement dans la liste de la succursale McGill.

Ce classement serait différent si le nombre de pages Web dans une langue autre que l'anglais était incorporé aux résultats. Les langues autres que l'anglais seraient au second rang à Metrotown et au cinquième rang à Kingsway et à McGill. Cela démontre l'importance de l'accès au matériel non anglophone fourni par Internet.

La diversité des langues non anglaises dans les succursales reflète sans doute le niveau de diversité des utilisateurs dans chaque succursale. C'est à Metrotown que le nombre et la diversité de pages Web non anglaises visionnées par l'entremise des terminaux Internet sont les plus élevés. Kingsway et McGill présentent un éventail de langues légèrement moins vaste, puisque cinq langues sont représentées. La succursale Cameron présente le pourcentage le moins élevé de pages Web non

anglaises et l'éventail de langues le plus restreint. Deux langues y figuraient et elles sont attribuables à deux séances sur Internet. L'espagnol est la seule langue qui apparaît systématiquement dans toutes les succursales.

On a également constaté que l'utilisation d'Internet dans les succursales plus petites visait un nombre restreint de domaines comparativement à la diversité constatée à Metrotown. Les chercheurs se sont demandé si l'utilisation d'Internet était influencée par le nombre de terminaux mis à la disposition des utilisateurs dans chaque succursale. Étant donné que Metrotown dispose de quatre terminaux, les contraintes de temps et les pressions dues aux files d'attente sont probablement des facteurs moins importants dans le cas de Metrotown.

Une autre différence entre les succursales se rapporte au pourcentage élevé de matériel à caractère sexuel qui est visionné à la succursale Metrotown. La proportion de matériel à caractère sexuel consulté s'établit à 7,0 % à Metrotown, tandis qu'elle atteint 4,4 % à Kingsway, 2,2 % à McGill et 1,1 % à Cameron. Ces différences sont peut-être dues à l'emplacement des terminaux Internet dans chaque succursale. En effet, les terminaux de Metrotown sont situés dans une aire peu fréquentée, loin du service de référence. Les terminaux des autres succursales sont situés dans des endroits très passants. Le profil démographique des utilisateurs d'Internet, notamment l'âge et le sexe, pourrait également varier d'une succursale à l'autre. Il faudrait mener d'autres études pour tirer des conclusions solides.

En général, l'utilisation d'Internet était semblable d'une succursale à l'autre. Les chercheurs estiment cependant que pour obtenir des résultats plus concluants, ils

faudrait mener une étude prolongée basée sur un échantillon plus important. Il est difficile d'évaluer dans quelle mesure les résultats ont pu être influencés par les séances inhabituellement longues par un seul utilisateur ou par d'autres circonstances anormales.

## **X. Problèmes liés à la recherche**

---

La plupart des problèmes qui sont survenus au cours de cette étude exploratoire ont été mentionnés dans les sections précédentes. Les paragraphes suivants fournissent des commentaires plus généraux sur les problèmes auxquels les chercheurs ont été confrontés lors de l'étude.

### *Surreprésentation*

— Un des principaux problèmes auxquels les chercheurs ont été confrontés en manipulant les données des relevés est la surreprésentation de certains types de pages due à la nature des pages ou à la façon dont elles sont utilisées. Le courrier électronique et les pages contenant du matériel à caractère sexuel sont les exemples les plus frappants de ce problème. On en a d'ailleurs parlé dans l'explication de ces deux catégories. Les études subséquentes devront tenir compte de ce problème et il faudra effectuer d'autres études pour tenter de trouver une façon de rééquilibrer cette représentation plutôt inégale ou de déterminer à quel point elle affecte les autres données. On devrait peut-être consulter les analyses statistiques effectuées par les propriétaires de serveurs Web afin de déterminer s'il est possible d'extraire les lignes représentant une visite à une page dans les relevés bruts plutôt que d'effectuer une analyse ligne par ligne. Le champ du relevé qui

enregistre l'heure de chaque transaction pourrait également servir à vérifier la durée d'utilisation dans certains domaines tels que le courrier électronique et les sites à caractère sexuel plutôt que représenter un simple nombre de lignes dans le relevé.

### *Données inclassables*

- \_ On doit également tenir compte du grand nombre de lignes qui produisent des données inclassables, même lorsqu'on a éliminé ce genre de fichier des relevés. Étant donné la nature des relevés, les chercheurs ne s'attendaient pas à pouvoir classer toute l'information recueillie, mais la proportion de données inclassables était plus importante qu'ils n'auraient voulu. Il est encore possible d'apporter des améliorations dans ce domaine. Ainsi, les techniques servant à cerner les pages où figurent une bannière permettraient de retirer les fichiers inclassables des relevés avant de choisir l'échantillon et, à ce titre, elles devraient être explorées. Les pages Web qui disparaissent ou qui sont protégées par un mot de passe poseront toujours un problème.

### *Pages Web qui disparaissent ou sont modifiées*

- \_ La disparition et la révision des pages Web soulève un autre problème. En effet, les chercheurs sont souvent incapables de déterminer si la page qu'ils observent est la même que celle que l'utilisateur de la bibliothèque a consultée au cours de l'étude. À titre d'exemple, le site Web de la bibliothèque de Burnaby a été révisée entre le moment où les relevés ont été recueillis et celui de l'analyse. Les liens

offerts aux clients ont donc été modifiés.

### *Durée de l'étude*

- \_ Il est important de tenir compte de la période au cours de laquelle l'étude a été menée. Des données non scientifiques suggèrent que l'usage varie en fonction de la période de l'année. Ainsi, au printemps, un plus grand nombre d'étudiants utilisent les terminaux Internet pour consulter les anciens examens provinciaux. On peut également présumer que les événements qui surviennent à certains moments de l'année affectent l'utilisation d'Internet. Les études subséquentes devront tenir compte de cet aspect, particulièrement celles qui porteront sur deux bibliothèques ou plus.

### *Nature des relevés*

- \_ La classification des sites a été lente parce que de nombreuses lignes représentent des parties de pages qui ne peuvent être classées qu'après que le chercheur est revenu sur ses pas pour déterminer où se trouve le fichier représenté dans la ligne de relevé. Comme on l'a mentionné dans la section sur la méthode, de nombreux fichiers représentant une ligne dans le relevé emmènent le chercheur à une icône, à un fichier .jpg ou à un élément d'une fenêtre et il est difficile de les classer sans poursuivre les recherches. Les chercheurs pouvaient habituellement classer environ 100 sites par heure. Cependant, dans le cas des pages qui n'étaient pas en anglais ou portaient sur plusieurs sujets, et des fichiers mentionnés ci-haut, le processus était plus lent et les chercheurs ne pouvaient classer que 30 sites par

heure.

### *Subjectivité de la classification*

- \_ En plus de ralentir le travail de classification, les pages portant sur plus d'un sujet s'exposent à des interprétations multiples et par le fait même à des classifications multiples. Il est très important de définir chaque catégorie avec précision. Toutes les personnes qui participent à la classification doivent se consulter avant et pendant l'analyse des relevés afin d'assurer un bon niveau d'uniformité. Tout en tentant de minimiser les irrégularités de la classification, les chercheurs qui ont mené la présente étude sont demeurés conscients de la subjectivité du processus de classification. Ils ont minimisé les irrégularités en communiquant assidûment pendant le travail de classification des relevés.

### *JavaScript*

- \_ Un problème auquel les chercheurs ont été confrontés en classant les pages Web est dû à la présence sur certaines pages de JavaScript qui ouvrait plusieurs navigateurs dès qu'ils avaient accédé à une page. Cela s'est notamment produit lors du visionnement de certaines pages à caractère sexuel. Le JavaScript était habituellement actionné lorsque les chercheurs tentaient de fermer le navigateur ou de cliquer sur le bouton de la page de départ pour quitter le site. Lorsque cela se produisait, le chercheur devait attendre que toutes les fenêtres se soient ouvertes avant de pouvoir les refermer. Le JavaScript n'était pas activé dans les terminaux de la BPB pendant la durée de l'étude, mais les études subséquentes devront tenir

compte de ce phénomène. En effet, chacune des fenêtres de navigateur serait enregistrée sur une ligne différente du relevé, ce qui hausserait les chiffres relatifs au matériel à caractère sexuel.

### *Amélioration de certaines catégories*

\_ On a consacré beaucoup de temps à l'élaboration d'une liste de catégories pouvant servir à classer les pages Web dans les relevés. Bien qu'elle ait été fort utile, cette liste pourrait être améliorée. Un des objectifs de cette étude pilote était de cerner les problèmes de méthode. Les sujets qui ont posé des problèmes étaient affaires et lettres et sciences humaines/sciences sociales. Ces sujets ont fait l'objet d'une discussion dans les sections qui précèdent. Des améliorations s'imposent, cependant, pour que l'utilisation d'Internet puisse être mieux comprise. Par exemple, la catégorie affaires – produits et services s'est avérée utile mais a servi de catégorie polyvalente. Une grande partie des renseignements sur les produits et services ont été consignés dans d'autres catégories, notamment dans la catégorie ordinateurs, lorsque les produits étaient reliés aux ordinateurs, et dans la catégorie sports, lorsqu'il s'agissait d'articles de sport, tels que les chaussures de courses. Ces catégories devraient probablement être subdivisées, par exemple sports (pointage, information sur les équipes, etc.) et sports – industrie (matériel d'athlétisme) pour accommoder les éléments commerciaux des pages Web ainsi que le domaine général. Un examen des catégories utilisées par les répertoires de catégories Internet pourrait s'avérer utile à la lumière des connaissances acquises

dans le cadre de l'étude.

- \_ Une étude plus vaste, menée sur une plus longue période, fournirait des données plus solides dans tous les domaines. Un ensemble plus vaste de données permettrait peut-être d'éliminer certaines anomalies qui se sont produites au cours de l'étude et de présenter une image plus claire des habitudes d'utilisation publique d'Internet dans les bibliothèques publiques telles que la Burnaby Public Library.

## **XI. Recommandations pour les études à venir**

---

Les chercheurs considèrent que cette étude pilote a atteint ses objectifs. Au cours de cette étude et de l'analyse des résultats, on a été en mesure de discerner clairement les problèmes et les défis que pose la réalisation d'un tel travail. En dépit de certains problèmes que les chercheurs ont rencontrés au cours de la réalisation de l'étude, leur impression concernant la méthode qu'il ont employée pour étudier l'utilisation d'Internet est globalement positive. Les problèmes rencontrés et les erreurs commises pendant l'étude ne sont pas insurmontables. Une fois qu'on y aura remédié, les chiffres qui en seront tirés donneront une idée de l'utilisation d'Internet dans les bibliothèques publiques telle que le souhaitent les chercheurs. Toute étude à venir devrait être menée pendant plus longtemps afin de pouvoir fournir des indications plus rigoureuses quant aux habitudes d'utilisation.

Une étude comparative à l'échelle du pays sur l'utilisation d'Internet dans les bibliothèques publiques serait d'un intérêt particulier. Elle nécessiterait toutefois une meilleure organisation et devrait tenir compte des différents types d'accès qu'offrent les



bibliothèques publiques. Même une comparaison des bibliothèques qui se trouvent dans le Lower Mainland de Vancouver devrait tenir compte des différents types d'accès à Internet qui sont fournis. Par exemple, la Burnaby Public Library n'utilise aucun logiciel de filtrage ou de blocage d'information, mais fournit des écrans de protection de la vie privée alors qu'on trouve l'inverse dans les systèmes de la Vancouver Public Library qui filtre l'accès offert à la section consacrée aux enfants, empêche l'accès aux sites de courrier électronique très en vogue et ne fournit pas d'écrans de protection de la vie privée. Toutes ces différences ainsi que les nombreuses autres qui existent devraient être prises en compte.

L'importance d'une étude à l'échelle du pays tient à la multitude de renseignements qu'elle produirait. Une telle étude étofferait également les résultats de toutes les autres études d'envergure plus modeste et permettrait d'établir s'il existe véritablement des tendances (corrigées de l'influence que peut avoir l'emplacement des bibliothèques) dans les habitudes d'utilisation d'Internet par les clients des bibliothèques publiques. Les différences locales que ferait ressortir une telle étude seraient également intéressantes.

## Annexe 1 Domaines

courrier électronique	ordinateurs et industrie de l'informatique
bavardage, babillards électroniques	sciences
	santé et médecine
sexe - information et articles	éducation
sexe - érotisme	lieux d'intérêt et voyages
sexe - pornographie	
	histoire, religion et autres lettres et sciences humaines
Moteurs de recherche/répertoires de thèmes	peinture, danse, musique et autres activités artistiques
	anthropologie, psychologie, et autres sciences sociales
affaires - immobilier	langue et littérature
affaires - finances personnelles	
affaires - industrie	sports et industrie du sport
affaires - services et produits	industrie du divertissement - divers
emploi	industrie du divertissement - musique
	industrie du divertissement - télévision/cinéma
journaux	industrie du divertissement - radio
journaux - canadiens (anglais/français)	jeux
journaux - canadiens (non anglais/non français.)	
journaux - américains	passé-temps
journaux - autres pays	pages web personnelles
	achats
autres médias d'information	petites annonces
politique et actualité	
	répertoires
magazines	organismes à but non lucratif, de bienfaisance/ONG
publications électroniques	
	langues étrangères
gouvernement - fédéral	404 - client
gouvernement - provincial	404 - chercheur
gouvernement - municipal	404 - protégé par mot de passe
gouvernement - étranger	inclassable
droit	bannière

**sous-domaines:**




### Références consultées

---

Connaway, Lynn Silipigni, John M. Budd, and Thomas R. Kochtanek. "An Investigation of the Use of an Online Catalog: User Characteristics and Transaction Log Analysis."

Flaherty, Patricia. "Transaction Logging Systems: A Descriptive Summary." *Library Hi Tech* 11, no. 2 (1993): 67-78.

Harvey, Kathy, and Tom Horne. "Surfing in Seattle: What Cyber-Patrons Want." *American Libraries* 26 (1995): 1028-30.

Jul, Erik. "Cataloging Internet Resources: Survey and Prospectus." *Bulletin of the American Society for Information Science* 24, no. 1 (1997): 6-9.

Kaske, Neal K. "Research Methodologies and Transaction Log Analysis: Issues, Questions, and a Proposed Model." *Library Hi Tech* 11, no. 2 (1993): 79-86.

Kurth, Martin. "The Limits and Limitations of Transaction Log Analysis." *Library Hi Tech* 11, no. 2 (1993): 98-103.

Lazinger, Susan S., Judit Bar-Ilan, and Bluma C. Peritz. "Internet Use by Faculty Members in Various Disciplines: A Comparative Case Study." *Journal of the American Society for Information Science* 48, no. 6 (1997): 508-18.

McClure, Charles R. "User-Based Data Collection Techniques and Strategies for Evaluating

Perry, Clifford. "Travelers on the Internet: a Survey of Internet Users." *Online* 19, no. 2 (1995): 29-34.

Peters, et al. "An Introduction to the Special Section on Transaction Log Analysis."

*Library Hi Tech* 11, no. 2 (1993): 38-40.

Peters, Thomas A. "The History and Development of Transaction Log Analysis." *Library Hi Tech* 11, no. 2 (1993): 41-66.

Simpson, Charles W. "Opac Transaction Log Analysis: the First Decade." In *Advances in Library Automation and Networking: A Research Annual*. Vol. 2, ed. Joe A. Hewitt, 35

"The Statue of Libertines?" *The Library Association Record* 99 (1997): 296

Stout, Rick. *Web Site Stats: Tracking Hits and Analyzing Traffic*. Berkeley: Osborne/McGraw-Hill, 1997.

Tillotson, Joy, Joan Cherry, and Marshall Clinton. "Internet Use Through the University of Toronto Library: Demographics, Destinations, and Users' Reactions." *Information Technology and Libraries* 14, no. 13 (1995): 190-98.