

Construction automatique de ressources dans Internet à partir de données numériques : le cas des statistiques sur les bibliothèques publiques du Québec

Automatic Construction of Resources on the Internet Using Numeric Data: The Case of Public Library Statistics in Québec

Desarrollo automático de los recursos en el Internet a partir de datos numéricos: el caso de las estadísticas sobre las bibliotecas públicas de Quebec

Yves Marcoux et Gabriela Gavrilit

Volume 43, numéro 1, janvier–mars 1997

URI : <https://id.erudit.org/iderudit/1033104ar>

DOI : <https://doi.org/10.7202/1033104ar>

[Aller au sommaire du numéro](#)

Éditeur(s)

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

ISSN

0315-2340 (imprimé)

2291-8949 (numérique)

[Découvrir la revue](#)

Citer cet article

Marcoux, Y. & Gavrilit, G. (1997). Construction automatique de ressources dans Internet à partir de données numériques : le cas des statistiques sur les bibliothèques publiques du Québec. *Documentation et bibliothèques*, 43(1), 23–30. <https://doi.org/10.7202/1033104ar>

Résumé de l'article

Dans le but d'assurer une diffusion plus large et plus rapide des statistiques annuelles relatives aux bibliothèques publiques et aux centres régionaux de services aux bibliothèques publiques, le ministère de la Culture et des Communications du Québec a convenu avec l'École de bibliothéconomie et des sciences de l'information (EBSI) de rendre disponibles ces données sur les sites Gopher et Web de l'EBSI. Le présent article décrit la procédure automatique de transfert élaborée à cette fin et les résultats obtenus.

Construction automatique de ressources dans Internet à partir de données numériques : le cas des statistiques sur les bibliothèques publiques du Québec

Yves Marcoux

Gabriela Gavrilut

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

Dans le but d'assurer une diffusion plus large et plus rapide des statistiques annuelles relatives aux bibliothèques publiques et aux centres régionaux de services aux bibliothèques publiques, le ministère de la Culture et des Communications du Québec a convenu avec l'École de bibliothéconomie et des sciences de l'information (EBSI) de rendre disponibles ces données sur les sites Gopher et Web de l'EBSI. Le présent article décrit la procédure automatique de transfert élaborée à cette fin et les résultats obtenus.

Automatic Construction of Resources on the Internet Using Numeric Data: The Case of Public Library Statistics in Québec

In order of ensure a wider and faster circulation of statistics pertaining to public libraries and regional library service centres, the ministère de la Culture et des Communications asked the École de bibliothéconomie et des sciences de l'information (EBSI) of the Université de Montréal to make the statistics available on the Gopher and Web sites of the school. This article describes the automatic transfer procedure used and the results.

Desarrollo automático de los recursos en el Internet a partir de datos numéricos: el caso de las estadísticas sobre las bibliotecas públicas de Quebec

Con el propósito de asegurar una difusión más amplia y más rápida de las estadísticas anuales relacionadas con las bibliotecas públicas y con los centros regionales de servicios a las bibliotecas públicas, el ministerio de la Cultura y de las Comunicaciones de la provincia de Quebec ha convenido con la Escuela de Biblioteconomía y de las Ciencias de la Información de la Universidad de Montreal (EBSI) para hacer disponibles esos datos en los sitios Gopher y Web de la EBSI. El presente artículo describe el proceso automático de transferencia elaborado con este fin y los resultados obtenidos.

Genèse du projet

En raison de son rôle, l'École de bibliothéconomie et des sciences de l'information (EBSI) de l'Université de Montréal s'intéresse à l'essor fulgurant des autoroutes de l'information et plus particulièrement de l'Internet. Ces nouveaux modes d'accès à l'information ont le potentiel de transformer radicalement l'environnement informationnel et par voie de conséquence de modifier en profondeur le rôle traditionnel du spécialiste de l'information. Pour ces raisons, l'EBSI s'est impliquée très tôt dans les nouveaux modes d'accès à l'information proposés par Internet, notamment en créant un forum de discussion (la liste EBSI-L), un site Gopher et un site Web¹. À l'été 1994, le ministère de la Culture et des Communications du Québec propose à l'EBSI la

diffusion, sur le site Gopher de l'EBSI, des statistiques colligées annuellement sur les bibliothèques publiques et les Centres régionaux de services aux bibliothèques publiques (CRSBP) par le Ministère. Les objectifs visés par le Ministère étaient d'assurer une diffusion très large des données auprès de la communauté des bibliothécaires francophones de tous les pays; d'assurer la publication intégrale de toutes les données recueillies par le Ministère pour toutes les bibliothèques; d'offrir au public une plus grande facilité de consultation des données.

Pour l'EBSI, le projet constituait une occasion d'augmenter sa visibilité auprès de la communauté bibliothéconomique francophone; de confirmer sa position de leadership en matière d'autoroutes de l'information; d'atteindre un des objectifs

de son site Gopher, à savoir la diffusion d'informations de source québécoise reliées aux sciences de l'information (Paquet 1994; Sévigny 1995). En outre, la publication des statistiques allait enrichir les ressources d'information d'Internet en langue française dans le domaine de la bibliothéconomie.

1. Le site Gopher de l'EBSI (*gopher://tornado.ere.umontreal.ca:7072*) a été conçu à l'hiver 1994 par André Paquet (Paquet 1994; Sévigny 1995). Il est actuellement en abandon progressif. Le site Web (*http://tornado.ere.umontreal.ca/~carmellu/ebsi/stats_BP/*) a été conçu par Gabriela Gavrilut, Martin Sévigny et Lucie Carmel à l'hiver 1995.

Le Ministère enregistre les statistiques colligées dans une base de données dBASE IV. Après examen, l'idée de simplement transférer ponctuellement les données de cette base dans le site Gopher de l'EBSI a été écartée au profit de l'élaboration d'une procédure *automatique* de transfert pouvant être appliquée chaque année aux nouvelles données colligées. Lors de la mise en service du site Web de l'EBSI, à l'hiver 1996, la procédure de transfert a été modifiée de façon à pouvoir intégrer les statistiques tant dans le site Web que dans le site Gopher.

Les statistiques sont donc disponibles, depuis mai 1996, autant sur le Web de l'EBSI que sur son Gopher. La figure 1 présente un exemple de tableau de données de la version Web. Il est à noter que, tant que la structure de la base de données dBASE IV ne changera pas, les procédures de transfert automatique mises au point continueront de fonctionner sans modification.

La modification des procédures automatiques pour verser les données sur le site Web plutôt que sur le site Gopher n'a posé aucun nouveau défi et a été réalisée très facilement. Pour cette raison, le présent article relate surtout la mise au point de la procédure de transfert vers le site Gopher.

La publication des données sur les sites Internet de l'EBSI ne rend pas désuet le document imprimé de diffusion des statistiques produit annuellement par le Ministère². En effet, celui-ci présente certaines données globales, selon différents regroupements de bibliothèques, ainsi qu'un tableau de comparaison entre les données de l'année courante et celles de l'année précédente qui ne sont pas inclus sur les sites de l'EBSI.

Éléments originaux de la recherche

Dès le début du projet, nous avons effectué des recherches documentaires

pour repérer des articles relatant des expériences de transfert de données statistiques, ou numériques en général, dans un Gopher ou d'autres services d'Internet. Malheureusement aucune source pertinente n'a été trouvée. Il ne nous a donc pas été possible de nous inspirer de travaux similaires et nous avons dû adopter une méthode générale d'analyse³.

Bien que les principaux résultats issus du projet de recherche soient des procédures automatiques de transfert des statistiques sur les bibliothèques publiques de la base dBASE IV du Ministère dans le Gopher et le Web de l'EBSI, le travail a comporté une étape importante d'analyse visant à élaborer une chaîne complète de traitement dans laquelle s'intégreraient les procédures de transfert. L'originalité du projet réside autant dans la chaîne de traitement élaborée que dans les procédures de transfert elles-mêmes. Nous croyons que ces deux éléments du projet pourraient suggérer des façons de faire pour d'autres projets impliquant le transfert de données statistiques ou numériques dans un site Gopher ou Web.

Deux aspects originaux méritent d'être soulignés. D'abord, la procédure de transfert a la particularité suivante: le point de départ est un programme dBASE IV qui fonctionne sur un micro-ordinateur IBM dans lequel la base du Ministère a été préalablement installée. Ce programme transforme les données de la base dBASE IV en un fichier texte ASCII étendu. Cependant, ce fichier texte n'est pas destiné à être directement inclus dans un site Gopher ou Web. Il contient plutôt des commandes Unix que le serveur Unix exécutera pour créer les répertoires et fichiers appropriés pour le site Gopher ou Web. Ainsi, le programme dBASE IV crée un programme Unix lequel est transféré puis exécuté sur le serveur Unix pour créer la «branche» appropriée du site Gopher ou Web.

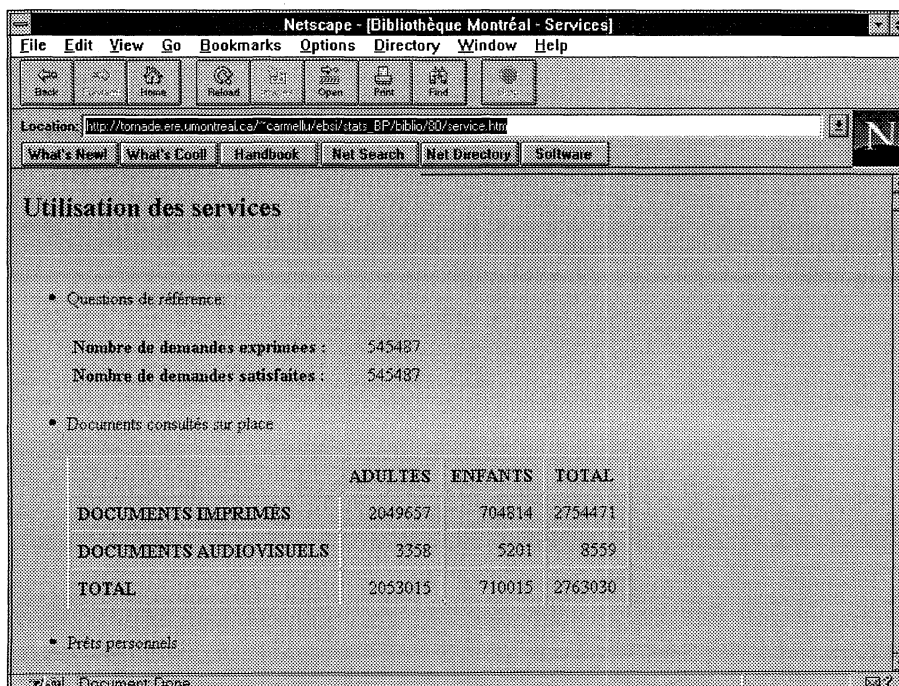


Figure 1. Exemple de tableau statistique affiché sur le Web

2. Québec (Province). Ministère de la Culture et des Communications. *Bibliothèques publiques - Statistiques* (Québec: Le Ministère).

3. Essentiellement, l'analyse fonctionnelle et organique. In Timothy J. O'Leary et Linda J. O'Leary, *Éléments d'informatique*, 2^e éd. (Montréal: Chenelière/McGraw-Hill, 1995).

Second aspect important d'originalité: il a été décidé, dès le début du projet, que des explications sur la signification exacte des données présentées seraient incluses dans le Gopher et le Web permettant à toute personne consultant les statistiques d'interpréter le mieux possible les données affichées. Cette approche est tout à fait en accord avec la philosophie de développement de systèmes d'information préconisée à l'EBSI, soit de toujours placer les explications sur l'interprétation des données le plus près possible des données elles-mêmes. Dans le cas de données colligées par questionnaire, une des meilleures sources d'explications sur l'interprétation des données que l'on puisse imaginer est le questionnaire lui-même, incluant les questions et les directives à l'intention des répondants. C'est donc du questionnaire que nous avons tiré la quasi-totalité des explications. Nous avons de plus adopté une structure de Gopher et de Web telle que chaque répertoire offrant l'accès à des données brutes offre également l'accès aux explications sur ces données. Ainsi, les explications sur les données sont très voisines des données elles-mêmes.

Problématique

Sites Gopher et sites Web

Un site Gopher est un ensemble structuré d'informations portant sur un thème donné et rendu accessible au public par Internet, grâce à un des protocoles clients-serveur normalisés d'Internet: le «protocole Gopher». Un serveur Gopher est un ordinateur relié à Internet sur lequel fonctionne un logiciel serveur Gopher et où se trouvent un ou plusieurs sites Gopher. Le logiciel serveur répond à toutes les requêtes d'accès provenant de n'importe quel logiciel client Gopher dans le monde pour les sites Gopher résidant sur ce serveur.

La structure générale de l'information que l'on peut placer sur un site Gopher est une arborescence hiérarchique de répertoires et de fichiers, analogue à celle que l'on retrouve sur la plupart des systèmes d'exploitation courants (ex.: DOS, Windows, MacOS). Une différence importante, cependant, est la possibilité de placer dans un répertoire des «liens» (ou «pointeurs») *non hiérarchiques* vers des

répertoires ou fichiers qui se trouvent n'importe où sur le serveur, ou même sur d'autres serveurs Gopher n'importe où dans le monde. Les sites Gopher peuvent ainsi se donner accès mutuellement, ce qui rend naturelle l'utilisation d'un site Gopher pour offrir un accès centralisé à différentes sources d'information dans un domaine donné.

Un site Web est analogue à un site Gopher du point de vue de la structure de l'information que l'on peut y placer, mais il offre plus de flexibilité. En effet, les liens (hiérarchiques ou non) n'ont pas à être regroupés en répertoires, ils sont inclus directement dans le texte des documents qui constituent le site. Ainsi, on est en présence d'une véritable structure hypertextuelle de l'information, où les liens peuvent émaner de mots ou d'expressions spécifiques dans un texte. Une autre différence est qu'il est beaucoup plus facile de mettre des contenus multimédias sur un site Web. Comme pour un site Gopher, l'accès à un site Web s'effectue au moyen d'un protocole client-serveur normalisé d'Internet: HTTP (*HyperText Transfer Protocol*)

Dans les deux sites (Gopher et Web), l'utilisateur «navigue» d'un élément d'information à un autre (à partir d'un point de départ de son choix) en activant les liens rencontrés qui lui semblent intéressants. L'analogie avec une «balade» improvisée est très naturelle. Il faut noter que pratiquement tous les logiciels clients Web développés jusqu'à présent gèrent également le protocole Gopher et peuvent donc accéder non seulement à n'importe quel serveur Web au monde, mais également à n'importe quel serveur Gopher. Dans ce contexte, le Web peut être vu comme une généralisation stricte du Gopher.

Les statistiques sur les bibliothèques publiques du Québec

Le ministère de la Culture et des Communications du Québec collige chaque année par questionnaire un certain nombre de renseignements sur les bibliothèques publiques du Québec et les CRSBP. Ces renseignements concernent autant les collections des bibliothèques que leur milieu, leur utilisation, leur personnel et leur budget. Remplir ce ques-

tionnaire est obligatoire. Pour fins de traitement interne, le Ministère stocke les données colligées dans une base de données dBASE IV⁴. Il existe actuellement au Québec 159 bibliothèques publiques autonomes et 11 CRSBP.

Chaque année, le Ministère extrait de ces renseignements des statistiques qu'il publie dans un document imprimé. Dans ce document, les bibliothèques sont regroupées par régions et par nombre d'habitants. Le document contient également des données globales par type de municipalité, type de bibliothèque ainsi qu'une brève comparaison entre les données de l'année courante et celles de l'année précédente. Bien que les données individuelles recueillies par le Ministère aient toujours été disponibles sur demande, seules les statistiques globales étaient publiées.

Pour les fins du projet, le ministère de la Culture et des Communications du Québec a mis à notre disposition le matériel suivant:

1. Le document *Bibliothèques publiques - statistiques 1993* (Ministère de la Culture et des Communications du Québec).
2. Le document *Enquête annuelle sur les bibliothèques publiques* contenant le questionnaire adressé chaque année aux bibliothèques publiques du Québec ainsi que les directives et définitions du Ministère à l'intention des répondants.
3. Trois bases de données créées à l'aide du logiciel dBASE IV, contenant les réponses aux questionnaires pour les années 1991, 1992 et 1993. En avril 1996 s'est ajoutée la base contenant les données de 1994.

Le questionnaire est divisé en 11 sections principales: Renseignements d'ordre général; Documents de la bibliothèque; Utilisation des services; Usagers;

4. Les données sur les CRSBP ne sont cependant pas stockées en format ordinaire. Elles ont donc dû faire l'objet d'un traitement spécial décrit plus loin dans l'article.

Locaux de la bibliothèque; Nombre d'heures d'ouverture de la bibliothèque; Personnel de la bibliothèque; Promotion et animation; Activités de coopération; Revenus; Dépenses. Trois autres sections sont présentes dans le questionnaire: Plan de développement, Collections spéciales, Questions complémentaires, mais la base de données dBASE IV ne prévoit aucun champ pour ces rubriques.

Il est à noter que le Ministère n'enregistrant pas les données des CRSBP sur support magnétique, ces dernières n'étaient donc pas incluses dans la base dBASE IV. Cependant, nous avons quand même pu inclure certaines informations sur les CRSBP dans le Gopher et le Web en nous basant sur quatre tableaux qui se retrouvent dans le document du Ministère. Nous avons retranscrit ces tableaux dans des fichiers que les procédures automatiques se chargent d'intégrer dans le Gopher et dans le Web.

Bien que la base de données du Ministère soit créée et gérée à l'aide du SGBD (système de gestion de bases de données) relationnel dBASE IV, sa structure en est plutôt une de «fichier plat» puisqu'elle n'exploite pas les caractéristiques spécifiques du modèle relationnel. En fait, elle est constituée de trois tables distinctes, mais celles-ci sont conceptuellement trois fragments d'une seule et même table. Le seul lien qui unit ces trois tables est une clé primaire commune.

Objet du projet

Physiquement, les sites Gopher et Web de l'EBSI résident sur un serveur Silicon Graphics des Services informatiques de l'Université de Montréal. Ils sont constitués de fichiers et de répertoires Unix localisés dans le compte de l'utilisateur assumant le rôle d'administrateur du site correspondant⁵. Quant au logiciel dBASE IV, il est disponible en environnement micro-ordinateur dans les laboratoires d'informatique documentaire de l'EBSI.

L'objet du projet comportait donc les éléments suivants:

- Déterminer la meilleure façon d'organiser les données «plates» de la base dBASE IV en une structure principalement hiérarchique pour les intégrer dans un site Gopher ou Web;

- Déterminer quelle partie du traitement des données doit être effectuée sur micro-ordinateur (avec dBASE IV ou tout autre logiciel) et quelle partie doit être effectuée en environnement Unix sur le serveur hébergeant les sites Gopher et Web;
- Déterminer la meilleure façon de transférer les données de l'environnement micro-ordinateur des laboratoires d'informatique documentaire de l'EBSI dans le serveur Unix;
- Mettre au point, tester et documenter les procédures de traitement et de transfert.

Solutions considérées

Transfert des données

Toute solution retenue pour le transfert des données de la base de données dBASE IV dans le serveur des Services informatiques doit forcément suivre le scénario général suivant: (1) installer la base dBASE IV sur un micro-ordinateur disposant du logiciel dBASE IV, (2) exporter les données vers un ou plusieurs fichiers sur le micro-ordinateur, puis (3) transférer ce ou ces fichiers dans le serveur et, éventuellement, (4) effectuer un post-traitement des données sur le serveur. Théoriquement, l'ordre des opérations peut se faire dans un sens ou dans l'autre. À un extrême, on pourrait avoir un programme d'exportation qui crée lui-même tous les répertoires et fichiers du Gopher (ou du Web), lesquels devraient alors simplement être transférés sur le serveur Unix sans autre post-traitement. À l'autre extrême, on pourrait simplement transférer les fichiers dBASE IV sur le serveur Unix et développer un programme de post-traitement qui lise directement ces fichiers et crée toute la structure Gopher (ou Web) correspondante.

La solution que nous avons adoptée est intermédiaire; elle est aussi un peu surprenante en ce que le programme d'exportation crée de toutes pièces le programme de post-traitement lequel contient en lui-même toutes les données de la base sous forme de constantes de caractères. Le programme de post-traitement est simplement un programme de commandes Unix (un *shell script*) constitué d'opérations de création de répertoires et d'écriture de fichiers. Le programme

d'exportation lit les données de la base dBASE IV, les formate et les insère dans le programme de post-traitement qu'il est en train de créer, sous la forme de constantes de caractères placées en argument à des commandes d'écriture de fichiers.

Le programme d'exportation est écrit dans le langage de programmation de dBASE IV et nous l'appellerons *Crée_Prog*. Il fonctionne sur le micro-ordinateur et crée le programme de post-traitement, un script de commandes Unix que nous appellerons *Crée_Stat*, lequel est ensuite transféré par FTP (*File Transfer Protocol*) sur le serveur Unix pour créer les fichiers et répertoires de la nouvelle branche du site Gopher ou Web.

Accès aux données par les serveurs

Avant de pouvoir présenter l'ensemble des choix qui s'offraient à nous pour rendre les statistiques du Ministère accessibles sur le Gopher et le Web de l'EBSI, il nous faut aborder une fonctionnalité des logiciels serveur Gopher et Web dont nous n'avons pas encore parlé; il s'agit de l'exécution dynamique de commandes (pour un serveur Web, cette fonctionnalité est habituellement désignée sous le nom de *scripts CGI* (Common Gateway Interface)).

Le scénario de base d'une interaction entre un serveur Gopher ou Web et un utilisateur est le suivant: le logiciel client de l'utilisateur établit une connexion avec le logiciel serveur puis lui transmet l'identificateur d'un élément d'information (un fichier ou un répertoire) spécifique auquel l'utilisateur désire avoir accès⁶. Sur réception de cet identificateur, le logiciel serveur lit l'élément désiré de son disque

5. Jusqu'à ce jour, l'administrateur du site Gopher est un étudiant de seconde année de maîtrise et l'administratrice du site Web est la responsable des laboratoires d'informatique documentaire de l'EBSI, Mme Lucie Carmel.

6. Notons que cet identificateur peut avoir été tapé explicitement par l'utilisateur ou choisi par lui dans sa liste de signets personnels; il peut encore provenir d'un lien se trouvant dans un élément d'information préalablement visité, lien que l'utilisateur aurait choisi d'activer.

dur puis le transmet au logiciel client lequel l'affiche sur l'écran de l'utilisateur.

Il existe une variante de ce scénario dans laquelle l'élément désiré n'est pas simplement lu du disque dur, mais est plutôt généré dynamiquement par un logiciel fonctionnant sur la même machine que le logiciel serveur, mais distinct de ce dernier. Dans ce nouveau scénario, l'identificateur d'élément désiré transmis au logiciel serveur contient en fait de façon «codée» une commande du système d'exploitation hôte. Sur réception d'un tel identificateur, le logiciel serveur décode la commande puis la passe au système d'exploitation hôte pour exécution et c'est le résultat de l'exécution de cette commande qui est retourné au logiciel client en réponse à l'ensemble de l'interaction. Étant donné que n'importe quel logiciel installé sur le serveur peut être appelé par une commande du système d'exploitation, on peut théoriquement utiliser n'importe quel logiciel disponible sur le serveur pour produire dynamiquement des éléments d'information que l'on désire inclure dans un site (Web ou Gopher).

Pour exploiter cette fonctionnalité, un concepteur de site Web ou Gopher n'a qu'à insérer dans les éléments d'information de son site des liens qui contiennent non pas des identificateurs d'éléments statiques, mais bien des commandes codées à l'intention d'un des logiciels installés sur le serveur.

Étant donné cette possibilité, nous aurions pu choisir de transférer les données dans un format pour un logiciel de traitement «évolué» disponible sur le serveur, par exemple, un SGBD relationnel, un logiciel statistique ou un tableur. Chacun de ces types de logiciels était disponible sur le serveur. Alors, les différents éléments de nos répertoires auraient été associés à des commandes du logiciel «évolué» pour extraire certaines informations des données brutes, des statistiques par exemple. Cette approche aurait eu l'avantage de permettre des requêtes complexes, ce qui aurait rendu possible l'obtention dynamique d'un grand nombre de statistiques différentes sans qu'il soit nécessaire de les précalculer au préalable. Malgré tout, nous n'avons pas retenu cette solution. La principale raison fut le facteur temps puisqu'il aurait fallu créer

des interfaces entre de nombreux outils hétérogènes et développer une grande variété de requêtes statistiques, tâches que le temps disponible ne nous permettait pas de réaliser. Une autre raison est qu'il nous semblait que les données disponibles étaient satisfaisantes même sans la possibilité de formuler des requêtes statistiques «sur mesure». Nous avons donc retenu l'approche de ne présenter que les données «brutes», telles qu'elles figuraient dans la base de données dBASE IV. Essentiellement, donc, les sites Gopher et Web permettent de visualiser les réponses inscrites sur le questionnaire par chaque bibliothèque.

Présentation des données

Même après la décision de ne présenter que les données brutes, deux avenues s'offraient à nous: nous aurions pu placer dans des fichiers les caractères numériques provenant de la base de données sans aucun intitulé correspondant et faire placer dynamiquement ces caractères numériques dans des bordereaux appropriés au moment de l'affichage seulement (nous aurions utilisé pour cela le mécanisme, décrit ci-dessus, d'exécution dynamique de commandes). L'autre possibilité était de préformater toutes les données, incluant les intitulés, dans des fichiers texte (ASCII) qu'il suffisait de placer aux bons endroits dans la hiérarchie de répertoires du site Gopher ou Web. Le mécanisme d'exécution de commande n'a pas alors à être invoqué. L'avantage de la première approche aurait été principalement d'épargner de l'espace-disque sur le serveur. En effet, les intitulés n'auraient pas été répétés dans les données de chaque bibliothèque. Par contre, elle était d'une plus grande complexité que la seconde approche puisqu'il aurait fallu développer, en environnement Unix, un programme pour insérer dynamiquement les caractères numériques dans les bordereaux. Après calculs, il s'est avéré que l'espace-disque requis pour la seconde approche ne causerait pas de problème (au moins, pour les données d'une année) et c'est donc celle que nous avons choisie.

Organisation logique des données

Il fallait aussi décider de l'organisation logique des données sur les sites. Nous avons opté pour une organisation

très simple, conçue spécialement pour le Gopher, mais facile à adapter au Web. Nous ne la décrivons que dans son implantation sur le Gopher, l'implantation sur le Web étant, à des détails près, identique.

Au plus haut niveau hiérarchique de la branche «Statistiques» du Gopher se trouvent trois répertoires: un répertoire «explic» contenant les explications sur les données, un répertoire «biblio» contenant les données sur les bibliothèques autonomes et un répertoire «CRSB» contenant les données sur les CRSBP. Le répertoire «biblio» contient un sous-répertoire pour chaque bibliothèque autonome, lui-même subdivisé en dix sous-sous-répertoires correspondant aux différentes sections du questionnaire du Ministère (les sections «locaux» et «heures d'ouverture» ont été regroupées dans un même répertoire). Chacun de ces dix sous-sous-répertoires contient deux éléments: un fichier texte (ASCII) contenant les données avec intitulés, prêt à être affiché, de même qu'un lien (non hiérarchique) avec le fichier du répertoire «explic» pertinent pour cette partie des données. Le répertoire «explic» est donc structuré de la même façon que chacun des sous-répertoires de «biblio» cependant, en plus des dix fichiers correspondant aux sections du questionnaire, un fichier additionnel offre de l'information sur la branche «Statistiques» elle-même. À l'intérieur du répertoire «biblio», les sous-répertoires correspondant aux bibliothèques autonomes sont présentés par ordre alphabétique de nom de municipalité. Le répertoire «CRSB» contient simplement quatre fichiers texte (ASCII) directement affichables qui correspondent à quatre tableaux sur les CRSBP du document imprimé du Ministère. Un fragment de la structure logique de la branche «Statistiques» apparaît dans la figure 2.

Recherche dans les données

Il restait à traiter la question de la recherche d'information dans la branche «Statistiques». Il existe certains logiciels qui indexent le contenu d'un site (Web ou Gopher) et qui permettent ensuite à l'utilisateur d'effectuer dynamiquement des recherches sur le site par l'entremise du mécanisme d'exécution de commandes décrit précédemment. WAIS et Jughead sont des exemples de tels logiciels.

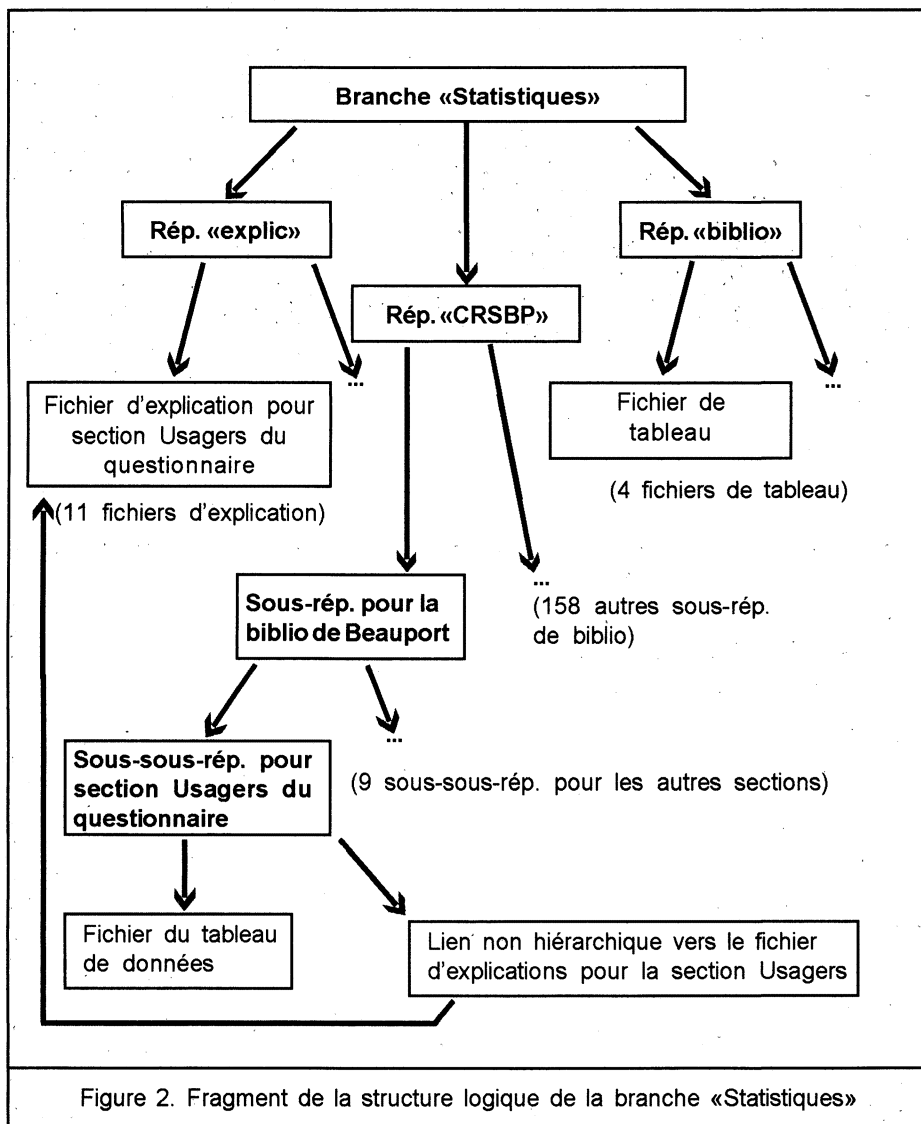


Figure 2. Fragment de la structure logique de la branche «Statistiques»

Dans notre cas, cependant, les seuls contenus susceptibles d'être recherchés sont les noms de municipalité, le reste étant constitué de caractères numériques et des intitulés correspondants. Or, la plupart, sinon la totalité, des logiciels clients Gopher et Web offre une fonction de recherche dans le texte d'un élément d'information et, en particulier, dans les intitulés d'éléments d'un répertoire. Comme les noms de municipalité sont utilisés comme intitulés dans le répertoire «biblio», la recherche par nom de municipalité peut s'effectuer par cette fonction. Il n'y avait donc aucun besoin de recourir à un outil d'indexation spécialisé.

Réalisation de la solution retenue

La réalisation de la solution retenue a été très similaire dans le cas du Gopher et du Web. Le Web fait intervenir un troisième langage: HTML (*Hypertext Markup Language*), langage utilisé pour créer les pages d'un site Web. Cependant, dans l'ensemble, c'est le cas du Gopher qui posait le plus de difficultés particulières; de plus, c'est celui qui a été réalisé en premier lieu. Pour ces raisons, nous ne présentons que le cas du Gopher.

La réalisation de la solution choisie a comporté plusieurs étapes dont la liste

apparaît dans la figure 3. Le reste de cette section fait état des principaux problèmes rencontrés ainsi que des correctifs apportés.

- 1- Conception des formats d'affichage
 - 2- Conception des fichiers de données et d'explications
 - 3- Création des fichiers contenant les tableaux des CRSBP
 - 4- Conception détaillée de la structure du Gopher
 - 5- Conception du programme de commandes Unix (Crée_Stat)
 - 6- Conception et création du programme d'exportation en langage dBASE IV (Crée_Prog)
 - 7- Transfert des données dans le Gopher
- Figure 3. Étapes de réalisation de la solution retenue

Parmi les quelques formats d'affichage considérés, la présentation sous un format identique à celui employé dans les questionnaires adressés aux bibliothèques nous est apparue comme étant non seulement la plus explicite mais également la plus appropriée à la présentation sur un Gopher. Nous avons donc adopté l'approche de reproduire le plus fidèlement possible les diverses sections du questionnaire, sous forme de fichiers texte (ASCII) qui seraient affichés sur le Gopher. Dans certains cas, cependant, nous avons modifié légèrement la forme utilisée dans le questionnaire, de façon à rendre la présentation plus claire ou mieux adaptée à un Gopher.

Le mode de fonctionnement général de la procédure automatique consiste en l'extraction de données de la base dBASE IV et en leur transformation sous une forme directement affichable reprenant exactement le questionnaire soumis aux bibliothèques. Seuls certains points mineurs ont exigé un traitement particulier. Par exemple, le questionnaire prévoyait certaines rubriques pour afficher la somme de divers éléments, cependant ces totaux n'étaient pas stockés dans la base de données. Nous avons décidé que la procédure automatique de transfert des données exécuterait les calculs

nécessaires et rendrait les résultats accessibles sur le Gopher.

Tel que mentionné précédemment, il n'y a pas de données sur les CRSBP dans la base de données du Ministère. Des fichiers contenant les données globales disponibles dans le document imprimé ont été créés manuellement et leur mise-à-jour devra être faite de la même façon à chaque année si les données ordinolingues ne sont toujours pas disponibles.

L'organisation logique des données a été décrite plus haut. Physiquement, sur le serveur, les relations hiérarchiques de cette organisation se traduisent simplement par des relations de répertoires à sous-répertoires dans le système de fichiers Unix. Les liens hiérarchiques sont donc automatiquement établis par la simple création d'une structure de répertoires et de sous-répertoires correspondant à l'organisation désirée. Le seul point auquel il a fallu faire attention est que, normalement, un répertoire ne contient que des noms de fichiers qui sont soumis à des règles syntaxiques strictes excluant entre autres les espaces. Heureusement, une caractéristique du logiciel serveur Gopher (le mécanisme des répertoires «.cap») permet d'attribuer aux éléments figurant dans un répertoire les intitulés que l'on veut. Ceci nous a permis, par exemple, d'utiliser les noms de municipalités comme intitulés.

L'organisation logique des données prévoit aussi des liens non hiérarchiques entre les sous-sous-répertoires du répertoire «biblio» et les fichiers du répertoire «explic», de façon à pouvoir facilement passer des éléments d'information spécifiques à chaque bibliothèque aux explications relatives à ces éléments. Pour introduire ces liens non hiérarchiques, nous avons utilisé le mécanisme prévu à cet effet dans le logiciel serveur Gopher, celui des fichiers «.link». Évidemment, il a fallu prévoir l'inclusion, dans le programme *Crée_Stat*, des commandes nécessaires pour créer les répertoires «.cap» et les fichiers «.link» requis. L'étape 5 (conception du programme *Crée_Stat*) consistait à construire «à la main» un programme de commandes Unix capable de créer les fichiers et répertoires contenant l'information relative à une bibliothèque. Ce programme qui consiste

essentiellement en des commandes de création de fichiers et de répertoires a ensuite servi de canevas pour la conception du programme «*Crée_Prog*» à l'étape suivante. Le programme «*Crée_Prog*», écrit dans le langage de programmation de dBASE IV, extrait des données de la base de données et les combine avec des chaînes de caractères constantes pour créer le programme de commandes Unix «*Crée_Stat*». La création du programme «*Crée_Prog*» a exigé beaucoup d'attention, car nous ne pouvions pas le tester fréquemment; en effet, chaque test complet nécessitait le transfert puis l'exécution du programme «*Crée_Stat*» résultant sur le serveur Unix. Les lecteurs intéressés par les détails du programme «*Crée_Prog*» sont priés d'entrer en contact avec les auteurs⁷.

Le transfert proprement dit des données dans le Gopher a consisté dans: 1) l'exécution du programme dBASE IV «*Crée_Prog*», lequel crée le programme Unix «*Crée_Stat*», 2) le transfert du programme «*Crée_Stat*» vers le compte de l'administrateur du Gopher, 3) l'élimination des caractères accentués du fichier exécutable «*Crée_Stat*» et 4) l'exécution du programme «*Crée_Stat*» qui crée les répertoires, fichiers et liens correspondants de la branche «Statistiques» du Gopher.

L'élimination des caractères accentués était requise parce que dès la mise sur pied du Gopher de l'EBSI, il avait été décidé de n'y inclure aucun caractère accentué⁸. Par contre les données provenant de la base dBASE IV contenaient parfois de tels caractères. Il fallait donc les éliminer avant de placer l'information sur le Gopher. Cette élimination aurait pu être faite à la source, par le programme «*Crée_Prog*», mais il s'avérait plus simple de le faire après le transfert de «*Crée_Stat*», à l'aide de l'utilitaire GNU *recode*, disponible sur le serveur Unix. Pour les données de 1993, le programme «*Crée_Stat*» obtenu avait une taille d'environ 4,5 méga-octets. Son traitement sur le serveur Unix a pris approximativement 20 minutes.

Conclusion

André Paquet, créateur du site Gopher de l'EBSI, soulignait dans son rapport (Paquet 1994) qu'un site Gopher

est un ensemble dynamique de documents qui doit croître et se développer continuellement, propos qui s'appliquent évidemment aussi à un site Web. Malgré la grande popularité d'Internet, beaucoup d'organisations continuent et continueront sûrement pendant plusieurs années de produire de l'information avec des outils traditionnels sous une forme non directement intégrable dans Internet. La croissance et les développements dont parlait Paquet obligeront les organisations pendant encore longtemps de créer des interfaces entre des ressources traditionnelles d'information et Internet; n'importe quel responsable de site Web aujourd'hui peut d'ailleurs confirmer l'ampleur des demandes de cette nature. En ce sens, le cas décrit dans cet article relève d'une problématique très actuelle et peut servir d'exemple ou de source d'inspiration pour des projets similaires.

Même dans le contexte actuel d'abandon progressif du site Gopher de l'EBSI, le développement de la procédure automatique de transfert dans le Gopher était selon nous pleinement justifié, d'une part parce qu'elle a permis de rendre accessibles les statistiques du Ministère avant la mise sur pied du site Web de l'EBSI, d'autre part parce qu'elle a pu être adaptée au nouveau contexte du site Web avec très peu d'efforts. Ceci démontre, à notre avis, la robustesse de la procédure développée et laisse présager qu'elle pourrait s'adapter à un autre format si nécessaire.

Les sites créés par nos procédures automatiques de transfert laissent beaucoup de place à des améliorations qui constitueront, nous l'espérons, autant de thèmes de recherche pour de futurs étudiants en informatique documentaire.

Une des améliorations possibles serait la diminution de l'espace-disque

7. Adresse électronique des auteurs :
{marcoux|gavrilug}@ere.umontreal.ca.

8. Le laboratoire d'informatique documentaire ne disposait pas à ce moment d'un environnement capable d'afficher correctement les caractères accentués provenant d'un Gopher; il aurait donc été peu logique que l'EBSI en place sur son propre site.

occupé par les sites. En raison des capacités limitées du serveur qui héberge les sites, l'espace actuel d'environ 4,5 mégaoctets par site apparaît relativement important. Une solution envisageable serait la compression des fichiers de données avec décompression «à la volée» au moment de l'affichage. Une autre solution, mentionnée précédemment, serait de remplacer les fichiers pleinement formatés par des commandes Unix qui liraient les caractères numériques à afficher de fichiers non formatés et les placeraient dynamiquement aux bons endroits dans des tableaux prédéfinis. L'approche la plus naturelle serait à notre avis de créer directement une interface entre le site Web et la base de données relationnelle. Il existe aujourd'hui de nombreux outils spécialisés capables d'établir un tel lien dynamique sans exiger le développement de scripts CGI originaux; l'Oracle

WebServer d'Oracle et Sapphire/Web de Bluestone en sont des exemples. Une telle solution aurait le double avantage de diminuer l'espace-disque et de permettre l'extraction dynamique de statistiques non précalculées.

Sources consultées

Allard, J. and Steven Sinofsky. 1994. Getting wired into the Internet: a crash course on ftp, Gopher, Web and more. *Microsoft Systems Journal* 9 (September): 53.

Arthur, Lowell Jay. *UNIX Shell Programming*. 1990. New York: John Wiley & Sons.

Dagorn, François et Claude Gross. *World-Wide Web*. <http://www.urec.fr/docs/WWW/WWW.html>

Martinet, Frédéric et Olivier Berthier. *Introduction au World-Wide Web et HTML*. <http://www-ensimag.imag.fr/ENSIMAG/Administration/Les.Cours/Exposes.Reseaux/Introweb/>

Mcgilton, Henry and Rachel Morgan. 1983. *Introducing the UNIX System*. New York: McGraw-Hill Book Company.

Michalski, Jerry. 1994. Internet tools 101: the building blocks. *RELease* 1.0 94 (1): 14.

O'Leary, Timothy J. et Linda I O'Leary. 1995. *Eléments d'informatique*. 2^e éd. Montréal: Chenelière/McGraw-Hill.

Paquet, André. 1994. *Implantation d'un Gopher à l'EBSI*. Montréal: École de bibliothéconomie et des sciences de l'information, Université de Montréal.

Québec (Province). Ministère de la Culture et des Communications du Québec. 1994. *Bibliothèques publiques - Statistiques 1993*. Québec: Le Ministère.

Sévigny, Martin. 1995. Le gopher de l'EBSI, un an plus tard. *Documentation et bibliothèques* 41 (4): 225-233.

Snyder, Joel. 1994. Internet integration. *LAN Magazine* 9 (5): S45.



• PERIODICA abonnements

10 000 titres (magazines, périodiques et journaux du monde entier)

• PERIODICA Vidéo

1 500 titres (arts, sciences, littérature, voyage, jeunesse, cinéma de répertoire)

• PERIODICA Multimédia

Plus de 300 CD-ROM multimedia de langue française

• BIBLIORAMA

Tous les livres disponibles de langue française distribués au Canada

• Partenaire CEDROM-SNI

500 banques de données sur CD-ROM (actualité, affaires, sciences et technologie, santé, médecine, éducation)

Accès direct à plus de 1 250 sources d'information électronique

PERIODICA, un guichet central d'accès aux produits d'information

PERIODICA INC.

**Case Postale 444 Outremont
Québec Canada H2V 4R6**

Tel.: (514) 274-5468

Fax: (514) 274-0201

Tout le Canada: 1-800-361-1431