

DEUX OUTILS POUR LES BIBLIOTHÈQUES DISTRIBUÉES

LA NORME Z39.50 ET LE PROTOCOLE HTTP

L'association américaine de normalisation (ANSI, American National Standards Institute) vient de valider la version 3 de la norme Z39.50 (1).

Héritière des grands projets bibliographiques en réseau des années 1980 (par exemple le *Linked System Project*), cette norme a pour objectif de permettre à un utilisateur d'effectuer des recherches documentaires dans des bases de données distantes à travers une interface unique.

La nouvelle version a été approuvée au terme de longs travaux qui ont mobilisé les experts des principales institutions bibliographiques américaines (Bibliothèque du Congrès, OCLC, Melvyl, RLG, etc.)². Compte tenu de l'influence exercée par ses promoteurs, elle semble destinée à un grand avenir dans le domaine de l'informatisation des bibliothèques.

Cependant, avec la diffusion très rapide des outils Internet, on peut se demander si elle ne risque pas d'être étouffée par les normes de fait qui sont déjà appliquées tous les jours sur le réseau mondial. Des interfaces universelles existent déjà, par exemple celle des *browsers* du *World Wide Web* (Mosaic, Netscape, etc.). Ces outils communiquent à l'aide de pro-

tocoles simples mais robustes qui semblent satisfaire la majorité des utilisateurs.

Dans ces conditions, on peut se demander pourquoi l'ANSI juge bon de prôner une nouvelle interface de communication spécifique aux bibliothèques. Sans prétendre répondre à cette question, nous essayons de donner ici quelques éléments de comparaison.

La norme Z39.50

Z39.50 est un protocole de communication dédié aux applications de recherche documentaire. Il permet de spécifier les procédures nécessaires pour effectuer des recherches dans des bases de données distantes et le rapatriement sur un poste local du résultat de ces recherches.

Si l'on se réfère à la division en couches utilisée par l'OSI (Open Systems Interconnection), Z39.50 se situe dans la couche application.

Z39.50 est basé sur le modèle client/serveur. La norme suppose en effet qu'une application cliente établit une connexion avec une application serveur et qu'elle lui demande des services. La norme Z39.50 utilise plutôt les termes *origin* et *target*, qui correspondent respectivement aux portions du client et du serveur uniquement dédiées aux tâches de communication (à l'exclusion par exemple de l'interface utilisateur).

Le client (*origin*), situé sur un système local, initie une connexion avec le serveur et lui envoie des requêtes. Le serveur (*target*), associé à la base de données distante, répond aux requêtes

1. Il faut signaler qu'il existe des normes internationales très proches de Z39.50, les normes ISO 10162/3 (5). Ces normes ont été élaborées par le groupe de travail TC46/SC4/WG4 de l'ISO. Elles constituent un sous-ensemble de Z39.50 et devraient disparaître si, comme cela semble se concrétiser, la version 3 de Z39.50 est reprise en norme internationale.

2. Ces experts ont travaillé dans le cadre d'un forum très actif, le *Z39.50 Implementors group* (ZIG).

FRANÇOIS ROLE

Direction de l'information
scientifique et technique
et des bibliothèques
role@distb.mesr.fr

provenant du client et transfère les enregistrements correspondants.

L'interaction entre le client et le serveur s'effectue à travers un échange de messages normalisés (requête et réponse). Les formats et les procédures à appliquer pour le transfert de ces messages sont également définis dans la norme (*Protocol specification*).

La version 3 définit de nombreux services qu'il serait fastidieux d'énumérer ici. Le cœur de la norme est constitué par quatre opérations de base : *Init*, *Search*, *Retrieve*, *Close*.

Init et *Close* décrivent l'établissement et la fermeture de la connexion entre le client et le serveur.

Les opérations *Search* et *Retrieve* ont donné leur surnom aux normes Z39.50 et ISO 10162/10163, souvent appelées normes SR.

Le service *Search* permet à l'origine d'interroger des bases de données distantes par l'intermédiaire du serveur et de recevoir les résultats de cette recherche.

L'utilisateur saisit une requête sur son système local. Le client traduit cette requête dans un formalisme indépendant de toute base de données et compris par tous les serveurs conformes à la norme Z39.50³. La requête traduite est envoyée à un serveur Z39.50 associé à la base de données que l'utilisateur souhaite consulter. À réception de cette requête, le serveur la traduit dans une syntaxe compréhensible par la base de données et la soumet à cette dernière. Les résultats de la recherche sont ensuite communiqués au client par le serveur. Le client (et l'utilisateur) n'ont ainsi pas à se soucier de la structure et de la syntaxe propres à chacune des bases de données auxquelles ils veulent accéder.

Le résultat de la recherche crée sur le serveur un *result set* qui peut être référencé par le client pendant toute la durée de la connexion.

Le service *Present* permet à l'origine de demander le transfert sur le poste local des enregistrements correspondant à un *result set* donné.

Le protocole HTTP

HTTP, *Hypertext Transfer Protocol* (2), est le protocole de communication utilisé par les outils du *World Wide Web* (WWW). C'est un protocole dédié au transfert d'informations, non à l'interrogation des bases de données. Basé sur un paradigme requête/réponse, il définit le contenu et le format des messages échangés au cours d'une transaction entre un client WWW et un serveur WWW.

Le *Web* a déjà fait l'objet de nombreuses présentations dans la presse professionnelle. Nous nous limiterons donc ici à une brève description fonctionnelle.

L'utilisateur qui utilise un poste client (souvent appelé *browser* compte tenu des origines hypertextuelles de WWW) consulte un document contenant des références hypertextuelles. En sélectionnant une de ces références, il provoque une connexion entre le client et le serveur distant qui possède ce document. Dans la plupart des implémentations actuelles, la connexion est établie par le client avant chaque requête et fermée par le serveur après l'envoi de la réponse.

Les requêtes émises par le client spécifient des « méthodes » à appliquer à

des « ressources » désignées par des identificateurs Internet nommés URI (*Uniform Resource Identifier*) (7).

Dans le cas de transfert de document évoqué plus haut, c'est la méthode GET suivie de l'URI du document qui est envoyée au serveur lorsque l'utilisateur sélectionne une référence hypertextuelle.

Le protocole prévoit également d'autres méthodes pour la mise à jour de serveurs à distance, l'invocation de scripts sur le serveur, la création de liens, etc. Seules certaines de ces fonctionnalités supplémentaires sont implémentées sur les clients les plus courants.

Quelques éléments de comparaison

Z39.50 propose des mécanismes sophistiqués pour l'établissement et la libération de la connexion. Des procédures d'identification et d'authentification sont prévues par la norme. Les services disponibles, les ressources, les jeux de caractères à utiliser, la taille des messages peuvent être négociés avant que l'échange ne commence entre le client et le serveur.

En ce qui concerne la recherche, la norme Z39.50 définit des ensembles d'attributs très complets correspon-

3. La *Reverse Polish Notation* (RPN) par exemple.

Interrogation de serveurs Z39.50 à partir de clients W3 : l'utilisateur peut utiliser pour sa recherche les ensembles d'attributs normalisés définis par Z39.50

dant aux applications de bibliothèque (*attribute set* 'bib1').

Par ailleurs, le *result set* produit par une question peut être référencé et servir d'opérande dans des questions posées ultérieurement. Le fait qu'il gère une véritable session donne au client Z39.50 une certaine « intelligence ».

Cette « intelligence » du client est encore perceptible dans la façon dont les résultats sont affichés. L'affichage des résultats est sous le contrôle du client : des résultats de recherche venant de plusieurs serveurs différents peuvent être affichés de façon strictement identique. Cette caractéristique distingue également le client Z39.50 du client WWW.

En ce qui concerne le transfert des résultats de la recherche, des procédures de segmentation sophistiquées ont été définies pour supporter le transfert de documents volumineux.

Ce point est à souligner : bien qu'issue du monde des bibliothèques, la norme Z39.50 n'a pas été prévue pour rechercher uniquement des références bibliographiques. Elle a été conçue pour permettre l'accès à des bases contenant tout type d'informations, y compris des documents primaires.

Comparé à la norme Z39.50, HTTP est un protocole de communication très simple, dédié avant tout au transfert de documents et non à la consultation de bases de données. Un document est demandé par l'utilisateur. Dès qu'il est transféré, la connexion est interrompue. Si l'on veut un autre document situé sur le même serveur, il faut une nouvelle connexion.

Il n'est pas question ici de réutiliser le résultat d'une recherche antérieure (*result set*). Aucun ensemble spécifique d'attributs (*attribute set*) n'est non plus défini.

Cependant, HTTP est très souple. Grâce à un mécanisme d'invocation de script particulièrement simple, le client peut envoyer des paramètres au serveur et demander à ce dernier d'effectuer des opérations qui vont au-delà du simple transfert de documents au coup par coup, par exemple des recherches documentaires.

En ce qui concerne le transfert proprement dit, HTTP gère efficacement la plupart des formats de documents existants (les types MIME).

La simplicité et la souplesse de HTTP ont permis au WWW de se développer très rapidement. C'est aussi ce caractère simpliste qui pourrait à terme freiner son essor⁴.

Cependant, malgré les différences importantes qui viennent d'être évoquées ces protocoles ont pu être utili-

4. Le consortium W3, créé depuis cette année par l'INRIA et le MIT, a pour mission d'apporter à HTTP le caractère professionnel qui lui fait défaut.

sés dans le cadre d'applications de bibliothèques distribuées dont nous donnons ci-après quelques exemples.

L'évolution vers les systèmes distribués

Dans un système informatique distribué, l'utilisateur accède à des ressources et à des données éparpillées géographiquement, tout en ayant l'impression d'interagir avec un système logique unique.

On a déjà dit dans la section précédente qu'un des atouts de Z39.50 était sa capacité à masquer l'hétérogénéité géographique et structurelle des systèmes (hétérogénéité des interfaces ; hétérogénéité des langages de

une recherche peut être lancée en parallèle sur plusieurs serveurs.

Notons que le système IRIS a été conçu à l'origine pour permettre la fourniture rapide de documents. Dès qu'il a localisé un ouvrage, l'utilisateur peut, sans sortir du système, en effectuer la commande.

Implémenté dans le cadre du *Computer Science Technical report project (CSTR)*, le protocole Dienst (8) est utilisé pour donner accès par le réseau aux rapports de recherche en informatique, produits par les principales universités américaines. Il peut être vu comme une spécialisation de HTTP : des messages spécifiques à la recherche documentaire sont encapsulés dans HTTP et permettent au client Dienst de demander des services documentaires aux serveurs distants, dont l'ensemble constitue une véritable bibliothèque virtuelle. Les deux services principaux sont :
 - *Index* : ce service permet au client d'obtenir des méta-informations sur un ou des documents stockés sur les serveurs Dienst (auteur, titre, nombre de pages, formats disponibles, etc.) ;
 - *Repository* : ce service permet d'obtenir les documents eux-mêmes. Les critères de recherche sont ceux de la RFC 1357 et, comme le système maintient une vue à la fois de la structure physique et de la structure logique des documents, il est possible de demander une page ou une section d'un document plutôt que le document complet.

Combiner les avantages des deux systèmes

Au terme de cette présentation schématique, nous voudrions insister sur les quelques points suivants.

A l'heure actuelle, Z39.50 est un protocole spécialisé, difficile à mettre en œuvre, mais adapté à la construction de systèmes d'informations répartis complexes. Lié à des systèmes de circulation, à des services de dépouillement de sommaires, il ouvre également des perspectives intéressantes dans le domaine de la fourniture de documents. D'une manière générale, il peut aider le monde des bibliothèques à faire ses premiers pas dans la marche irréversible vers les systèmes distribués.

HTTP est un protocole généraliste, fruste mais simple à mettre en œuvre. Il n'est pas adapté à des tâches documentaires complexes, mais peut être étendu et est utilisé tous les jours par des millions d'utilisateurs.

Dans ces conditions, l'idéal serait de combiner les avantages des deux systèmes en permettant d'accéder à des serveurs Z39.50 à partir d'un client *Web*. C'est l'orientation à moyen terme que semblent prendre de grandes institutions américaines (Bibliothèque du congrès, Stanford, OCLC, MIT, etc.).

A long terme tout dépend de la façon dont évolueront les protocoles Internet...

Juillet 1995

BIBLIOGRAPHIE

1. ANSI/NISO Z39.50-1995 *Information retrieval : Application Service Definition and Protocol Specification*.

2. Berners-Lee, T. ; Fielding, R. T. ; Nielsen, H. E. - *Hypertext Transfer Protocol, HTTP/1.0* . - Internet Draft, décembre 1994.

<URL:ftp://nic.nordu.net/internet-drafts/draft-ietf-http-v10-spec-00.txt>

3. Dempsey, Lorcan. - « Distributed Library and Information Systems : the Significance of Z39.50 ». - *Managing Information*, juin 1994, vol. 1, n° 6, p. 41-43.

4. Holm, Liv A. - *Models for Open System Protocol Development : a Technical Report* . - Ottawa : IFLA UDT, 1994.

5. Holm, Liv A. - « The Current Development of SR and European Participation ». - *VINE*, décembre 1994, n° 97, p. 9-13.

6. Lynch, Clifford A. - « Information Retrieval as a Network Application ». - *Library HI TECH*, 1990, vol. 8, n° 4, p. 57-72.

7. Role, François. - « Les systèmes d'identification de documents ». - *Documentaliste - Sciences de l'information*, février 1995, vol. 32, n° 1, p. 43-47.

8. Lagoze, Carl. - « Dienst : an Architecture for Distributed Document Libraries ». - *Communication of the ACM*, avril 1995, vol. 38, n° 3, p. 47.

Z39.50 PEUT AIDER LE MONDE DES BIBLIOTHÈQUES À FAIRE SES PREMIERS PAS DANS LA MARCHÉ IRRÉVERSIBLE VERS LES SYSTÈMES DISTRIBUÉS

requête). Cette norme semble donc constituer un outil de choix pour réaliser un catalogue collectif « logique » ou « virtuel » (il ne s'agit ici que des aspects consultation). Une première expérience a été réalisée en Irlande, en 1994, dans le cadre du système IRIS. Utilisant Z39.50, le système IRIS permet, à travers une interface unique, d'effectuer des recherches bibliographiques dans les catalogues de six bibliothèques universitaires irlandaises utilisant des systèmes informatiques différents.

Le client Z39.50 présente à l'utilisateur une vision complètement unifiée de tous les catalogues. Par ailleurs,