

SRU et SRW en quelques mots

Après le protocole de communication OpenUrl, nous présentons aujourd'hui SRW et SRU, deux protocoles de recherche qui utilisent un langage standardisé permettant à deux sources distribuées de dialoguer.

Dans un article paru dans le numéro précédent de cette revue (pages 12-13), nous parlions de l'OpenURL, la spécification technique qui permet d'étendre une recherche à d'autres ressources ou applications que celle utilisée au départ. Mais qu'en est-il de la question elle-même qui initie le processus ?

C'est ici qu'entrent en lice SRW (Search/Retrieve Web Service) et SRU (Search/Retrieve via URL Web Service) [1], deux protocoles également fondés sur les technologies de l'Internet et du Web. Ces deux protocoles sont des évolutions de la norme Z 3950 et sont comme lui maintenus par la Bibliothèque du Congrès. Ils proposent un langage formel permettant à deux sources distribuées de dialoguer entre elles.

Le contexte : architecture distribuée et web service

La nécessité technique de normaliser une recherche (SRU et SRW) ou de l'étendre (OpenURL) tient à la conception des systèmes documentaires actuels.

Une architecture correspond à la façon dont des ressources sont réparties et reliées entre elles. Dans les systèmes documentaires traditionnels, cette architecture était totalement centralisée. Avec une architecture client-serveur, par exemple, un seul logiciel ou application centralisant toutes les données et toutes les fonctions permettait de produire les notices, de les interroger, de gérer les prêts des utilisateurs, de consulter les documents numériques. Outre la nécessité de faire évoluer les applications, les ressources, pour pouvoir être interrogées, devaient toutes être traitées selon un même format de description.

Les technologies de l'Internet sont bâties sur une architecture

informatique différente : l'architecture distribuée sans nœud central, reposant sur la possibilité d'utiliser des ressources (données ou applications) qui s'exécutent sur des machines réparties sur le réseau et qui communiquent par messages. Comme précédemment, le Client émet une demande et le Serveur y répond, mais sans nœud central, ce qui offre une plus grande souplesse pour construire son offre documentaire.

Le programme informatique lui-même permettant l'échange de données entre applications et systèmes hétérogènes dans des environnements distribués se nomme un *service web*. Ces programmes fonctionnent par et pour des applications (service de prêt, utilisation d'un thésaurus, etc.) ou des machines sans intervention humaine et en temps réel.

SRW et SRU, protocoles de recherche

SRU et SRW sont donc tous deux des protocoles de services web pour la recherche documentaire. La distinction entre les deux vient de la technologie utilisée pour encapsuler requête et réponse. SRU encapsule les éléments des échanges à l'intérieur d'une URL en utilisant l'architecture REST (Representational State Transfer) et le protocole HTTP, alors que SRW utilise le protocole SOAP (Simple Object Access Protocol). Celui-ci est utilisable avec d'autres protocoles qu'HTTP comme e-mail, telnet, etc. Il comporte ainsi moins de restrictions, en particulier sur la longueur ou les jeux de caractères.

Protocole de recherche et langage de requête

En abordant ces protocoles, il est nécessaire de faire la distinction entre *protocole de recherche* et

langage de requête. Un protocole de recherche intègre une série de messages entre Agent et Service utilisés exclusivement par les machines. La requête qui contient la question n'est qu'un des éléments de ce protocole. Pour élaborer la requête, on utilise un langage de requête, le CQL (Context Query Language, ou langage de requête contextuel, anciennement Common Query Language), qui fournit la syntaxe de formulation selon des paramètres précis et des éléments de contexte.

SRU pour le dialogue entre systèmes

Le protocole de recherche SRU permet de transmettre une requête embarquée dans une URL (d'où son nom), à l'instar des moteurs de recherche. Requête et réponses véhiculées dans une URL, nécessairement en XML, sont envoyées sur le réseau. Sur le plan fonctionnel, SRU n'est pas différent de Z3950 : lancement d'une requête par le client, traduction en format standard, interrogation par le serveur, traduction dans une syntaxe qu'il peut comprendre, transmission au client et affichage dans les formats requis.

Mais, à la différence de Z3950, SRU ne propose que trois services de base :

- Explain (Expliquer) permet à la Cible d'indiquer au Demandeur ce qu'elle lui permet de faire. La spécification n'impose pas que ce document Explain existe, mais le recommande fortement ;
- Search/Retrieve (Rechercher/Récupérer). Cette fonction gère l'envoi de la requête élaborée par le Demandeur à la Cible selon les préférences indiquées, puis la réponse. La syntaxe utilisée est conforme au langage CQL ;
- Scan (Balayer) permet au Demandeur de consulter les termes proposés par une cible autour d'un terme spécifié dans les paramètres, ainsi que le nombre d'occurrences de chacun de ces termes. Scan est utilisé pour sélectionner des termes avant une recherche, ou pour vérifier

visuellement les résultats d'une recherche.

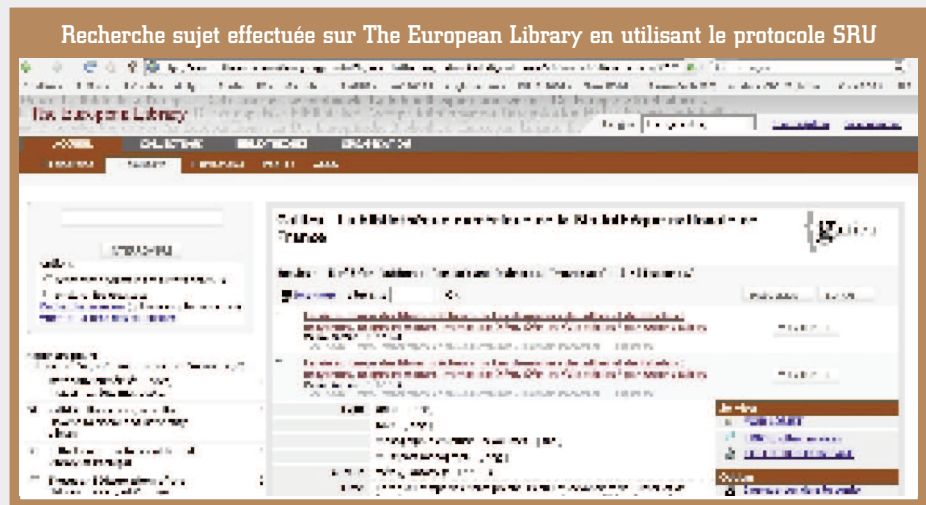
CQL, pour la formulation de la requête et l'exploitation des résultats

Le langage de requête CQL utilisé avec SRU permet de formuler aussi bien des requêtes simples, dans des fonds non structurés avec un ou plusieurs mots, que des requêtes plus complexes sur des fonds structurés avec plus ou moins de métadonnées. Ce langage à la fois intuitif et très expressif permet l'utilisation des valeurs d'index ou de vocabulaires contrôlés récupérées par ce mécanisme. Une synthèse entre la richesse de Z3950 et la simplicité des technologies du Web. CQL intègre aussi des informations sur le contexte comme les préférences utilisateur : forme des réponses, ordre de tri, etc.

L'aspect intuitif de ce langage par rapport à Z3950 va rappeler d'étranges souvenirs aux spécialistes des serveurs professionnels : booléen, tri (*sort*), proximité (*prox*), diverses relations. Des jeux d'affichage ont été développés pour des contextes différents : Dublin Core, Musique, zThes, MARC, GILS, etc. Une feuille de style XSLT peut être ajoutée en sortie pour l'affichage, et SRU permet aussi de gérer des paramètres de données supplémentaires, par exemple pour proposer à l'utilisateur des requêtes additionnelles pertinentes si sa requête initiale ne donne pas de réponse, ou pour limiter sa requête dans le cas contraire.

SRU, une spécification normalisée

Le langage CQL enrichit considérablement SRU par rapport à OpenSearch [3], autre protocole de recherche via les URL, développé par A9, une branche d'Amazon, qui est proposé en *common licence* et utilisé très largement par des sociétés comme Microsoft ou Yahoo!. Cette expressivité, associée à une diffusion par les URL, intéresse une communauté plus large.



C'est ainsi qu'un comité technique appelé « OASIS Search Web Services Technical Committee » a été constitué fin 2007 au sein d'OASIS [4a], consortium en charge d'un ensemble de protocoles et schémas de métadonnées qui s'imposent comme des normes dans le monde de l'Internet et du Web. En quelques mois, ce groupe a établi deux avant-projets de documents publiés le 30 juin 2008, l'un sur un protocole de recherche [4b], l'autre sur le langage de recherche [4c]. Ces spécifications intègrent plus complètement les caractéristiques du protocole REST et ont été enrichies d'autres fonctionnalités, comme celles implantées dans

OpenSearch : le support d'ATOM pour la diffusion.

Le site officiel SRU [1] répertorie des outils pour la plupart en libre accès, et des applications. Mentionnons également The European Library [5] qui exploite le protocole SRU pour interroger simultanément les collections de trente-deux bibliothèques nationales d'Europe [voir la figure ci-dessus].

Avec un protocole et un langage de recherche standardisés et riches, le principe même de la recherche fédérée ou des métamoteurs va se trouver grandement facilité. •

Sylvie Dalbin, sylvieatd@aol.com
Odile Giraud, ogiraud@docandco.fr

Références

- [1] Site de référence de SRU (version 1.2) : www.loc.gov/standards/sru
- [2] *Un point sur les nouveaux protocoles de recherche d'information : SRU, OpenSearch/A9, CQL et Xquery* / Sally H. McCallum, IFLA 2006, Seoul. Traduction française : www.ifla.org/IV/ifla72/papers/102-McCallum_trans-fr.pdf
- [3] OpenSearch 1.1 de A9 : www.opensearch.org/Specifications/OpenSearch/1.1/Draft_3
- [4a] OASIS : www.oasis-open.org/specs
- [4b] *Search Web Services - searchRetrieve Operation: Binding for OpenSearch Version 1.0*, Committee Draft 01, 30 June 2008, <http://docs.oasis-open.org/search-ws/v1.0/opensearch-v1.0.html>
- [4c] *CQL 1.2: The Contextual Query Language Version 1.0*, Committee Draft 01, 30 June 2008, <http://docs.oasis-open.org/search-ws/june08releases/cql-1-2-V1.0-cd-01.doc>
- [5] The European Library : <http://search.theeuropeanlibrary.org>
Voir aussi : *Des bibliothèques 2.0 : SRU/SRW Webographie/webography*, mars 2007, <http://bibliotheque20.wordpress.com/2007/03/16/srusrw-webographie-webography>