

Ouverture de DATA.IDREF.FR

Lors de la première conférence du World Wide Web Consortium (W3C) organisée en 1994, Tim Berners-Lee, l'inventeur du Web, traçait déjà le possible futur du web et proposait, dès 1998, une véritable «feuille de route» vers le web sémantique.

L'idée maîtresse pour favoriser interopérabilité et traitements automatisés : intégrer à l'espace documentaire que constituait le web dans sa dimension initiale une nouvelle logique basée sur les relations entre les différentes entités (lieu, personne, concept, oeuvre...).

À cet effet, le W3C a développé RDF¹, langage de modélisation conçu pour décrire tous les types de relations entre les ressources elles-mêmes et leurs métadonnées selon la forme canonique de triplets «sujet, prédicat, objet».

Prenons l'exemple d'une série d'assertions, dont chacun des éléments possède une adresse pérenne, URI composée de préférence à partir de référentiels ou de vocabulaires standardisés : «Edgard Allan Poe» (sujet) «est l'auteur de» (prédicat 1) «Histoires extraordinaires» (objet 1) et «Aventures d'Arthur Gordon Pym» (objet 2) qui «ont pour traducteur» (prédicat 2) «Charles Baudelaire» (objet 2) et «Charles Baudelaire (sujet) «est l'auteur de» (prédicat 1) «Les fleurs du mal» (objet 3). À partir de cette logique relationnelle, se forme un graphe de connaissances présentant par exemple la liste de tous les titres ayant pour auteur ou pour traducteur Charles Baudelaire.

Au delà de certaines réalisations phare comme *Wikidata*², les réservoirs de données structurées en RDF se mettent progressivement en place dans le monde de la documentation révélant toute la richesse et le potentiel des métadonnées produites par les bibliothèques, comme en témoignent des initiatives comme *data.bnf.fr*³, *data.persee.fr*⁴ et maintenant *data.idref.fr*⁵.

Des autorités Sudoc à data.idref.fr

Depuis 2010, les principes du web sémantique ont guidé plusieurs réalisations de l'Abes, notamment en ce qui concerne la valorisation des données d'autorité.

En effet, jusqu'à la mise en service

d'IdRef⁶ en 2010, les notices d'autorité Sudoc (personnes, familles, titres et auteur/titre, collectivités, congrès, lieux, concepts) n'étaient visibles que par les professionnels du réseau Sudoc. IdRef a non seulement renforcé leur visibilité mais a également permis d'en diversifier les usages, tout d'abord au service des autres environnements documentaires de l'Abes (Calames et theses.fr) puis au service de plusieurs projets nationaux.

En accord avec les principes du web sémantique, les données d'autorité ont été dotées d'adresses pérennes et d'une modélisation RDF exprimée dans les vocabulaires standards, ce qui garantit leur interopérabilité et une bonne exposition sur le Web en vue de leur réutilisation.

Toutefois, pour que les utilisateurs puissent bénéficier pleinement de la richesse des données, il est apparu nécessaire de mettre à leur disposition un Triple Store⁷, base conçue spécifiquement pour accueillir des données modélisées en RDF et interrogeables avec le langage SPARQL⁸. C'est chose faite avec la mise en service fin 2018 de *data.idref.fr*.

Faciliter l'exploration des données

Dès son ouverture, *data.idref.fr* est riche d'environ 110 millions de triplets – ce qui correspond aux 3,5 millions de notices d'autorités IdRef associés aux 33 millions de liens générés avec les 11 millions de références bibliographiques Sudoc.

Ainsi, *data.idref* accroît les possibilités d'exploration et d'exploitation des données Sudoc : il est désormais possible de surfer sur les relations entre documents et autorités ou de relier les autorités entre elles par l'intermédiaire des documents...

Illustration de la richesse de la modélisation en graphe, les liens sémantiques ont une densité particulière puisqu'ils font le lien « classique » entre les auteurs et leurs références bibliographiques, mais portent également l'information sur le sens et la nature des relations. Ainsi, il

est désormais possible de distinguer les documents dont Baudelaire est l'auteur de ceux pour lesquels il est traducteur ou préfacer.

Passer du calcul et de l'affichage à la volée des liens et des ressources au calcul stocké de ces mêmes objets ne devait toutefois pas se traduire par une perte de « fraîcheur » des données. Petite originalité, tant technique que fonctionnelle, par rapport à d'autres Triple Stores en service, *data.idref.fr* est branché directement sur les environnements de production des catalogues. Les données sont donc modifiées en temps réel, pour des résultats parfaitement à jour.

Seconde originalité de *data.idref.fr* : pour élargir l'accès à ce formidable gisement de données et faciliter la navigation dans les triplets, trois interfaces de recherche sont proposées, s'adaptant ainsi aux usages néophytes comme plus aguerris. La navigation peut commencer par une interrogation en langage naturel ouvrant sur des rebonds de lien en lien et un affichage par facettes. Avec l'interface YASGUI⁹, la rédaction des requêtes est facilitée grâce aux nombreuses requêtes-exemples proposées. Enfin, pour des usages experts, un éditeur SPARQL est disponible.

FRANÇOIS MISTRAL

Abes - Service Référentiels et Autorités
francois.mistral@abes.fr

[1] Resource Description Framework w3.org/RDF

[2] wikidata.org

[3] data.bnf.fr

[4] data.persee.fr

[5] data.idref.fr

[6] www.idref.fr

[7] fr.wikipedia.org/wiki/Triplestore

[8] SPARQL Protocol and RDF Query Language est un langage d'interrogation spécifique pour les données RDF.

[9] about.yasgui.org