

Web sémantique et big data : rapport de projet

Une ontologie sur les livres de fiction

Ana Armenta

Alexane Jouglar

Sorbonne Université

Dans le cadre de notre enseignement de Web sémantique, nous avons travaillé au développement d'une véritable ontologie. Nous avons repris le sujet choisi lors de notre premier travail (wiki), à savoir : les littératures de l'imaginaire. Lors de l'élaboration, notre travail a évolué vers une ontologie de livres de fiction avec un intérêt plus prononcé pour les littératures de l'imaginaire. Le temps nous étant compté, il était en effet nécessaire de restreindre notre objectif et de ne nous intéresser qu'à un un pan de la littérature. Dans ce rapport, nous réaliserons un bref état de l'art des ontologies avant d'évoquer le processus d'élaboration de la nôtre.

Etat de l'art

L'ontologie (du grec *ontos*, ce qui existe, et *logos*, le discours) désignait d'abord un domaine de la philosophie appartenant à la branche de la métaphysique. L'ontologie s'intéresse à ce qui existe dans le monde et aux relations entre ces existants. En informatique, dès les années 1970, des informaticiens ont commencé à concevoir des schémas de classes dans le but d'implémenter une application avec des concepts clairs. Ceci dit, la notion d'ontologie n'a été acceptée par la communauté informatique qu'à partir des années quatre-vingt-dix et ne désignait plus alors l'étude de l'être mais un objet (on ne parle plus de l'ontologie mais d'une ontologie) contenant des choses ayant

des liens entre elles. D'ailleurs dans Protégé, nous pouvons bien voir que toute nouvelle classe découle de Thing qui est la classe ultime : tout est une chose. La création des classes provient du principe de généralisation : ce sont des abstractions ayant des réalisations concrètes dans notre monde. Ces réalisations sont dites instances. Les relations précédemment évoquées entre les êtres dans le monde sont dites propriétés dans une ontologie. Enfin, les instances peuvent avoir des caractéristiques qui ne sont pas liées à leur appartenance à une classe : ces caractéristiques sont des attributs.

« L'introduction d'une ontologie dans un système d'information vise à réduire [...] la confusion conceptuelle et terminologique et à tendre vers une compréhension partagée pour améliorer la communication, le partage, l'interopérabilité et le degré de réutilisation possible. Une ontologie informatique offre un cadre unificateur et fournit des « primitives », des éléments de base pour améliorer la communication entre les personnes, entre les personnes et les systèmes, et entre les systèmes. » Gandon (2006)

Aujourd'hui, le Web actuel s'étend peu à peu vers ce qui est appelé le Web sémantique et qui fait l'objet de l'enseignement dont nous avons bénéficié : le Web sémantique utilise des systèmes ontologiques. L'information enregistrée dans les bases de données a un sens clairement défini, ce qui permet de traiter l'information disponible en ligne de manière plus efficace et plus exacte. Pour certains, la généralisation du Web sémantique ne dépasse pas le domaine de l'utopie. Pourtant de plus en plus de contenu y est ajouté. Par ailleurs, le Web sémantique offre des avantages non négligeables qui finiront sûrement par avoir raison des réfractaires (Domingue et al. (2011)) :

- réutilisation des données plus simple
- mise à jour automatique lorsque la page Web est modifiée

- la recherche de document sémantique permet aux utilisateurs de trouver tous les documents qui mentionnent une ou plusieurs instances de l'ontologie et / ou des relations.

D'ailleurs, certaines applications web utilisent déjà des ontologies. On pourrait citer ici Amazon puisque lorsque l'entreprise a été créée en 1994, son activité initiale concernait la vente de livres.

Pour le développement de l'ontologie, nous nous sommes inspirés de la méthodologie *Méthontologie* qui a été développée par le groupe d'ingénierie ontologique de l'Universidad Politécnica de Madrid Corcho (2003). Cette méthodologie permet la construction d'ontologies au niveau des connaissances, et trouve ses racines dans les principales activités identifiées par le processus de développement logiciel de l'Institute of Electrical and Electronics Engineers (IEEE) et dans d'autres techniques d'ingénierie des connaissances. *Méthontologie* donne un guide sur la manière de développer une ontologie à travers la spécification, la conceptualisation, la formalisation, l'implémentation et la maintenance de l'ontologie. Pour construire des taxonomies de concepts, on doit choisir des termes qui sont des concepts dans le glossaire des termes. Ce guide propose également d'utiliser les quatre relations taxonomiques :

- Sous-classe de
- Décomposition disjointe
- Décomposition exhaustive
- Partition

Elaboration de notre ontologie

Dans cette partie, nous présenterons les méthodes utilisées pour construire notre ontologie avant d'indiquer les requêtes faites sur celle-ci et leur résultat.

Construction de notre ontologie : méthode

Nous avons donc commencé par préparer le glossaire des termes afin de créer nos classes : 25 classes au total dont 14 sous-classes. Une de ces classes se nomme "style". Nous avons ajouté à "style" une seule sous-classe : "Littératures de l'imaginaire". Cela ne veut pas dire qu'il n'y ait qu'un seul style. Nous aurions pu ajouter : "Littérature contemporaine", ou encore "Littérature historique" mais il s'agissait avant tout de répondre à une exigence : construire une ontologie qui se tienne et qui nous permette de formuler quelques requêtes. Puis nous avons créé des propriétés entre ces classes (20 relations sont répertoriées) : c'est ici que les taxonomies de concept nous ont aidées. Chacune de ces relations a été analysée pour déterminer les classes concernées et leurs relations ainsi que les complémentarités entre ces relations (c'est-à-dire les relations binaires).



Figure 1. Exemple de relation binaire

Note. 1 Toutes les relations ne sont pas nécessairement binaires

Comme notre ontologie portait particulièrement sur les littératures hispanophones et francophones, les étiquettes ont été indiquées dans ces deux langues. L'avant-dernière étape de cette construction reposait sur l'ajout des attributs : nous avons créé 10 attributs au total. La dernière étape consistait alors à ajouter des instances de livres, d'auteurs, d'imprimeurs, etc. C'est-à-dire des instances pour nos classes.

Pour vérifier la consistance de notre ontologie, nous avons régulièrement utilisé Her-miT. Une fois que notre ontologie nous a semblé correcte et propre à tester quelques requêtes SPARQL, nous nous y sommes consacrées.

Requêtes SPARQL

Les deux pages suivantes présentent sous forme de tableau les requêtes SPARQL que nous avons pu construire en fonction du résultat recherché.

Table 1: Requêtes

Recherche	Requête
Ceux qui ont écrit quelque chose	SELECT ?Author ?Book WHERE {?Author <http://www.semanticweb.org/alexane_ana/projet_semantiqueweb#wrote> ?Book}
Les livres avec leur auteur	PREFIX livres: <http://www.semanticweb.org/alexane_ana/projet_semantiqueweb#> SELECT ?subject ?object WHERE { ?subject livres:wrote_by ?object }
Les dates de publications (limitées à trois résultats)	PREFIX livres: <http://www.semanticweb.org/alexane_ana/projet_semantiqueweb#> SELECT ?subject ?object WHERE { ?subject livres:date_of_publication ?object } LIMIT 3
Livres en Espagnol qui n'ont pas comme Style Science-fiction	PREFIX Livres: <http://www.semanticweb.org/alexane_ana/projet_semantiqueweb#> SELECT ?Author ?Speculative_fiction WHERE { ?Author Livres:its_language Livres:Spanish . OPTIONAL { ?Author Livres:has_style ?Speculative_fiction . FILTER (?Speculative_fiction != Livres:Science-fiction)}}}
Livres en Espagnol qui ont comme Style Gothique et Horreur	PREFIX Livres: <http://www.semanticweb.org/alexane_ana/projet_semantiqueweb#> SELECT ?Author ?Speculative_fiction WHERE { ?Author Livres:its_language Livres:Spanish . OPTIONAL { ?Author Livres:has_style ?Speculative_fiction . FILTER (?Speculative_fiction = Livres:Science-fiction ?Speculative_fiction = Livres:Horreur)}}}
Les livres et leur année de publication, les résultats par année descendante	PREFIX Livres: <http://www.semanticweb.org/alexane_ana/projet_semantiqueweb#> SELECT * WHERE { ?Book Livres:date_of_publication ?year . } ORDER BY DESC(?year)

Table 2: Requêtes

Recherche	Requête
Les livres qui ont "silence" dans leur titre	PREFIX rdf:<http://www.semanticweb.org/alex-ane_ana/projet_semantiqueweb#> SELECT ?auteur ?livre WHERE { ?auteur rdf:wrote ?livre FILTER regex(str(?livre), "silence") }
Les auteurs qui ont écrit deux livres	PREFIX rdf:<http://www.semanticweb.org/alex-ane_ana/projet_semantiqueweb#> SELECT ?auteur WHERE { ?auteur rdf:wrote ?livre } GROUP BY ?auteur HAVING (count(?livre) = 2)
Les livres qui sont en format Poche	PREFIX rdf:<http://www.semanticweb.org/alex-ane_ana/projet_semantiqueweb#> SELECT ?livre ?format WHERE { ?livre rdf:has_format ?format FILTER regex(str(?format), "Poche") }
Les livres qui coûtent plus de quinze euros	PREFIX rdf:<http://www.semanticweb.org/alex-ane_ana/projet_semantiqueweb#> SELECT ?livre ?prix WHERE { ?livre rdf:cost ?prix FILTER (?prix > 15.00) }
Les livres qui coûtent moins de dix euros	PREFIX rdf:<http://www.semanticweb.org/alex-ane_ana/projet_semantiqueweb#> SELECT ?livre ?prix WHERE ?livre rdf:cost ?prix FILTER (?prix < 10.00)
Les livres publiés en Argentine	PREFIX rdf:<http://www.semanticweb.org/alex-ane_ana/projet_semantiqueweb#> SELECT ?livre ?lieu WHERE { ?livre rdf:has_publishing_place ?lieu FILTER regex(?lieu, "Argentine", "i") }

Conclusion

Dans ce rapport nous avons fait un rapide tour des ontologies avant de présenter notre travail. Ce travail consistait en la construction d'une ontologie portant sur les livres de fiction avec un intérêt tout particulier pour les littératures de l'imaginaire. Ce travail nous a permis d'en apprendre davantage sur le fonctionnement des ontologies et sur les possibilités qu'elles offrent. Nous pouvons tout à fait imaginer un système basé sur une telle ontologie dans une librairie qui permettrait aux libraires de pouvoir renseigner leurs clients avec rapidité et précision sans avoir à fournir par exemple un titre ou un numéro ISBN complets. Il suffirait de quelques informations seulement pour trouver le livre recherché.

Remerciements

Nous remercions les enseignants nous ayant accompagné lors de ce semestre pour l'enseignement de Web sémantique.

References

- Corcho, O; Fernández-López, M. (2003). Building legal ontologies with methontology and webode. *Lecture Notes in Computer Science*.
- Domingue, J., Fensel, D., and Hendler, J. A., editors (2011). *Handbook of Semantic Web Technologies*. Springer.
- Gandon, F. (2006). Ontologies informatiques. *Interstices.info*.