

Focus 1 page (p.34), 3.500 signes

Bio

Raphaël Troncy est maître de conférences à l'institut Eurecom où il anime l'axe de recherche *Multimedia semantics*. Docteur en informatique à l'université de Grenoble en 2004, il a travaillé au CNR à Pise en 2005 et au CWI à Amsterdam jusqu'en 2009. Il dirige le groupe de travail *Media Fragments* au sein du W3C et est spécialiste dans les formats de métadonnées pour le multimédia et leur interopérabilité.

chapo

Dans le contexte du web sémantique, les ontologies fournissent le vocabulaire propre à un domaine et fixent – avec un degré de formalisation variable – le sens des concepts et des relations entre ceux-ci. RDFS, et mieux encore OWL, permettent de formaliser ces ontologies.

OWL, un « chouette » langage pour représenter des ontologies

Texte : 3398 s.

Exprimer de la connaissance sur le Web est l'ambition du web sémantique. Au-delà de ce simple mot d'ordre, diffuser des ontologies sur le Web est le moyen de permettre à d'autres de se les approprier, de les étendre et de les réutiliser. Le W3C a standardisé au fil du temps différents langages, d'expressivité (et de complexité) croissante, pour représenter et publier ces ontologies.

Le langage RDF Schema (RDFS) a tout d'abord été développé pour étendre RDF en décrivant plus précisément les ressources utilisées pour étiqueter les graphes. Pour cela, il fournit un mécanisme permettant de spécifier les classes dont les instances sont des ressources, comme les propriétés.

RDFS ne fournissant que des mécanismes très primitifs pour spécifier ces classes, OWL permet de contraindre plus précisément leurs descriptions (en les décrivant comme union, intersection, complémentaire d'autres descriptions ou comme l'ensemble d'un certain nombre d'individus), des domaines de relations (en spécifiant le type de leurs valeurs) ou des relations (en les déclarant transitives, symétriques ou en spécifiant leur inverse).

En 2009, une nouvelle version du langage OWL intitulée OWL2 ajoute de nouveaux constructeurs tels que la possibilité de construire des chaînes de propriétés ou d'introduire des clés. De plus, OWL2 généralise la notion de profils (EL, QL, RL) correspondant à des sous-ensembles pour lesquels la complexité des raisonnements est bien maîtrisée.

Des ontologies pour tous les goûts

Au delà des langages de représentation, un certain nombre d'ontologies, généralement petites,

se sont imposées pour couvrir à peu près tous les domaines : de la description d'un profil utilisateur (FOAF) à son activité sur un réseau social (OGP) ou sur un système documentaire collaboratif (SIOC) ; de la notice bibliographique (DublinCore) à la description de produits (GoodRelations), des actualités (rNews) ou des événements (Lode) ; de la culture (Cidoc) aux sciences de la vie et à la médecine (UMLS).

La multiplication des ontologies impose leur catalogage. Plusieurs portails sont ainsi apparus pour référencer ces ontologies et fournir des exemples d'utilisation. Signalons à ce stade deux initiatives intéressantes : le vocabulaire VOAF qui permet de décrire les dépendances entre les ontologies du web sémantique et le jeu de données LOV¹ associé ; et le service prefix.cc qui permet de trouver le vocabulaire le plus populaire pour un préfixe donné. À terme, l'objectif est de pouvoir trouver automatiquement quelle ontologie est à la fois largement utilisée et couvre au mieux un besoin donné. Un tel service est en cours de développement dans le projet Datalift².

L'ontologie rêvée par les moteurs de recherche

Le 2 juin 2011, les trois plus gros moteurs de recherche du Web (Google, Yahoo!, Bing) ont conjointement annoncé la sortie de schema.org, un vocabulaire général pour structurer les données des pages web. Cette ontologie est tantôt perçue comme la reconnaissance de l'utilité des technologies sémantiques pour l'indexation des ressources documentaires et leur adoption par les plus grands acteurs du Web, ou comme l'échec de la construction d'ontologies diverses aux dépens d'une seule ontologie monolithique, développée secrètement et sans consensus, pour ne refléter au final qu'une vue parcellaire et biaisée des connaissances. ●

¹<http://labs.mondeca.com/dataset/lov/>

²<http://datalift.org/>