

Du raisonnement avec OWL et NoSQL?

IT Sécurité / Directeur de thèse: Prof. Dr. Olivier Biberstein
Expert: Prof. Dr. Gilles Falquet

Notre but est de découvrir des connections ou des relations entre sites Web hébergeant des informations illicites pour permettre l'étude et la compréhension de l'organisation en réseaux de ces derniers. La représentation des sites est décrite au moyen de Web Ontology Language (OWL). Un «raisonneur» est employé afin de déduire de nouvelles relations entre les sites à partir de l'information OWL. Parallèlement, plusieurs types de bases de données NoSQL (Not Only SQL), mieux adaptées dans ce contexte, sont étudiés comme alternatives aux bases de données relationnelles.

L'objectif de ce travail est d'offrir des moyens techniques pour mettre en évidence des liens entre sites Web qui hébergent des informations illicites (utilisation abusive de marques déposées, vente de produits contrefaits ou prohibés, etc.) et de comprendre comment ces sites sont organisés entre eux. Pour ce faire l'information extraite des sites Web est exprimée sous forme d'ontologie au moyen de Web Ontology Language (OWL). A partir de cette information, sous forme d'ontologie, un raisonneur est employé pour inférer les relations entre sites.

Raisonnement

Une partie importante du raisonnement réside en réalité dans la description de l'information. Pour qu'un raisonneur puisse interpréter la structure de l'information et

l'associer à des concepts, chacun de ces concepts doit être spécifié. OWL permet d'exprimer des concepts et d'associer des données à ces concepts. Il est aussi possible de définir des concepts équivalents. Les concepts équivalents offrent la possibilité d'observer ou d'analyser l'information sous différents points de vue.

Un raisonneur basé sur OWL permet des raisonnements liés au typage de l'information. Une fonction du raisonneur consiste à trouver des relations entre les différents concepts. Des équivalences de concepts ou des relations d'héritages peuvent alors être découvertes. Un raisonneur détecte aussi les données satisfaisant la description de concepts. Ces données sont alors reconnues comme étant des instances

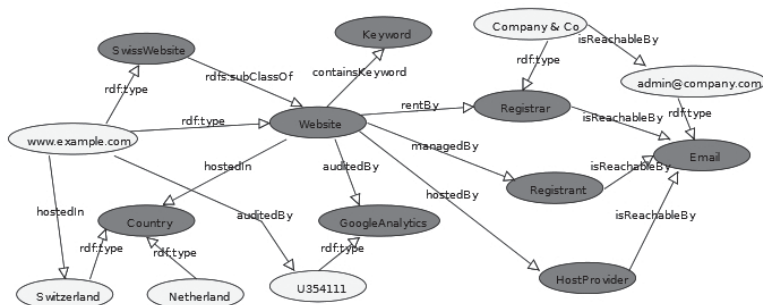
de ces concepts. Le raisonneur signale également toute incohérence découverte.

NoSQL

Les bases de données relationnelles sont adaptées pour la grande majorité des applications. Néanmoins ces bases de données persistent l'information sous une structure relativement rigide limitant l'évolution de cette dernière. Il n'est pas simple non plus de recomposer une donnée répartie sur plusieurs tables. L'approche Not Only SQL (NoSQL) apporte des solutions à cette problématique. Certaines de ces bases de données persistent l'information sous forme de graphes, d'autres sous forme de documents. D'autres encore persistent l'information sous forme de triplets. Nous avons évalué trois implémentations de bases de données que sont Neo4j une base de données à graphes, AllegroGraph une base de données à triplets et MongoDB une base de données à documents. Notre étude traite des langages de requêtes mises à disposition et les possibilités de migration des données vers OWL.



José Beuret



Exemple de représentation de sites Web au moyen d'OWL.