

Plateforme pour l'élaboration de réseaux sociaux spécialisés reposant sur des bases de données graphe

David Fernandez*, Benjamin Billet*, Didier Parigot*

*Équipe-projet ZENITH, Inria
{prenom.nom}@inria.fr

Introduction En partenariat avec la startup BEEPEERS¹, nous concevons une plateforme pour le développement rapide de *réseaux sociaux spécialisés* (RSS), à destination de communautés ciblées (p. ex. fédérées autour d'un évènement ou d'une activité sportive). Cette plateforme exploite les *bases de données graphe* (BDG), de façon à implémenter efficacement des RSS manipulant des données connectées (p. ex. liste d'amis, de publications ou d'activités). L'objectif de cet article consiste à présenter les fonctionnalités et les caractéristiques qui devraient être assurées par une BDG pour réaliser rapidement de nouveaux RSS. Par exemple, le typage et de sous-typage des sommets et des arcs est requis pour faciliter la réutilisation des modèles de données, tout comme la recherche de motifs dans le voisinage d'un sommet (p. ex. les amis en communs). Cependant, les BDG destinées au monde industriel ne répondent pas à ces besoins, ou partiellement (Labute et Dombroski, 2014). Pour résoudre ce problème, nous introduisons une couche logicielle permettant d'abstraire les BDG existantes et de fournir les fonctionnalités nécessaires pour le développement rapide de RSS. En pratique, cette couche logicielle permet à BEEPEERS de développer rapidement de nouveaux RSS à la demande, par simple configuration des schémas et des requêtes à appliquer sur la BDG.

Élaboration de réseaux sociaux spécialisés La mise en place de nouveaux RSS passe par la définition d'un modèle de données générique et sa spécialisation en modèles de données spécifiques à chaque communauté ciblée. Par exemple, tous les réseaux sociaux partagent le concept d'entité capable de publier et d'évaluer du contenu. Pour l'organisateur d'une exposition, ce concept se traduira par des exposants qui publient des actualités sur leurs stands et des visiteurs qui partagent, discutent et évaluent ces actualités. Pour ce faire, un mécanisme de *typage* et de *sous-typage* est requis, permettant (i) d'exprimer les concepts sociaux génériques sous la forme de sommets (des entités sociales) et d'arcs (des relations sociales) et (ii) de spécialiser ces concepts pour une communauté. Cette hiérarchie de type forme alors le schéma de la BDG et contraint les types de sommets que les différents types d'arcs peuvent connecter.

Dans un réseau social, les requêtes usuelles ne parcourent qu'une portion limitée du graphe, au voisinage d'un petit nombre de sommets ; par exemple, rechercher les commentaires associés à une publication ou les dernières activités d'une personne. Les systèmes d'analyse destinés à recommander du contenu aux utilisateurs peuvent eux-aussi se décomposer en sous-tâches consistant à évaluer le voisinage d'un contenu (recommander du contenu similaire) ou le

1. www.beepeers.com

Plateforme pour l'élaboration de RSS reposant sur des BDG

voisinage d'un individu (recommander du contenu apprécié par les individus similaires) (Bobadilla et al., 2013). Aussi, un mécanisme de *requête par motif* est requis, permettant (i) d'analyser le voisinage d'un sommet et d'y extraire les sous-graphes qui satisfont un motif donné puis (ii) de collecter, de transformer et d'agréger les informations contenues dans ces sous-graphes.

À l'heure actuelle, trois moteurs de BDG sont populaires² dans le monde industriel : *Neo4j*, *OrientDB* et *Titan*³. Toutefois, aucun d'entre eux ne possède les caractéristiques nécessaires pour l'implémentation de RSS, soit par l'absence d'un mécanisme de typage satisfaisant, soit par l'absence d'outil pour rechercher des motifs dans le voisinage d'un sommet (Labute et Dombroski, 2014). À cette fin, nous introduisons un cadre de développement uniforme pour la conception de RSS. Celui-ci se base sur la couche d'abstraction *Blueprint*⁴, un ensemble d'interfaces de manipulation de graphe, supportée par les trois BDG considérées. Notre plateforme complète cette abstraction avec les fonctionnalités que nous avons évoquées : typage et requêtes par motifs. Ainsi, une fois qu'un développeur a écrit une fonctionnalité générique, il peut alors la spécialiser rapidement sans écrire de code supplémentaire. Pour ce faire, seuls des fichiers de configuration doivent être fournis : (i) le schéma du RSS, qui étend le schéma générique grâce au mécanisme de sous-typage, et (ii) les motifs spécifiques à rechercher dans le graphe de ce RSS, p. ex. pour répondre aux requêtes des utilisateurs ou recommander du contenu. Du fait de cette approche par configuration, le nouveau RSS peut être rapidement mis en œuvre ou modifié avec une interruption de service minimale.

Conclusion Nous venons de présenter les fonctionnalités requises par une BDG pour l'implémentation rapide de nouveaux RSS, p. ex. les mécanismes de typage ou la recherche de motifs complexes. Ces fonctionnalités sont à priori requises pour d'autres applications, telles que l'évaluation de contenu à grande échelle (Labute et Dombroski, 2014). Outre notre implémentation actuelle basée sur les données réelles fournies par BEEPEERS, nous prévoyons d'étudier d'autres besoins tels que la mise en cache automatique des données dans les BDG.

Remerciements Nous remercions Mickaël Jurret de BEEPEERS pour sa collaboration active.

Références

- Bobadilla, J., F. Ortega, A. Hernando, et A. Gutiérrez (2013). Recommenders systems survey. *Knowledge-Based Systems* 46.
- Labute, M. X. et M. J. Dombroski (2014). Review of graph databases for big data dynamic entity scoring. Technical report, Lawrence Livermore National Laboratory.

Summary

This paper analyzes the functions required by industrial graph databases in order to apply them to fast development of dedicated social networks for micro-communities. As a solution, we introduce a framework which provides the required functions on top of these databases.

2. Voir l'historique de popularité : db-engines.com/en/ranking_trend/graph+dbms (03/09/2015)

3. neo4j.com, orientdb.com, thinkaurelius.github.io/titan (03/09/2015)

4. blueprints.tinkerpop.com (03/09/2015)