

---

# Répondre à des requêtes dans un système distribué à base de connaissances

Faisal Alkhateeb — Antoine Zimmermann

INRIA Rhône-Alpes

{faisal.alkhateeb, antoine.zimmermann}@inrialpes.fr

---

*RÉSUMÉ. Un système distribué à base de connaissances comportent un ensemble d'ontologies, reliée entre elles par des relations sémantiques. Nous nous intéressons aux réponses à une requête posée en termes d'une ontologie d'un tel système. Ces réponses peuvent comporter des individus de différentes ontologies. Pour évaluer ces réponses, nous présentons deux méthodes avec leurs avantages et leurs inconvénients.*

*ABSTRACT. A knowledge-based distributed system comprises a set of ontologies, interconnected through semantic relationships. We are interested in answering queries posed in terms of an ontology of such system. The answers may contain individuals from different ontologies. In order to evaluate these answers, we present two methods as well as their advantages and drawbacks.*

*MOTS-CLÉS : Réponses à des requêtes, base de connaissances distribuée, logiques de description*

*KEYWORDS: Query answering, distributed knowledge base, description logics*

---

## 1. Introduction

L'émergence du Web Sémantique a porté l'attention sur le développement de systèmes dans lesquels les connaissances peuvent être partagées dans un environnement hétérogène et distribué. Par ailleurs, répondre à des requêtes sur des bases de connaissances expressives est une tâche difficile. Par conséquent, lorsque la description des connaissances est répartie, ce problème devient d'autant plus complexe. Nous nous intéressons à des systèmes formant un réseau de bases de connaissances (ou ontologies) reliées entre elles par des relations sémantiques. Notre problème consiste à répondre à des requêtes sur un tel système, posées en terme d'une ontologie lui appartenant, appelée l'*ontologie cible*. Nous distinguons les réponses locales (non distribuées), déterminées par une évaluation de requêtes sur l'ontologie cible, et les réponses distribuées déduites de l'ensemble des ontologies en exploitant les relations sémantiques. Cette notion de réponse distribuée est décrite en Sect. 2.

À cet effet, nous souhaitons exploiter au mieux les outils de raisonnement locaux supposés existants pour chaque nœud. Deux approches sont envisagées : (1) la première s'appuie sur une méthode d'extension de l'ontologie cible en compilant localement des connaissances des autres ontologies (Sect. 3.1). Ainsi les réponses sont obtenues par une simple évaluation de requête sur cette ontologie étendue. (2) la seconde est fondée sur la diffusion de requêtes en les transformant grâce aux relations sémantiques (Sect. 3.2). Cette approche est inspirée de travaux mettant en jeu des relations non directionnelles, pour des langages spécifiques et pour des données plutôt homogènes (Goasdoue *et al.*, 2003; Halevy *et al.*, 2003; Nejdil *et al.*, 2002; Cruz *et al.*, 2004). Cette méthode permet d'exploiter les outils d'évaluation de requêtes associés à chaque ontologie. Les deux approches peuvent être combinées (Sect. 3.3). L'application de la première méthode aux logiques de description distribuées (Distributed Description Logics (Borgida *et al.*, 2002)) est étudiée dans (Alkhateeb *et al.*, 2007).

## 2. Réponses à des requêtes sur une base de connaissances distribuée

Nous ne nous intéressons ici qu'aux réponses à des *requêtes conjonctives* (plus précisément des *requêtes ciblées*) écrites en termes d'une ontologie donnée, appelée l'*ontologie cible*<sup>1</sup>, mais en accord avec les connaissances décrites dans l'ensemble des ontologies et des relations sémantiques. Les requêtes conjonctives s'écrivent  $q(\bar{x}) \leftarrow \text{conj}(\bar{x}, \bar{y})$  où  $\bar{x}$  est un tuple de *variables distinguées*,  $\bar{y}$  est un tuple de *variables non-distinguées* qui sont quantifiées existentiellement, et *conj* est le corps de la requête. Celui-ci est une conjonction d'atomes de la forme  $C(\tau)$  ou  $R(\tau_1, \tau_2)$ , où  $C$  est un concept,  $R$  est un rôle et  $\tau, \tau_1, \tau_2$  sont soit des individus, soit des variables. Une *requête ciblée*  $q_t$  est une requête conjonctive où tous les concepts, rôles et individus apparaissant dans les atomes sont écrits en termes de l'ontologie cible  $O_t$ .

---

1. Les autres ontologies du système sont appelées *ontologies extérieures*.

Une *réponse* à une requête ciblée  $q_t$  sur une base de connaissances distribuée  $S$  est un *assignement*, *i.e.*, une fonction des variables distinguées de  $q_t$  vers les termes des ontologies de  $S$ , telle qu'en substituant les valeurs assignées aux variables de  $q_t$ , les atomes résultants sont des conséquences logiques du système complet (Alkhateeb *et al.*, 2007).

### 3. Évaluer une requête sur une base de connaissances distribuée

Nous proposons deux méthodes d'évaluation de requêtes ciblées. La Sect. 3.1 présente une méthode étudiée pour le cas de DDL (Borgida *et al.*, 2002) dans (Alkhateeb *et al.*, 2007). La Sect. 3.2 présente une méthode de réécriture de requêtes, inspirée de (Calvanese *et al.*, 2005). La Sect. 3.3 présentent les avantages et inconvénients des deux méthodes ainsi que les bénéfices d'une méthode combinée.

#### 3.1. Étendre l'ontologie cible

Cette méthode consiste à étendre l'ontologie cible en y ajoutant (1) d'une part tous les termes des ontologies extérieures, et (2) d'autre part en complétant les axiomes avec ceux déduits des connaissances extérieures. Ce procédé, qui est détaillée dans le cas de DDL dans (Alkhateeb *et al.*, 2007), permet de transformer le problème d'évaluation de requête distribuée en un problème d'évaluation locale. Néanmoins, cette stratégie ne permet pas d'obtenir toutes les réponses distribuées dans le cas général.

#### 3.2. Réécriture de requêtes

Afin de répondre à des requêtes sur des systèmes distribués, il est possible de diffuser la requête vers les ontologies extérieures en la traduisant dans le vocabulaire de chacune. Plus précisément, on utilise les relations sémantiques entre l'ontologie cible et chaque ontologie extérieure  $O$  pour transformer la requête initiale en des requêtes posées en termes de  $O$ , dans le but d'obtenir des réponses distribuées en répondant localement aux requêtes transformées. C'est ainsi que fonctionne PEPSINT (Cruz *et al.*, 2004), SomeWhere (Goasdoue *et al.*, 2003) et Piazza (Halevy *et al.*, 2003), mais ceux-ci utilisent une sémantique distribuée restreinte qui n'est pas applicable dans le cas général (par exemple, DDL).

#### 3.3. Discussion et conclusion

Bien que les deux approches présentées ci-dessus permettent de trouver des réponses distribuées que l'on ne trouve pas par simple évaluation locale, elles ont chacune des avantages et des inconvénients :

D'une part, étendre l'ontologie cible donne en général un plus grand nombre de réponses. Cependant, c'est un processus lourd et qui peut dans certains cas ne pas terminer. Pour limiter ce problème, il est possible de retourner des réponses tandis que le processus d'extension est en cours. Toutefois, le passage à l'échelle est un facteur limitant pour cette méthode. En outre, si le contenu des ontologies extérieures n'est pas accessible (contenu privé ou caché), elle est inapplicable. D'autre part, bien que la réécriture donne moins de réponse, elle permet le passage à l'échelle puisque les outils d'évaluation de chaque ontologie peuvent être employés en parallèle. Par ailleurs, lorsque le contenu est inaccessible depuis la cible, il suffit de transmettre les requêtes transformées, de les évaluer en interne sans rompre les mesures de sécurité, pour obtenir les réponses d'une ontologie extérieure. Par contre, la qualité de cette méthode dépend beaucoup de la pertinence des relations sémantiques entre ontologies.

Pour tirer profit des avantages des deux approches, il est possible de les combiner, c-à-d diffuser les requêtes tout en étendant l'ontologie cible. Bien que certaines réponses peuvent être redondantes, cela permet de trouver des réponses qui ne sont pas obtenues par transformation, surtout si des heuristiques focalisent le raisonnement sur les connaissances relatives aux termes locaux. Aussi, lorsqu'on ne sait rien des raisonneurs locaux extérieurs, cette approche assure une meilleure maîtrise de l'évaluation.

Enfin, selon le formalisme de représentation de connaissances distribuées utilisé, ces méthodes offrent des avantages différents. Par exemple, le formalisme DDL ne permet pas la composition des relations sémantiques, ce qui implique la non-transitivité de la transformation de requêtes. Néanmoins, ces approches sont suffisamment génériques, et elles peuvent être appliquées à divers formalismes.

#### 4. Bibliographie

- Alkhateeb F., Zimmermann A., « Query Answering in Distributed Description Logics », *Proc. NTMS'2007*, LNCS, Springer, 2007. to appear.
- Borgida A., Serafini L., « Distributed Description Logics: Directed domain correspondences in federated information sources », *OTM'2002: Confederated International Conferences CoopIS, DOA, and ODBASE 2002*, vol. 2519 of LNCS, Springer, 2002.
- Calvanese D., De Giacomo G., Lembo D., Lenzerini M., Rosati R., « DL-Lite: Tractable description logics for ontologies », *Proceedings of the 20th National Conference on Artificial Intelligence (AAAI 2005)*, p. 602-607, 2005.
- Cruz I., Xiao H., Hsu F., « Peer-to-Peer Semantic Integration of XML and RDF Data Sources », *3rd International Workshop on Agents and P2P Computing (AP2PC 2004)*, July, 2004.
- Goasdoue F., Rousset M.-C., « Querying Distributed Data through Distributed Ontologies: A Simple but Scalable Approach », *IEEE Intelligent Systems*, vol. 18, n° 5, p. 60-65, 2003.
- Halevy A., Ives Z., Mork P., Tatarinov I., « Piazza: data management infrastructure for semantic web applications », *Proc. of WWW'2003*, ACM Press, p. 556-567, 2003.
- Nejdl W., Wolf B., Qu C., Decker S., Sintek M., Naeve A., Nilsson M., Palmer M., Risch T., « EDUTELLA: A P2P Networking Infrastructure Based on RDF », *Proc. of WWW'2002*, 2002.