

Synthèse de concepts formels par réécriture à partir d'une ontologie client

Fadhela Kerdjoudj^{*,**}, Olivier Curé^{*}

^{*} Université Paris-Est, LIGM, CNRS UMR 8049, France.

^{**} Société GeolSemantics, 32 rue Brancion 75015, Paris, France.
{fadhela.kerdjoudj,ocure}@univ-mlv.fr

Résumé. A l'issue de traitements linguistiques, les outils de la société GeolSemantics fournissent des descriptions d'entités analysées. Celles-ci peuvent être volumineuses et risquent d'introduire un bruit qu'un expert peut difficilement gérer. Nous proposons une solution de réécriture, à partir de traitements sur une ontologie du domaine et un de ces sous-ensembles. Nous présentons une syntaxe et une sémantique permettant de gérer notre approche de réécriture pour le langage de description $\mathcal{AL}\mathcal{E}\mathcal{H}\mathcal{O}$.

1 Introduction

La multiplication de sources textuelles plus ou moins structurées sur le Web nécessite de considérer leur caractère informationnel. Il convient alors de faire collaborer des approches linguistiques à des méthodes issues des traitements sémantiques pour apporter un cadre d'interprétation des concepts extraits. Dans ce contexte, la société GEOLSemantics a pour objectif de concevoir de nouveaux outils d'assistance aux tâches d'intelligence collective. Une extraction reposant sur une ontologie OWL est effectuée. Elle décrit les principaux concepts et rôles (Personne, Événement, Date...). Dès lors, il devient possible de générer une description des principales caractéristiques des entités décrites dans un ensemble de documents. Notre approche se base sur le fait qu'une réécriture doit nécessairement être adaptée aux exigences de l'expert. Il convient donc que celui-ci puisse spécifier ses points d'intérêts. Les profils sont représentés dans une logique de description. Un profil correspond donc à un concept.

2 Réécriture de concepts

2.1 Syntaxe

Notre méthode de *réécriture* de concepts se base sur la fonction suivante : $RE(O, C, U)$ où O et C sont respectivement une ontologie et un concept décrit dans le formalisme de $\mathcal{AL}\mathcal{E}\mathcal{H}\mathcal{O}$ comme le montre l'exemple suivant :

Exemple 1 Soit l'ontologie O suivante :

$H \sqsubseteq P, F \sqsubseteq P, H \sqsubseteq \neg F, BD \sqsubseteq I, Res \sqsubseteq I, Prog \sqsubseteq I, aFils \sqsubseteq aEnfant$

Nous définissons un concept C comme suit :

$C \equiv H \sqcap \exists enseigne. \top \sqcap \forall enseigne. (I \sqcap \neg Res) \sqcap \exists aFils. \{n\}$

Synthèse de concepts formels par réécriture

U spécifie les contraintes définies par le client et se présente de la manière suivante : $U = \langle U_{obl}, U_{opt} \rangle$. U_{obl} est la liste des concepts que l'utilisateur souhaite absolument retrouver dans la réécriture, alors que U_{opt} correspond à une liste optionnelle de concepts à inclure dans la réécriture.

2.2 Sémantique

Nous proposons maintenant une sémantique pour la fonction RE dans la définition 1.

Définition 1 Soit $RE(O, C, U)$, nous considérons que les axiomes de O et U ne contiennent que des concepts satisfaits dans $\mathcal{AL}\mathcal{E}\mathcal{H}\mathcal{O}$, que les termes U sont strictement inclus dans les termes de O . Nous définissons le concept C comme une conjonction de concepts de la forme $C \equiv C_1 \sqcap \dots \sqcap C_n$. Le résultat D de $RE(O, C, U)$ a les caractéristiques suivantes : (i) D est concept $\mathcal{AL}\mathcal{E}\mathcal{H}\mathcal{O}$, (ii) D ne comporte que des termes de U , (iii) D est une conjonction de concepts de la forme $D \equiv D_1 \sqcap \dots \sqcap D_m$ tel que pour chaque $1 \leq i \leq m$, il existe un j tel que $D_i \sqsupseteq C'_j$ avec $C'_j = \prod C_l$ pour $1 < l \leq n$ (où \prod représente un ensemble de conjonctions sur les concepts C_l).

L'exemple 2 caractérise cette sémantique dans le cadre de notre cas d'utilisation.

Exemple 2 Considérons $U_3 = \langle \{H, \exists enseigne.DB\}, \{\exists pratique.Tennis, \exists enseigne.\top\} \rangle$. Nous obtenons $D = RE(O, C, U_3) = H \sqcap \exists enseigne.DB$ qui satisfait bien l'identité $D_1 = H \sqsupseteq H = C_1$ et $D_2 = \exists enseigne.DB \sqsupseteq \exists enseigne.\top \sqcap \forall enseigne.(I \sqcap \neg Res) = C_2 \sqcap C_3$.

2.3 Simplification de U

Il est important de réaliser un certain nombre de contrôles sur les concepts décrits dans U_{obl} et U_{opt} pour garantir une réécriture efficace. Ces contrôles peuvent se répartir en catégories : structurels (e.g., détection de doublons...) et sémantiques (e.g., détection d'inconsistance...). De manière générale, les opérations de simplification de U s'inspirent fortement des optimisations des raisonneurs basés sur la méthode des tableaux.

3 Conclusion

Dans cet article, nous avons présenté un contexte pour la réécriture de concepts décrits en logique de description. La réécriture est établie à partir d'une spécification d'un expert du domaine qui correspond à deux listes de concepts. Une première liste obligatoire, qui stipule que si le concept à réécrire ne comporte pas l'un de ces concepts, le résultat de la réécriture sera vide. La seconde liste de concepts optionnels, c'est-à-dire qu'ils peuvent ou pas apparaître dans la réécriture suivant le contenu du concept à synthétiser. Nous sommes actuellement dans une phase de développement d'un prototype. Celui-ci nous permettra d'évaluer notre démarche sur des cas concrets. De manière plus théorique, nos perspectives sont d'étudier des langages plus expressifs qui intégreront des constructeurs comme la négation complète (et donc la disjonction de concepts) ainsi que les cardinalités.

Summary

Donner la traduction anglaise du résumé dans le préambule avec la commande `\summary{Your abstract ...}`