

Adaptation de noms de cocktails : saison 2

Sujet de projet tutoré, master SCA, M1, UFR Math-Info, Université de Lorraine

Jean Lieber (Jean.Lieber@loria.fr)

Emmanuel Nauer (Emmanuel.Nauer@loria.fr)

LORIA (CNRS, Université de Lorraine) / Inria Nancy Grand Est

année 2016-2017

Ce projet fait suite à un projet tutoré effectué l'an dernier par un binôme d'étudiants du master SCA. Ce projet a conduit à une publication dans un congrès international [1].

Préliminaire. L'abus d'alcool est dangereux pour la santé.

Le projet de l'an dernier avait pour objectif de « débroussailler » le problème de l'adaptation de cocktails et de proposer un modèle sur la base d'exemples que les étudiants ont collectés. Ce problème est le suivant :

Étant donné un cocktail C dont le nom est N et une modification de ce cocktail par substitution d'ingrédients σ , comment modifier le nom N en un nom pour le nouveau cocktail $\sigma(C)$?

Un modèle a été élaboré avec les étudiants, sous la forme d'un ensemble de stratégies d'adaptation (décrites principalement au niveau algorithmique).

À titre d'exemple, considérons l'*Irish coffee* et la substitution du whisky irlandais par de la tequila et du café par du chocolat chaud. En s'appuyant sur des connaissances du domaine ce modèle conduit à nommer *Mexican hot chocolate* le nouveau cocktail. Une stratégie qui permet de résoudre ce problème part d'une représentation de la *dépendance* entre la recette et son nom (le fait qu'elle contient du café et du whisky, qui est un ingrédient irlandais) et cherche à généraliser *a minima* cette dépendance en une dépendance pouvant s'instancier dans la recette adaptée. Cette généralisation *a minima* se fait par exploration d'un espace d'états (utilisant par exemple l'algorithme A*).

RDFS. Le modèle d'adaptation des cocktails s'appuie sur des technologies liées au formalisme de représentation des connaissances RDFS qui est une recommandation du W3C [2] et est un langage du Web sémantique. RDFS est une logique dont les formule sont des triplets tels que :

```
<whiskyIrlandais aPourPaysDeProduction Irlande> <Irlande adjectifAnglais "Irish">  
<tequila aPourPaysDeProduction Mexique> <Mexique adjectifAnglais "mexican">  
<whiskyIrlandais subClassOf BoissonAlcoolisée> <téquila subClassOf BoissonAlcoolisée>  
<BoissonAlcoolisée subClassOf Boisson>
```

Les inférences en RDFS sont définies généralement sur la base de règles d'inférences, par exemple :
 $\frac{\langle C \text{ subClassOf } D \rangle \quad \langle D \text{ subClassOf } E \rangle}{\langle C \text{ subClassOf } E \rangle}$, qui permet d'inférer le triplet $\langle \text{téquila subClassOf Boisson} \rangle$ des triplets précédents.

Pour interroger et manipuler des bases de connaissances écrites en RDFS, il existe des outils disponibles gratuitement en ligne, par exemple le système Corese/KGRAM [3].

Travail à réaliser. Pour rendre ce projet pleinement opérationnel, deux tâches doivent être menées à bien. La première fait l'objet de ce projet tutoré. La deuxième est optionnelle et ne sera abordée que quand la première tâche aura été achevée.

Cette première tâche est le développement d'outils implantant les stratégies d'adaptation des cocktails définies dans [1] en utilisant un outil tel que Corese/KGRAM.

La deuxième tâche vise à enrichir la base de connaissances en faisant appel aux données libres liées (*linked open data*, LOD), à savoir aux données publiées en RDFS sur le Web, contenant entre autres les triplets de DBPedia (qui sont extraits automatiquement de Wikipédia).

Échéancier. L'échéancier qui suit concerne la première tâche (celle qui n'est pas optionnelle) :

- Lecture de documents en relation avec le sujet (l'article [1], le rapport des étudiants de l'an dernier, des articles sur RDFS) ;
- Apprentissage des technologies liées à RDFS ;
- Apprentissage d'un outil de manipulation de bases RDFS ;
- Spécification d'un outil *générique* pour l'adaptation de bases RDFS, dont l'adaptation des cocktails soit une application ;
- Implantation et test de cet outil.

De façon générale, l'organisation du suivi du projet se fera de la façon suivante :

- Réunions régulières (réunion en début de projet puis, environ une fois toutes les deux semaines) ;
- Construction au fur et à mesure du rapport (comme outil de communication encadrants-étudiants et de préparation des réunions).

Ce travail, s'il se déroule dans de bonnes conditions, fera l'objet d'une soumission à un congrès international.

Compétences utiles. Les étudiants devront avoir des compétences en programmation (si c'est en Java, c'est encore mieux : la plupart des outils pour RDFS sont en Java). La connaissance de technologies liées au Web (p. ex., XML) serait un plus.

Pour en savoir plus. N'hésitez pas à nous contacter par courriel, ne serait-ce que pour trouver une date de rendez-vous téléphonique ou au Loria.

Références

- [1] N. KIANI, J. LIEBER, E. NAUER et J. SCHNEIDER : Analogical Transfer in RDFS, Application to Cocktail Name Adaptation. In A. K. Goel et AL., éditeur : *Case-Based Reasoning Research and Development, Proceedings of the 24th ICCBR*, Lecture Notes in Computer Science, Atlanta, Georgia, 2016. Springer International Publishing.
- [2] <https://www.w3.org/TR/rdf-schema/>. Dernière consultation : octobre 2016.
- [3] <http://wimmics.inria.fr/corese>. Dernière consultation : octobre 2016.