

Illustration d'un scénario d'assistance à la conception d'une ontologie à partir d'un modèle semi-formel

Michel Héon¹, Gilbert Paquette¹, Josianne Basque¹

¹ Centre de recherche LICEF, Télé-Université, Montréal, Canada
(michel.heon, gilbert.paquette, josianne.basque)@licef.ca

Résumé : Cet article présente une illustration du scénario d'assistance fournie par un système expert à un ingénieur de la connaissance lors d'un processus de transformation d'un modèle semi-formel en une ontologie de domaine. Trois étapes composent le scénario soit: 1) conception d'un modèle semi-formel, 2) formalisation, 3) validation. La progression du scénario est suivie pas à pas à l'aide d'un exemple permettant de démontrer l'applicabilité de la méthodologie proposée pour formaliser en ontologie OWL des connaissances déclaratives, procédurales et stratégiques et d'illustrer comment des erreurs d'ordre sémantique commises lors de la première étape peuvent être identifiées au cours des étapes de formalisation et de validation .

Mots-clés : Formalisation de connaissances, Ingénierie des connaissances, Ingénierie ontologique, Ontologie de transformation, Système expert, Assistance à la formalisation.

Introduction

Cet article vise à illustrer l'application de la méthodologie de formalisation instrumentée par un système expert d'assistance à la formalisation présentée dans Héon *et al.* (2009). Nous illustrons le déroulement de la méthodologie de formalisation par le biais d'un exemple fictif. La méthodologie se déploie en trois étapes itératives : (1) production du modèle semi-formel; (2) formalisation du modèle semi-formel en ontologie; (3) validation de l'ontologie. Cette dernière étape entraîne soit l'arrêt de la construction de l'ontologie, soit une nouvelle itération de conception. En plus de présenter l'application de la méthodologie, l'article démontre comment sont traitées non seulement les connaissances conceptuelles mais également les connaissances procédurales et stratégiques en vue de les formaliser dans l'ontologie. Quelques-uns des aspects fonctionnels de l'assistant à la formalisation sont aussi présentés.

Étape 1: Production d'un modèle semi-formel de connaissances

Cette étape a pour objectif d'élaborer un modèle de connaissances semi-formel. L'étape est pilotée par un ingénieur de la connaissance en interaction avec un ou plusieurs experts de contenu ou encore, s'il s'agit d'une situation qui vise également

le transfert d'expertise, avec des groupes impliquant des experts et des novices (Basque *et al.*, 2008). Cependant, pour fins d'illustration du scénario d'assistance proposé et de son pouvoir de détection des erreurs de nature sémantique, nous avons produit nous-mêmes un exemple de modèle semi-formel de connaissances (sur le thème des corps célestes), en y introduisant délibérément différents types d'erreurs ou d'imprécisions sémantiques. L'exemple apparaît à la **fig. 1**. Le modèle est créé dans le langage de Modélisation par Objets Typés (MOT), dont on trouve une description abrégée dans Héon *et al.*(2009) et détaillée dans Paquette (2007). Les erreurs de modélisation introduites dans le modèle sont inspirées du concept d'*antipattern* proposé par Allemang et Hendler (2008). Des étiquettes (identifiées par une lettre dans la figure) servent de guide pour le traitement de ces points dans le texte qui suit.

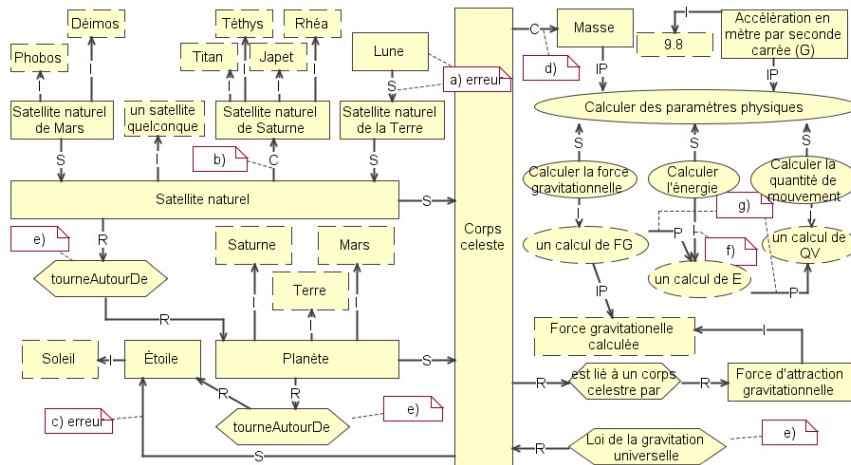


Fig. 1: Exemple d'un modèle de connaissances semi-formel

Voici en quelques lignes la pensée qui aurait pu conduire à une telle modélisation. *Le corps céleste est régi par le principe de la loi de la gravitation universelle. Il a pour attribut une masse et est lié à d'autres corps célestes par la force d'attraction gravitationnelle (FG). Cette force est calculée à partir de la masse et de l'accélération gravitationnelle (G) qui est de 9.8 m/s². D'autres paramètres physiques peuvent être calculés tels que l'énergie (E) et la quantité de mouvement (QV). Un calcul de la FG suivi d'un calcul de E et d'un calcul de QV sont des exemples spécifiques de calcul de paramètres physiques. Les étoiles, les planètes et les satellites naturels sont des corps célestes. Les planètes tournent autour des étoiles et les satellites naturels tournent autour des planètes. Saturne, Mars et la Terre sont des planètes, Déimos, Phobos sont des satellites naturels de Mars; Titan, Téthys, Japet et Rhéa sont des satellites naturels de Saturne. Finalement, la Lune est un satellite naturel de la Terre. Il existe aussi, quelque part, un satellite quelconque.*

Avant d'entreprendre plus avant la formalisation du modèle expérimental, nous désirons soulever quelques remarques concernant le modèle. L'étiquette **a** illustre un exemple d'erreur de modélisation couramment observée. Il s'agit d'une erreur de

Scénario de conception assistée d'une ontologie à partir d'un modèle semi-formel

classification du niveau d'abstraction des "choses" représentées. Il arrive qu'un modélisateur confonde la représentation d'un objet concret avec l'abstraction de l'objet. Il est vrai que le concept de *Lune* est une sorte de *Satellite naturel de la Terre*, mais, de manière formelle, ce qui est représenté ici devrait se lire *l'objet de type Lune est un exemple (une instance par le lien I) des "satellites naturels" de la Terre*. L'étiquette **b** illustre une autre erreur de représentation, soit la confusion dans l'utilisation du *LienS* (sorte de [is-a]) et du *LienC* (composé ou [part-of]). Nous pourrions être tentés d'exprimer dans le modèle la proposition que les *Satellites naturels* se composent de *Satellite naturel de Saturne* pour exprimer la relation entre ces deux concepts, alors qu'il s'agit plutôt d'une relation de spécialisation car les satellites naturels de Saturne forment un sous-ensemble de la classe de tous les satellites naturels. Nous verrons à l'étape de validation comment cette erreur pourra être relevée et présentée à l'expert. Finalement, une autre erreur commise est l'inversion de l'orientation du *LienS* (voir l'étiquette **c**). Nous en verrons les conséquences à l'étape de validation. Les étiquettes **d** à **g** soulignent, quant à elles, des points de précisions qui seront traités plus loin.

Étape 2: Formalisation

Cette étape est pilotée par l'ingénieur et sert à produire une version formelle — sous la forme d'une ontologie — du modèle semi-formel. La **fig. 2** présente l'interface de l'assistant à la formalisation (appelé « eLi ») utilisé par l'ingénieur et développé dans le cadre de nos travaux. Au centre de la moitié supérieure, on trouve le schéma du modèle semi-formel de la figure 1 alors qu'à gauche, la vue en graphe de composite en présente la taxonomie. La moitié inférieure droite présente un tableau de bord permettant à l'ingénieur de lancer les processus nécessaires à la formalisation

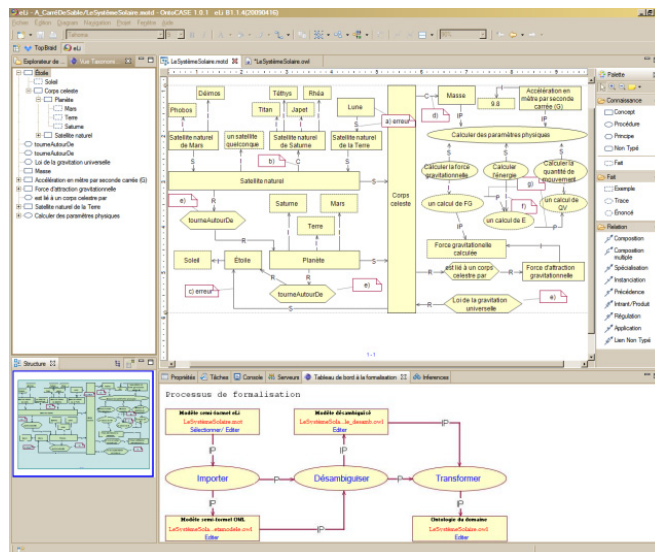


Fig. 2: Interface utilisateur d'eLi, l'outil d'assistance à la formalisation

(importer, désambiguïser et transformer) et d'éditer au besoin, chaque modèle après le déclenchement du processus.

Le processus automatique d'importation sert à traduire le modèle semi-formel dans le langage ontologique choisi sans le modifier, avec ses erreurs et ambiguïtés. Dans cet exemple, le format du modèle source est le *XML Metadata Interchange* (XMI) qui reproduit le modèle MOT de la figure 1 et le format cible est l'*Ontology Web Language* (OWL pour l'ontologie à construire).

Le processus semi-automatique de désambiguïstation sert à donner une interprétation unique aux entités semi-formelles du modèle. Certaines entités (comme le *LienI* et le *LienS*) sont univoques (signifiant qu'une entité semi-formelle correspond à une seule entité formelle). Il s'agit alors d'une désambiguïstation automatique de type *typologique*. D'autres ambiguïtés nécessitent un traitement automatique basé sur des patrons de modélisation de type *topologique*. Par exemple, les *Principes* indiqués par l'étiquette **e** subissent ce type de désambiguïstation puisqu'un principe entre deux *Concepts* est interprété comme une *owl:ObjectProperty* alors qu'un *Principe* lié à un *Concept* par un *LienR* sera interprété en tant que *owl:class* de catégorie *AGENT CONTRAINTE NORME*. Il en va de même pour les *LienIP* qui sont désambiguïsés en *owl:ObjectProperty* de catégorie *INTRANT* ou *PRODUIT* selon l'orientation des liens. Lorsqu'un objet ne peut être désambiguïté de façon typologique ou topologique, alors il doit l'être de façon manuelle selon la sémantique du domaine. Dans l'exemple, le *LienC* de l'étiquette **b** et celui de l'étiquette **d** n'ont pas la même signification. Dans le premier cas, il est interprété comme une relation d'agrégation (*la classe des satellites naturels se composent de la classe des satellites de la Terre*) alors que celui de l'étiquette **d** s'interprète comme une relation d'attribution (*le corps céleste a pour attribut une masse*).

La *transformation* en ontologie du domaine est le dernier processus de l'étape de formalisation. Étant donné que le modèle semi-formel comprend des connaissances procédurales et stratégiques, il est nécessaire de procéder à une transformation qui mènera à une représentation de ces connaissances selon un format déclaratif dans l'ontologie cible. Une ontologie contenant des méta-classes et des méta-propriétés de domaines, aussi appelée *ontologie des méta-connaissances de domaine* (Héon et al. 2009) est importée par l'ontologie du domaine. L'ontologie des méta-connaissances de domaine permet de regrouper les connaissances formalisées sous trois grandes catégories: *MD_Declarative*, *MD_Procedurale* et *MD_Strategique*. Les propriétés du domaine sont regroupées sous six groupes de méta-propriétés soit: *A-POUR-COMPOSANT*, *A-POUR-DÉPENDANCE*, *EST-LA-PARTIE-DE*, *INTRANT-PRODUIT*, *PERMET-DE*, *A-POUR-ATTRIBUT*; *REGIT*; et finalement *EST-REGI-PAR*. Certaines méta-propriétés sont les inverses les unes des autres; par exemple, *A-POUR-COMPOSANT* est une propriété inverse de *EST-LA-PARTIE-DE*. Nous verrons à l'étape de validation que cette attribution de propriété inverse sera fort utile pour l'interprétation des connaissances de l'ontologie.

Étape 3: Validation

Nous ne traiterons ici que de la validation sémantique (la validation syntaxique étant elle, traitée dans un autre article (Héon, Paquette et Basque, 2009)). L'idée

Scénario de conception assistée d'une ontologie à partir d'un modèle semi-formel

maîtresse de la validation sémantique est de comparer les conclusions inférées humainement (à partir du modèle semi-formel) avec les conclusions inférées automatiquement (à partir de l'ontologie du domaine). La première phase de la validation est l'analyse de l'ontologie. On retrouve à la figure 3 un portrait des classes, instances et propriétés qui composent l'ontologie du domaine. Dans la taxonomie de classes, le *Corps-céleste* est subsumé par *Etoile*, ce qui voudrait dire que la *Terre*, qui est une sorte de *Planète*, serait considérée comme une *Étoile*. Cette erreur de classification provient de l'inversion de la direction du *LienS* (étiquette c de la figure 1). Selon le cas de figure représenté par l'étiquette b, on pourrait interpréter qu'un *satellite quelconque* se compose d'un *satellite naturel de Saturne* tel que *Titan*. Or, une inférence automatique nous indique (selon ce sous-modèle) que *Titan EST-LA-PARTIE-DE un satellite quelconque* alors qu'il est plus exact de dire que *Titan et un satellite quelconque SONT* tous les deux des *satellites naturels*. La formalisation du cas de figure présenté à l'étiquette a, indique que la classe *Lune* est une sorte de

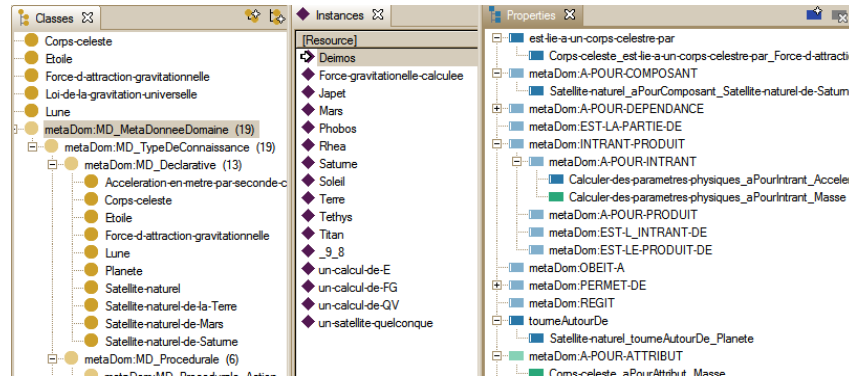


Fig. 3: Classes, instances et propriétés composant l'ontologie du domaine

satellite naturel de la Terre. Mécaniquement, nous pourrions instancier des individus de type *Lune*, ce qui, en termes représentationnels, est faux puisque la lune est un objet et non une classe d'objets. L'étiquette d présente le cas où un *LienC* est formalisé en tant qu'attribut (voir la propriété *Corps-celeste_aPourAttribut_Masse* de catégorie *A-POUR-ATTRIBUT*), ce qui correspond bien à la sémantique initiale. L'étiquette e présente deux cas d'utilisation du *Principe*. Le premier des cas, *Loi de la gravitation universelle*, est désambiguïté et considéré en tant que *owl:class* de catégorie *Agent_Contrainte_Norme*, alors que le deuxième cas, *tourneAutourDe*, est formalisé en *owl:ObjectProperty* dont le domaine et l'image correspondent à la classe source et cible du *Principe*. L'étiquette f indique que les entités *Traces: un calcul de E* est une instance de la *Procédure: calculer l'énergie* qui elle-même est une composante de la *Procédure: calculer des paramètres physiques*. En ce qui concerne l'Exemple: "9.8", le lecteur notera qu'il s'agit d'une valeur de l'attribut G (un *Concept*) et que ce n'est pas une connaissance procédurale même si elle est utilisée par la procédure de *calcul des paramètres physiques*. L'étiquette g présente un cas de figure fort intéressant puisqu'il concerne le traitement ontologique de connaissances procédurales. Avant l'inférence, l'ontologie nous indique que *un calcul de QV EST-*

LE-SUIVANT-DE un calcul de E (voir la figure 4, ligne A). Grâce aux méta-propriétés définies par l'ontologie des méta-connaissances de domaines, le moteur d'inférence permet de conclure (voir la figure 4, ligne B) que *un calcul de QV A-POUR-DÉPENDANCE* l'exécution des étapes *un calcul de FG* et *un calcul de E*.

| | |
|----|---|
| A) | : un-calcul-de-QV a metaDom:MD_Procedurale_Procedure , :Calculer-la-quantite-de-mouvement ; metaDom:EST-LE-SUIVANT-DE :un-calcul-de-E . |
| B) | :un-calcul-de-QV a metaDom:MD_Procedurale_Procedure , :Calculer-la-quantite-de-mouvement ; metaDom:A-POUR-DEPENDANCE :un-calcul-de-E , :un-calcul-de-FG ; metaDom:EST-LE-SUIVANT-DE :un-calcul-de-E . |

Fig. 4: Représentation en OWL-N3 de *calculer l'énergie* : A) avant et B) après inférence.

C'est grâce à la comparaison entre les conclusions humaine et automatique que le processus est soit repris à partir de l'étape 1, soit considéré comme terminé parce que l'ontologie produite est jugée satisfaisante

Conclusion

Par cet article, nous avons voulu illustrer l'utilisation de notre méthodologie de formalisation d'une ontologie du domaine à partir d'un modèle semi-formel. Nous avons soumis cette méthodologie à la formalisation d'un modèle semi-formel fictif contenant plusieurs erreurs ou imprécisions sémantiques. Plus spécifiquement, nous nous sommes attardés à démontrer comment les connaissances procédurales et stratégiques pouvaient être représentées dans une ontologie. Nous avons aussi présenté comment l'assistant à la formalisation pouvait être utilisé pour, d'une part, formaliser le modèle de connaissances et, d'autre part, assister l'ingénieur et l'expert dans leur tâche de validation de la sémantique du domaine. Pour l'heure, la méthodologie et l'assistant à la formalisation sont à l'état de prototype. Nos travaux actuels sont concentrés sur la finalisation de la formulation des règles de désambiguïsation et de formalisation, plus spécifiquement celles concernant les opérateurs de négation et "inverse de", la représentation de règles d'implication ou encore les quantificateurs universel et existentiel. Dans une autre étape, nous envisageons la mise à l'échelle de la méthode par des expérimentations en laboratoire puis un déploiement en entreprise.

Bibliographie

- ALLEMANG D. & HENDLER, J. (2008). *Semantic Web for the Working Ontologist: Effective Modeling in RDFS and OWL*. Coll. «Semantic Web / Web programming». Burlington MA: Morgan Kaufmann, 330 p.
- BASQUE J., PAQUETTE, G., PUDELKO, B. & LÉONARD, M. (2008). Collaborative Knowledge Modeling with a Graphical Knowledge Representation Tool: A Strategy to Support the Transfer of Expertise in Organizations. In, Alexandra Okada, Simon Buckingham Shum et Tony Sherborne Eds. *Knowledge Cartography. Mapping Techniques and Software Tools*. London: Springer-Verlag.
- HÉON M., PAQUETTE, G. & BASQUE, J. (2009). «Conception assistée, d'une ontologie à partir d'une conceptualisation consensuelle exprimée de manière semi-formelle». In *Conférence d'Ingénierie des connaissances* (25 au 29 mai 2009).
- PAQUETTE G. (2007). «Graphical Ontology Modeling Language for Learning Environments». *Technology, Instruction., Cognition and Learning* vol. 5, p. 133-168.

Scénario de conception assistée d'une ontologie à partir d'un modèle semi-formel