

Heuristiques pour l'adaptation des mappings entre ontologies dynamiques

Julio Cesar Dos Reis^{*,**}, Cédric Pruski^{*}, Chantal Reynaud-Delaître^{**}

^{*}Centre de Recherche Public Henri Tudor,
29, avenue John F. Kennedy L-1855 Luxembourg-Kirchberg
{julio.dosreis,cedric.pruski}@tudor.lu,

^{**}Université Paris-Sud, Laboratoire de Recherche en Informatique
Bât 650 Ada Lovelace, Université Paris Sud, 91405 Orsay Cedex France
chantal.reynaud@lri.fr

Résumé. Les correspondances sémantiques entre ontologies (mappings) jouent un rôle essentiel dans les systèmes d'information. Cependant, en vertu de l'évolution des connaissances, les éléments ontologiques sont sujets à modification invalidant potentiellement les alignements préalablement établis. Des techniques de maintenance sont donc nécessaires pour maintenir la validité des mappings. Dans cet article, nous présentons un ensemble d'heuristiques guidant leur adaptation. Notre approche s'appuie sur l'explication des mappings existants, les informations provenant de l'évolution des ontologies ainsi que les adaptations possibles applicables aux mappings. Nous proposons une validation expérimentale à partir d'ontologies du domaine médical et des mappings qui leur sont associés.

1 Introduction

L'utilisation d'ontologies s'est montrée très efficace dans bien des domaines. La taille et la dynamique des domaines considérés demande toutefois l'exploitation combinée de plusieurs ontologies, d'où la nécessité d'établir des correspondances sémantiques, ou mappings (Euzenat et Shvaiko, 2007), entre ontologies. Ainsi, la qualité des résultats produits par les systèmes utilisant des ontologies dépend de la validité des mappings entre ontologies, ce qui oblige des experts du domaine à réviser leur définition lorsque les ontologies évoluent. Si cette maintenance peut être effectuée manuellement sur de petits ensembles de mappings, une approche plus automatique est nécessaire lorsqu'ils sont volumineux, comme dans le domaine médical.

L'existence de mappings erronés est souvent due à l'évolution des ontologies, les erreurs d'alignement mises à part (Dos Reis et al., 2014). Il est alors fondamental de comprendre l'évolution des ontologies pour pouvoir agir sur les mappings afin de garantir leur validité. Ce faisant, nous avons proposé un ensemble de patrons de changement permettant de caractériser l'évolution des concepts d'une ontologie en analysant les changements dans la définition des concepts (Dinh et al., 2014). Nous avons également observé sur des jeux de données réelles le comportement des mappings dans le temps, ce qui nous a permis d'identifier un ensemble d'actions d'adaptation pouvant s'appliquer aux mappings pour les faire évoluer (Dos Reis et al.,

2013). L'objet de cet article formalise, sous forme d'heuristiques, le lien entre les patrons de changement et les actions d'adaptation pour faire évoluer les mappings lorsque les ontologies liées évoluent. Après avoir introduit les concepts de notre approche (Section 2), en particulier les patrons de changement et les actions d'adaptation, nous présentons les heuristiques proposées (Section 3) et le cadre expérimental emprunté au domaine biomédical pour les évaluer (Section 4) et les discuter par rapport à l'existant (Section 5) avant de conclure (Section 6).

2 Préliminaires

Nous présentons ici les notions et notations utilisées pour définir notre approche. Soit $O_X^j = (C_X, R_X, A_X)$ et $O_Y^j = (C_Y, R_Y, A_Y)$ deux ontologies distinctes à l'instant j , avec $j \in \mathbb{N}$, où C_X et C_Y sont deux ensembles de concepts, R_X et R_Y sont deux ensembles de relations entre concepts, A_X et A_Y sont deux ensembles d'attributs caractérisant les concepts (correspondant à des propriétés d'annotations dans le langage OWL, par exemple le label, les termes dénotant le concept). Un mapping $m_{st}^j = (c_s^j, c_t^j, semType)$ relie deux concepts $c_s \in C_X$ et $c_t \in C_Y$ par la relation sémantique $semType$, telle que $semType \in \{\equiv, <, >, \approx\}$ (cf. tableau 1).

TABLE 1 – Notations pour la formalisation des heuristiques

Notation	Description
$C(O_X^j)$	Ensemble des concepts de l'ontologie O_X au temps j
$CT(c_k^j)$	Contexte du concept c_k au temps j comprenant l'ensemble des super, sous, et concepts frères de c_k
$A(c_k^j)$	Ensemble des valeurs des attributs $\{a_1, a_2, \dots, a_p\}$ d'un concept c_k au temps j (e.g., label, termes)
$A_{ct}(c_k^j)$	Ensemble des attributs de tout le contexte de c_k^j
$topA(c_s, c_t, n)$	Les n attributs expliquant le mapping m_{st}
$sim(a_i, a_e)$	Valeur de la similarité entre les valeurs d'attributs a_i et a_e

2.1 Le problème de l'adaptation des mappings

Considérons un ensemble de mappings M_{XY}^j entre O_X^j et O_Y^j , ainsi que la nouvelle version de O_X^{j+1} et O_Y^{j+1} au temps $j + 1$. L'évolution des ontologies remet en cause les mappings M_{XY}^j dans la mesure où les changements apportés à c_s^j ou c_t^j peuvent modifier leur sens et par conséquent celui de $semType$ reliant c_s^j à c_t^j . Dans ce contexte, l'adaptation des mappings fait référence au problème de la modification des composants d'un mapping (i.e. $semType$, c_s ou c_t) en fonction des changements menant O_X^j et/ou O_Y^j à leur nouvelle version respective.

2.2 Patrons de changement caractérisant l'évolution d'ontologies

Nos travaux ont montré que l'évolution des ontologies rend nécessaire d'adapter les mappings. Nous pensons qu'une compréhension précise de cette évolution va nous renseigner sur la

façon d'adapter les mappings au cours du temps. Pour caractériser l'évolution des ontologies, nous avons proposé un ensemble de patrons de changement (Dinh et al., 2014). Contrairement à ceux de la littérature (Djedidi et Aufaure, 2009), (Javed et al., 2013), (Gröner et al., 2010), nos patrons considèrent les changements syntaxiques et sémantiques au niveau des attributs des concepts. Ce choix a été motivé par les résultats expérimentaux obtenus montrant que la définition des mappings repose sur certaines valeurs d'attributs (Dos Reis et al., 2013).

Les Patrons de changement lexicaux (LCP) décrivent les changements lexicaux pouvant affecter la valeur d'un attribut d'une version à la suivante par rapport à un concept et son voisinage. Sur la base de nos observations, nous avons défini 4 types de LCP : *Copie Totale (TC)*, *Copie Partielle (PC)*, *Transfert Total (TT)* et *Transfert Partiel (PT)*.

Une *Copie Totale* caractérise le changement à travers lequel une valeur d'attribut devient également la valeur d'un attribut d'un autre concept. Par exemple, un attribut a_1 d'un concept c_1 a pour valeur "portal systemic encephalopathy" au temps j . Au temps $j + 1$, a_1 a la même valeur, mais un attribut a_2 d'un concept c_2 aura également cette valeur.

Une *Copie Partielle* est une copie d'une partie de la valeur d'un attribut. Un attribut a_1 d'un concept c_1 a la valeur "familial hyperchylomicromenia" au temps j . Au temps $j + 1$, a_1 garde cette valeur, mais un attribut a_2 aura "familial chylomicromenia" comme nouvelle valeur.

Un *Transfert Total* correspond au transfert de la totalité de la valeur d'un attribut à un autre attribut. Contrairement au cas TC, la valeur originale de l'attribut n'est pas conservée.

Un *Transfert Partiel* définit le transfert d'une partie de la valeur d'un attribut. Par exemple, un attribut a_1 peut valoir "eye swelling" au temps j . Au temps $j + 1$, cette valeur sera supprimée partiellement de a_1 mais un attribut a_2 vaudra "head swelling" (*i.e.*, le terme "swelling" est déplacé de a_1 vers a_2 entre j et $j + 1$).

Les Patrons de changement sémantiques (SCP) s'intéressent à l'évolution de la sémantique de la valeur des attributs au cours du temps. Nos observations ont montré qu'à travers leurs évolutions successives, les concepts pouvaient devenir plus généraux, plus spécifiques ou rester équivalents, modifiant ainsi la relation sémantique des mappings. Les 4 SCP que nous proposons sont : *Equivalent (EQV)*, *Plus Spécifique (MSP)*, *Moins Spécifique (LSP)* et *Recouvrement Partiel (PTM)*.

Equivalent stipule que les changements syntaxiques au niveau de la valeur de l'attribut ne modifient pas sa sémantique. Par exemple, la valeur d'un attribut peut passer de "Diabetes type I" à "Diabetes type I" sans en affecter le sens.

Plus Spécifique identifie un changement rendant un concept plus spécifique que sa nouvelle version. Le changement menant de "kappa light chain disease" à "kappa chain disease" rend le premier plus spécifique du fait de la suppression du qualificatif "light". *Moins Spécifique* décrit l'effet inverse.

Recouvrement Partiel identifie un changement au niveau de la sémantique ne pouvant être caractérisé par les autres SCP. Considérons la valeur originale "focal atelectasis" et son évolution "helical atelectasis". Ces deux valeurs font toutes deux référence à la notion de "atelectasis", mais ne peuvent être déclarées ni équivalentes, ni plus ou moins spécifiques.

2.3 Les actions d'adaptation de mappings

Les actions d'adaptation de mappings (MAAs) sont des traitements appliqués aux mappings afin d'obtenir la version leur correspondant au temps $j + 1$. Comme précédemment,

Heuristiques pour l'adaptation de mappings

nous avons défini ces MAAs sur la base d'une analyse empirique de l'évolution d'un corpus de mappings officiels entre ontologies du domaine biomédical (Dos Reis et al., 2013). On distingue six MAAs : *AdditionM*, *RemoveM*, *MoveM*, *DeriveM*, *ModSemTypeM* et *NoAction*.

AdditionM(m_{st}) et *RemoveM*(m_{st}) sont des actions atomiques ajoutant ou supprimant un mapping d'une version de mapping M_{XY}^j à la suivante. Ces deux actions sont des composants des autres actions dites complexes.

MoveM(m_{st} , c_{cand}^1) est observée lorsqu'un concept source c_s^0 d'un mapping est remplacé par un autre c_{cand}^1 du fait des changements affectant c_s^0 .

DeriveM(m_{st} , c_{cand}^1) diffère de *MoveM* par le fait que le mapping original est conservé, mais qu'un nouveau mapping est créé avec c_{cand}^1 différent de c_s^0 , la même relation sémantique *semType* et le même concept cible c_t .

ModSemTypeM(m_{st} , *semType*) consiste en la modification de *semType* (la relation sémantique) à cause des modifications sur les concepts sources (c_s) et/ou cibles (c_t).

NoAction(m_{st}) est appliquée lorsque les modifications sur les concepts sources et/ou cibles n'entraînent pas de changements sur la sémantique des mappings.

3 Heuristiques d'adaptation des mappings

Nous présentons dans cette section des heuristiques indiquant sous quelles conditions adapter les mappings afin qu'ils restent valides dans le temps. Ces heuristiques ont été établies expérimentalement à partir de l'observation de l'évolution des mappings entre des jeux de données réelles et de l'analyse de l'impact des patrons de changements sur leur évolution. Des heuristiques ont été définies pour chaque type d'adaptation.

3.1 Heuristiques pour les mappings de type Move et Derive

Soit *Cand* l'ensemble des concepts dits candidats regroupant les concepts du contexte de c_s en $j + 1$ pour lesquels il existe un changement de type LCP entre un de leurs attributs et un attribut de c_s . Soit la fonction *topA*(c_s , c_t , n) retournant les n attributs expliquant le mieux le mapping entre c_s et c_t . Soit la fonction *SLCP*(a_1 , a_2) retournant VRAI s'il existe un changement de type LCP entre a_1 et a_2 et FAUX sinon. Soit la fonction *SCP*(a_1 , a_2) retournant les changements de type SCP entre a_1 et a_2 ou \emptyset si aucun changement de type SCP n'a été identifié.

$$Cand = \left\{ c_{cand}^1 \mid \begin{array}{l} \exists c_{cand}^1 \in CT(c_s^1), \\ \exists a_q^1 \in A(c_{cand}^1), \\ \forall a_i \in topA(c_s^0, c_t^0, n), \forall a_z^1 \in A(c_s^1) | LCP(a_i, a_q^1), \\ SCP(a_i, a_z^1) = \emptyset \end{array} \right\}$$

MoveM. Expérimentalement, nous avons observé que l'adaptation d'un mapping de type *MoveM* correspondait à l'existence de changements de type LCP entre attributs. Plus précisément, nous avons observé que lorsque l'adaptation du mapping est de type *MoveM*, il n'existe qu'un seul attribut de c_s au temps j expliquant le mapping avec un changement de type LCP avec un attribut du contexte de c_s au temps $j + 1$. De ce fait, nous appliquons l'adaptation de type *MoveM* lorsqu'il n'existe qu'un seul concept candidat avec un changement de type LCP

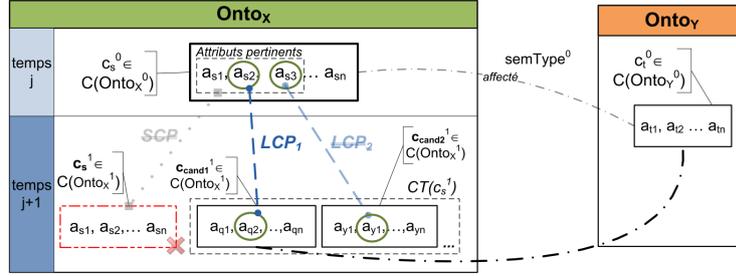


FIG. 1 – Formalisation de l’heuristique associée aux actions d’adaptation de type *MoveM*

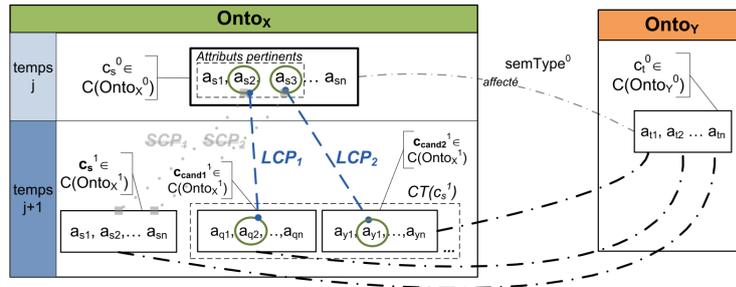


FIG. 2 – Formalisation de l’heuristique associée aux actions d’adaptation de type *DeriveM*

et pas de changements de type SCP avec c_s^1 (cf. Figure 1). Intuitivement, cela signifie que le mapping suit l’évolution des attributs qui l’expliquent.

$$\left. \begin{array}{l} \|Cand\| = 1, \\ \forall a_i \in topA(c_s^0, c_t^0, n), \\ \forall a_z^1 \in A(c_s^1), SCP(a_i, a_z^1) = \emptyset \end{array} \right\} \implies c_{cand}^1 \in Cand, MoveM(m_{st}, c_{cand}^1) \quad (1)$$

Soit les concepts ‘128829008’ “Acute myeloid leukemia, 11q23 abnormalities (morphologic abnormality)” et ‘C6924’ “Acute Myeloid Leukemia with 11q23 MLL Abnormalities” issus de SNOMED CT et de NCI. Deux changements de type LCP se sont produits lors de l’évolution de ‘C6924’. Un transfert total s’est produit sur l’attribut qui expliquait le mieux le mapping (“Acute Myeloid Leukemia with 11q23 Abnormalities”). Une copie totale a eu lieu pour un autre attribut qui expliquait aussi le mapping. Ces deux changements concernent le même concept candidat ‘C82403’ “Acute Myeloid Leukemia with t 9 11 p22 q23 MLL3-MLL” dans la nouvelle version de l’ontologie, mais deux attributs différents expliquant le mapping. Suite à l’évolution de ‘C6924’, le mapping relie dorénavant les concepts ‘128829008’ et ‘C82403’.

DeriveM. De façon similaire à l’action *MoveM*, *DeriveM* est appliquée lorsque plusieurs changements entre attributs de type LCP sont reconnus suite à une évolution. Le mapping

Heuristiques pour l'adaptation de mappings

original est préservé et c_s existe dans la nouvelle version de l'ontologie. Le fait que plusieurs changements de type LCP concernent les attributs pertinents de c_s et qu'il existe donc plusieurs concepts candidats conduit à la création de nouveaux mappings entre ces candidats et c_t (cf. Figure 2). Formellement, nous définissons cette heuristique de la façon suivante :

$$\left. \begin{array}{l} \|Cand\| > 1, \\ c_s^1 \in C(Onto_X^1) \end{array} \right\} \implies \forall c_{cand}^1 \in Cand, DeriveM(m_{st}, c_{cand}^1) \quad (2)$$

En guise d'illustration, considérons le mapping '*plus spécifique que* (<) entre '41452004' – "Uterus acollis (disorder)" dans SNOMED CT et '752.3' – "Other anomalies of uterus" dans ICD-9-CM. L'analyse de l'évolution de 752.3 permet d'observer plusieurs changements de type LCP concernant des concepts candidats différents et également plusieurs changements de type SCP. Par exemple, il y a une copie totale (TC) de l'attribut "Other anomalies of uterus" expliquant le mapping vers un attribut de '752.33' qui fait apparaître une équivalence. En plus, l'attribut "Bicornuate uterus" est totalement transféré (TT) dans '752.34' faisant également apparaître une relation d'équivalence. L'action d'adaptation à réaliser, selon notre heuristique, est *DeriveM* en l'appliquant aux deux concepts candidats concernés par les changements de type LCP et en gardant la même relation sémantique que la relation du mapping original.

3.2 Heuristiques associées à la modification de relation sémantique

L'application de l'action *ModSTR* dépend des changements de types SCP trouvés entre attributs. Nous avons identifié deux scénarios différents. Le premier concerne la modification de la relation du mapping original m_{st}^0 alors que c_s ne change pas (cf. Equation 3). Le second scénario concerne la modification de la relation du mapping original suite à un *MoveM* ou *DeriveM* (cf. Equation 4). Dans le premier cas, la nouvelle relation sémantique lie le concept source (en terme de contenu) au temps $j + 1$ et le concept cible alors que dans le second cas, c_s est remplacé par un concept candidat (un concept appartenant au contexte de c_s). Le type de la relation sémantique après évolution est obtenu en combinant le type de la relation du mapping original *semType* et le type d'un changement de type SCP détecté entre attributs. Soit la fonction *getSemType* qui fournit la relation sémantique obtenue en combinant le type d'un mapping original *semType* (x) avec des relations identifiées par des patrons de type SCP, SCPs (y). Par exemple, si la relation sémantique entre c_s^0 et c_t^0 est une relation d'équivalence, et qu'un changement *Less Specific* (>) est détecté entre un attribut de c_s^0 et un autre attribut de c_s^1 , la relation sémantique résultant de la combinaison sera *More Specific* (<).

$$getSemT : \{ '=, '<', '>', '\approx' \} \times SCP \longrightarrow \{ '=, '<', '>', '\approx' \}$$

$$(x, y) \longrightarrow \begin{cases} '= & \text{si } x = '= \wedge y = '= \\ '<' & \text{si } (x = '<' \wedge y = '>') \\ & \vee (x = '>' \wedge y = '=') \\ & \vee (x = '= \wedge y = '>') \\ '>' & \text{si } (x = '>' \wedge y = '<') \\ & \vee (x = '<' \wedge y = '=') \\ & \vee (x = '= \wedge y = '<') \\ '\approx' & \text{sinon} \end{cases}$$

Voici les heuristiques associées au premier et deuxième scénario :

$$\left. \begin{array}{l} \exists a_i \in \text{top}A(c_s^0, c_t^0, n), \exists c_s^1 \in C(O_X^1), \\ \exists m_{st}^1 \in \mathcal{M}_{ST}^1, \\ \exists a_z^1 \in A(c_s^1), SCP(a_i, a_z^1) \neq \emptyset, \\ \exists \text{sem}T_{st}^1 \in \{\equiv, <, >, \approx\}, \\ \text{sem}T_{st}^1 \in \text{getSem}T(\text{sem}T_{st}^0, SCP(a_i, a_z^1)), \\ \text{sem}T_{st}^0 \neq \text{sem}T_{st}^1 \end{array} \right\} \implies \text{ModSTR}(m_{st}, \text{sem}T_{st}^1) \quad (3)$$

$$\left. \begin{array}{l} \exists a_i \in \text{top}A(c_s^0, c_t^0, n), \exists c_{ct}^1 \in CT(c_s^1), \\ \exists m_{ct} \in \mathcal{M}_{ST}^1, \\ \exists a_q^1 \in A(c_{ct}^1), SCP(a_i, a_q^1) \neq \emptyset, \\ \exists \text{sem}T_{ct}^1 \in \{\equiv, <, >, \approx\}, \\ \text{sem}T_{ct}^1 \in \text{getSem}T(\text{sem}T_{st}^0, SCP(a_i, a_q^1)), \\ \text{sem}T_{st}^0 \neq \text{sem}T_{ct}^1 \end{array} \right\} \implies \text{ModSTR}(m_{ct}, \text{sem}T_{ct}^1) \quad (4)$$

Considérons le mapping : “268838005” – “*Neonatal dacryocystitis or conjunctivitis due to Escherichia coli (disorder)*” de SNOMED CT plus spécifique que (<) ‘041.4’ – “*Escherichia coli [E.coli]*” de ICD-9-CM. L’attribut “*Escherichia coli [E.coli]*” a été entièrement copié (TC) dans “*Escherichia coli [E. coli] NOS*”, qui est un nouvel attribut du concept ‘041.49’ appartenant au contexte de 041.4. Un changement de type SCP (>) a également été détecté, ce qui, selon nos heuristiques, nous conduit à proposer les actions *MoveM* et *ModSTR*.

3.3 Heuristiques associées à l’action *Removal* et à l’absence d’actions

Mapping *Removal*. L’action *RemoveM* est appliquée à un mapping lorsqu’aucun patron de changement ne s’applique entre les attributs expliquant le mapping et un des concepts du contexte de c_s au temps $j + 1$. L’équation 5 est la formalisation de cette heuristique. Dans cette équation, $A_{ct}(c_s^1)$ correspond à l’ensemble des attributs des concepts du contexte de c_s au temps $j + 1$. On voit aussi qu’il n’existe aucun patron de type SCP entre les attributs expliquant le mapping et c_s au temps $j + 1$ (quand il existe toujours). Quand $c_s^1 \notin C(\text{Onto}_X^1)$ ou que c_s devient obsolète, nous supprimons tous les attributs de c_s . Lorsque $c_s^1 \notin C(\text{Onto}_X^1)$, nous sélectionnons les concepts candidats au temps $j + 1$ à partir du contexte de c_s au temps j . Les expérimentations à partir desquelles cette heuristique a été proposée ont montré que la similarité entre les attributs expliquant un mapping et les attributs des concepts du contexte de c_s après évolution était très faible lorsque le mapping était supprimé (Dos Reis et al., 2014). Ceci a conduit à introduire une condition portant sur la similarité dans l’heuristique proposée.

$$\left. \begin{array}{l} \forall a_i \in \text{top}A(c_s^0, c_t^0, n), \forall a_q^1 \in A_{ct}(c_s^1), \neg LCP(a_i, a_q^1), \\ SCP(a_i, a_q^1) = \emptyset, \\ \exists \alpha \in \mathbb{R}_{>0}, \text{sim}(a_i, a_q^1) \leq \alpha, \\ \forall a_z^1 \in A(c_s^1), SCP(a_i, a_z^1) = \emptyset, \\ \exists a_i \in \text{top}A(c_s^0, c_t^0, 1), a_i \notin A(c_s^1) \end{array} \right\} \implies \text{RemoveM}(m_{st}) \quad (5)$$

No action. L’heuristique pour *NoAction* considère que des patrons de type LCP et SCP ne sont pas observables (cf. Equation 6). Elle s’applique dans des cas où les changements portant

Heuristiques pour l'adaptation de mappings

sur un concept lié par un mapping n'ont pas d'effet sur les attributs expliquant ce mapping, ou lorsque la similarité avec de nouveaux attributs du contexte est faible (Dos Reis et al., 2014).

$$\left. \begin{array}{l} \forall a_i \in \text{top}A(c_s^0, c_t^0, n), \forall a_q \in A_{ct}(c_s^1), \neg LCP(a_i, a_q^1), \\ SCP(a_i, a_q^1) = \emptyset, \\ \exists c_s^1 \in C(\text{Onto}_X^1), \forall a_z^1 \in A(c_s^1), SCP(a_i, a_z^1) = \emptyset, \\ \left[\begin{array}{l} a_i \in A(c_s^1) \\ \vee \\ \exists \alpha \in \mathbb{R}_{>0}, \text{sim}(a_i, a_q^1) \leq \alpha \end{array} \right. \end{array} \right\} \implies \text{NoAction}(m_{st}) \quad (6)$$

4 Evaluation

Cette section porte sur l'évaluation des heuristiques proposées. Nous étudions si les changements identifiés dans une ontologie conduisent à des adaptations correctes (i.e. MAAs) des mappings affectés. Cette évaluation a été réalisée sur plusieurs versions d'ontologies biomédicales (SNOMED CT et ICD-9-CM) et plusieurs versions de mappings associés.

4.1 Protocole d'expérimentation

Nous avons comparé les mappings adaptés par application de nos heuristiques aux mappings de référence dont nous disposons (la version des mappings correspondant à l'ontologie après évolution). Pour cela, nous avons défini une protocole expérimental en cinq étapes :

1. Nous avons calculé le *diff* entre deux versions successives d'une ontologie en utilisant des outils existants (Hartung et al., 2013). Nous avons alors pu identifier les mappings affectés par l'évolution de l'ontologie (uniquement ceux affectés par l'évolution de c_s ou de c_t). Nous avons également recherché les actions d'adaptation qui avaient été faites sur les mappings modifiés. Ces actions correspondent à celles qui sont attendues.
2. Nous avons adapté les mappings en appliquant nos heuristiques. Le seuil de similarité entre attributs α a été fixé à 0.7. La similarité a été calculée en utilisant la méthode n-gram. Cette étape crée l'ensemble des actions d'adaptation des mappings proposées.
3. Nous comparons les ensembles de mappings attendus et l'ensemble des mappings proposés et calculons, pour chaque type d'action, les mesures standards d'évaluation qui sont la précision, le rappel et la F-mesure.
4. Deux types de mesure, représentées respectivement par \blacktriangle and ∇ , sont proposées pour évaluer de manière rigoureuse les actions proposées (cf. Résultats obtenus dans le tableau 2). Cette distinction est bien appropriée à l'évaluation des actions *MoveM*, *DeriveM* and *ModSTR*. Le symbole \blacktriangle exprime la précision, le rappel et la F-mesure lorsque les actions proposées sont correctes par rapport aux actions attendues, en plus du concept candidat ou de la relation sémantique. Le symbole ∇ correspond aux cas où seul le type d'action est correct, le concept candidat ou la relation sont erronés (par exemple, *MoveM* ou *DeriveM* est proposé mais pas avec le bon concept candidat à j+1). La mesure \blacktriangle est plus contrainte, elle pourrait conduire à des valeurs plus faibles. Ainsi, si nous observons un *MoveM* avec un concept c_{obs} du contexte et si notre mécanisme propose

un *MoveM* avec le même concept c_{obs} , nous choisissons de mesurer la précision, le rappel et la F-mesure de \blacktriangle (action d'adaptation, concept lié et relation sémantique exacts). En revanche, si seule l'action *MoveM* est correcte, nous faisons le calcul correspondant à ∇ . Nous procédons de la même façon pour *ModSTR*, en considérant en plus de l'action d'adaptation du mapping, le type de la relation sémantique.

- Enfin, nous effectuons un calcul global qui calcule la précision, le rappel et la F-mesure pour chaque jeu de données, indépendamment des types d'action d'adaptation.

4.2 Résultats et discussion

Les résultats de l'adaptation des mappings basés sur nos heuristiques sont présentés dans la table 2 par type d'adaptation. Les tests ont été réalisés sur SNOMED CT et ICD-9-CM.

TAB. 2 – Résultats de l'évaluation des heuristiques d'adaptation des mappings

MAA type	Dataset	#MAA	Precision	Recall	F-Measure
<i>MoveM</i>	SCT	54	0.15 ∇ ;0.13 \blacktriangle	0.60 ∇ ;0.52 \blacktriangle	0.23 ∇ ;0.20 \blacktriangle
	ICD9	436	0.90 ∇ ;0.76 \blacktriangle	0.55 ∇ ;0.47 \blacktriangle	0.68 ∇ ;0.58 \blacktriangle
<i>DeriveM</i>	SCT	713	0.69 ∇ ;0.52 \blacktriangle	0.52 ∇ ;0.40 \blacktriangle	0.59 ∇ ;0.45 \blacktriangle
	ICD9	0	N/A	N/A	N/A
<i>ModSemTypeM</i>	SCT	218	0.65 ∇ ;0.39 \blacktriangle	0.67 ∇ ;0.39 \blacktriangle	0.66 ∇ ;0.39 \blacktriangle
	ICD9	73	0.09 ∇ ;0.05 \blacktriangle	0.14 ∇ ;0.07 \blacktriangle	0.11 ∇ ;0.06 \blacktriangle
<i>RemoveM</i>	SCT	136	0.66 ∇ ;0.66 \blacktriangle	0.75 ∇ ;0.75 \blacktriangle	0.70 ∇ ;0.70 \blacktriangle
	ICD9	51	0.41 ∇ ;0.41 \blacktriangle	0.79 ∇ ;0.79 \blacktriangle	0.54 ∇ ;0.54 \blacktriangle
<i>NoAction</i>	SCT	5 892	0.94 ∇ ;0.94 \blacktriangle	0.94 ∇ ;0.94 \blacktriangle	0.94 ∇ ;0.94 \blacktriangle
	ICD9	3 139	0.93 ∇ ;0.93 \blacktriangle	0.94 ∇ ;0.94 \blacktriangle	0.93 ∇ ;0.93 \blacktriangle
<i>Global</i>	SCT	7 013	0.88 ∇ ;0.86 \blacktriangle	0.88 ∇ ;0.86 \blacktriangle	0.88 ∇ ;0.86 \blacktriangle
	ICD9	3 699	0.85 ∇ ;0.84 \blacktriangle	0.87 ∇ ;0.86 \blacktriangle	0.86 ∇ ;0.85 \blacktriangle

Mapping Move. Pour ce type d'action, les résultats varient selon que l'évolution concerne ICD-9-CM ou SNOMED CT uniquement pour la précision. Ainsi, nos heuristiques peuvent générer des adaptations plus ou moins correctes selon les jeux de données. Les différences de résultats entre \blacktriangle et ∇ étant très faibles, on remarque que, la plupart du temps, lorsque notre système propose un *MoveM*, c'est avec un bon concept candidat du contexte.

Mapping Derive. Les résultats obtenus sont plus élevés que ceux de *Move* pour SNOMED CT. L'action *DeriveM* est complexe car l'action propose plusieurs *DeriveM* par mapping. Aucune action *DeriveM* n'a été observée pour ICD-9-CM. Là encore, les conditions d'application des heuristiques sont différentes selon les jeux de données. L'explication peut provenir du processus de maintenance, de matching ou de la granularité des ontologies alignées.

Modification de la relation sémantique. L'évaluation de cette heuristique a été difficile car les mappings dont nous disposions contenaient peu de relations différentes (la plupart était des équivalences). Néanmoins, les résultats montrent que les nouvelles relations proposées quand une action de type *ModSemTypeM* est détectée, sont pertinentes (surtout pour SNOMED

CT). Ce type d'adaptation est généralement combiné avec un *MoveM* ou un *DeriveM* ce qui oblige à choisir la bonne action avant de changer la relation. Parfois, les changements dans l'ontologie ne sont pas l'unique raison pour changer la relation. Cela influe négativement sur nos résultats. Enfin, considérer la nouvelle version des mappings comme référence peut aussi poser problème. C'est un ensemble de mappings qui a subi des évolutions mais ce n'est pas à proprement parler un jeu de données de référence. Une relation sémantique proposée peut être considérée comme fautive par notre processus d'évaluation alors qu'elle peut être correcte d'un point de vue sémantique. L'intervention d'un expert serait nécessaire pour y remédier.

Mapping Remove. La précision et le rappel sont relativement bons. Ils ne varient que légèrement selon l'ontologie. La F-mesure minimale est de 0.54. La prise en compte du retrait de l'attribut expliquant le mieux le mapping dans l'heuristique semble être une bonne décision.

Application d'aucune action. Ce type d'adaptation couvre le plus grand nombre de cas que ce soit en valeur absolue (5892 attendus dans SNOMED CT et 3139 dans ICD-9-CM) ou relative. Les résultats montrent une grande efficacité des heuristiques. On constate que, bien que les concepts source ou cible évoluent, si les attributs expliquant les mappings qui les concernent restent inchangés, ces mappings n'évolueront pas non plus. Les très bons résultats obtenus avec l'application de cette heuristique permettent aux experts de se concentrer sur l'étude d'une toute petite partie des mappings (en comparaison à l'ensemble initial), ces mappings étant plus difficiles à adapter et pouvant nécessiter une intervention humaine.

Résultats globaux. Nous avons analysé les résultats en combinant tous les types d'adaptation pour avoir une idée de la qualité du processus général d'adaptation. La F-mesure est élevée (un min de 0.85 dans ICD-9-CM). L'action d'adaptation *NoAction* influe beaucoup sur ces résultats. Néanmoins les résultats sont globalement acceptables et prometteurs même si certaines actions sont difficiles à appliquer. Ceci montre que les conditions d'application des heuristiques que nous proposons sont adaptées. L'amélioration des résultats liés à certaines actions nécessiterait de rechercher quels autres éléments influencent les types d'adaptation et d'étudier comment les prendre en compte dans les heuristiques.

5 Etat de l'art

Nous distinguons trois principaux types d'approches pour la maintenance des mappings due à l'évolution d'ontologies. La première consiste à identifier et réparer les mappings invalides (Meilicke et al., 2008) (Ivanova et Lambrix, 2013). Ces approches effectuent des raisonnements logiques pour identifier les mappings produisant des incohérences logiques. Ce mécanisme ne s'applique que sur des ontologies formelles, qui n'existent pas toujours.

Les travaux s'inscrivant dans la deuxième catégorie reposent sur des techniques de réalignement total ou partiel d'ontologies. Si le premier type d'approche ne considère aucune information provenant de l'évolution, le second recoupe l'ensemble des concepts modifiés au niveau des ontologies avec ceux impliqués dans des mappings pour ne réaligner que ce sous-ensemble de concepts (Khattak et al., 2012). Ces approches sont coûteuses en temps de calcul et de validation lorsque les ontologies sont volumineuses (Shvaiko et Euzenat, 2013).

La troisième catégorie, inspirée du monde des bases de données (Velegrakis et al., 2003), fait référence à des approches qui profitent au maximum des informations provenant de l'évolution d'ontologies pour éviter de les réaligner totalement. Tang et Tang (2010) ont proposé une méthode visant à trouver l'impact minimal de la propagation des changements au niveau

d'une ontologie. Martins et Silva (2009) suggèrent d'adapter les mappings en procédant de la même façon que lors de la modification des concepts de l'ontologie qui évolue. Cependant, dans leur approche, les mappings ne sont adaptés que lorsque les concepts sont supprimés.

L'originalité de nos travaux, par rapport à cet état de l'art, réside dans : (i) l'importance donnée aux modifications des ontologies sous-jacentes, (ii) à leur caractérisation et (iii) à la prise en compte du changement au niveau de la relation sémantique des mappings, ce dernier point étant souvent négligé dans les approches existantes qui ne considèrent qu'un seul type de relation (Dos Reis *et al.*, 2012).

6 Conclusion

Dans cet article, nous avons proposé un ensemble d'heuristiques guidant l'adaptation des mappings entre ontologies. Ces dernières formalisent le lien existant entre les changements identifiés au niveau des éléments ontologiques et les actions d'adaptation à appliquer sur les composants des mappings pour préserver leur validité. L'approche décrite a été validée expérimentalement sur des données réelles du monde biomédical. Le manque de données de référence nécessite l'implication d'experts du domaine mais le travail réalisé permet de réduire considérablement le temps nécessaire aux experts pour valider les adaptations proposées.

Remerciements

Ces travaux ont été financés par le Fonds de la Recherche Luxembourgeois (FNR).

Références

- Dinh, D., J. C. Dos Reis, C. Pruski, M. Da Silveira, et C. Reynaud-Delaître (2014). Identifying change patterns of concept attributes in ontology evolution. In *The Semantic Web : Trends and Challenges*, pp. 768–783. Springer International Publishing.
- Djedidi, R. et M.-A. Aufaure (2009). Change management patterns (cmp) for ontology evolution process. In *Proceedings of the 3rd International Workshop on Ontology Dynamics (IWOD 2009) in ISWC*.
- Dos Reis, J., D. Dinh, C. Pruski, M. Da Silveira, et C. Reynaud-Delaître (2014). The influence of similarity between concepts in evolving biomedical ontologies for mapping adaptation. In *European Medical Informatics Conference (MIE)*, pp. 1003–7.
- Dos Reis, J., C. Pruski, M. Da Silveira, et C. Reynaud-Delaître (2013). Characterizing semantic mappings adaptation via biomedical kos evolution : A case study investigating snomed ct and icd. In *Proc. of the AMIA Symposium*, pp. 333–342.
- Dos Reis, J. C., D. Dinh, C. Pruski, M. Da Silveira, et C. Reynaud-Delaître (2013). Mapping adaptation actions for the automatic reconciliation of dynamic ontologies. In *Proceedings of the 22nd ACM international conference on Conference on information & knowledge management*, pp. 599–608. ACM.

Heuristiques pour l'adaptation de mappings

- Dos Reis, J. C., C. Pruski, M. Da Silveira, et C. Reynaud-Delaître (2014). Understanding semantic mapping evolution by observing changes in biomedical ontologies. *Journal of biomedical informatics* 47, 71–82.
- Dos Reis, J. C., C. Pruski, M. Da Silveira, et C. Reynaud-Delaître (2012). Analyzing and Supporting the Mapping Maintenance Problem in Biomedical Knowledge Organization Systems. In *Proc. SIMI Workshop at ESWC*, pp. 25–36.
- Euzenat, J. et P. Shvaiko (2007). *Ontology matching*. Springer.
- Gröner, G., F. S. Parreiras, et S. Staab (2010). Semantic recognition of ontology refactoring. In *The Semantic Web–ISWC 2010*, pp. 273–288. Springer.
- Hartung, M., A. Groß, et E. Rahm (2013). Conto-diff – generation of complex evolution mappings for life science ontologies. *Biomedical Informatics* 46, 15–32.
- Ivanova, V. et P. Lambrix (2013). A unified approach for aligning taxonomies and debugging taxonomies and their alignments. In *ESWC 2013*, pp. 1–15. Springer.
- Javed, M., Y. Abgaz, et C. Pahl (2013). Ontology change management and identification of change patterns. *Journal on Data Semantics* 2(2-3), 119–143.
- Khattak, A., Z. Pervez, K. Latif, et S. Lee (2012). Time efficient reconciliation of mappings in dynamic web ontologies. *Knowl.-Based Syst.* 35.
- Martins, H. et N. Silva (2009). A User-Driven and a Semantic-Based Ontology Mapping Evolution Approach. In *Proc. of ICEIS*.
- Meilicke, C., H. Stuckenschmidt, et A. Tamilin (2008). Reasoning Support for Mapping Revision. *Journal of Logic and Computation* 19(5).
- Shvaiko, P. et J. Euzenat (2013). Ontology Matching : State of the Art and Future Challenges. *IEEE Trans. Knowl. Data Eng.* 25(1), 158–176.
- Tang, F. et R. Tang (2010). Minimizing Influence of Ontology Evolution In Ontology-based Data Access System. In *IEEE Int. Conf. on Progress in Informatics and Computing (PIC)*, pp. 10–14.
- Velegarakis, Y., R. J. Miller, et L. Popa (2003). Mapping adaptation under evolving schemas. In *Proc. of the 29th Int. Conf. on Very large data bases - Volume 29, VLDB '03*, pp. 584–595. VLDB Endowment.

Summary

Ontologies and their associated mappings play a central role in several decision support systems. However, by virtue of knowledge evolution, ontology entities are modified over time impacting mappings and potentially turning them invalid. This requires methods to maintain such semantic correspondences up-to-date at ontology evolution time. In this article, we present the design and formalization of original heuristics that drive the adaptation of ontology mappings. Our approach takes into account the definition of established mappings, the evolution of ontologies and the possible changes that can be applied to mappings. This study experimentally evaluates the proposed heuristics on realistic case studies borrowed from the biomedical domain by assessing their validity on official set of mappings.