

Typologie de changements et leurs effets sur l'évolution de Ressources Termino-Ontologiques

Anis Tissaoui

Laboratoire IRIT, Université Paul Sabatier
Anis.Tissaoui@irit.fr

Résumé : L'évolution des ressources termino-ontologiques (RTO) est en lien avec l'évolution de leur domaine et leur utilisation. Cette évolution est définie comme une adaptation opportune de cette RTO aux changements se produisant. Dans cet article, nous étudions les différents types de changements pouvant se produire au moment de l'évolution et leurs conséquences sur la RTO.

Mots-clés : Ressource termino-ontologique, Évolution d'ontologies, Typologie de changement.

1 Introduction

Une ressource termino-ontologique (RTO) est une ressource comportant une composante conceptuelle, l'ontologie, et une composante lexicale, la terminologie. Plusieurs travaux comme (Aussenac-Gilles et al., 2006), (Maedche, 2002) et (Cimiano, 2006) évoquent la nécessité d'associer un lexique indépendant à une ontologie. La RTO contient alors non seulement les concepts du domaine, mais aussi les termes associés (termes désignant les concepts) qui permettent d'annoter ou indexer des documents dans le cadre d'une annotation sémantique.

L'évolution d'une RTO est le processus de changement impliqué par des modifications du domaine de connaissance qu'elle représente, de sa conceptualisation ou de son usage (par exemple, la réutilisation de l'ontologie pour des tâches différentes demande sa modification). Ce processus se traduit en réalité par l'exécution d'un ou plusieurs changements pour passer d'une version de l'ontologie (VN) à une version nouvelle (VN+1), tout en préservant la consistance et les rôles de la RTO. Plusieurs auteurs mettent en évidence l'importance majeure de l'évolution d'ontologies et proposent des approches pour en gérer l'évolution (Stojanovic, 2004), (Klein & Noy, 2003), (Luong, 2007), (Rogozan, 2008). Tous ces travaux, ne prennent en compte que l'ontologie et mettent de côté la terminologie. Dans (Aussenac-Gilles et al., 2006), (Maedche, 2002) et (Cimiano, 2006), la problématique liée à l'évolution de RTO est évoquée, mais aucune méthodologie d'évolution n'est proposée.

Dans cet article, nous proposons une typologie de changements. Nous définissons un certain nombre de caractéristiques afin de mieux spécifier le type et l'effet de chaque changement sur la RTO. Ce travail est un travail préliminaire et il devra être poursuivi pour étudier les effets sur les artefacts dépendant de cette RTO (les ontologies dépendantes, les applications logicielles utilisant l'ontologie, ...).

Après une présentation de l'état de l'art sur les méthodologies d'évolution d'ontologie, nous présentons notre typologie de changements pour le cas des ontologies à composante terminologique (RTO). Nous étudions ensuite les conséquences des différents types de changement sur la RTO et la manière de les prendre en compte avant de conclure sur les perspectives de ce travail.

2 Évolution d'ontologie

Plusieurs approches d'évolution d'ontologie ont été proposées dans la littérature. Stojanovic définit diverses notions nécessaires pour gérer l'évolution d'une ontologie (Stojanovic, 2004). Elle décrit un processus d'évolution en six phases et introduit aussi une taxonomie de changements qui contient trois niveaux de classification : les changements élémentaires, composites et complexes. Le processus global assure le passage de l'identification des changements à leur interprétation et leur mise en œuvre. Cette méthodologie demeure toutefois limitée car elle ne conserve pas l'historique des évolutions. Elle se contente de représenter l'ontologie dans sa version la plus récente.

Une autre méthode (Klein & Noy, 2003) repose sur la gestion et la conservation des versions différentes de l'ontologie en évolution. Un modèle des relations rend compte des changements effectués entre les versions. Cependant la méthode n'indique pas comment les versions de l'ontologie peuvent être créées.

Une approche qui englobe les deux précédentes a été proposée et intègre un formalisme de représentations de l'évolution des ontologies sous forme de graphes orientés (Eder & Koncilia, 2004). Elle se fonde sur les concepts et les techniques développées pour les bases de données temporelles, l'évolution de schéma et le versioning des bases de données. Cette approche ne permet pas de découvrir les liens entre les entités ontologiques de différentes versions car l'historique de modification d'une entité ontologique n'existe pas.

D'autres auteurs traitent le problème de l'évolution d'une ontologie en lien avec son utilisation. (Luong, 2007) présente une nouvelle approche qui gère à la fois l'évolution de l'ontologie et des annotations sémantiques produites avec cette ontologie, dans le cadre d'un web sémantique d'entreprise (une approche particulière de la Gestion des Connaissances d'une entreprise). Luong élabore une ontologie d'évolution qui a pour but de modéliser d'une manière formelle les types de changements et les informations concernant le processus d'évolution. (Rogozan, 2008) intègre les deux méthodologies présentées ci-dessus (travaux de Klein et de Stojanovic) dans une méthodologie unifiée appliquée aux systèmes d'apprentissage en ligne. Il présente aussi une ontologie des changements applicables à des ontologies OWL-DL. Cette méthode élargit la conceptualisation de (Klein, 2004) par l'ajout d'un certain nombre de caractéristiques comme : (a) une typologie des changements complexes ; (b) une description des effets de changements sur l'annotation sémantique (la représentation formelle du contenu d'une ressource) ; (c) une clarification de la terminologie utilisée pour caractériser la complexité d'un changement. Les deux auteurs (Luong et Rogozan) gèrent les annotations sémantiques mais ils n'utilisent pas

les termes de la terminologie pour annoter les documents. De plus, Luong et Rogozan traitent des ontologies OWL, et leurs modèles sont dépendants de ce langage.

3 Modèle de Ressource Termino-Ontologique

La définition de formalismes pour représenter une RTO est l'objet de recherches récentes, de sorte qu'il n'existe pas de standard à proprement parler. TMF (Terminological Markup Framework) permet de décrire tous les éléments d'une terminologie avec un langage formel (Romary, 2001). Reymonet définit dans (Reymonet et al., 2007) un modèle pour manipuler une RTO en OWL. Les termes y sont représentés comme des classes à part entière reliées aux concepts par une relation de dénotation. Cimiano mentionne la nécessité d'associer un lexique indépendant à une ontologie et définit un modèle pour manipuler la RTO (Cimiano, 2006). Nous choisissons de rester indépendant du langage utilisé pour représenter l'ontologie et généraliser le modèle de (Reymonet et al., 2007) afin de gérer les évolutions d'ontologies de différents formats. Nous définissons un modèle de RTO décrit dans la figure 1.

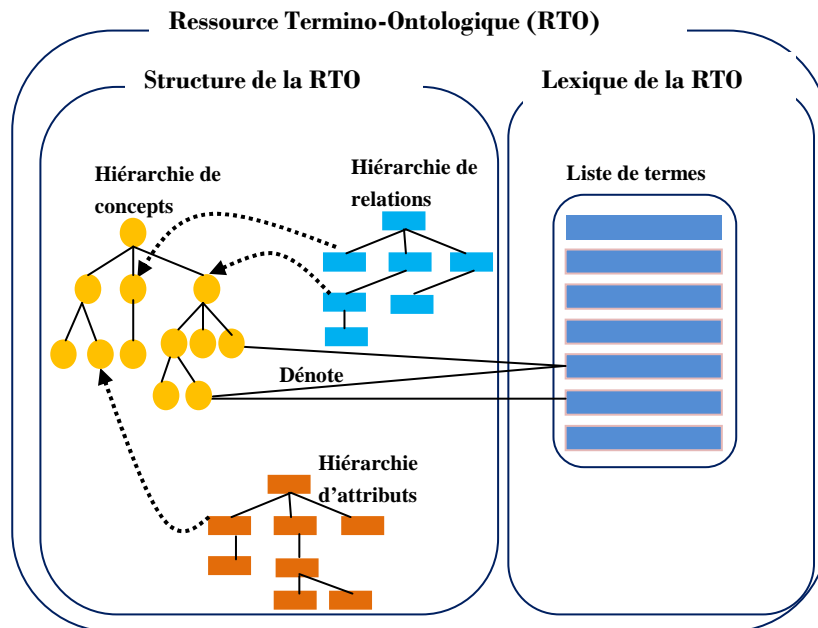


Fig. 1 – Modèle de ressource termino-ontologique

Les différents modèles existants qui sont présentés dans l'état de l'art sont insuffisants pour représenter correctement la partie terminologique associée à une ontologie. Le modèle de Ressource Termino-Ontologique (RTO) que nous proposons enrichit la composante conceptuelle, l'ontologie, par une composante lexicale, la terminologie. La RTO contient alors :

- Un ensemble de concepts organisés en une hiérarchie où les concepts sont reliés par une relation de spécialisation (subsomption).
- Un ensemble de relations binaires organisées sous forme d'une hiérarchie, chaque relation spécifiant des couples de concepts définissant le domaine (concept d'où part la relation) et le co-domaine (concept cible de la relation).
- Un ensemble d'attributs organisés aussi sous forme d'une hiérarchie, les attributs spécifiant les propriétés, les fonctionnalités, les caractéristiques ou les paramètres que les concepts peuvent posséder et partager.

Il a été choisi d'associer un lexique indépendant à une ontologie, de manière à étiqueter les concepts, le tout formant une ontologie à composante lexicale aussi appelée RTO. Il faut noter aussi qu'un terme n'est présent dans la RTO s'il est relié à au moins un concept qu'il désigne. Toutefois, nous considérons que l'ontologie est une représentation qui évolue avec le temps : l'apparition de nouveaux textes peut entraîner l'ajout ou la modification de termes ou de concepts, ce qui aura pour conséquence la modification (réorganisation) des relations entre termes et concepts. Un terme désignant à l'origine un concept c sera par exemple associé à un nouveau concept $c1$; il est alors nécessaire de traiter cette mise à jour et ses conséquences sur l'évolution de la RTO d'un côté et sur l'annotation sémantique d'un autre côté.

4 Typologie de changements

Une fois les changements identifiés (à partir de besoins liés à l'application utilisant la RTO ou d'évolutions du domaine), qu'ils soient explicites ou implicites, les utilisateurs choisissent ceux qui leur semblent pertinents et produisent une liste de changements à apporter à la RTO. En partant du modèle de RTO ci-dessus et en tenant compte des typologies existant dans l'état de l'art, nous proposons une typologie des changements applicables à une RTO définie selon ce modèle. Cette typologie élargit la conceptualisation de (Stojanovic, 2004), par l'ajout d'un certain nombre de caractéristiques dont la plus notable est la présence d'un lexique de RTO qui peut évoluer à cause des changements portant sur les termes et les liens de dénnotations. Deux types de changements peuvent être identifiés : les changements simples ou élémentaires et les changements composés. Ces changements permettent d'aider ensuite l'ontologie à faire évoluer une RTO en lui indiquant les conséquences de chaque changement (les effets additionnels).

4.1 Définition 4 – Changement élémentaire

Un changement élémentaire est un changement qui porte sur une seule entité de la RTO (ajoute, modifie ou supprime) (Stojanovic, 2004). Nous présentons ici quelques fonctions de changement élémentaire :

- $CreateTerme(t)$: Ajouter le nouveau terme t à T^C .
- $CreateDenotationLinkConcept(t, c, LD)$: Ajouter le couple (t, c) à l'ensemble des liens de dénnotation LD .

4.2 Définition 5 – Changement composé

Un changement composé est un changement qui affecte deux ou plusieurs entités de la RTO.

Les changements composés s'expriment donc en combinant des changements élémentaires. Ils présentent plusieurs avantages. Tout d'abord, ils permettent l'évolution de l'ontologie sans perdre des données. Par exemple, on va définir *MoveConcept* de manière à préserver les instances du concept déplacé, ce qui ne serait pas nécessairement assuré en effectuant la succession des changements élémentaires *DeleteConcept* puis *AddConcept*. De plus manipuler une seule opération de changement complexe (comme *MergeConcept*), permet aux acteurs de formuler plus aisément la modification qu'ils souhaitent effectuer sur l'ontologie. Nous présentons ici quelques fonctions de changement composé :

- *AddTerme* (t, c, LD) : Créer le nouveau terme t et lui faire dénoter un concept existant c.
- *ModifyRelationDomain* (r, c, c1) : Modifier le domaine d'une relation r, en échangeant le concept c faisant partie du domaine par un autre concept c1.

Chaque type de changement peut générer des changements additionnels sur les autres parties de la RTO, qui sont les conséquences de cette évolution. Nous présentons un exemple qui permet d'expliquer la définition d'un changement additionnel et les effets d'une évolution sur la RTO. Si l'on supprime un concept c de la hiérarchie de concepts, plusieurs questions se posent pour garder la RTO dans un état cohérent :

- Quels sont les effets sur les fils du concept c ?
- Quels sont les effets sur les propriétés du concept c ?
- Quels sont les effets sur les propriétés dont le domaine ou le co-domaine contient c ?
- Quels sont les effets sur les instances de c et les instances d'autres concepts qui sont en relation avec les instances de c ?
- Quels sont les effets sur les termes associés au concept c ?

Ainsi, pour chaque type d'évolution dans l'ontologie, il est possible de définir un ensemble de changements supplémentaires (changements additionnels) qui peuvent être effectués pour garder l'ontologie dans un état cohérent. Lorsque plusieurs effets sont possibles pour une même évolution un choix doit être effectué, soit arbitrairement par le système de gestion des évolutions dans lequel on aura fixé un des effets possibles, soit par l'ontographe. Prenons l'exemple de la suppression d'un concept et les effets sur ses sous-concepts. L'utilisateur peut (a) attacher les sous-concepts au concept père, (b) attacher au top de la hiérarchie, (c) supprimer les sous-concepts. La manière de faire intervenir l'utilisateur pour formuler le choix d'un de ces effets (au cas par cas, une fois pour toutes, ...) revient à fixer une stratégie de résolution des évolutions au sens de (Stojanovic, 2004).

5 Conclusion et perspectives

Nous avons étudié de manière générale les types de changements pouvant intervenir sur les composants d'une ressource termino-ontologique. Cet article justifie de distinguer changements élémentaires et composés et en présente quelques exemples ainsi que leurs effets sur la RTO. Nos futurs travaux concernent trois axes principaux liés à la gestion des évolutions d'une RTO dans le cas où elle est utilisée pour annoter sémantiquement des documents.

- Les effets de l'évolution d'une RTO sur les annotations sémantiques produites à l'aide de cette RTO.
- Les stratégies de résolution de différents types de changements pour la RTO et l'annotation sémantique.
- Inversement, les évolutions d'une annotation sémantique et leurs effets sur la RTO.

Références

- AUSSENAC-GILLES N. CONDAMINES A. SEDES F. (2006). *Evolution et maintenance des ressources termino-ontologiques : une question à approfondir*. Information - Interaction - Intelligence, Cépaduès Editions, Numéro spécial Ressources termino-ontologiques, Vol. Hors-série, p. 7-14, 2006.
- CIMIANO P. (2006). *Ontology Learning and Population from Text: Algorithms, Evaluation and Applications*, Springer Verlag, 2006.
- EDER J. & KONCILIA C. (2004). *Changes in Ontologies*. OTM Workshops 2004: 662-673
- KLEIN M. (2004). *Change Management for Distributed Ontologies*. Thèse de doctorat de l'université de Vrije Amsterdam., 2004.
- KLEIN M. & NOY N. (2003). *Ontology evolution: Not the same as schema evolution*. Knowledge and Information Systems, 2003.
- LUONG P. (2007). *Gestion de l'évolution d'un web sémantique d'entreprise*, Thèse de doctorat de Mines de Paris, 2007.
- MAEDCHE A. (2002). *Ontology learning for the semantic web*. Kluwer Academic, Boston; Dordrecht; London, 2002.
- REYMONET A., THOMAS J., AUSSENAC-GILLES N. (2007). *Modélisation de ressources termino-ontologiques en OWL*. Journées Francophones d'Ingénierie des Connaissances (IC 2007), Grenoble (F).
- ROGOZAN D. (2008). *Gestion de l'évolution des ontologies : méthodes et outils pour un référencement sémantique évolutif fondé sur une analyse des changements entre versions d'ontologie*. Thèse de doctorat de l'université du Québec à Montréal, 2008.
- ROMARY L. (2001). *An abstract model for the representation of multilingual terminological data: TMF – Terminological Markup Framework*. TAMA, Anvers, Belgium. 2001.
- STOJANOVIC L. (2004). *Methods and tools for ontology evolution*. Thèse de doctorat de l'Université de Karlsruhe, 2004.