

Produire une ontologie à partir d'un thésaurus – méthode et outil

Gilles Hubert, Josiane Mothe, Pascal Will

hubert@irit.fr, will@irit.fr, mothe@irit.fr

IRIT, 118 route de Narbonne, 31062 Toulouse, France.

Mots clefs :

Ressources pour l'indexation, thésaurus, ontologies, OWL

Keywords:

Indexing, thesaurus, ontology, OWL

Résumé

Le développement des technologies du web sémantique laissent entrevoir de nombreuses applications avec, parmi une des plus importantes la recherche d'information (RI) sémantique. La RI sémantique s'appuie en particulier sur la représentation de connaissances de domaines formalisée sous forme de ressources terminologiques, de thésaurus ou d'ontologies. L'interopérabilité des systèmes et la réutilisation des ressources impliquent cependant d'adopter un format commun, ce qui n'est actuellement pas le cas. Cet article apporte une première solution à cette problématique. Après avoir présenté une méthodologie de transformation de thésaurus en ontologie légère, nous proposons des solutions techniques pour la mettre en œuvre. Les thésaurus supportés peuvent être hétérogènes dans leur structure syntaxique pourvu qu'ils respectent les normes en vigueur.

1 Introduction

Le développement des technologies du web sémantique (Berners-Lee, 2001), laissent entrevoir de nombreuses applications avec, parmi les plus importantes la recherche d'information sémantique. Le concept de recherche d'information sémantique n'est pas nouveau puisque les systèmes documentaires basés sur un thésaurus la permettent depuis plusieurs décennies. En effet, depuis les premiers systèmes documentaires, les thésaurus sont utilisés pour permettre à un utilisateur de mieux définir son besoin en termes non ambigus. En amont, ils permettent une indexation homogène des documents au travers du langage documentaire qu'ils représentent. Ils impliquent cependant une analyse documentaire et un choix de descripteur manuels. Dans les années 70, avec l'augmentation des quantités de documents à gérer, des approches alternatives ont été étudiées. Ainsi, l'indexation par « sac de mots », c'est-à-dire à partir des termes issus des documents eux-mêmes, a été plébiscitée. Le juste retour à plus de sémantique, tout en gardant l'atout de l'automatisation, semble devenir possible grâce aux technologies du web sémantique. Il ne s'agit plus d'utiliser manuellement une ressource terminologique mais d'intégrer une ressource de connaissances (éventuellement orientée domaine) dans une application. Cette approche implique cependant que les ressources ontologiques de domaine existent ou soient construites. Comme cela est souligné dans (Chrisment et al., 2006), il nous paraît dommageable de ne pas tirer profit des ressources terminologiques existantes comme point de départ d'une ontologie.

Dans cet article, nous présentons les solutions techniques que nous avons retenues pour transformer un thésaurus en une ontologie OWL. Nous présentons également l'outil que nous avons développé permettant de réaliser cette tâche. Cet article correspond donc à une suite logique de l'article (Chrisment et al., 2006) puisqu'il propose des solutions techniques directement ré-applicables.

La suite de l'article est organisée comme suit. Dans la section 2, nous présentons ce que sont les thésaurus et les normes qui les régissent. Dans la section 3, nous présentons les ontologies légères de domaine. Les premiers correspondent aux points d'entrée des solutions que nous proposons, alors que les secondes correspondent aux points de sortie. La section 4 présente la méthode de transformation et son implantation pratique. La section 5 conclut cet article.

2 Thésaurus

2.1 Définitions

Selon (Fosket, 1980), un thésaurus fournit une cartographie d'un champ de connaissance en indiquant comment les concepts ou les idées associées sont reliés. Un thésaurus aide un indexeur ou une personne qui recherche de l'information à comprendre la structure d'un domaine.

Plus précisément, un thésaurus est un ensemble de termes normalisés et contrôlés, descripteurs et non-descripteurs obéissant à des règles terminologiques propres. Les termes d'un thésaurus sont reliés entre eux par des relations sémantiques (hiérarchiques, associatives, ou d'équivalence). Les termes non-descripteurs sont reliés aux descripteurs par la seule relation d'équivalence. Un thésaurus sert à traduire en un langage artificiel (qualifié de langage documentaire) dépourvu d'ambiguïté, des notions exprimées en langage naturel (Dégez & Ménillet, 2001). Un thésaurus peut être mono ou multilingue.

Dans les systèmes documentaires, un thésaurus a pour but d'aider à l'indexation manuelle de documents ; il permet à l'indexeur de traduire la description du contenu d'un document dans un vocabulaire contrôlé (Chaumier, 1988). Les règles d'indexation stipulent en particulier que seuls les termes descripteurs peuvent être choisis pour indexer un document (Rameau, 2006). Lors de la recherche, le thésaurus permet à l'utilisateur de traduire son besoin d'information dans le langage documentaire utilisé lors de l'indexation.

2.2 Normes

Différentes normes régissent la conception de thésaurus. Elles édictent les règles relatives à l'identification des descripteurs, relations entre eux, termes composés, à la présentation et la gestion du thésaurus mais également les modalités de maintenance.

Les principales normes internationales sont : ISO 2788 :1986 (ISO 2788, 1986) et ANSI/NISO Z39.19-2003 (ANSI/NISO Z39.19, 2003) pour les thésaurus monolingues, ISO 5964 (ISO 5964,) pour les thésaurus multilingues. Pour la France il y a les normes AFNOR NF Z47-100 à AFNOR NF Z47-103.

2.3 Structure sémantique définie par les normes

Les normes homogénéisent ce que nous appellerons la structure sémantique du thésaurus. En effet bien qu'elles définissent les composants d'un thésaurus et leur structure (les différents types de termes et de relations), elles ne précisent pas comment ces composants doivent être écrits ou formatés.

2.3.1 Descripteurs et non descripteurs

Il existe deux types de termes dans les normes : les descripteurs (ou termes préférés ou termes vedettes) et les non-descripteurs. Par défaut tout nouveau terme est non-descripteur. Les descripteurs quant à eux sont obtenus en regroupant les non-descripteurs synonymes et en choisissant celui qui représente le mieux le groupe (c'est la relation d'équivalence ou de représentation).

2.3.2 Relations

Les normes définissent également trois types de relations entre termes. Deux relations sont composées de deux prédicats ayant des sens symétriques liant des termes du domaine : la relation d'équivalence (« utiliser plutôt » / « utilisé pour ») et la relation hiérarchique (« est père de » / « est fils de »). La troisième relation est composée d'un unique prédicat. Il s'agit de la relation associative (« est lié à »).

La première relation à deux prédicats est la relation souvent nommée dans la littérature « relation d'équivalence ». Cette relation a pour essence de désigner le choix d'un terme descripteur comme représentant d'un ou plusieurs termes non-descripteurs. Ainsi, le terme descripteur (aussi appelé terme préféré ou terme vedette) est considéré comme équivalent aux autres non-descripteurs. La relation comporte deux prédicats symétriques :

|| *« Employé pour » « utilisé pour » lie un descripteur à un ensemble de non-descripteurs, et*
|| *« Employer » « Utiliser plutôt » lie un non-descripteur à son descripteur.*

La deuxième relation est la relation de hiérarchie, aussi nommée relation de subsumption, ou de taxonomie. Son rôle est de hiérarchiser les descripteurs. Elle possède le prédicat de généralisation :

|| *« est père de », lie un descripteur avec un ensemble de descripteurs du domaine.*

Ainsi le terme de gauche du prédicat sera plus générique que le(s) terme(s) de droite. Ce terme permettra de représenter la relation de l'ensemble de termes avec le reste du domaine, comme une sorte d'« encapsulation » de termes. Le deuxième prédicat de la relation de hiérarchie est le prédicat de sémantique inverse au précédent, donc

|| *« est fils de » lie deux descripteurs entre eux (un fils n'ayant qu'un père).*

Au sens des thésaurus, il représente un terme plus spécifique que le terme en partie droite du prédicat.

La relation à un prédicat correspond à la relation d'association définie par les normes. Elle lie deux descripteurs du domaine. Elle a pour essence de mettre en relation des descripteurs autrement que par la relation de taxonomie, sans préciser la nature du lien. Elle permet de naviguer dans le domaine sans passer par la relation hiérarchique et donc de prendre des « raccourcis » lors de la recherche dans l'arborescence. Il n'existe pas de restriction sur cette relation, elle peut lier deux descripteurs entre eux mais elle peut tout aussi bien lier un descripteur à un ensemble de descripteurs.

|| *« est lié à » lie deux descripteurs entre eux.*

Il existe d'autres relations à un prédicat dans certains thésaurus, même si celles-ci ne sont pas prévues par les normes. La relation de définition qui lie un terme (descripteur ou non-descripteur) du domaine à sa définition entre dans ce type. La relation de traduction qui lie un terme (descripteur ou non-descripteur) du domaine à sa traduction dans une autre langue en est un autre exemple.

2.4 Structure syntaxique non définie par les normes

Si les normes définissent bien ce que doit contenir un thésaurus et comment il doit être organisé, en revanche, elles ne précisent pas la structure syntaxique. Les normes ne définissent pas la manière de nommer ces éléments ni dans le cas de format électronique, la façon de les stocker. Ainsi, les noms : des descripteurs, des non-descripteurs, des relations, de la mise en page, etc. ne sont pas définis. Les normes ne garantissent donc pas l'interopérabilité. En effet, des thésaurus qui sont conformes à la même norme pourront avoir des mises en forme totalement différentes. Ceci ne facilite pas l'exploitation de différents thésaurus au sein d'une même application.

3 Ontologies légères

Une ontologie est définie par Gruber (Gruber, 1993) comme "une spécification explicite d'une conceptualisation". Une ontologie fournit la base de la communication entre les machines et entre humains et machines en définissant le sens des objets à travers les symboles (mots ou expressions) qui les désignent et les caractérisent et à travers une représentation structurée ou formelle de leur rôle dans le domaine (Aussenac, 2004).

(Bachimont, 2000) précise que le niveau de spécification formelle permettant de restreindre l'interprétation de chaque concept et ainsi d'en donner la sémantique (le « degré d'engagement sémantique ») distingue également les ontologies. (Heijst, 1997) distingue trois types d'ontologies en fonction de la structure de la connaissance contenue dans l'ontologie.

- Les ontologies terminologiques ou linguistiques spécifient les termes utilisés pour représenter la connaissance d'un domaine.
- Les ontologies de l'information spécifient la structure des enregistrements d'une base de données comme les schémas de base de données.
- Les ontologies pour la modélisation de la connaissance spécifient la conceptualisation de la connaissance.

Dans cet article, nous nous intéressons aux ontologies terminologiques.

3.1 Contenu d'une ontologie

3.1.1 Concepts

Un concept correspond un objet matériel, une notion ou une idée (Uschold, 1995). Un concept est constitué de la notion ou intention du concept, c'est-à-dire la sémantique liée au concept, définie via les propriétés et les attributs. Il est également constitué de ses labels ou termes synonymes qui désignent le concept dans un langage et de ses instances, c'est-à-dire de l'ensemble des objets couverts ou définis par le concept. Il faut noter que les concepts abstraits n'ont pas d'instance.

3.1.2 Relation sémantique

Une relation sémantique lie des concepts. Les relations les plus courantes dans la littérature sont les relations d'équivalence, taxonomiques, patronymiques, de dépendance, topologique, causale, fonctionnelle, chronologique (Gomes-Peres, 2000). Nous retrouvons ici des noms de relations communs avec celles des thésaurus, bien qu'elles ne portent pas sur les mêmes objets (concepts pour les ontologies vs termes pour les thésaurus).

3.1.3 Axiome

Les axiomes ont pour but de définir des connaissances n'ayant pas un caractère strictement terminologique (Staab, 2000) comme définir des restrictions sur la valeur des attributs, déduire de nouvelles informations, etc. Ils s'expriment dans un langage logique.

3.1.4 OWL

Le W3C recommande Ontologie Web Language (OWL) (McGuinness et van Harmelen, 2004) comme standard pour représenter les ontologies. OWL se veut plus représentatif du contenu du Web que XML, RDF et RDF-Schéma en apportant un nouveau vocabulaire avec une sémantique formelle. OWL comprend trois versions. OWL lite permet de définir une hiérarchie de concepts et des contraintes simples. OWL DL est conçu pour pouvoir supporter la logique de description. OWL Full garde la liberté de syntaxe de RDF et permet une plus ample expressivité. OWL DL et Full se basent sur OWL Lite.

3.2 Hétérogénéité syntaxique des thésaurus

L'hétérogénéité syntaxique évoquée plus haut se traduit par la nécessité d'identifier les chaînes de caractères utilisées dans le thésaurus pour désigner :

- Les termes de la terminologie ou du domaine représenté,
- Les relations entre termes de la terminologie,
- Les termes ou chaînes outils.

La deuxième difficulté est de traduire correctement les informations identifiées en OWL (McGuinness et van Harmelen, 2004).

L'hétérogénéité syntaxique entre thésaurus est illustrée par quelques extraits de thésaurus conformes aux mêmes normes. Le thésaurus suisse JuriVoc¹ est un thésaurus trilingue (français, italien, allemand), multi-fichiers dans le domaine du droit.

effet anticipé positif USE EFFET ANTICIPÉ
--

Dans cet extrait, le non-descripteur « effet anticipé positif » est en relation avec le descripteur « EFFET ANTICIPÉ ». Bien que cela ne soit pas précisé dans les normes pour un grand nombre des thésaurus, les descripteurs sont en majuscules et les non-descripteurs sont en minuscules, cela permet de faciliter leur lecture

¹ <http://www.bger.ch/fr/index/jurisdiction/jurisdiction-inherit-template/jurisdiction-jurivoc-home/jurisdiction-jurivoc-more.htm>

visuelle. Dans un traitement automatique, il est également possible de faire une première différenciation rapide des différents éléments et déduire les relations qui les lient. Cependant la mise en forme étant libre, les seules règles existantes portent sur des consensus obtenus sur les types de relations entre les éléments de description. A partir de l'exemple précédent nous pouvons déduire que la relation qui lie « effet anticipé positif » et « EFFET ANTICIPÉ » est la relation d'équivalence. Dans l'extrait proposé cette relation est nommée « USE » et est précédée d'un retour à la ligne. Nous prenons l'extrait d'un autre thésaurus suivant la même norme du sens des relations. Cet extrait provient du thésaurus sur l'astronomie¹.

Le thésaurus IAU² est un thésaurus dans le domaine de l'astronomie.

```
binaries
U  BINARY STARS
```

Dans cet extrait nous trouvons le terme « binaries » qui est en relation avec le descripteur « BINARY STARS ». Comme pour l'extrait précédent ils sont liés par une relation d'équivalence. Ici, le prédicat se nomme « U » précédée d'un retour à la ligne.

Nous avons illustré l'hétérogénéité au travers d'un exemple et sur la seule relation d'équivalence. Les hétérogénéités existent à différents niveaux : nom donné au prédicat, séparation entre les termes et les prédicats, casse des caractères.

4 Méthode de transformation

Dans cette partie nous allons présenter une réponse aux deux problématiques de transformation de thésaurus en ontologie légère de domaine.

Pour répondre à ces problématiques il faudra tout d'abord extraire l'information de mise en forme quelconque du thésaurus, puis traduire cette information en langage OWL. Cette traduction en langage OWL sera faite en deux parties. Dans la première partie vont être présentées les différentes règles portant sur la transformation théorique (indépendant de la programmation) de thésaurus en ontologie (Chrisment et al., 2006). Dans la deuxième partie nous présenterons la traduction de l'information en OWL.

4.1 Règles de transformations théoriques

Nous nous appuyons sur les règles définies par (Chrisment et al., 2006) qui portent sur les trois types de relations que l'on trouve dans les thésaurus normalisés.

Les prédicats de la relation d'équivalence obéissent aux règles suivantes :

```
Pour le prédicat « Employé pour » :
« t1 employé pour t2 » dans le thésaurus
=> t1 et t2 correspondent à deux labels d'un même concept. Le concept lui-même porte le nom t1 (terme préféré).
```

¹ <http://msowww.anu.edu.au/library/theseaurs/french/>

² <http://www.site.uottawa.ca:4321/astromony/index.html>

Pour le prédicat « Employer » :

« t3 employer t1 » dans le thésaurus

=> t1 et t3 correspondent à deux labels d'un même concept. Le concept lui-même porte le nom t1 (terme préféré).

De même, les prédicats de la relation binaire de taxonomie obéissent aux règles suivantes :

Pour le prédicat « est fils de » :

« t1 est fils de t2 » dans le thésaurus

=> soit le concept c1 (resp. c2) tel que t1 (resp. t2) label de c1 (resp. C2). c1 est « une sous-classe » de c2

Pour le prédicat « est père de »

« t1 est père de t2 » dans le thésaurus

=> soit le concept c1 (resp. c2) tel que t1 (resp. t2) label de c1 (resp. C2). C2 est « une sous-classe » de c1.

Enfin, concernant la relation d'association, nous avons :

Pour le prédicat « est associé à » :

« t1 est associé à t2 » dans le thésaurus

=> soit le concept c1 (resp. c2) tel que t1 (resp. t2) label de c1 (resp. C2). c1 est « associé à » c2.

(Chrisment et al., 2006) propose une méthode afin de désambiguïser cette relation entre concepts, en se basant sur un corpus de textes adhoc. Dans cet article, nous considérons que nous ne disposons que du thésaurus et nous intéressons à sa traduction.

4.2 Extraction de l'information

L'étude des thésaurus et des définitions des relations dans les normes montre que les thésaurus bien construits sont structurés en blocs. Un bloc représente les relations d'un élément (descripteur ou non-descripteur) avec d'autres éléments du domaine, ou des éléments extérieurs (définitions, traduction,...). Il arrive cependant que des blocs soient absents ou incomplets. Nous définissons un bloc de la manière suivante :

*terme*_(x)

pB₍₁₎

element₍₁₎

[...]¹

*element*_(1+i)]

[...]

avec $i \in [0, n-1]$

*pB*_(1+i)

element₍₁₎

[...]

[*element*]_(1+i)]²

*terme*_(x) est un terme quelconque du thésaurus. Il s'agit d'un descripteur ou d'un non-descripteur issu du thésaurus (ensemble de caractères).

element : peut être un terme ou une chaîne de caractères quelconque représentant une information utile.

pB : prédicat représentant une relation, il lie un *terme* à un *element*

n : est le nombre max de *element*_(i) liés par le prédicat *pB*_(i) à *terme*_(x)

Comme indiqué dans la section 4, la mise en forme étant libre, nous pouvons trouver dans un thésaurus le bloc :

term₁

pB

term₂

et dans un autre:

term₁

pB : est en relation : → **term₂**

Le sens des deux blocs est rigoureusement identique, la partie « : est en relation : → » et le caractère de tabulation placé avant le prédicat, représente la mise en forme.

¹ Les « ... » signifient zéro ou plusieurs occurrences

² Les « [] » signifient que ce qui est entre est optionnel

La structuration en blocs est la première étape permettant d'analyser le corps d'un thésaurus. Dans ces blocs nous pouvons identifier trois types de lignes : la ligne portant le terme(x) donc l'élément gauche du prédicat, la ligne prédicat-élément droit (elle peut contenir un retour chariot), et la ligne de séparation de bloc. Pour extraire l'information utile, il est nécessaire de connaître les expressions régulières définissant ces lignes. Ces expressions se calculent simplement, grâce à la structuration en blocs qui par récurrence se généralise à tout le thésaurus. Tout ceci permet le découpage de l'information, puisqu'il suffit que l'on fournisse à l'analyseur les chaînes de caractères nommant les prédicats (leur nombre est fini et petit) pour obtenir l'information utile (les éléments, et les liens entre eux). Ainsi les noms des prédicats permettent d'identifier les blocs, la connaissance des blocs permet de déduire les expressions régulières qui permettent d'extraire, quels terme(x) sont liés, les prédicats, quand se finit un prédicat et commence un autre, quels sont les éléments en partie droite des prédicats et quand se finit le bloc du terme(x).

4.3 Solutions techniques

Dans cette section, nous proposons une manière d'implanter les règles présentées précédemment pour transformer le contenu d'un thésaurus normé en une ontologie légère de domaine écrite en langage OWL.

4.3.1 Les termes

Les termes du thésaurus deviennent des labels dans l'ontologie. Ils seront des labels de termes si ce sont des non-descripteurs et des labels de concepts si ce sont des descripteurs. En OWL, ils seront des labels d'instances si ce sont des non-descripteurs et des labels de classes si ce sont des descripteurs. En effet en OWL les classes représentent des ensembles d'éléments possédant les mêmes propriétés. Ces éléments sont des instances.

La transformation des termes découle de la transformation de la relation d'équivalence distinguant les descripteurs des autres termes. La transformation de cette relation se traduit par la création d'une nouvelle classe portant comme label le nom du descripteur dans le thésaurus. Puis par la création d'autant d'instances, qu'il y a de non-descripteurs représentés par le descripteur label de la classe. Ces instances ont pour label le nom des non-descripteurs auxquelles elles correspondent dans le thésaurus.

Ci-dessous un exemple de traduction d'un extrait de thésaurus en code OWL :

Si nous avons le bloc :

ARBRE

```
R      Peuplier
      Pin
      Sapin
```

Avec R le prédicat binaire représentant la relation de « Employé pour ». Dans cet exemple, le terme ARBRE a été choisi comme descripteur pour les termes Peuplier, Pin, Sapin qui sont donc des non-descripteurs. Cela donne le code suivant en OWL :

```

<owl:Class rdf:ID="ARBRE">
  <rdfs:label xml:lang="fr">ARBRE</rdfs:label>
  <rdfs:subClassOf rdf:resource="# thing "/>
</owl:Class>
< ARBRE rdf:ID=" Peuplier ">
  <rdfs:label xml:lang="fr"> Peuplier </rdfs:label>
  </ ARBRE >
< ARBRE rdf:ID=" Pin ">
  <rdfs:label xml:lang="fr"> Pin </rdfs:label>
</ ARBRE >
< ARBRE rdf:ID=" Sapin ">
  <rdfs:label xml:lang="fr"> Sapin </rdfs:label>
</ ARBRE >

```

4.3.2 Les relations

Ce sont les éléments les plus importants à traduire car ils représentent la sémantique du thésaurus.

Relation de hiérarchie

La relation de hiérarchie représente la hiérarchie de descripteurs dans le thésaurus et la hiérarchie de concepts dans l'ontologie. Elle se traduit dans le langage OWL par la notion de sous-classe. Un exemple de traduction d'un extrait de thésaurus en code OWL est le suivant :

Le bloc :

```

ARBRE
EFIL      FORET

```

Avec EFIL le prédicat binaire représentant la relation « est fils de » entre deux descripteurs.

Se traduit en OWL :

```
<owl:Class rdf:ID=" FORET ">
  <rdfs:label xml:lang="fr"> FORET </rdfs:label>
  <rdfs:subClassOf rdf:resource="# thing "/>
</owl:Class>

<owl:Class rdf:ID=" ARBRE ">
  <rdfs:label xml:lang="fr"> ARBRE </rdfs:label>
  <rdfs:subClassOf rdf:resource="# FORET "/>
</owl:Class>
```

Relation d'association

Dans le thésaurus, la relation d'association lie deux termes descripteurs autrement que par la relation de hiérarchie. Dans l'ontologie, toutes les instances qu'ils représentent seront liées par cette relation. Dans un thésaurus, la relation entre deux descripteurs est unique. En revanche, dans une ontologie il peut en exister plusieurs. Cette relation sera traduite en OWL par une propriété commune aux deux classes. Ainsi les instances des deux classes posséderont cette propriété. Un exemple de traduction d'un extrait de thésaurus en code OWL est le suivant :

Le bloc :

```
ARBRE
ASSOC      PAPIER
```

Avec ASSOC le prédicat qui représente la notion d'association entre deux descripteurs.

donne le code OWL suivant :

(Ici nous aurons le code de déclaration de classe comme précédemment.)

```
<owl:ObjectProperty rdf:ID=" ARBRE--- PAPIER ">
  <rdfs:domain rdf:resource="# ARBRE"/>
  <rdfs:range rdf:resource="# PAPIER "/>
</owl:ObjectProperty>
```

L'identifiant de la propriété pourra ensuite être changé pour mieux prendre en compte la sémantique de l'association.

Autres relations (hors norme)

La *relation de définition* entre un terme du thésaurus et une chaîne de caractères est transformée en un commentaire spécial dans le fichier OWL. Nous proposons d'écrire un commentaire débutant par « Définition : », ce qui permettra par la suite de retrouver cette définition. Ainsi la transformation ne souffre d'aucune perte d'informations. Nous prenons l'exemple suivant dans lequel Peuplier est Non-descripteur.

Peuplier

DEF **Arbre élevé, de forme élancée, au bois blanc et au feuillage caduc très mobile d'un vert pâle, qui croît rapidement en des lieux humides.**

Avec DEF le prédicat unaire représentant la « définition » entre un non-descripteur ou un descripteur et sa définition (chaîne de caractères).

qui se traduit par le code OWL:

```
<owl:Class rdf:ID=" Peuplier ">
  <rdfs:label xml:lang="fr"> Peuplier</rdfs:label>
  <rdfs:comment xml:lang="fr">Définition : Arbre élevé, de forme élancée, au bois blanc et au feuillage caduc
très mobile d'un vert pâle, qui croît rapidement en des lieux humides.</rdfs:comment>
  <rdfs:subClassOf rdf:resource="#thing"/>
</owl:Class>
```

Cette transformation peut être généralisée à toute sorte de propriété autre que la définition.

Enfin, la *relation de traduction* lie un terme dans une langue à un terme synonyme dans une autre langue. Dans ce cas, une instance ou une classe va être construite pour tous ces termes synonymes de langues différentes, et autant de labels de langues différentes qu'il y a de traduction dans le domaine.

Par exemple :

ARBRE	
ANG	TREE
ESP	ÁRBOL
ALL	BAUM

donne en code OWL :

```
<owl:Class rdf:ID=" ARBRE ">
  <rdfs:label xml:lang="fr">ARBRE </rdfs:label>
  <rdfs:label xml:lang="en">TREE </rdfs:label>
  <rdfs:label xml:lang="es">ÁRBOL </rdfs:label>
  <rdfs:label xml:lang="de">BAUM </rdfs:label>
  <rdfs:subClassOf rdf:resource="#thing"/>
</owl:Class>
```

Arbre étant un descripteur d'une classe OWL qui a été construite, cette classe possède quatre labels de langues différentes.

5 Conclusions et perspectives

Cet article avec pour objectif de présenter une méthode de transformation d'un thesaurus en une ontologie, ainsi que son implantation. La transformation des thésaurus en ontologies est primordiale dans la mesure où il existe de nombreuses ressources terminologiques sous forme de thésaurus. Les applications issues du web sémantique vont se développer mais nécessitent d'avoir en amont des ontologies. Notre contribution permet donc la réutilisation de ressources existantes pour de nouvelles applications.

6 Bibliographie

- Aussenac-Gilles, N., Mothe J, (2004). Ontologies as Background Knowledge to Explore Document Collections, Actes de la Conférence sur la Recherche d'Information Assistée par Ordinateur (RIAO), pp 129-142.
- Bachimont B. (2000). Engagement sémantique et engagement ontologique : conception et réalisation d'ontologies en ingénierie des connaissances, Ingénierie des connaissances : évolutions récentes et nouveaux défis, pp 305–323, Eyrolles.
- Berners-Lee T., Hendler J., Lassila O., (2001). The Semantic Web, Scientific American, pp 28–37.
- Chaumier, J., (1988). Le traitement linguistique de l'information. 3e éd., Paris : Entreprise Moderne d'Édition.
- Chrisment C., Hernandez N., Genova F., Mothe J. (2006). D'un thesaurus vers une ontologie de domaine pour l'exploration d un corpus. AMETIST, INIST, Vol. 0, p. 59-92.
- Dégez, D., Ménillet, D., (2001). Thésauroglossaire des langages documentaires : un outil de contrôle sémantique, Sciences de l'information. Recherches et documents, ISSN 1159-7666.
- Foskett, D. J. (1980). Thesaurus. Encyclopedia of library and information science. New York.
- Gómez-Pérez A. , Moreno A., Pazos J., Sierra-Alonso A., (2000). Knowledge Maps: An essential technique for conceptualisation, Data Knowledge Engineering, 33(2):169-190.

- Gruber T.R. (1993). A translation approach to portable ontology specifications, *Knowledge Acquisition*, 5 (2), pp 199-220.
- van Heijst G., Schreiber G., Wielinga B., (1997). Using explicit ontologies for KBS development, *International Journal of Human-Computer Studies*, 42(2/3), pp 183-292.
- McGuinness D., van Harmelen F., (2004). OWL Web Ontology Language, Overview, WWW Consortium, Recommendation REC-owl-features-20040210, <http://www.w3.org/TR/2004/REC-owl-features-20040210>
- RAMEAU: Guide d'indexation RAMEAU (2006). 6e édition, ISBN 2-7177-2315-3.
- Staab S., Maedche A. (2000). Axioms are objects too: Ontology engineering beyond the modeling of concepts and relations, Research report 399, Institute AIFB, Karlsruhe.
- Uschold M., King M. (1995). Towards a Methodology for Building Ontologies. Basic Ontological Issues in Knowledge Sharing, Inter. Conf. on Artificial Intelligence (IJCAI).
- ISO 2788 :1986 Principes directeurs pour l'établissement et le développement de thésaurus monolingues, http://www.iso.org/iso/fr/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=7776
- ANSI/NISO Z39.19-2003 Lignes directrices pour la construction, la mise en forme et la gestion de thésaurus monolingues, http://www.techstreet.com/cgi-bin/detail?product_id=1171385
- ISO 5964 Principes directeurs pour l'établissement et le développement de thésaurus multilingues.
- AFNOR Z47, http://www.boutique.afnor.org/NEL5DetailNormeEnLigne.aspx?CLE_ART=FA015919&nivCtx=NELZNELZ1A10A101A107&ts=3671626