D'AGROVOC à l'Agricultural Ontology Service / Concept Server

Un modèle OWL pour la création d'ontologies dans le domaine de l'agriculture

A.C. Liang

Perot Systems, Inc. Tel : +1-240-478-5948 acliang@alum.mit.edu

Boris Lauser

Food and Agriculture Organization of the United Nations, Rome, Italie Tel : +390657054638 Fax : +390657054049 boris.lauser@fao.org

Margherita Sini

ORGANISATION DES NATIONS UNIES POUR L'ALIMENTATION ET L'AGRICULTURE (FAO), Rome, Italie Tel : +390657056805 Fax : +390657054049 margherita.sini@fao.org

Johannes Keizer

ORGANISATION DES NATIONS UNIES POUR L'ALIMENTATION ET L'AGRICULTURE (FAO), Rome, Italie Tel : +390657052729 Fax : +390657054049 johannes.keizer@fao.org

Stephen Katz

ORGANISATION DES NATIONS UNIES POUR L'ALIMENTATION ET L'AGRICULTURE (FAO), Rome, Italie Tel : +390657053774 Fax : +390657054049 stephen.katz@fao.org

Cet article décrit la conversion du thésaurus AGROVOC dans sa forme traditionnelle en un service d'ontologies dans le domaine de l'agriculture (Agricultural Ontology Service/Concept Server AOS/CS). Ce Serveur de concepts (AOS/SC) est un entrepôt multilingue de concepts dans ce domaine proposant des relations ontologiques et une terminologie étendue et sémantiquement riche. L'Organisation des Nations Unies pour l'alimentation et l'agriculture, FAO (Food and Agriculture Organization) a récemment développé le modèle pour ce nouveau système dans le langage d'ontologies : OWL. Dans cet article, nous décrivons l'objectif de cette conversion et l'utilisation du langage OWL ; nous mettons en évidence en particulier les principales fonctionnalités de ce modèle OWL. Nous avons voulu expliquer comment il évolue et diffère de l'approche traditionnelle des thesauri.

Ontologie, Thésaurus, Web sémantique, OWL, Schéma de classification, Métadonnée, AGROVOC, AOS

1 Contexte

Depuis 2003, l'Organisation des Nations Unies pour l'alimentation et l'agriculture (FAO) a eu à cœur de développer un nouveau modèle pour le thésaurus AGROVOC qui prenne en considération les relations sémantiques et lexicales de manière plus fine et plus précise, avec pour objectif de constituer une base multilingue de concepts dans le domaine de l'agriculture : le Serveur de concepts (SC).

Cet effort s'inscrit dans le projet global de la FAO de constitution d'un Service d'Ontologie dans le domaine de l'agriculture (Agriculture Ontology Service, AOS). Celui-ci doit fonctionner comme un outil qui permettrait de structurer et de standardiser la terminologie en agriculture en plusieurs langues afin de l'utiliser dans différents systèmes dans le monde. Il sera possible, à partir du SC d'exporter le thésaurus traditionnel AGROVOC ainsi que d'autres formes de systèmes d'organisation des connaissances (KOS)[1]. Il sera également possible d'en extraire des concepts ontologiques et de les utiliser pour construire des ontologies spécifiques par domaine de connaissances.

Au cours de cette étude, un certain nombre de modèles et d'approches ont été étudiés et proposés. Au départ, on a pensé qu'une base de données relationnelle représentait une solution de stockage avantageuse en raison de :

- sa gestion facile, ses performances et sa capacité à monter en charge ;
- sa similitude avec le format actuel et la capacité à assurer la rétrocompatibilité ;

[1] Les KOS sont des structures de connaissances, qui incluent des fichiers d'autorité, des systèmes de classification, des espaces de concepts, des dictionnaires, des listes contrôlées, des taxonomies, des référentiels géographiques, des glossaires, des ontologies, des autorités matières, des thesauri etc...

- l'utilisation de bases de données relationnelles pour stocker d'autres terminologies à intégrer dans AGROVOC, telles que FAOTERM ou le glossaire de la FAO.

Par la suite, on a étudié la possibilité d'utiliser le langage d'ontologies Web OWL pour représenter le modèle du SC. OWL suscite un intérêt croissant chez les chercheurs dans de multiples disciplines, notamment la médecine, la défense, l'agriculture, la biologie, les sciences de l'information et on développe des technologies toujours plus performantes et complexes de création et d'utilisation d'ontologies OWL. Bien qu'il soit nécessaire d'effectuer d'autres tests sur les bases OWL et les bases de stockage de triplets afin de déterminer leur performance et leur capacité à monter en charge, il semble que suffisamment d'arguments plaident en faveur d'une transition vers OWL plutôt que la création d'une nouvelle base de données relationnelle ad-hoc.

Premièrement, l'un des objectifs majeurs du service d'ontologies AOS est la promotion de standards et de l'interopérabilité des systèmes d'informations dans le domaine de l'agriculture. Concevoir un nouveau système propriétaire de gestion de terminologie serait contraire à cet objectif. En revanche, l'utilisation d'un standard reconnu tel qu'OWL permettra d'optimiser l'interopérabilité avec d'autres systèmes. Des outils en code source libre (*Protégé*, SWOOP, etc.) et procédés existants peuvent être utilisés pour gérer le modèle, et réutilisés et adaptés à des applications locales, limitant de ce fait les efforts de développement nécessaires.

Deuxièmement, un schéma de base de données personnalisé n'est pas directement interopérable avec d'autres solutions de stockage. Au contraire, un standard reconnu tel que OWL, basé sur XML/RDF, est déjà interopérable avec n'importe quelle base de triplets RDF, ce qui permet une intégration aisée d'autres sources de données en RDF/XML pour le stockage, et un traitement et une visualisation directs des données. Mais le langage OWL, c'est plus que du RDF. L'utilisation d'OWL permet de partager facilement des ontologies sur le Web, puisqu'il est possible de définir des équivalences entre des classes et des entités dans plusieurs terminologies. Il est également possible de vérifier la cohérence des ontologies ainsi reliées, d'identifier et de résoudre des conflits éventuels et de déduire des informations nouvelles à partir des relations exprimées.

Troisièmement, l'utilisation d'un standard reconnu tel que le modèle OWL limitera les besoins en formation. Il suffit de se référer à la documentation OWL publique, plutôt que d'avoir à créer de multiples documentations nouvelles pour un système propriétaire.

Enfin, OWL a désormais le statut de recommandation du W3C, ce qui signifie qu'il est devenu une spécification stable de grande qualité technique et qu'il est voué à un large déploiement au service de l'interopérabilité sur l'Internet.

Sur la base de ces considérations, nous avons développé un nouveau modèle OWL qui servira de squelette pour la création d'ontologies dans le domaine de l'agriculture. Dans cet article, nous allons présenter les éléments les plus importants de ce modèle, qui serviront de base pour le futur Serveur de concepts AOS/SC. Nous détaillons également les problèmes liés au multilinguisme et décrivons les solutions adoptées. Nous n'expliquerons cependant pas les principes de base d'OWL dans cet article, nous considérons que le lecteur connaît bien les ontologies et les concepts de base d'OWL. Pour plus d'informations sur OWL, le lecteur pourra se référer à (OWL, 2004). Nous avons utilisé *Protégé 3.2* comme outil de modélisation, un outil aujourd'hui largement utilisé ; c'est un éditeur d'ontologies en code source libre développé en Java à l'université de Stanford (*Protégé ontologie Editor*)[2]. Les captures d'écran utilisées pour illustrer cet article ont été créées à partir de *Protégé*.

2 Exprimer la sémantique d'AGROVOC en OWL

L'objectif de transformer AGROVOC en un modèle OWL avec une structure proche de celle d'une ontologie est de :

- faciliter son utilisation pour le développement de terminologies, y compris d'ontologies, dans le domaine de l'agriculture, et ainsi éviter d'avoir à reconstruire des terminologies à partir de rien ;
- permettre le développement d'applications utilisant les techniques sémantiques,
- permettre l'interopérabilité entre applications utilisant ces ontologies.

Nous avons pris comme source AGROVOC qui se prête particulièrement bien à une transformation en

[2] Protégé Ontology Editor. http://protege.stanford.edu/

ontologie, puisque, contrairement à d'autres glossaires ou listes de termes ordinaires, il contient la sémantique explicite d'une structure hiérarchique des descripteurs (termes représentant les concepts en agriculture). Il contient aussi des relations associatives génériques qui indiquent une relation sémantique entre deux entités et qui peuvent être redéfinies ultérieurement en relations plus spécifiques. Les listes d'espèces végétales et animales, les entités géopolitiques ainsi que les substances chimiques forment des taxonomies naturelles dont la sémantique peut aisément être exprimée en OWL. De même, les attributs (par exemple, le nombre de pattes, la superficie) et les relations non hiérarchiques (par exemple, l'appartenance à un groupe, les parties d'une plante) peuvent aussi être exprimés en OWL.

3 La question du multilinguisme

Dans la phase de préparation d'AGROVOC pour son utilisation en tant qu'ontologie, il est essentiel de représenter les concepts en minimisant les biais en faveur d'une langue ou famille de langues donnée. C'est-à-dire que, dans la mesure du possible, nous considérons le sens indépendamment de sa réalisation dans une langue particulière. Chaque langue serait donc capable d'exprimer les concepts du domaine pour lesquels elle a des lexicalisations et les concepts pour lesquels elle n'en a pas. Une terminologie qui traduirait simplement les termes dans une langue donnée, par exemple l'anglais, perdrait les concepts qui ne sont pas lexicalisés dans cette langue. Par exemple, le mot italien « *oculo* » qui désigne un « endroit qui contient un cercueil ou une urne funéraire » n'a pas d'équivalent en anglais. D'autres exemples similaires sont rencontrés comme dans les langues asiatiques les concepts liés au riz ou à la mangue. Une terminologie multilingue qui serait centrée sur l'anglais comme l'est dans une certaine mesure AGROVOC ne permettrait pas de rendre compte de ces concepts. Ainsi, la refonte proposée de la structure terminologique d'AGROVOC donnera un modèle du domaine qui sera conceptuellement plus riche qu'un modèle basé sur une seule langue et des traductions. Ce modèle devrait permettre non seulement de prendre en compte des concepts existants dans différentes langues (et, donc, dans différentes cultures), mais aussi de représenter les relations lexicales à la fois à l'intérieur d'une même langue et entre plusieurs langues. Cela permettrait d'établir des équivalences lexicales précises (p. ex. synonymes, traductions), de traiter de manière efficace les termes et concepts et d'optimiser la valeur de l'ontologie pour un grand nombre d'applications.

Les trois niveaux de représentation que nous souhaitons exprimer dans ce modèle sont :

- les concepts (la signification abstraite), par exemple le « *riz* » au sens de la plante ;
- les termes (formes lexicales spécifiques à une langue), par exemple 'Riz', 'Rice', 'Arroz', '稻米', 'ข้าว' ou 'Paddy' ;
- les variantes de terme (les différentes formes que peut prendre un terme), par exemple 'O. sativa' ou 'Oryza Sativa', 'Organisation' ou 'Organization' en anglais.

Les concepts abstraits constituent la hiérarchie et la structure sémantique de l'ontologie. Les termes ne sont plus organisés selon une structure hiérarchique ou par leurs relations sémantiques, comme c'est actuellement le cas dans AGROVOC. Chaque terme est une entité distincte dans chaque langue qui peut être liée à des concepts ou à d'autres termes et à d'autres variantes du même terme.

Ces distinctions nous permettent d'établir les relations hiérarchiques suivantes :

Concept vers terme	*has_lexicalization* (lie les concepts à leurs réalisations lexicales)
Terme vers sa variante	*has_acronym*, *has_spelling_variant*, *has_abbreviation* (lient les termes à leurs variantes)

Les variantes de terme ne constituent pas de nouveaux termes mais sont les formes variables du même terme.

Les relations internes à un niveau existent tant au niveau du concept qu'au niveau du terme, le tableau illustrant la notion :

Concept vers concept	*is_a* (indique une hiérarchie) ; ex : *pest_of*, *pest*, etc.
Terme vers terme	*is_synonym_of*, *is_translation_of*

4 Le modèle OWL

4.1 Les sous-langages ou espèces OWL

Dans cet article, nous présentons un modèle OWL qui permet de rendre compte des distinctions conceptuelles et lexicales présentées ci-dessus tout en conservant la complétude computationnelle. La conception des multiples niveaux de représentation lexicale présentés dans cet article (classes, propriétés, annotations) est donc effectuée dans la version d'OWL connue sous le nom de OWL DL[3].

4.2 Le modèle de base

Le nouveau modèle OWL est fondé sur trois concepts au sommet de la hiérarchie, comme indiqué dans la Figure 1. Chaque entité d'une ontologie OWL a un URI (identifiant uniforme de ressource)[4] unique. Dans la Figure 1, seule la dernière partie identifiante de l'URI est visible. Par convention, chaque URI d'entité dans notre modèle est composé d'un préfixe, c_ (pour les classes), r_ (pour les relations/propriétés[5]), i_ (pour les instances), suivi d'une séquence numérique ou alphanumérique.

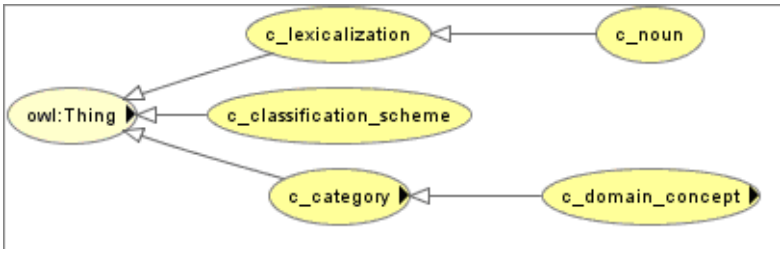

Fig. 1 : Concepts de niveau supérieur

La racine de tous les concepts d'un domaine est le concept « c_domain_concept ». Il constitue la structure hiérarchique centrale du Serveur de concepts AOS. Ce nœud a toutes les caractéristiques structurelles de l'ontologie de domaine, à savoir une hiérarchie de classe avec des classes et leurs instances, leurs relations, propriétés, axiomes, contraintes et annotations afférents au domaine de connaissance. Tous les termes du thésaurus AGROVOC ou plus précisément les descripteurs AGROVOC seront modélisés à l'intérieur de ce nœud.

La classe « *c_domain_concept* » est modélisée comme une sous-classe de « *c_category* », ce qui implique que tous les concepts du domaine sont également potentiellement des catégories. La classe distincte « *c_category* » a été créée pour pouvoir prendre en compte des catégories spécifiques qui ne sont pas des concepts de domaine. Les catégories sont organisées en schémas de classification représentés par la classe « *c_classification_scheme* ». Le thème des catégories et schémas de classification sera détaillé dans la partie 4.5.

Alors que la structure centrale de l'ontologie de domaine est modélisée sous « *c_domain_concept* », les lexicalisations de ces concepts apparaîtront comme des instances de la classe « *c_lexicalization* ». Cette approche a été choisie plutôt que l'utilisation du rdfs:label pour chaque concept afin de représenter sa lexicalisation dans une langue particulière. Elle permet de gérer la question du multilinguisme. La modélisation des lexicalisations comme des concepts distincts permettra d'établir des relations entre les différentes lexicalisations qui décrivent un concept. Elle fournira ainsi une sémantique plus riche.

---

[3] Voir aussi : http://www.w3.org/TR/owl-ref/#Sublanguages

[4] Nous n'aborderons pas les URI dans le détail . Vous pouvez vous référer à l'adresse suivante pour tout complément d'informations sur les URI http://www.w3.org/Addressing/

[5] Dans cet article, nous utilisons les termes propriété et relation comme synonymes. Propriété (*property*) est un terme utilisé dans le monde des ontologies et d'OWL alors que les relations sont plus courantes dans le monde des thesauri traditionnels.

4.3 La structure centrale hiérarchique

Dans un premier temps, les termes d'AGROVOC (plus précisément ses principaux descripteurs) constitueront la structure centrale hiérarchique du modèle. Tous les descripteurs d'AGROVOC seront modélisés comme des sous-classes de « *c_domain_concept* » en utilisant le code du terme AGROVOC pour former un URI de classe (par exemple « *c_208* » pour le concept 'Agriculture'). Les relations du thésaurus traditionnel « *Terme spécifique* » et « *Terme générique* » sont ensuite traduites en relations superclasse OWL et sous-classe OWL. AGROVOC permet ainsi de constituer la hiérarchie initiale du Serveur de concepts (SC).

4.3.1 Associer des concepts : l'interface de gestion des concepts

AGROVOC (tout comme d'autres thesauri[NDT] classiques) ne propose qu'un seul type de relations conceptuelles non hiérarchiques, le *terme associé.* Dans notre modèle, nous avons voulu qu'il soit possible d'associer des concepts qui ont des relations plus spécifiques. Nous introduisons à cet effet une hiérarchie de relations pour les relations entre concepts. Chaque relation conceptuelle spécifique (telle que « *is part of* », « *is infected by* », etc.) est modélisée comme sous-propriété de « *r_has_related_concept* » telle que présentée dans les quelques exemples de la Figure 2.

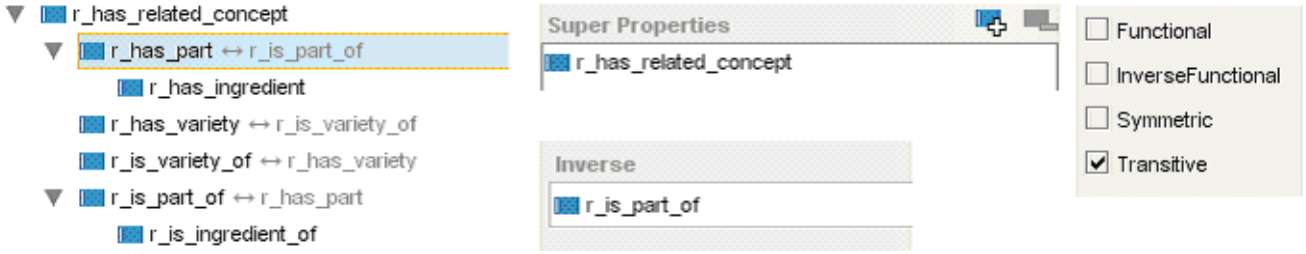

Fig. 2 : Organisation hiérarchique des relations entre concepts

Comme nous utilisons ces relations pour définir un concept, c'est-à-dire pour l'associer à d'autres concepts, le domaine et le type de toutes ces relations sont définis au niveau de « *c_domain_concept* ». La hiérarchie de relations est importante pour la rétrocompatibilité avec les exports de thesauri classiques, c'est-à-dire que toutes les relations entre concepts peuvent toujours être définies par la relation la plus générique, « *has related concept* » [a un concept associé] (équivalent pour un thésaurus à *terme associé*). Nous proposons une liste initiale de relations plus fines entre concepts, qui pourra être révisée par la suite[6].

De plus, nous introduisons « *r_domain_specific_relationship* » pour qu'il soit possible de créer des relations conceptuelles qui ne sont valables que dans un domaine spécifique. Cela peut être utile pour que des applications puissent filtrer de telles relations spécifiques. Toutes les propriétés de ce type sont à la fois des sous-propriétés de « *r_has_related_concep*t » et de « *r_domain_specific_relationship* ».

4.4 Les lexicalisations

Dans le chapitre précédent, nous avons introduit le modèle pour créer l'ossature conceptuelle du Serveur de concepts (SC). Nous devons à présent introduire les lexicalisations afin de représenter cette structure en plusieurs langues. Toutes ces informations lexicales sont incluses dans le concept « *c_lexicalization* ». Chaque terme (c'est-à-dire chaque lexicalisation ou mot)[7] qui décrit un concept dans une langue spécifique est modélisé comme une instance de ce concept.

L'instance URI est composée de *i_* suivi du code de langue à deux lettres ISO639 de ce terme, suivi par le terme lui-même (en utilisant des tirets bas pour remplacer les espaces et les caractères spéciaux). Si une forme particulière d'un mot s'avère avoir un homonyme dans une langue, un tiret bas est ajouté, suivi d'un chiffre, par exemple « *en_sole_1* » (la semelle, en anglais), « *en_sole_2* » (la sole, en anglais). L'annotation rdfs:label est utilisée pour fournir le label du terme à afficher. La Figure 3 montre une copie d'écran de *Protégé* avec quelques instances de « *c_lexicalization* ».

Les instances sont en réalité des instances de « *c_noun* », un sous-concept de « *c_lexicalization* ». Cela laisse le modèle suffisamment ouvert pour inclure d'autres formes telles que des verbes, des

[6] Relations proposées par le Serveur de concept CS : http://www.fao.org/aims/cs_relationships.htm

[7] Nous utiliserons ces trois formes comme synonymes, c'est-à-dire que c'est un terme/lexicalisation/mot qui représente un concept.

adjectifs, entre autres par la suite.

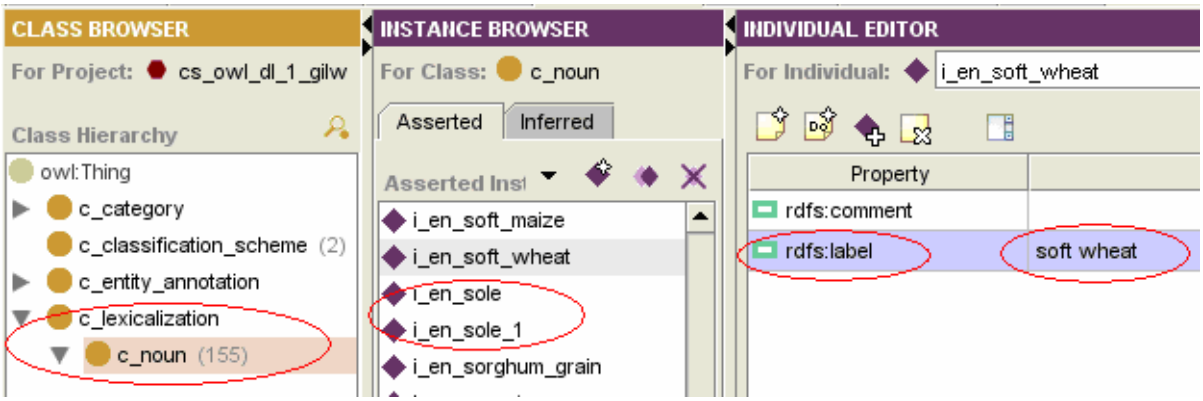

Fig. 3 : Représentation de la désambiguation des termes et URI

La décision de traiter les termes comme des instances plutôt que comme des annotations (p. ex., rdfs:label) est principalement motivée par le fait que les relations en OWL DL peuvent seulement être définies entre deux individus ou entre un individu et un littéral. Afin de pouvoir également exprimer des relations entre les termes (telles que les relations de traduction et de synonymie), les termes doivent être réalisés comme instances. Nous allons présenter brièvement les relations terme à terme, mais auparavant nous allons expliquer la manière d'associer les termes aux concepts du domaine.

4.4.1 Associer les lexicalisations à des concepts : l'interface de gestion des relations Concept - Terme

Les termes sont associés au concept dont ils lexicalisent le sens via deux propriétés d'objets OWL, « *r_has_lexicalization* » et la relation inverse, « *r_means* », représentées sur la Figure 4.

Nous avons modélisé les relations au niveau de « *c_category* » puisque nous traitons de la même façon les lexicalisations des catégories et des concepts du domaine. La classe « *c_domain_concept* » hérite des relations de « *c_category* ».

Chaque instance de « *c_lexicalization* » est liée à une et une seule instance de « *c_category* » ou de « *c_domain_concept* ». Une catégorie ou un concept de domaine sera généralement associé à plusieurs instances de « *c_lexicalization* » ; au moins une pour chaque langue dans laquelle il est disponible et d'autres pour les synonymes et noms scientifiques.

Il reste à déterminer quels effets cela aura sur la performance des applications qui utilisent la terminologie.

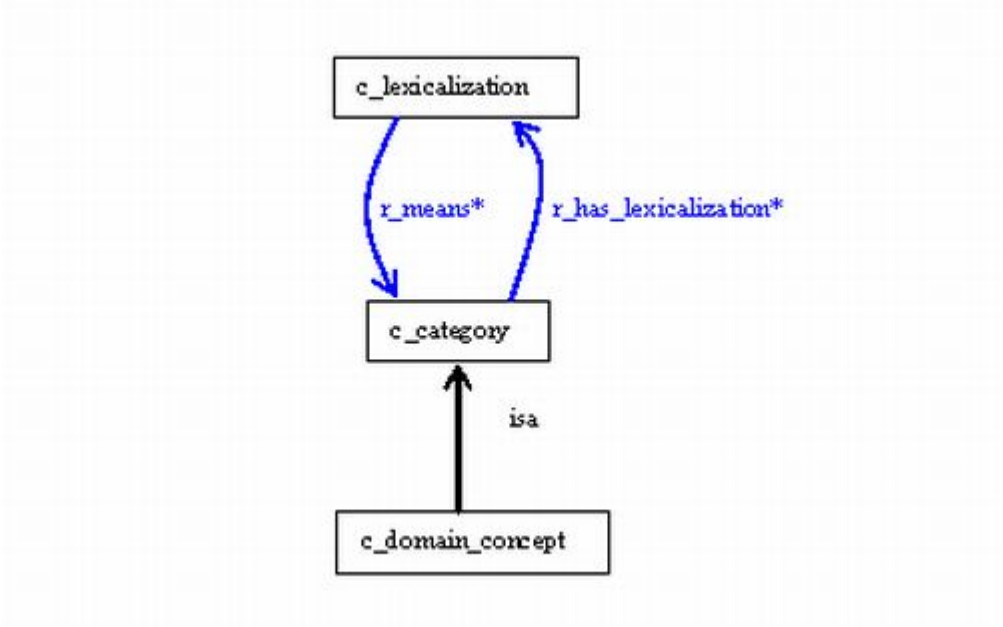

Fig. 4 : Associer des termes à des concepts

4.4.2 Associer les lexicalisations : l'interface de gestion des relations entre termes

Afin d'associer deux termes (ou lexicalisations) entre eux, nous introduisons la propriété « *r_has_related_term* ». Cette propriété est la super-propriété de toutes les relations entre termes. Il est important de noter que cette relation NE correspond PAS à la relation *terme associé* des thesauri classiques, puisqu'elle décrit une relation conceptuelle, pas une relation entre termes. Nous avons initialement identifié trois associations possibles entre termes. Un terme peut avoir :

- une ou plusieurs traductions ;
- un ou plusieurs synonymes par langue ;
- un ou plusieurs noms scientifiques.

La Figure 5 présente la hiérarchie des propriétés telle qu'elle est modélisée dans *Protégé*. Le domaine et le type OWL de toutes les propriétés sont définis au niveau de « *c_lexicalization* ».

« *r_has_synonym* » et « *r_has_translation* » sont des relations symétriques, alors que « *r_has_scientific_taxonomic_name* » est une relation unidirectionnelle pour laquelle nous avons introduit une propriété inverse. Les relations des thesauri traditionnels EMPLOYER et EMPLOYE POUR sont initialement traduites dans des relations « *r_has_synonym* » lors de la migration d'AGROVOC vers le nouveau modèle.

Ce modèle fournit une grande flexibilité au niveau lexical. Il est par exemple possible d'exprimer qu'un terme anglais « *corn* » a un synonyme en anglais « *maize* ». Le terme français « *maïs* » décrit le même concept mais n'est une traduction que du terme anglais « *maize* ». « *Corn* » n'a pas de traduction en français. Notre modèle est capable d'exprimer ce cas de figure mais aussi de fournir plusieurs traductions pour un terme.

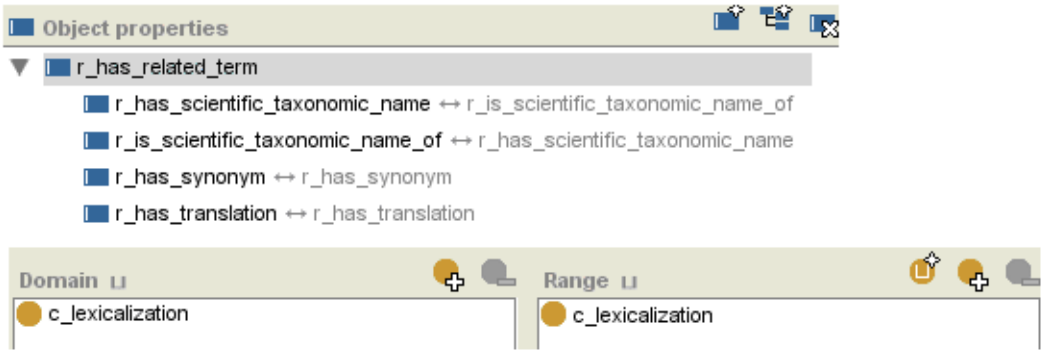

Fig. 5 : L'organisation hiérarchique des propriétés de termes

La Figure 6 représente l'ensemble du modèle, c'est-à-dire la relation du modèle lexical avec la structure centrale en utilisant l'exemple « *corn/maize* ». La représentation est réalisée avec *OntoViz* (un plug-in de *Protégé*). La partie supérieure de l'image montre le modèle conceptuel, alors que la partie inférieure affiche les instanciations avec leurs relations. Comme OWL DL permet seulement d'associer deux instances avec une relation, pour associer une instance de lexicalisation au concept qu'elle décrit, nous devons créer une instance du concept. L'URI de l'instance du concept du domaine est identique au nom du concept mais commence par *i_* au lieu de *c_*. Le graphique montre le concept « *corn/maize* » (qui correspond au code de terme AGROVOC 12332) associé à ses deux lexicalisations anglaises « *corn* » et « *maize* » via la paire de relations « *r_has_translation* / *r_means* » (interface de gestion concepts-termes). Les deux termes sont ensuite associés avec la relation symétrique « *r_has_synonym* » (interface de gestion des relations entre termes). Les flèches bleues sur la partie inférieure de l'image sont donc des instances des propriétés modélisées pour les concepts de la partie supérieure de l'image.

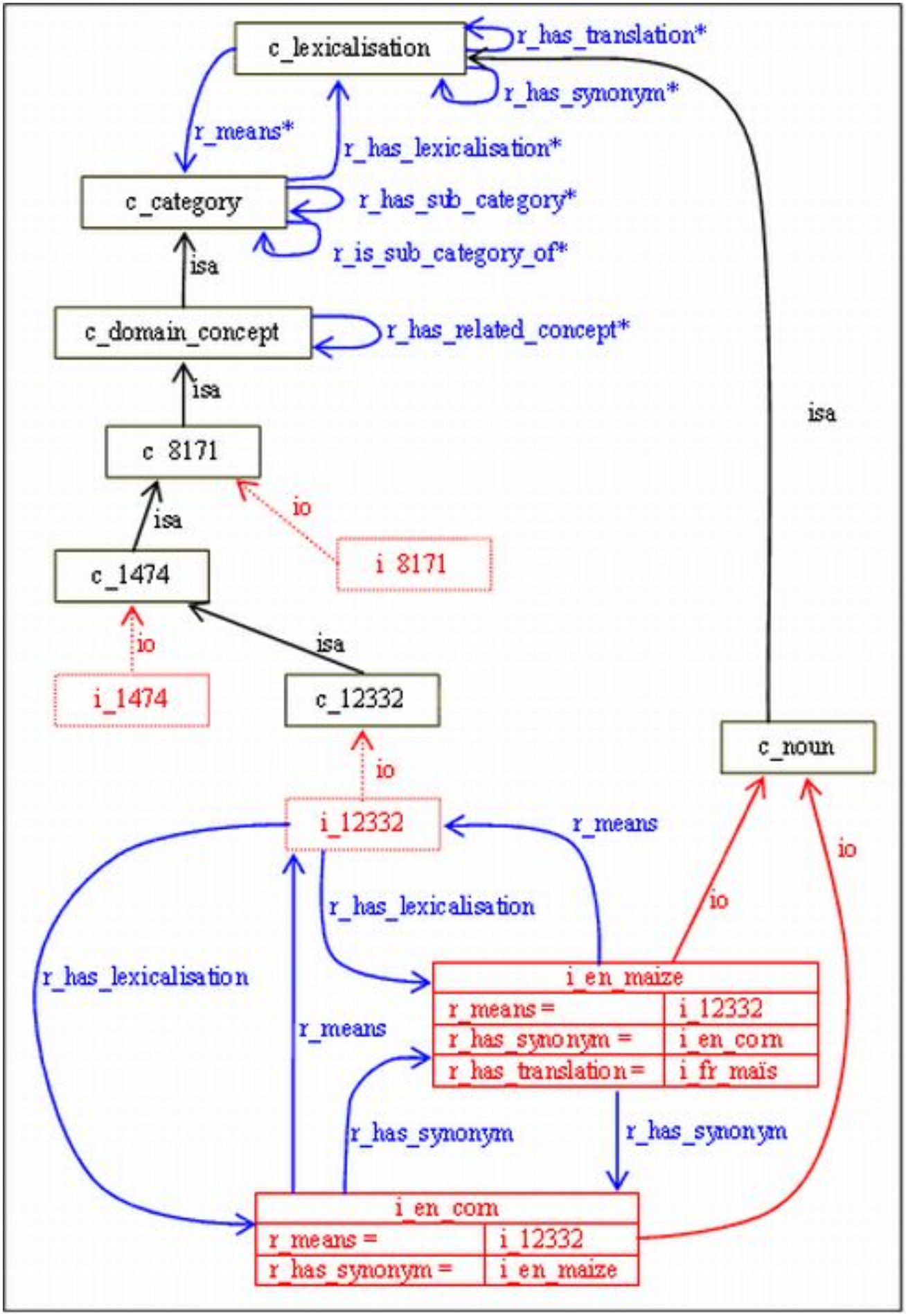

Fig. 6 : Représentation d'un concept avec deux synonymes

4.4.3 Gérer les variantes de termes : l'interface de gestion des relations entre termes et variantes

Les termes eux-mêmes peuvent être représentés de différentes manières. Par exemple, le terme « *université de Californie à Berkeley* » a les variantes suivantes :

- UCB (acronyme) ;
- Cal (forme raccourcie) ;
- UC Berkeley (abréviation) ;
- University of California at Berkeley (nom officiel).

Un terme est associé à ses variantes grâce à des propriétés de type de données telles que rdfs:label et à des propriétés spécifiques telles que « *has acronym* », « *spelling variant* » et « *abbreviation* ». Suivant notre organisation hiérarchique des propriétés, nous avons modélisé ces relations comme des sous-propriétés de la propriété de type de données « *r_has_term_variant* ».

Le domaine de ces propriétés est défini dans « *c_lexicalization* », alors que leur type est une simple variante. Cela implique qu'aucune nouvelle relation ne peut être établie entre des acronymes, des abréviations ou des variantes orthographiques. Jusqu'à présent, nous n'avons pas considéré cela comme une limitation à l'expressivité lexicale de notre modèle.

4.5 Les schémas de classification

Une autre partie importante de notre modèle est le concept de « *c_classification_scheme* ». Un schéma de classification est habituellement une hiérarchie peu profonde (souvent 2 niveaux seulement) de catégories de niveau supérieur. Un schéma de classification reconnu dans le domaine de l'agriculture est le schéma AGRIS/CARIS[8]. Les concepts du domaine peuvent être organisés dans un schéma de classification pour donner une vue d'ensemble sur les concepts du domaine. Des équivalents sont déterminés entre tous les termes AGROVOC et le schéma de classification AGRIS/CARIS. Notre modèle offre la possibilité d'avoir plusieurs schémas de classification mais également d'associer les catégories à des concepts du domaine.

La Figure 7 illustre ce modèle avec l'exemple du schéma de classification d'AGRIS/CARIS. Chaque catégorie est associée via la paire « *belongs_to_scheme* / *has_category* » à au moins un schéma de classification auquel il appartient (les catégories peuvent appartenir à plusieurs schémas de classification). La paire « *r_is_sub_category_of* / *has_subcategory* » est utilisée pour créer une hiérarchie à l'intérieur du schéma de classification. Nous introduisons ces relations spécifiques car nous voulons disposer d'un modèle suffisamment ouvert pour utiliser les concepts du domaine comme catégories. C'est la raison pour laquelle « *r_domain_concept* » est en réalité une sous-classe de « *r_category* ». La hiérarchie des concepts du domaine ne peut cependant pas être équivalente à une hiérarchie dans un schéma de classification particulier, c'est pourquoi, nous avons besoin d'une relation spécifique pour créer des hiérarchies de schémas de classification. Dans l'exemple, « *i_asc* » représente le schéma de classification AGRIS/CARIS et « *i_fao_pa* » un autre schéma appelé FAO Priority Areas. Le concept de domaine « *Education* » (*i_2488*) est en réalité une catégorie dans les deux schémas de classification, alors que la catégorie « *Education, Extension and Advisory Work* » (*asc:i_c*) est une catégorie thématique spécifique à AGRIS/CARIS. La relation de sous-catégorie n'existe donc que dans le schéma de classification AGRIS/CARIS. En pratique, cela apparaît dans le modèle grâce à l'utilisation de la propriété « *r_has_asc_sub_category* ». Dans l'outil de visualisation *Protégé*, la propriété de niveau supérieur, plus générique, « *r_has_sub_category* » a été utilisée. Il y aura ainsi une sous-propriété ou « *r_has_sub_category* » pour chaque schéma de classification pour pouvoir modéliser les différentes hiérarchies de schémas de classification avec les mêmes catégories ou concepts du domaine.

---

[8] Schéma de classification AGRIS/CARIS : http://www.fao.org/agris/Centre.asp?Content=DT&Menu_1ID=DT&Menu_2ID=DT1&Language=EN.

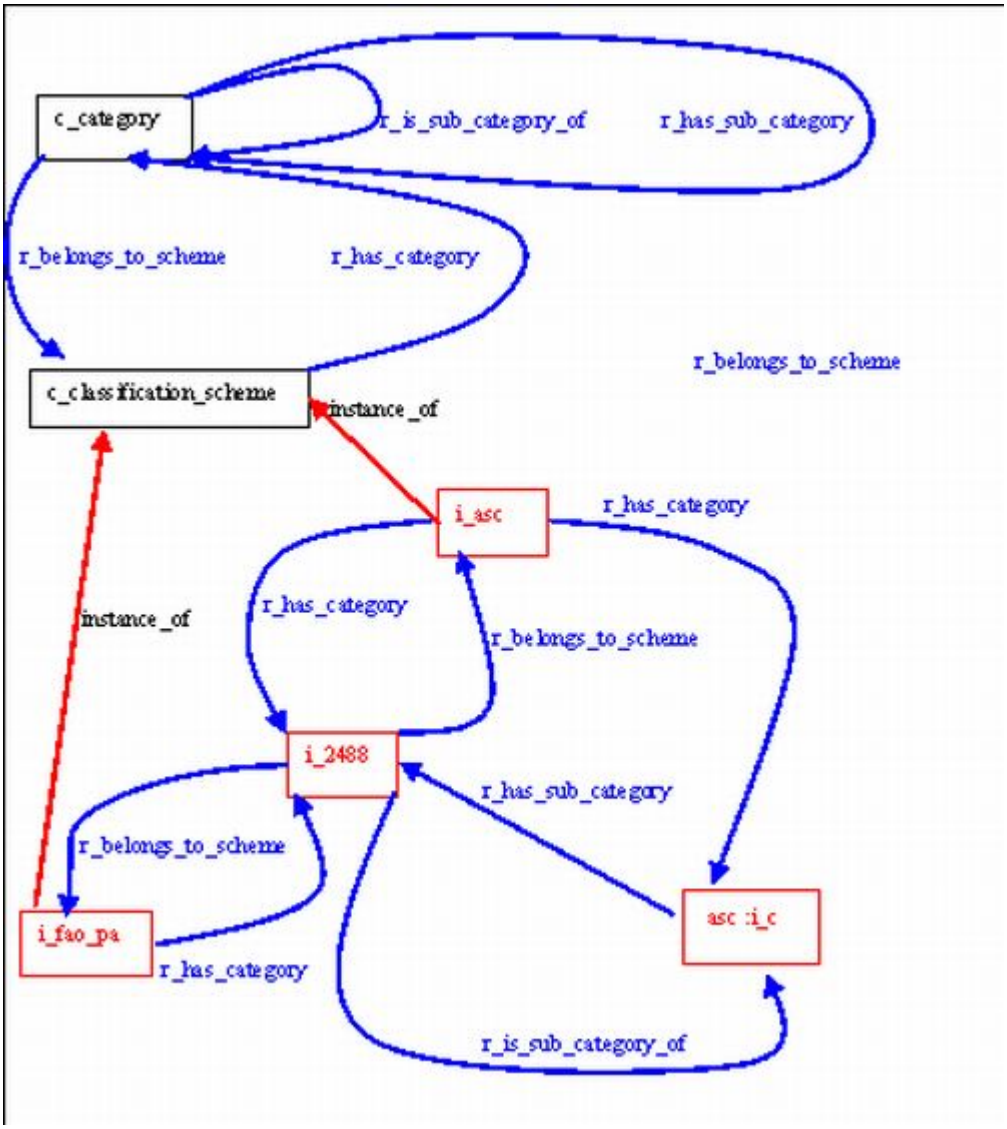

Fig. 7 : Représentation des schémas de classification et de leurs catégories

4.6 Annotations des concepts et des sous-vocabulaires

Les annotations des concepts sont des informations supplémentaires associées à des concepts du domaine ou à des catégories, certaines venant du monde des thesauri traditionnels comme les définitions, les commentaires, les notes d'application ou scope notes (précisent le concept d'un domaine), les images et l'historique (informations sur les modifications). Le modèle envisagé contient ces annotations des concepts modélisées comme des concepts séparés, associés à « *c_category* » ou *à « c_domain_concept ».* Il contient en outre de simples annotations telles que la date de création et de dernière mise à jour, le statut et la source (d'où le concept provient).

Une nouvelle notion émanant de la « note d'application » utilisée dans les thesauri est celle de sous-vocabulaires. Une sous-classe de « *c_domain_concept* » appelée « *c_geographic_concept* » a été introduite pour extraire les sous-structures géographiques spécifiques du Serveur de concepts (SC). En outre, « *c_scientific_name* », une sous-classe de « *c_lexicalization* », comme d'autres sous-classes « *c_taxonomic_name* » ou « *c_chemical_name* » permettront d'extraire du Serveur de concepts (SC) des taxonomies spécifiques ou des ensembles de noms chimiques avec leur hiérarchie de concepts et leur structure relationnelle. Nous désignons ces extractions comme des sous-vocabulaires.

4.7 Rétrocompatibilité

L'un des problèmes majeurs, qui se pose lors de la migration vers un nouveau système et de nouveaux formats, est la compatibilité avec les systèmes existants. Nous ne pouvons pas supposer que tous les utilisateurs d'AGROVOC vont spontanément arrêter d'utiliser le thésaurus dans sa version traditionnelle. Nous avons donc ajouté des annotations à notre modèle afin d'assurer une rétrocompatibilité totale lors de l'extraction du thésaurus AGROVOC tel qu'il est utilisé aujourd'hui à

partir du nouveau système.

5 La feuille de route : quel avenir ?

Maintenant que le modèle OWL pour le Serveur de concepts AOS/SC est réalisé, les terminologues doivent insérer les contenus avec la sémantique appropriée. Nous prévoyons de développer un outil de maintenance du système avec interface web, l'AOS Concept Server Workbench, qui pourra être utilisé par des experts et des terminologues du monde entier pour assurer l'enrichissement et la maintenance du système. Cet outil sera spécialement développé dans le but de modifier les structures terminologiques et conceptuelles complexes modélisées dans le Serveur de concepts AOS/SC. Il sera ainsi plus approprié que l'outil *Protégé* qui s'est avéré être trop lourd pour ce travail.

6 Conclusion

Le modèle OWL présenté dans cet article ainsi que la future interface AOS Workbench seront disponibles en code source libre et nous encourageons les terminologues du monde entier à utiliser le modèle OWL pour représenter leurs systèmes d'organisation des connaissances (KOS) et leurs systèmes terminologiques. Il est vrai qu'il existe d'autres normes et d'autres propositions pour les systèmes terminologiques et les thesauri. TermBase eXchange (TBX)[9]est une norme ISO pour représenter les terminologies en XML, pour leur échange et leur interopérabilité. Le format Simple Knowledge Organization System (SKOS)[10] est une proposition du W3C pour représenter les systèmes simples d'organisation des connaissances comme les thesauri qui incluent des hiérarchies entre les concepts. Notre modèle est différent des autres approches dans le sens où il combine les technologies nouvelles et émergentes du Web sémantique et l'environnement des systèmes terminologiques et des thesauri des bibliothèques traditionnelles. Notre modèle englobe les autres approches mentionnées, c'est-à-dire que nous fournirons des moyens de créer des extractions compatibles TBX ou SKOS à partir de notre système. Cependant, notre modèle offre en plus la possibilité de modéliser des structures ontologiques plus complexes qui peuvent être utilisées dans des systèmes plus élaborés. Prenez par exemple le cas d'un système d'alertes dans le domaine de la pêche, il est nécessaire de procéder à une modélisation conceptuelle détaillée afin de l'utiliser pour calculer des inférences et déduire de nouvelles informations à partir des modifications dynamiques effectuées dans l'ontologie.

Nous nous efforçons de faire du Serveur de concepts AOS/CS un premier point d'accès pour toute personne ayant besoin d'une ontologie ou d'un système terminologique complexe dans le domaine de l'agriculture ou dans les domaines proches.

[1] 1. OWL Web Ontology Language Reference, 2004. http://www.w3.org/TR/owl-ref.

[2] 3. Oxford-Paravia, 2002

[9] Site Internet TBX : http://www.lisa.org/standards/tbx/

[10] Site Internet SKOS : http://www.w3.org/2004/02/skos/