



HAL
open science

Modèle termino-ontologique pour la collaboration entre communautés hétérogènes

Benjamin Diemert, Marie-Helene Abel, Claude Moulin

► **To cite this version:**

Benjamin Diemert, Marie-Helene Abel, Claude Moulin. Modèle termino-ontologique pour la collaboration entre communautés hétérogènes. IC 2011, 22èmes Journées francophones d'Ingénierie des Connaissances, May 2011, Chambéry, France. pp.623-638. hal-00746728

HAL Id: hal-00746728

<https://hal.science/hal-00746728v1>

Submitted on 29 Oct 2012

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Modèle termino-ontologique pour la collaboration entre communautés hétérogènes

Diemert Benjamin, Marie-Hélène Abel, Claude Moulin

Université de Technologie de Compiègne, HeuDiaSyC UMR 6599
{benjamin.diemert, marie-helene.abel, claude.moulin}@utc.fr

Résumé : Dans cet article, nous traitons de la représentation de jargons métiers dans le cas d'une production collaborative de contenu audiovisuel. L'hétérogénéité de compréhension des concepts de la réalisation audiovisuelle aussi bien que les différences d'usages de la langue impliquent de documenter, d'illustrer et d'associer plusieurs terminologies à une même conceptualisation. Nous présentons un cas pratique servant à l'étude des modèles et normes existantes. Nous proposons un modèle termino-ontologique permettant de faciliter le partage et la gestion indépendante de terminologies et de modèles conceptuels.

Mots-clés : terminologie, ontologie, production collaborative

1. Introduction

Les nouvelles possibilités offertes par la mise en réseau des organisations facilitent l'échange de documents numériques. Pour autant l'échange de documents ne suffit pas à transmettre l'information et partager les connaissances qu'ils renferment. La collaboration entre des communautés d'horizons différents posent ainsi le double problème d'un consensus sur l'usage de la langue (trouver les mots pour se parler) et la conceptualisation sous-jacente (s'entendre sur la signification de ces mots).

L'utilisation de systèmes d'organisation des connaissances (SOC) permet de mettre en place des pratiques de partage d'informations normalisées. Parmi l'ensemble des pratiques existantes observées par Zacklad & Giboin (2010), on notera l'émergence de ressources termino-ontologiques associant des dénominations ou termes à des concepts. La distinction d'une dénomination de référence par rapport à ses variantes dans une même langue facilite ainsi la

normalisation de l'usage terminologique pour désigner un concept.

Lorsque la collaboration entre communautés est occasionnelle et l'usage des terminologies est déséquilibré ou propre à chaque organisation, la normalisation ne semble pas nécessaire et peut se ramener à une transposition d'un vocable à un autre. Par exemple, la production collaborative de contenu entre amateurs et professionnels aboutit à un dialogue entre une communauté au vocabulaire bien défini et presque formalisé conceptuellement avec une communauté sans aucun consensus linguistique ou conceptuel. Par ailleurs, la collaboration entre différents milieux professionnels aux pratiques linguistiques différentes mais partageant un fond conceptuel commun nécessite des moyens d'intégrer ou d'associer des terminologies. La langue n'est plus le seul élément de caractérisation des termes et il s'agit alors d'étendre cette caractérisation à d'autres codes d'écriture afin de couvrir les usages spécifiques à une communauté liée par un métier ou une organisation. Il est possible par exemple de modéliser l'appartenance d'un terme à un jargon ou à une notation métier et de proposer à des membres d'une autre communauté une variante, une explication voire une illustration graphique.

Notre contribution se concentre sur la modélisation de terminologies métiers attachées au même fond conceptuel et les relations nécessaires pour les associer, les expliciter ou les illustrer par des ressources graphiques. Pour faciliter le partage de ces ensembles conceptuels et terminologiques, nous proposons un modèle considérant chaque ensemble de manière distincte, ce qui permet de les gérer de manière indépendante.

A partir d'une application d'assistance au tournage, nous développons les motivations du problème de recherche abordé et explicitons les besoins en modélisations (section 2.). Nous étudions ensuite des modèles candidats et montrons ce qu'il leur manque pour satisfaire à ces besoins (section 3.). Ensuite, nous détaillons notre modélisation multi-jargon et discutons ses apports par rapports aux précédents modèles (section 4.) avant de conclure.

2. Motivations

Le travail présenté dans cet article a été réalisé dans le cadre du projet de MediaMap¹. Ce projet développe des solutions innovantes dans le domaine de la production audiovisuelle et en particulier pour l'amélioration de la qualité

1. MediaMap est un projet Celtic-Initiative, voir <http://www.mediamaproject.org/>. Ce papier n'engage que la responsabilité de ses auteurs et ne peut pas être considéré comme le reflet de l'opinion de Celtic-Initiative ou d'un autre partenaire du projet MediaMap.

des contenus générés par les utilisateurs. L'enjeu d'un tel projet est de faciliter l'intégration de contenus amateurs dans des productions professionnelles mais également de faciliter la production collaborative entre professionnels. L'application d'assistance au tournage que nous présentons ci-après correspond à l'un des résultats du projet (sous-section 2.1.). Nous dégageons de cette application des besoins de modélisation correspondant à des problèmes de recherche plus généraux (sous-section 2.2.). Enfin, nous présentons un cas qui illustre nos besoins et nous permettra par la suite d'étudier des modèles candidats (sous-section 2.3.).

2.1. Application d'assistance au tournage

La motivation d'une application d'assistance au tournage amateur émerge des constats suivants :

- la généralisation d'équipements électroniques capable de capturer des vidéos de bonne qualité (téléphone, appareil photo numérique, etc.) permet au grand public de capter des événements exclusifs ou avant l'arrivée d'équipes de journalistes professionnels.
- l'intégration de ces contenus dans des productions professionnelles peut s'avérer difficile du fait de l'inexpérience des amateurs dans la pratique du tournage (zooms intempestifs, mouvements brusques, etc.)

Dès lors, le développement d'assistant de tournage pour guider les amateurs paraît souhaitable pour amener le contenu filmé à un niveau de qualité exploitable. Il faut cependant faire la distinction entre les propositions de dépôt spontané de contenu (comme le pratique une chaîne d'information telle que BFM²) et les appels à contribution où le professionnel passe commande auprès d'amateurs en détaillant ses exigences.

Dans ce dernier cadre, on souhaite fournir un plan de tournage au caméraman afin de guider sa prise de vue. Le plan de tournage est construit à partir de recommandations rédigées par un réalisateur (position, cadrage, lumière, etc.) utilisant un vocabulaire métier. L'originalité de l'application est d'adapter l'information présentée au caméraman suivant ses capacités (amateur, professionnel) ou son employeur si c'est un professionnel travaillant dans une tierce organisation. On suppose ainsi que malgré quelques variations dans le vocabulaire utilisé, les professionnels de l'audiovisuel utilisent les mêmes

2. La rubrique témoins BFM permet à un utilisateur de déposer des photos ou vidéos sur le site. Après modération, le contenu est diffusé et peut même faire l'objet d'une vente. Voir <http://temoins.bfmtv.com/>

concepts pour décrire le contenu. Par exemple, la notion de cadrage fait appel à des concepts de valeurs de plan indiquant la portion visible d'un personnage à l'écran, voir figure 1. Un *plan américain* indique ainsi que le personnage principal est cadré de la tête jusqu'au dessus des genoux. Le terme est utilisé en Europe en rappel à son emploi caractéristique dans les films américains des années 1910-1940, notamment dans les westerns où il permettait de montrer l'ensemble du pistolet à la ceinture des personnages³. Ce cadrage est aussi appelé *plan 3/4* et en anglais *3/4 shot*, *medium-long shot* ou *american shot* pour traduire l'expression popularisée en Europe. Si le terme utilisé varie suivant le lieu et la littérature de référence, la définition de ce type de cadrage est sans équivoque.

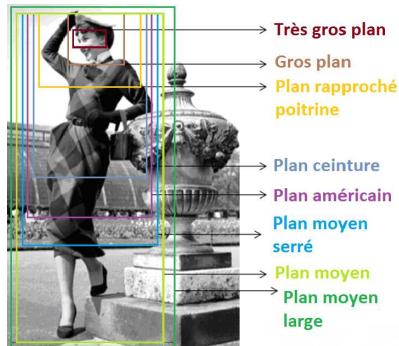


FIGURE 1: Différentes valeurs de plan pour le cadrage d'un personnage à l'écran

Les amateurs quand à eux ignorent ces concepts et n'ont pas été initiés à ces pratiques. Ils ont donc besoin d'explications et d'illustrations pour comprendre les recommandations du réalisateur. Dans le cas du cadrage, une illustration graphique est d'autant plus pertinente. L'enjeu se situe donc dans la collaboration entre un prescripteur et un opérateur qui doivent s'accorder sur le contenu à produire malgré leurs différences de vocabulaire.

2.2. Besoins en modélisation

La mise en place d'une telle application nécessite de représenter le vocabulaire de la réalisation audiovisuelle dans toutes ses variations possibles et

3. Roger Boussinot, l'Encyclopédie du Cinéma, Bordas.

de le documenter suffisamment afin de le rendre compréhensible pour des novices. Cet objectif nous amène à considérer la construction d'une ressource termino-ontologique. L'ontologie permet de représenter les concepts partagés par les professionnels de la réalisation audiovisuelle et la terminologie permet de capturer les différentes formes d'expression associées à ces concepts.

La spécificité de notre problématique est de considérer la collaboration de communautés hétérogènes par leur degré de compréhension des concepts ou leur utilisation de la terminologie. Il nous faut donc dépasser l'usage classique de la terminologie comme un moyen d'associer à des éléments ontologiques (concept, relation, instances) une chaîne lexicale ou des ressources média. L'enjeu de notre problème n'est pas simplement de normaliser les pratiques linguistiques entre les membres d'une même organisation. Il s'agit en plus de cela de fixer la manière de s'adresser à d'autres communautés. Ainsi, notre originalité se situe dans le fait que chaque chaîne ou ressource s'adresse en particulier à une communauté dont les membres partagent une capacité d'interprétation commune. Il n'existe donc plus une terminologie de référence par langue, mais des terminologies pour chaque communauté d'utilisateurs.

Par ailleurs, les types de réalisations sont divers et nécessitent des concepts spécifiques pour être décrits. Une fiction se structure en séquences et en scènes alors que les documentaires ou magazines d'information se composent de sujets. La variabilité des types de contenu à filmer implique donc de pouvoir étendre le fond conceptuel initial pour représenter de nouveaux usages. De la même manière, la collaboration avec de nouveaux partenaires nécessite de pouvoir ajouter de nouvelles terminologies au fond conceptuel existant. Ontologie et terminologie doivent se gérer de manière indépendante. A partir de ces besoins, nous définissons maintenant les exigences en terme de modélisation.

Nos besoins en modélisation peuvent être exprimés par les assertions suivantes :

- (A1) la variabilité des pratiques linguistiques des organisations et des communautés implique d'associer plusieurs termes à un même concept. Il n'y a pas de choix des termes préférés par une communauté mais une *correspondance* entre les termes d'une ou plusieurs communautés, quels que soient la langue et le code d'écriture utilisé.
- (A2) la variabilité de compréhension des communautés implique d'associer des explications (chaîne lexicale) ou des illustrations (ressource média) aux concepts afin d'en enrichir la *documentation*.

(A3) la variabilité des cas de collaboration implique de pouvoir étendre la conceptualisation initiale ou la terminologie pour s'adapter à de nouvelles pratiques ou de nouvelles communautés. Cela implique une gestion et une *évolution* indépendante de l'ontologie et de la terminologie.

2.3. Cas d'étude

Considérons le cas d'étude suivant issu du projet MediaMap. Ce cas nous permet d'étudier différents modèles. Il met en jeu trois communautés en collaboration en vue de réaliser des reportages sur des événements culturels de type concert :

- la RTBF (Radio Télévision Belge Francophone) établit des commandes de contenu dans un jargon métier propre. Son objectif est d'externaliser tant que possible la réalisation de la commande. Cela implique une compréhension commune sur le contenu à réaliser qui passe par un accord sur la manière de décrire la commande.
- le contenu commandé est tourné soit par la VRT (Radio-Télévision Flamande) qui utilise un jargon différent de la RTBF, soit des amateurs qui ne connaissent pas les concepts de la réalisation audiovisuelle. Dans le premier cas, la conceptualisation est commune, seuls les termes changent. Dans le second cas, il s'agit d'expliquer et d'illustrer les concepts utilisés

Dans le cas d'une demande de cadrage en plan américain, la demande est d'abord exprimée dans le jargon de la RTBF puis traduite dans le jargon de la VRT (plan américain pour la RTBF, plan 3/4 pour la VRT) [A1]. Ensuite, pour les amateurs, la terminologie est enrichie par des illustrations [A2]. Enfin, un nouveau concept de cadrage est ajouté (plan américain large ou plan moyen serré) [A3] en vue d'une nouvelle coopération avec la VRT. En plus de cela, le problème de la langue (français et flamand) s'ajoute à la question des jargons métiers.

3. Modèles candidats

Les modèles étudiés se limitent à désigner des termes préférés et des variantes orthographiques pour une communauté partageant un même fond conceptuel. Les termes sont rattachés à un concept et caractérisés par leur appartenance à une langue naturelle. Ainsi, plusieurs termes préférés peuvent être

rattachés au même concept, pourvu qu'ils se rapportent à des langues différentes. Parmi les modèles existants nous proposons d'étudier en particulier le standard SKOS Core du fait de sa large adoption pour représenter différents types de SOC et SKOS-XL du fait des perspectives d'extension qu'il offre. Ensuite, nous nous intéressons au méta-modèle de ressources termino-ontologiques proposé par Vandebussche & Charlet (2009) et qui semble servir de base au développement en cours de la norme ISO 25964. Cette dernière se concentre sur la création et la maintenance des thésaurus.

3.1. SKOS et SKOS-XL

SKOS⁴ est un modèle RDF/OWL visant à représenter tout type de SOC en vue de le publier sur le web de données, liées et ouvertes (Linked Open Data). En témoigne les travaux et méthodes de conversion proposés par Summers *et al.* (2008) ou Van Assem *et al.* (2006) ainsi que des applications de gestion de thésaurus développés autour de SKOS, comme celle de Schandl & Blumauer (2010). On notera que l'objectif de publication semble pousser vers une représentation minimale mais extensible de SOC déjà construits.

Concept et Etiquette

SKOS centre son modèle sur les concepts (`skos:Concept`) et considère les étiquettes comme des propriétés de ces derniers. On distingue trois types d'étiquettes :

- les étiquettes préférées (`skos:prefLabel`) qui sont uniques par langue et servent de référence.
- les étiquettes alternatives (`skos:altLabel`) qui servent de synonymes pour l'étiquette de référence.
- les étiquettes cachées (`skos:hiddenLabel`) qui servent à la récupération d'erreurs de frappes les plus courantes ou de variantes graphiques.

Chacune de ces propriétés est formalisée comme une instance de `owl:AnnotationProperty`, ce qui permet de l'attacher dans les faits à tout type d'éléments ontologiques, et pas seulement à des concepts. Les valeurs lexicales portées par ces étiquettes sont formalisées comme des `rdf:PlainLiteral`, ce qui permet de spécifier la langue et l'alphabet utilisés. Par exemple, la chaîne "higashi"@ja-Latn correspond à un mot japonais écrit avec l'alphabet latin. Il est à noter que l'utilisation des étiquettes de langues peut être étendu pour

4. Voir <http://www.w3.org/TR/skos-primer/> et <http://www.w3.org/TR/skos-reference/>

gérer les jargons métiers⁵, par exemple "Plan Américain"@fr-x-rtbf. Cependant, cette classification est unique ce qui ne permet pas d'intégrer un terme dans plus d'un jargon. De plus, cette solution s'éloigne d'une représentation sémantique des jargons.

Documentation

Différentes notes de documentation existent afin de :

- définir un concept (`skos:definition`), expliciter son contexte d'usage (`skos:scopeNote`) ou donner des exemples (`skos:example`)
- spécifier l'historique de sa signification (`skos:historyNote`), les changements effectués (`skos:changeNote`) ou à faire (`skos:editorialNote`)

L'ensemble de ces notes est défini comme une spécialisation de `skos:note`, formalisé comme une `owl:annotationProperty`. De cette manière, les notes peuvent servir à porter de la documentation écrite (comme les étiquettes), pointer vers des ressources RDF ou des documents identifiés par une URI. Cela permet ainsi de prévoir l'extension de ces notes à des besoins plus spécifiques.

Groupes et relations entre Concepts

Les concepts peuvent être regroupés dans des schémas conceptuels (`skos:-ConceptScheme` et relation `skos:inScheme`) et structurés par différentes relations :

- des relations de structuration hiérarchiques (`skos:broader`, `skos:narrower`) ou associatives (`skos:related`)
- des relations de correspondances entre concepts de schémas différents, soit une relation d'équivalence exacte (`skos:exactMatch`), une équivalence approximative (`skos:closeMatch`), des relations hiérarchiques (`skos:broadMatch`, `skos:narrowMatch`) ou associative (`skos:relatedMatch`).

SKOS-XL

SKOS-XL est une extension de SKOS développée courant 2008 pour proposer une modélisation alternative au vocabulaire de base et favoriser des extensions plus fines. Dans SKOS-XL les termes ne sont plus portés par les concepts mais deviennent des éléments à part entière (`skosxl:Label`). Les relations d'attachement entre termes et concepts sont analogues aux attributs de

5. Voir BCP 47 - Tags for Identifying Languages de l'IETF.

SKOS mais les formalisent comme des instances de `owl:objectProperty`. Le domaine de ces relations n'est pas défini ce qui permet de les associer à n'importe quelle ressource RDF, et donc en particulier aux concepts SKOS mais également à des `ConceptScheme`. Si cette dernière possibilité permet de créer des groupes de termes, elle introduit une confusion sur la sémantique des `ConceptScheme` (groupe de concepts, de termes, de triplets ?). Les étiquettes portent une seule chaîne lexicale avec les mêmes possibilités que pour SKOS grâce à l'attribut `skos:literalForm`. L'indépendance des étiquettes permet également de spécifier des relations entre eux comme la synonymie, la traduction, etc. Pastor *et al.* (2009). Cette possibilité est ouverte par la relation générique `skosxl:labelRelation`.

```
//===== CONCEPTS =====//
C-PlanAmericain rdf:type skos:Concept ;
  skosxl:prefLabel T-PlanAmericain ;
  skosxl:altLabel T-Plan34 ;
  skos:inScheme RTBF-Scheme ;
  skos:example ex:PlanAmericain.png ; //Illustration//
  skos:exactMatch C-Plan34 .
C-Plan34 rdf:type skos:Concept ;
  skosxl:prefLabel T-Plan34 ;
  skosxl:altLabel T-PlanAmericain ;
  skos:example ex:PlanAmericain.png ; //Illustration//
  skos:inScheme VRT-Scheme .

//===== TERMES =====//
T-PlanAmericain rdf:type skosxl:Label ;
  skosxl:literalForm "Plan américain"@fr .
T-Plan34 rdf:type skosxl:Label ;
  skosxl:literalForm "Plan 3/4"@fr .
```

FIGURE 2: Modélisation avec SKOS-XL de termes de jargons différents correspondant au même concept de cadrage

Dans SKOS, la gestion de plusieurs jargons métiers dans une même conceptualisation [A1] est rendue difficile par la caractérisation simple des termes par la langue. Ainsi, même si on accroche plusieurs étiquettes au même concept, on ne sait pas les sélectionner pour les présenter à l'une ou l'autre communauté. Cela implique un dédoublement des concepts et donc des schémas conceptuels nécessaires (un par communauté). Avec le découplage terme-concept de SKOS-XL, on peut gérer terme et concept de manière séparés sans pour autant avoir de primitives spécifiques pour regrouper les termes par jargon ou code d'écriture, voir figure 2. Une solution consisterait à regrouper les termes dans des `ConceptScheme` (SKOS-XL l'autorise). Les `ConceptScheme`

serviraient alors à la fois à regrouper les concepts (AV-Scheme) et les termes spécifiques aux organisations (RTBF-Scheme, VRT-Scheme). Cependant, si cette modélisation permet de gérer vocabulaire métier et conceptualisation de manière séparés c'est au prix d'un flottement sur la sémantique de Concept-Scheme. Le support d'un nouveau jargon peut donc se faire sans toucher à la conceptualisation [A3] grâce à la permissivité de l'extension SKOS-XL. L'extension ou la mise en correspondance de la conceptualisation est facilitée par les relations sémantiques entre concepts. Enfin, du point de vue de la documentation [A2], les définitions des concepts sont prévues et l'ajout d'illustration possible même si aucune relation n'est définie dans ce sens.

3.2. ISO 25964-1

Cette norme propose une modélisation terme-concept similaire à SKOS-XL mais se concentre sur la représentation des thésaurus. Elle se fonde sur des modèles pré-existants, le méta-modèle Vandenbussche & Charlet (2009) ainsi que la norme BS 8723. L'originalité par rapport à SKOS est de considérer la composition de termes ou de concepts et d'enrichir la description des éléments du modèle par des attributs Dublin Core.

Les termes se distinguent entre termes préférés simples ou composés et termes non préférés simples. La caractérisation des termes porte également sur l'appartenance à une langue à laquelle s'ajoutent des attributs de dates, une définition ainsi que des notes d'historique et de révision. Des relations sémantiques entre termes sont également considérées en particulier l'équivalence composée, la synonymie, l'abréviation, l'acronyme, etc. Le thésaurus est considéré comme l'élément central décrit par l'ensemble des quinze attributs originaux du Dublin Core ainsi que des notes d'historique pour la maintenance. Sur les questions de documentation et de groupement de concepts, il existe une similarité importante avec les primitives de SKOS (note, groupe de concepts, etc.).

Les apports de la norme ISO 25964 par rapport à SKOS concernent davantage les pratiques de création et de maintenance de thésaurus que la gestion des jargons métiers [A1] et l'illustration de concepts par des ressources média [A2]. Ainsi, les manques par rapport à nos besoins sont similaires. L'attention portée sur les détails de description de chaque élément du modèle est tout à fait convaincante. L'écart avec l'aspect épuré et synthétique de SKOS s'explique certainement par l'écart entre leurs objectifs. Alors que SKOS vise la

publication de tout type de SOC, l'ISO 25946 se concentre sur la construction et l'évolution des seuls thésaurus. La norme ne s'est pas encore attelée aux questions d'interopérabilité et de correspondance avec d'autres vocabulaires. Ce travail est en cours et sera dévoilé dans la seconde partie de la norme (ISO 25946-2). De notre point de vue, il manque toujours un moyen de regrouper des termes indépendamment des concepts pour ajouter des jargons à une conceptualisation existante [A3].

3.3. Bilan

L'étude des standards et normes de références précédentes ne semble pas apporter de solution complètement satisfaisante pour l'ensemble de nos besoins. En effet, les approches restent centrées sur le concept et sa structuration auquel on intègre (SKOS) ou rattache (SKOS-XL, ISO 25946-1) ensuite les termes. Ces derniers sont représentés de manière plus ou moins fine (gestion des compositions dans ISO 25946 absente de SKOS). Dans tous les cas, la préférence d'un terme s'établit uniquement sur l'appartenance à une langue et non par rapport à une communauté de jargon. De ce fait, ces modèles se limitent à représenter un seul jargon de référence par thésaurus et suppose ainsi l'existence d'une communauté homogène dans sa compréhension et dont on cherche à normaliser l'usage linguistique. Nous proposons dans la suite un modèle Multi-Jargon afin d'associer plusieurs jargons métiers et des explications à une conceptualisation commune.

4. Modèle Multi-Jargon

Le modèle Multi-Jargon que nous proposons peut être décomposé en deux parties. La première propose une représentation concept-terme qui étend la caractérisation des termes aux jargons dont ils font partie et codes d'écriture utilisés (sous-section 4.1.). La seconde partie propose des concepts d'encapsulation des terminologies et des conceptualisations afin de faciliter leur gestion indépendante (sous-section 4.2.).

4.1. Modèle Concept-Terme

Le modèle concept-terme que nous avons développé étend la représentation de la terminologie à son contexte d'accès, voir Diemert *et al.* (2010). Dans un premier temps, on associe à chaque concept un ou plusieurs termes,

une définition voire une illustration. Chaque terme est caractérisé par son appartenance à une ou plusieurs notations qui représentent une langue, un vocabulaire métier, un code d'écriture, etc. Les fichiers peuvent être caractérisés de manière équivalente, voir figure 3.

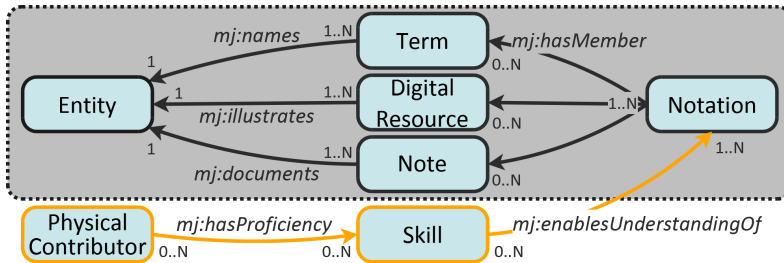


FIGURE 3: Modèle Concept-Terme

On distingue ainsi les concepts suivants :

- Entity représente un concept auquel on peut associer soit un terme (*mj:names/mj:isNamedBy*) soit une ressource média (*mj:illustrates/isIllustratedBy*) ou une note de documentation (*mj:documents/isDocumentedBy*).
- Term représente un terme portant une valeur lexicale.
- Note représente une chaîne lexicale dans un but de documentation. Par exemple, la définition d'un terme (*DefinitionNote*).
- DigitalResource représente un fichier numérique. Par exemple, un fichier texte, une photo ou bien un son.
- Notation qui représente les caractéristiques d'un terme (*mj:hasMember/mj:isMemberOf*) ou d'un fichier. Une Notation permet de définir un groupe de termes ou de fichiers, ceux partageant la même caractéristique. On distingue quatre types de Notation :
 - NaturalLanguage représente les langues humaines.
 - SyntaxEncodingScheme représente un codage de l'information comme l'encodage des caractères ou bien un format de fichier.
 - AuthorityList représente une liste d'autorité composé de termes.
 - VocabularyEncodingScheme (VES) représente un SOC ou un jargon propre à une organisation ou un métier.

Si l'on reprend le cas d'étude, le concept de plan américain peut être associé à deux termes issus du jargon de la réalisation audiovisuelle mais se distinguant par la communauté qui les utilise, Plan américain pour la RTBF

et Plan 3/4 pour la VRT. Chaque jargon est représenté dans notre modèle comme un VocabularyEncodingScheme (le vocabulaire de réalisation audiovisuelle professionnelle englobe une variante propre à la RTBF et une autre propre à la VRT). Chaque terme est décrit par rapport à son appartenance à un ou plusieurs jargons (VES) et par sa langue (NaturalLanguage). On peut également raffiner la description du terme en précisant l'encodage et l'alphabet utilisé (SES) ou attacher une définition voire une illustration, voir figure 4.

```
//===== INSTANCE =====//
pop:C-PlanAmericain rdf:type mj:Entity ;
  mj:isNamedBy pop:T-PlanAmericain ;
  mj:isNamedBy pop:T-Plan34 ;
  mj:isDocumentedBy pop:Definition-PlanA ; //Definition//
  mj:isIllustratedBy pop:Picture-PlanA . //Illustration//

//===== TERMES =====//
pop:T-PlanAmericain rdf:type mj:Term ;
  mj:value "Plan Américain" ;
  mj:isMemberOf pop:NL-French, pop:VES-Audiovisual,
  pop:VES-RTBF .
pop:T-Plan34 rdf:type mj:Term ;
  mj:value "Plan 3/4" ;
  mj:isMemberOf pop:NL-French, pop:VES-Audiovisual,
  pop:VES-VRT .

//===== DEFINITION, ILLUSTRATION =====//
pop:Definition-PlanA rdf:type mj:definitionNote ;
  mj:value "Un plan dans lequel le personnage filmé apparaît de
  la tête jusqu'au dessus des genoux." ;
  mj:isMemberOf pop:NL-French .
<pop:Picture-PlanA> rdf:type mj:DigitalResource .
```

FIGURE 4: Modélisation du couplage concept plan américain et des termes associés en RDF N3

La description de l'utilisateur nous permet de sélectionner le terme ou le fichier le plus adapté à ses capacités de compréhension. On introduit alors les concepts suivants :

- `PhysicalContributor` représente une personne physique ayant accès à la conceptualisation et aux termes associés.
- `Skill` représente les capacités d'action et de compréhension de notation (`mj:enablesUnderstandingOf`) associées aux personnes (`mj:hasProficiency`), voir figure 3.

Par rapport au cas d'étude, la capacité de maniement d'une caméra attribuée à un caméraman professionnel implique la compréhension des concepts

de cadrage utilisés pour donner des instructions de tournage. La compétence donne alors accès au jargon de la réalisation audiovisuelle. La description peut être raffinée pour décrire l'employeur et donc distinguer entre le jargon RTBF ou VRT. Pour l'amateur, l'absence de compétences amène à sélectionner la définition et l'illustration plutôt qu'un des termes du jargon professionnel.

4.2. Modèle termino-ontologique

Le principe de notre modèle termino-ontologique est de séparer conceptualisation et assertions de manière similaire au formalisme des Logiques de Description développé par Baader & Nutt (2003). On groupe les instances dans des `Population` correspondants aux `ABox`, et la définition des conceptualisations dans des `ConceptualModel` correspondants aux `TBox`. Enfin, un `Profile` correspondant à une base de connaissances permet de regrouper concepts et instances. Ainsi, on considère conceptualisation et terminologie comme des ressources que l'on combine dans des profils suivant l'évolution des besoins, voir figure 5. Les concepts et relations de notre modèle sont :

- `ConceptualModel` : peut se construire à partir d'un modèle existant dont il spécialise un ou plusieurs éléments (`mj:specializes`).
- `Population` : représente un ensemble d'instances créées à partir d'un ou plusieurs modèles conceptuels (`mj:createdFrom`). On distingue deux types de population :
 - la population d'initialisation (`SetupPopulation`) qui transporte des termes, des ressources média, des notations nécessaires à l'utilisation du modèle conceptuel dont ces instances sont issues.
 - la population d'assertion (`StatementPopulation`) qui transporte des assertions créées à partir d'un ou plusieurs modèles conceptuels.
- un `Profile` représente une base de connaissances composée d'un ou plusieurs modèles conceptuels et populations. La construction d'un profil peut se faire à partir de profils existants (`mj:requiresProfile`) en fonction des besoins de descriptions.

Suivant le cas représenté figure 4, le plan américain est en réalité une instance appartenant à une liste d'autorité de type de cadrage. Notre mécanisme de nommage s'applique identiquement à tous les éléments ontologiques (concept, relation, attribut, instance). Ainsi, la population d'initialisation (identifiée par le préfixe `pop`) contient d'une part les instances des concepts énumérés et d'autre part les jargons audiovisuels qui correspondent. La définition de nouveaux termes pour une nouvelle langue ou un nouveau jar-

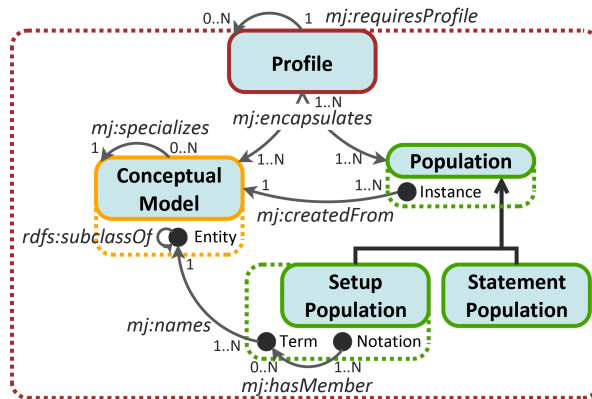


FIGURE 5: Modèle termino-ontologique

gon se fait simplement par leur définition dans une population. L'ajout d'un concept par spécialisation se fera de manière similaire que les termes mais au sein d'un modèle conceptuel identifié dans notre exemple par le préfixe *mj*. L'attachement d'un modèle conceptuel (avec de nouveaux concepts) ou d'une population (avec les termes d'un nouveau jargon) est réalisé en les intégrant à un profil additionnel qu'on raccroche au profil initial (par la relation *mj:requiresProfile*).

5. Conclusion

Dans cet article nous avons discuté de la nécessité d'attacher plusieurs terminologies à un même fond conceptuel afin de représenter des jargons métiers de communautés hétérogènes. Les disparités entre ces communautés portent sur leur pratiques linguistiques ou leur niveau de compréhension d'un domaine d'application. Cela implique de pouvoir documenter et illustrer la conceptualisation afin de s'assurer de sa compréhension commune. Nous avons illustré notre propos au moyen d'une application d'assistance au tournage développé dans le cadre du projet MediaMap.

Les modèles existants ne permettent de distinguer les termes que par la langue ce qui oblige à créer des conceptualisations redondantes pour gérer plusieurs jargons métiers. Ceci nous a amené à développer un modèle multi-jargon qui repose sur une caractérisation plus fine des termes, notes de documentation et fichiers médias associés à une conceptualisation. Ces informations supplémentaires servent comme autant de points d'attaches qui per-

mettent de constituer non pas des groupes de concepts, mais des ensembles de termes partageant des traits communs. Cela permet de définir des jargons métiers qui sont ensuite rendus disponibles à l'utilisateur par l'intermédiaire des compétences qu'on lui attribue. Ce modèle est actuellement utilisé dans l'application d'assistance au tournage du projet MediaMap. Notre modèle termino-ontologique donne la possibilité de grouper conceptualisation et terminologie de manière indépendante afin de pouvoir les manipuler et de les combiner plus facilement. Ces concepts répondent à une perspective de gestion qui n'est pas fourni par un langage comme OWL. En effet, la primitive `owl:Ontology` est un ensemble qui regroupe autant des concepts que des instances. Dans notre cas, la distinction entre les uns et les autres est significative pour gérer indépendamment chaque jargon et les combiner à la demande dans des profils avec les conceptualisations qu'elles nomment, décrivent et illustrent.

Références

- BAADER F. & NUTT W. (2003). The description logic handbook. chapter Basic description logics, p. 43–95. New York, NY, USA : Cambridge University Press.
- DIEMERT B., ABEL M.-H. & MOULIN C. (2010). Modélisation des dénominations ontologiques. In *Terminologie et Ontologie Théorie et pratiques (TOTh)*.
- PASTOR J., MARTINEZ F. & J.V. R. (2009). Advantages of thesaurus representation using SKOS compared with proposed alternatives. *Information Research*, **14**(4).
- SCHANDL T. & BLUMAUER A. (2010). PoolParty : SKOS Thesaurus Management Utilizing Linked Data. *The Semantic Web : Research and Applications*, **6089**, 421–425.
- SUMMERS E., ISAAC A., REDDING C. & KRECH D. (2008). LCSH, SKOS and Linked Data. *CoRR*.
- VAN ASSEM M., MALAÏSÉ V., MILES A. & SCHREIBER G. (2006). A Method to Convert Thesauri to SKOS. In *The Semantic Web : Research and Applications*, number C, p. 95–109.
- VANDEBUSSCHE P.-Y. & CHARLET J. (2009). Méta-modèle général de description de ressources terminologiques et ontologiques. In *Actes d'IC*, p. 193–204.
- ZACKLAD M. & GIBOIN A. (2010). Qu'est-ce qu'un système d'organisation des connaissances (SOC) ? *Document numérique*, **13**(2), 7–12.