



HAL
open science

Construction d'une ontologie commune par alignement flou d'ontologies d'un domaine

Konstantin Todorov, Peter Geibel, Céline Hudelot

► **To cite this version:**

Konstantin Todorov, Peter Geibel, Céline Hudelot. Construction d'une ontologie commune par alignement flou d'ontologies d'un domaine. RFIA 2012 (Reconnaissance des Formes et Intelligence Artificielle), Jan 2012, Lyon, France. pp.978-2-9539515-2-3. hal-00656491

HAL Id: hal-00656491

<https://hal.science/hal-00656491v1>

Submitted on 17 Jan 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.

Construction d'une ontologie commune par alignement flou d'ontologies d'un domaine

Konstantin Todorov¹

Peter Geibel²

Céline Hudelot¹

¹ Laboratoire MAS, Ecole Centrale Paris

² TU Berlin, Sekr. FR 5-8, Fakultät IV

MAS, ECP, Grande Voie des Vignes, F-92295 Chatenay-Malabry, France
konstantin.get@gmail.com, info@peter-geibel.de, celine.hudelot@ecp.fr

Résumé

Ce papier¹ traite du problème de la construction d'une ontologie commune pour un ensemble d'ontologies de domaine afin de permettre leur partage et leur intégration dans une plate-forme collaborative. Nous proposons un nouvel algorithme hiérarchique pour représenter les concepts comme des ensembles flous à l'aide d'une ontologie de référence. Contrairement aux représentations originales des concepts basées sur des instances, cela permet l'application de méthodes de raisonnement flou dans le but de caractériser et de mesurer le degré des relations entre les concepts des ontologies de domaine. Nous proposons une application de l'approche dans le domaine du multimédia.

Mots Clef

Alignement flou, ontologies, construction d'ontologie commune

Abstract

This paper deals with the problem of building a common knowledge body for a set of domain ontologies in order to enable their sharing and integration in a collaborative framework. We propose a novel hierarchical algorithm for concept fuzzy set representation mediated by a reference ontology. In contrast to the original concept representations based on instances, this enables the application of methods of fuzzy logical reasoning in order to characterize and measure the degree of the relationships holding between concepts from different ontologies. We present an application of the approach in the multimedia domain.

Keywords

Fuzzy Ontology Matching, Constructing a Common Ontology

1 Introduction

Dans un contexte collaboratif, il est très fréquent de devoir mettre en correspondance plusieurs ontologies, créées

¹. Le papier est une traduction d'un papier des mêmes auteurs publié dans URSW Worskhop (ISWC) 2011.

de manière indépendante, afin de permettre leur interopérabilité. En effet, en raison des hétérogénéités provenant de la nature décentralisée de leur acquisition, des différences dans leur application ou encore des différences dans leur syntaxe et leur terminologie, l'interopérabilité sémantique est un problème difficile.

Dans ce papier, nous présentons une nouvelle approche pour construire un corps de connaissance commun, représentée par une ontologie floue, pour un ensemble d'ontologies d'un domaine particulier. Cette ontologie a pour but de capturer et d'exposer les diverses relations existant entre les concepts dans ontologies du domaine telles que les relations de similarité, de subsumption ou encore de complémentarité. Cela peut être très utile dans un certain nombre de scénarios réels, en particulier dans des plate-formes collaboratives. Imaginons un projet qui comprend plusieurs partenaires, dont chacun a son propre vocabulaire structuré de termes pour décrire son domaine d'activité. Le cadre proposé permettrait à chaque partenaire de consulter le corps de connaissance commun à chaque fois que la collaboration est nécessaire, tout en gardant son propre vocabulaire. Des exemples de telles requêtes peuvent être : "Quel est le concept de mon partenaire P_1 le plus proche à mon concept A ?", ou "Me donner les concepts de tous mes partenaires qui sont à une distance égale de mon concept B ", ou "Quel est le concept de mon partenaire P_2 qui subsume fortement mon concept C ?", ou "Quelles sont les spécificités et les points communs entre mon concept A et les concepts de mes partenaires ?".

Le cadre flou, qui permet de traiter des informations hétérogènes est au coeur de la méthode proposée. En effet, nous proposons de représenter chaque concept d'une ontologie du domaine par un ensemble flou sur un univers de discours composé des concepts d'une ontologie de référence. Ceci peut être vu comme une projection de tous les concepts des ontologies du domaine ciblé vers un espace sémantique commun, où la distance et les relations entre deux concepts peuvent être exprimées selon des critères fixes. Contrairement au cas où les concepts sont représentés de manière extensionnelle par leurs instances, l'ap-

proche floue permet l'application de méthodes de raisonnement flou afin de caractériser les relations entre les concepts d'ontologies différentes. En outre, les représentations floues permettent de quantifier le degré de vérité d'une relation existant entre deux concepts.

L'article est structuré comme suit. La section 2 rappelle les bases sur la théorie des ensembles flous ainsi que dans le domaine de l'alignement d'ontologies. Les travaux antérieurs se rapportant à notre travail sont présentés dans la section 3. Nous présentons notre algorithme de floutage dans la section 4, avant de présenter la méthode de construction du corps de connaissance commun dans la section 5. Des résultats expérimentaux et des conclusions sont présentées dans les sections 6 et 7, respectivement.

2 Avant propos

Dans cette section, nous introduisons les bases de la théorie des ensembles flous et de l'alignement d'ontologies.

2.1 Ensembles flous

Un sous-ensemble flou \mathcal{A} est défini sur un domaine d'objets X par sa fonction

$$\mu_{\mathcal{A}} : X \mapsto [0, 1], \quad (1)$$

qui exprime le degré d'appartenance de chaque élément de X à \mathcal{A} en attribuant à chaque $x \in X$ une valeur de l'intervalle $[0, 1]$ [20]. L'ensemble des parties d'un ensemble X , dénoté par $\mathcal{F}(X, [0, 1])$, représente l'ensemble de tous les sous-ensembles flous de \mathcal{A} complètement équivalents à leur fonction d'appartenance $\mu : X \mapsto [0, 1]$.

Comme dans le cas des ensembles classiques, les opérations d'union, d'intersection et de complémentation peuvent être définies sur les ensembles flous. Dans la suite, nous définissons ces opérations selon la sémantique de Gödel.

L'intersection de deux ensembles flous \mathcal{A} et \mathcal{B} est donnée par une fonction, dite une t -norme, définie par : $T(\mu_{\mathcal{A}}(x), \mu_{\mathcal{B}}(x)) = \min(\mu_{\mathcal{A}}(x), \mu_{\mathcal{B}}(x))$. L'union de \mathcal{A} et \mathcal{B} est définie par $S(\mu_{\mathcal{A}}(x), \mu_{\mathcal{B}}(x)) = \max(\mu_{\mathcal{A}}(x), \mu_{\mathcal{B}}(x))$ où S est une t -conorme. Le complément d'un ensemble flou \mathcal{A} , dénoté par $\neg\mathcal{A}$, est défini par la fonction d'appartenance $\mu_{\neg\mathcal{A}}(x) = 1 - \mu_{\mathcal{A}}(x)$. Nous considérons la définition de Gödel d'implication floue :

$$\mu_{\mathcal{A} \rightarrow \mathcal{B}}(x) = \begin{cases} 1, & \text{if } \mu_{\mathcal{A}}(x) \leq \mu_{\mathcal{B}}(x), \\ \mu_{\mathcal{B}}(x), & \text{otherwise.} \end{cases} \quad (2)$$

2.2 Ontologies, hétérogénéité et mise en correspondance

Une ontologie consiste en un ensemble de concepts sémantiquement liés. Elle permet de représenter de manière explicite et formelle de la connaissance au sujet d'un domaine d'intérêt donné [8]. Dans ce papier, nous nous intéressons aux ontologies, dites peuplées, dont les concepts sont associés à un ensemble d'instance. Ces ontologies sont définies de la manière suivante :

Definition 1 (Ontologie Classique) Soit C un ensemble de concepts, $is_a \subseteq C \times C$ une relation de subsomption entre les concepts, R un ensemble de relations sur $C \times C$, I un ensemble d'instances, et $g : C \rightarrow 2^I$ une fonction qui associe à chaque concept de C un sous-ensemble d'instances de I . Nous exigeons que is_a et g soient compatibles, i.e. $is_a(A', A) \leftrightarrow g(A') \subseteq g(A)$ est vrai pour tout $A', A \in C$. Particulièrement, cela implique que is_a est une relation d'ordre partiel. Ainsi, la collection

$$O = (C, is_a, R, I, g)$$

forme une ontologie classique, dite peuplée.

Ci-dessus, un concept est modélisé intentionnellement par ses relations à d'autres concepts, et extensionnellement par l'ensemble des instances qui lui sont assignées par la fonction g . Par hypothèse, chaque instance peut être représentée comme un vecteur des valeurs réelles, défini par un nombre fixe de variables d'un certain type (identiques pour toutes les instances dans l'ensemble I).

L'hétérogénéité au niveau des ontologies se produit lorsque deux ou plusieurs ontologies sont créées indépendamment les unes des autres sur des domaines similaires. L'hétérogénéité peut être observée au niveau linguistique (terminologique), conceptuel ou extensionnel [8].

La mise en correspondance d'ontologies est définie comme le processus de construction des relations entre les éléments de deux ou plusieurs ontologies hétérogènes. Différentes techniques de mise en correspondance ont été introduites au cours des dernières années afin de répondre aux différents types d'hétérogénéité [11].

La mise en correspondance dite basée-instances, ou extensionnelle regroupe un ensemble d'approches autour de l'idée que les concepts d'une ontologie peuvent être représentés comme des ensembles d'instances et la similarité mesurée sur ces ensembles reflète la similarité sémantique entre les concepts que ces instances peuplent.

2.3 Mesures de similarité entre concepts classiques

Considérons deux ontologies $O_1 = (C_1, is_a_1, R_1, I_1, g_1)$ et $O_{ref} = (X, is_a_{ref}, R_{ref}, I_{ref}, g_{ref})$. Nous nous basons sur l'idée simple que déterminer la similarité $sim(A, x)$ entre deux concepts $A \in C_1$ et $x \in X$ consiste à comparer leurs ensembles d'instances $g_1(A)$ et $g_{ref}(x)$. Pour cela, nous avons besoin d'une mesure de similarité entre deux instances (ou entre la représentation vectorielle de ces deux instances) \mathbf{i}^A et \mathbf{i}^x , où $\mathbf{i}^A \in g_1(A)$ et $\mathbf{i}^x \in g_{ref}(x)$. Nous avons utilisé la mesure classique du produit scalaire et le cosinus $s(\mathbf{i}^A, \mathbf{i}^x) = \frac{\langle \mathbf{i}^A, \mathbf{i}^x \rangle}{\|\mathbf{i}^A\| \|\mathbf{i}^x\|}$. Basée sur la similarité des éléments, la similarité des ensembles peut être définie en calculant la similarité entre les vecteurs moyens correspon-

dants aux classes prototypes [12] :

$$sim_{proto}(A, x) = s \left(\frac{1}{|g_1(A)|} \sum_{j=1}^{|g_1(A)|} i_j^A, \frac{1}{|g_{ref}(x)|} \sum_{k=1}^{|g_{ref}(x)|} i_k^x \right). \quad (3)$$

Notons que d'autres approches de similarité conceptuelle peuvent être également appliquées, telles que la méthode basée sur la sélection des variables [18]. Dans notre approche, nous avons utilisé la méthode qui est à la fois la plus efficace et la plus simple. La mesure de similarité a été appliquée de manière hiérarchique pour deux ontologies, comme cela est décrit dans [18].

3 Travaux antérieurs

La théorie des ensembles flous permet le traitement de données imprécises et vagues [20]. Plusieurs travaux proposent d'étendre les ontologies pour la représentation et la gestion de connaissances imparfaites parmi lesquels les travaux de Sanchez, Calegari et leurs collègues [5], [6], [15] peuvent faire figure de référence. Dans ces travaux, chaque concept est défini comme un ensemble flou sur le domaine des instances et les relations entre des instances et des concepts sont définies comme des relations floues. Une revue étendue est proposée dans [3].

Les travaux portant sur la mise en correspondance d'ontologies floues peuvent être classés en deux familles : (1) les approches étendant la mise en correspondance classique aux ontologies floues et (2) les approches prenant en compte l'imprécision de la mise en correspondance des concepts (flous ou classiques). Parmi les approches de la première famille, nous pouvons citer les travaux de Xu *et al.* [19] qui ont proposé une approche pour l'alignement de concepts flous en se basant sur les travaux sur le mapping approximatif de Stuckenschmidt [17] et Akahani *et al.* [1]. Avec une idée similaire, Bahri *et al.* [2] proposent une approche pour définir des mesures de similarité entre des composants ontologiques flous. Comme exemple de la deuxième famille d'approches, nous pouvons citer [9] où une approche floue pour traiter l'incertitude de la mise en correspondance est proposée. Une méthode d'alignement basée sur les graphes et des règles conceptuelles floues est proposée dans [4]. Afin de définir de nouvelles mesures de similarité des concepts intra-ontologie (i.e. venant d'une même ontologie), Cross *et al.* [7] modélisent un concept comme un ensemble flou de ses concepts pères et de lui-même, à l'aide d'une fonction d'appartenance basée sur la mesure du contenu d'information d'un concept à l'égard de son ontologie.

La mise en correspondance classique d'ontologies dite extensionnelle est basée sur l'idée que la similarité entre les instances de deux concepts reflète la similarité intentionnelle entre ces concepts [8]. Nous citons en particulier l'approche Caiman dans laquelle l'estimation de similarité entre deux concepts est basée sur la distance entre les vecteurs-moyens des classes d'instances qui peuplent ces concepts [12].

L'originalité de notre approche et ce qui la différencie des approches existantes consiste en (1) l'utilisation d'une ontologie de référence pour flouter les concepts, (2) la construction d'une ontologie commune basée sur ces représentations floues des concepts (au coeur de notre approche), (3) l'application d'une méthode floue aux ontologies peuplées avec des instances. Notre approche propose un changement de perspective ; l'on entre dans le domaine du raisonnement flou où l'on n'a plus besoin des instances. La construction d'une ontologie commune dans notre cadre est algorithmiquement très efficace.

4 Un algorithme hiérarchique de floutage des concepts

Soit $\Omega = \{O_1, \dots, O_n\}$ un ensemble d'ontologies comme définies dans la Déf. 1 que nous dénommerons par la suite l'ensemble d'ontologies source. L'ensemble de concepts $C_\Omega = \bigcup_{i=1}^n C_i$ sera appelé l'ensemble des concepts source. Les ontologies de l'ensemble Ω sont censées partager des caractéristiques communes tout en étant hétérogènes dans le sens d'un des types d'hétérogénéité décrits dans la Section 2. Une certaine complémentarité de ces ressources peut être assumée : elles peuvent être définies dans le contexte d'une même application par exemple, mais à des niveaux différents, traitant d'aspects différents et complémentaires du même problème réel.

Soit $O_{ref} = (X, is_{a_{ref}}, R_{ref}, I_{ref}, g_{ref})$ une ontologie appelée ontologie de référence dont les concepts seront appelés les concepts de référence. Contrairement aux ontologies sources, cette ontologie est supposée être moins dépendante du contexte d'application et donc plus générique en tant que source de connaissance. Selon la définition 1, les ontologies de l'ensemble Ω et l'ontologie O_{ref} sont peuplées.

La procédure de floutage que nous proposons se base sur la similarité entre un concept source et les concepts de l'ontologie de référence. En particulier, un score est calculé pour chaque concept source et chaque concept de référence à l'aide de la mesure définie dans l'équation (3). Le concept source A sera représenté alors par la fonction suivante :

$$\mu_A(x) = score_A(x), \forall x \in X, \quad (4)$$

où $score_A(x)$ est la similarité entre le concept source A et le concept de référence x . Puisque $score$ prend ses valeurs entre 0 et 1, (4) définit un ensemble flou. Nous ferons référence à un tel ensemble flou comme le concept A flouté, dénoté par \mathcal{A} .

Afin de flouter (i.e. rendre flous) les concepts d'une ontologie source O_1 , nous proposons l'algorithme hiérarchique suivant : tout d'abord, nous assignons des vecteurs-scores, i.e. des fonctions d'appartenance, à tous les concepts-feuilles de l'arbre hiérarchique de O_1 . Pour tout autre concept ne contenant pas d'instances qui lui sont propres, un score correspondant au maximum des scores de ses enfants, pour tout $x \in X$, lui est assigné. Si au contraire, un

concept non-feuille, contient des instances qui lui sont propres, alors nous calculons d'abord son score par rapport à ces instances à l'égard de l'ontologie de référence, et après nous calculons un score final en prenant le maximum des scores de ses enfants et celui de lui même.

Plus formellement, soit A le concept d'une ontologie source et A' et A'' ces enfants dans cette ontologie selon la relation is_a . Soit $g^*(A)$ l'ensemble des instances propres au concept A . Nous calculons les scores de similarité :

$$\text{score}_A(x) = \max\{\text{score}_{A'}(x), \text{score}_{A''}(x), \text{score}_{g^*(A)}(x)\}, \forall x \in X.$$

Ci-dessus, $\text{score}_{g^*(A)}(x)$ dénote la similarité obtenue pour le concept A et le concept de référence x en tenant compte seulement des documents propres à A . Un pseudo-code de cet algorithme est donné dans Alg. 1.

Notons que l'attribution du \max de tous les enfants au parent pour tout x conduit à une convergence vers l'uniformité des fonctions d'appartenance pour les noeuds les plus hauts de la hiérarchie. Naturellement, il est attendu que les fonctions d'appartenance associées aux concepts se situant en haut de la hiérarchie soient moins précises que celles associées aux concepts de niveaux inférieurs. Un concept dans une structure hiérarchique peut être vu comme l'union de ses descendants, et une union correspond au \max comme définie ci-dessus (une approche qui est en phase avec la stratégie de liaison unique (single link strategy) utilisée dans le clustering).

La procédure de scoring hiérarchique a l'avantage que, étant donné un concept référence x , chaque score sera plus grand pour un noeud parent que pour ses enfants. Du point de vue algorithmique, la procédure qui n'associe des scores qu'aux concepts peuplés est moins coûteuse, par rapport au scoring de tous les noeuds un par un (l'algorithme est d'une complexité globale de $O(|C| * \max(|C|, |X|))$).

5 Construction d'une ontologie commune

La construction d'un corps de connaissance ou ontologie commune pour un ensemble d'ontologies source vise à rendre explicites les relations entre les concepts de ces différentes ontologies. Pour cela, nous proposons d'exploiter les représentations floues des concepts construites comme cela est expliqué dans la section précédente. Dans ce qui suit, nous considérons deux ontologies source O_1 et O_2 mais il faut noter que toutes les définitions peuvent être facilement étendues à plusieurs ontologies.

Soit $\mathcal{C}_\Omega = \{\mathcal{A}_1, \dots, \mathcal{A}_{|C_1|}, \mathcal{B}_1, \dots, \mathcal{B}_{|C_2|}\}$ la réunion des ensemble des concepts de O_1 (les concepts \mathcal{A}) et O_2 (les concepts \mathcal{B}). Nous introduisons plusieurs relations et opérations qui peuvent être calculées sur $\mathcal{C}_\Omega \times \mathcal{C}_\Omega$ et qui seront utilisées pour construire l'ontologie commune.

5.1 Relations floues entre les concepts

L'implication $A' \rightarrow A$ tient pour tout A' et A tels que $\text{is_a}(A', A)$. Nous fournissons une définition de subsomp-

Function score(concept A , ontology O_{ref} , sim. measure sim)

begin

for $i = 1, \dots, |X|$ **do**

$\text{sim}[i] = \text{sim}(A, x_i) // x_i \in X$

return sim

end

Procedure hierarchicalScoring(ontology O , ontology O_{ref} , sim. measure sim)

begin

1. Let C be the list of concepts in O .

2. Let L be a list of nodes, initially empty

3. Do until C is empty :

 (a) Let L' be the list of nodes in C that have only children in L

 (b) $L = \text{append}(L, L')$

 (c) $C = C - L'$

4. Iterate over L (first to last), with A being the current element :

if $\text{children}(A) = \emptyset$ **then**

$\text{score}(A) = \text{score}(A, O_{ref}, \text{sim})$

else

if $g^*(A) \neq \emptyset$ **then**

$\text{score}(A) =$

$\max\{\max_{B \in \text{children}(A)} \text{score}(B), \text{score}(A, O_{ref}, \text{sim})\}$

else

$\text{score}(A) = \max_{B \in \text{children}(A)} \text{score}(B)$

return $\text{score}(A), \forall A \in C$

end

Algorithme 1 : Algorithme hiérarchique de floutage des concepts sources.

tion floue de deux concepts flous \mathcal{A}' et \mathcal{A} basée sur l'implication floue comme définie dans (2).

Definition 2 (Subsorption Floue) La subsorption " \mathcal{A}' est un \mathcal{A} " est définie et dénotée de la manière suivante :

$$\text{is_a}(\mathcal{A}', \mathcal{A}) = \inf_{x \in X} \mu_{\mathcal{A}' \rightarrow \mathcal{A}}(x) \quad (5)$$

Ainsi, la subsorption floue est définie comme le degré (une valeur entre 0 et 1) avec lequel un concept subsomme un autre. Il peut être démontré que is_a , analogiquement à sa version classique, est une relation réflexive et transitive (i.e. est un quasi-ordre). En outre, la procédure de floutage des concepts introduite dans la section précédente assure que $\text{is_a}(A', A) = 1$ est vraie pour toute paire "enfant-parent", ou autrement dit, la subsorption classique est préservée par le processus de floutage.

En reprenant l'exemple d'une plate-forme collaborative de l'introduction, le calcul de la relation is_a floue entre deux concepts permet de répondre à une requête utilisateur concernant la généralité et la spécificité d'un nombre de concepts donnés par rapport à un concept cible.

Nous en déduisons donc la définition d'une ontologie floue découlant directement de la procédure de floutage de concepts proposée et de la relation de subsomption définie ci-dessus.

Definition 3 (Ontologie Floue) Soit \mathcal{C} un ensemble de concepts flous, $\mathbf{is_a} : \mathcal{C} \times \mathcal{C} \rightarrow [0, 1]$ une relation *is_a* floue, \mathcal{R} un ensemble de relations floues définies sur \mathcal{C} , i.e., \mathcal{R} contient des relations $r : \mathcal{C}^n \rightarrow [0, 1]$, où n est l'arité de la relation (pour des raisons de présentation, l'on ne considère que des relations binaires), \mathcal{X} un ensemble d'objets, et $\phi : \mathcal{C} \rightarrow \mathcal{F}(\mathcal{X}, [0, 1])$ une fonction qui associe une fonction d'appartenance à chaque concept flou dans \mathcal{C} . Nous exigeons que $\mathbf{is_a}$ et ϕ soient compatibles, i.e., $\mathbf{is_a}(\mathcal{A}', \mathcal{A}) = \inf_x \mu_{\mathcal{A}' \rightarrow \mathcal{A}}(x)$ est vraie pour tout $\mathcal{A}', \mathcal{A} \in \mathcal{C}$. Particulièrement, il peut être démontré que cela implique que $\mathbf{is_a}$ est un quasi-ordre flou. Ainsi, la collection

$$\mathcal{O} = (\mathcal{C}, \mathbf{is_a}, \mathcal{R}, \mathcal{X}, \phi)$$

forme une ontologie floue.

Ci-dessus, l'ensemble \mathcal{X} est défini comme un ensemble d'objets abstraits. Dans notre contexte, ce sont les concepts de l'ontologie de référence ($\mathcal{X} = X$). L'ensemble \mathcal{C} est un sous-ensemble donné de \mathcal{C}_Ω . Dans le cas où $\mathcal{C} = \mathcal{C}_1$, où \mathcal{C}_1 est l'ensemble des concepts flous de l'ontologie O_1 , \mathcal{O} définit une version floue de l'ontologie classique O_1 . Dans le cas où $\mathcal{C} = \mathcal{C}_\Omega$, \mathcal{O} définit une ontologie commune pour les deux ontologies source. Notons que les valeurs d'appartenance des concepts de référence impliquent des valeurs d'appartenance des documents qui les peuplent. Cependant, nous travaillons directement avec les scores des concepts dans ce qui suit.

En se basant sur la relation de subsomption définie ci-dessus, nous définissons l'équivalence floue entre deux concepts de la manière suivante.

Definition 4 (θ -Equivalence Floue) Une θ -équivalence entre le concept \mathcal{A} et le concept \mathcal{B} , dénotée par $\mathcal{A} \sim_\theta \mathcal{B}$ est vraie si et seulement si $\mathbf{is_a}(\mathcal{A}, \mathcal{B}) > \theta$ et $\mathbf{is_a}(\mathcal{B}, \mathcal{A}) > \theta$, où θ prend ses valeurs entre 0 et 1.

La relation d'équivalence permet de définir des classes d'équivalence sur l'ensemble \mathcal{C}_Ω . Dans le contexte de la plate-forme collaborative du début, ceci peut être utilisé pour faire des requêtes sur les concepts qui sont équivalents (à un certain degré près) à un concept donné.

5.2 Mesures de similarité entre des concepts flous

Nous proposons plusieurs mesures de similarité entre deux concepts flous \mathcal{A} et \mathcal{B} . Nous commençons par l'introduction d'une mesure de base :

$$\rho_{\text{base}}(\mu_{\mathcal{A}}, \mu_{\mathcal{B}}) = 1 - \max_{x \in X} |\mu_{\mathcal{A}}(x) - \mu_{\mathcal{B}}(x)|. \quad (6)$$

Ensuite, nous considérons une similarité basée sur la distance Euclidienne :

$$\rho_{\text{eucl}}(\mu_{\mathcal{A}}, \mu_{\mathcal{B}}) = 1 - \|\mu_{\mathcal{A}} - \mu_{\mathcal{B}}\|_2, \quad (7)$$

où $\|x\|_2 = (\sum_{x \in X} |x|^2)^{1/2}$ est la norme ℓ^2 . Plusieurs mesures de compatibilité d'ensembles flous peuvent aussi être appliquées. La correspondance partielle de Zadeh est définie par :

$$\rho_{\text{sup-min}}(\mu_{\mathcal{A}}, \mu_{\mathcal{B}}) = \sup_{x \in X} T(\mu_{\mathcal{A}}(x), \mu_{\mathcal{B}}(x)). \quad (8)$$

Finalement, le coefficient de Jaccard est donné ainsi :

$$\rho_{\text{jacc}}(\mu_{\mathcal{A}}, \mu_{\mathcal{B}}) = \frac{\sum_x T(\mu_{\mathcal{A}}(x), \mu_{\mathcal{B}}(x))}{\sum_x S(\mu_{\mathcal{A}}(x), \mu_{\mathcal{B}}(x))}. \quad (9)$$

Il est nécessaire qu'au moins l'une des fonctions $\mu_{\mathcal{A}}$ où $\mu_{\mathcal{B}}$ prenne une valeur non-nulle pour un x donné.

Les mesures de similarité données ci-dessus fournissent des informations différentes par rapport aux relations introduites dans la sous-section précédente. Subsomption et équivalence ont pour but de caractériser une relation structurelle entre les concepts, alors que la similarité exprime la proximité entre des éléments d'un ensemble. Les deux types d'informations peuvent être utilisées de manière complémentaire dans le cadre de collaboration.

5.3 Quantifier la généralité et la spécificité relatives des concepts

L'union de deux concepts flous peut être décomposée en trois composantes, chacune quantifiant, respectivement, les points communs de ces deux concepts, la spécificité de l'un de ces concepts par rapport à l'autre et vice versa de la manière suivante :

$$S(\mathcal{A}, \mathcal{B}) = (\mathcal{A}\mathcal{B}) + (\mathcal{A} - \mathcal{B}) + (\mathcal{B} - \mathcal{A}). \quad (10)$$

Ces composantes sont définies comme suit. $\mathcal{A}\mathcal{B} = T(\mathcal{A}, \mathcal{B})$ représente ce qui est commun pour les deux concepts ; $\mathcal{A} - \mathcal{B} = T(\mathcal{A}, \neg\mathcal{B})$ décrit ce qui est caractéristique pour \mathcal{A} ; et $\mathcal{B} - \mathcal{A} = T(\mathcal{B}, \neg\mathcal{A})$ représente ce qui est caractéristique pour \mathcal{B} .

Plusieurs options de fusion peuvent être proposées à l'utilisateur par rapport aux valeurs de ces trois composantes. Dans le cas où $\mathcal{A}\mathcal{B}$ est sensiblement supérieur $\mathcal{A} - \mathcal{B}$ et $\mathcal{B} - \mathcal{A}$, les deux concepts peuvent être remplacé par leur union. Dans le cas où soit $\mathcal{A} - \mathcal{B}$ soit $\mathcal{B} - \mathcal{A}$ est plus grand que les deux autres composantes, les concepts peuvent être remplacés soit par \mathcal{A} , soit par \mathcal{B} .

6 Applications dans le domaine du multimédia

Pour valider notre approche, nous proposons son application dans le domaine du multimédia, en s'intéressant au rapprochement de deux ontologies hétérogènes complémentaires contenant des images annotées. Nous avons

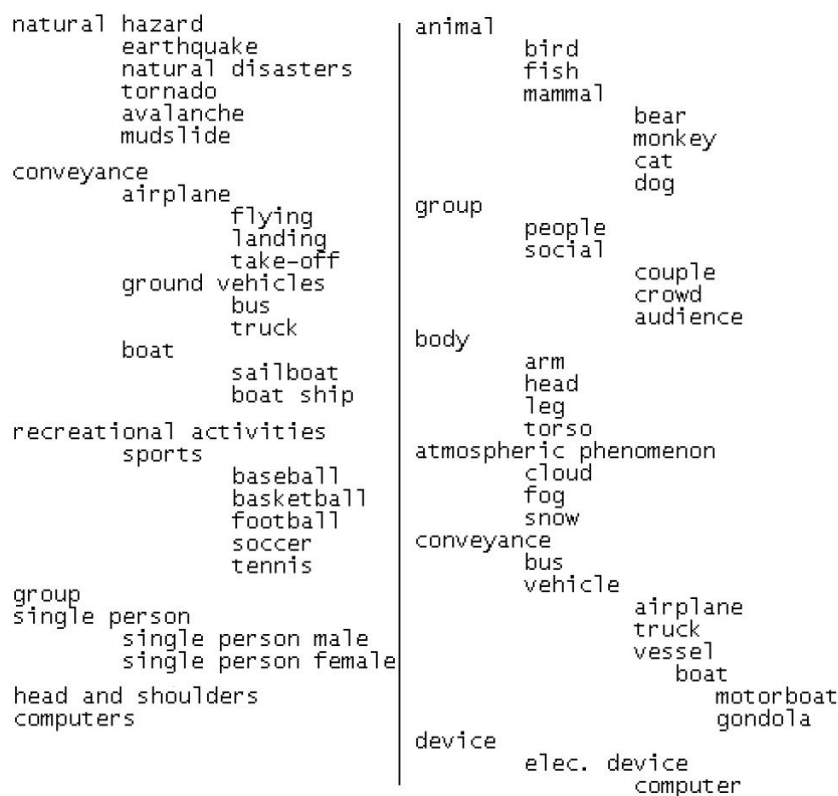


FIGURE 1 – Les ontologies LSCOM (à gauche) et LabelMe (à droite).

choisi d'un côté LSCOM [16] qui a été originalement conçue dans le cadre du challenge TRECVID², peuplée par la base d'images de TRECVID 2005. Puisque cet ensemble contient des images de journaux télévisuels, LSCOM est particulièrement adaptée à l'annotation de ce type de contenu. En conséquence, cette ontologie contient des concepts abstraits et spécifiques (e.g. SCIENCE_TECHNOLOGY, INTERVIEW_ON_LOCATION). De l'autre côté, nous avons utilisé WordNet [13] peuplée par la base de données LabelMe [14], et que l'on appellera par la suite l'ontologie LabelMe. Contrairement à LSCOM, cette ontologie est très générale, peuplée par des photographies du quotidien et contient des concepts tels que CAR, COMPUTER, PERSON, etc. Les parties de ces ontologies que nous avons utilisées dans nos expériences sont montrées dans la Figure 1.

Un document texte a été généré pour chaque image peuplant un des concepts visés dans les deux ontologies, en prenant le nom de tous les concepts qui sont contenus dans l'annotation de l'image, ainsi que les définitions (textuelles) de ces concepts (les définitions LSCOM pour les images de TRECVID ou les glosses WordNet pour les images LabelMe). Un exemple d'une instance d'un concept visuel multimédia et de la description textuelle construite est donnée dans la Figure 2. Plusieurs problèmes liés à cette représentation sont à considérer. Les descriptions des mot-clés (concepts) de LSCOM utilisent parfois



107_Standing One or more people standing up. 227_Bus Shots of a bus. 224_Outdoor Shots of Outdoor locations. 217_Person Shots depicting a person. The face may be partially visible. 202_Crowd Shots depicting a crowd. 181_Adult Shots showing a person over the age of 18. 104_Male_Person One or more male persons. 290_Daytime_Outdoor shots that take place outdoors during the day. 316_Group We defined a group as 3-10 people. 109_Windows An opening in the wall or roof of a building or vehicle fitted with glass or other transparent material.

FIGURE 2 – Le concept Bus de LSCOM et une instance visuelle et textuelle.

2. <http://www-nlpir.nist.gov/projects/tv2005/>

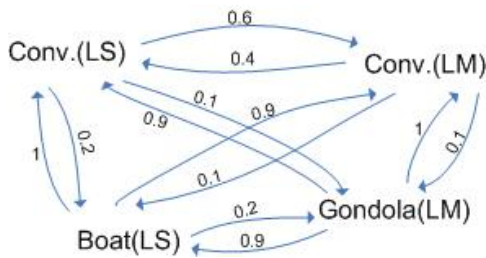


FIGURE 3 – Un fragment de l’ontologie commune construite pour (LS) et LabelMe (LM).

la négation et l’exclusion ce qui est difficile à manipuler dans une simple approche “<sac-de-mots”>. Prendre les glosses de WordNet pour les termes dans LabelMe introduit des problèmes liés à la polysémie et la synonymie. De plus, une scène se compose souvent de plusieurs objets, qui souvent ne sont pas sémantiquement liés à l’objet prédominant dans l’image. Dans de tels cas, les autres objets qui apparaissent dans l’image agissent comme du bruit.

Afin de flouter nos concepts source, nous avons appliqué l’algorithme de scoring hiérarchique (décrit dans la section 4) indépendamment pour chacune des ontologies source. Comme une ontologie de référence, nous avons utilisé une version étendue de la classification des catégories principales de Wikipédia, contenant plus de 30 classes. Pour chaque catégorie, nous avons inclus un ensemble de documents qui correspondent à cette catégorie à partir du corpus Inex 2007 [10].

La nouvelle ontologie commune a été construite en prenant d’abord l’union de tous les concepts source flous. Pour chaque paire de concepts, nous avons calculé leur relation de subsomption, ainsi que le degré de leur similarité (en appliquant les mesures décrites dans la section 5.2 et la mesure standard du cosinus). Outre la définition classique de la subsomption (5), nous avons considéré une version de celle-ci qui prend la moyenne sur tous les x au lieu de la plus petite valeur, c’est à dire : $is_a^{mean}(\mathcal{A}', \mathcal{A}) = \text{avg}_{x \in X} \mu_{\mathcal{A}' \rightarrow \mathcal{A}}(x)$. Les résultats obtenus pour plusieurs concepts qui proviennent de la même ontologie ou d’ontologies différentes sont donnés dans le tableau 1. La Figure 3 montre un fragment de l’ontologie commune floue construite pour LSCOM et LabelMe. Les étiquettes des arêtes du graphe correspondent aux valeurs de la subsomption floue entre les concepts. Finalement, la figure 4 compare la performance des mesures de similarité floues avec la mesure classique.

Nous soulignerons quelques lacunes qui doivent être adressées dans les travaux futurs. Certaines relations méronymiques (part-of) sont exprimées en tant que des relations de subsomption (e.g. *torso is_a person*). Ceci est dû à la représentation des instances par un document textuel comme expliqué au début de la section. En effet, la représentation textuelle des images doit être améliorée en prenant en compte les limitations discutées précédemment. Notons que le calcul de l’ontologie commune floue est peu

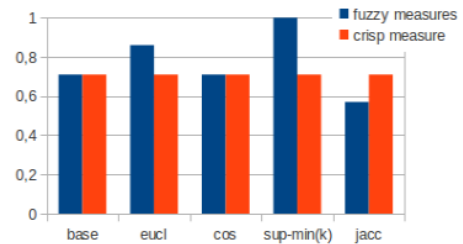


FIGURE 4 – Comparaison de la précision des mesures d’alignements floues avec des mesures non floues. La précision est calculée à la base d’un alignement manuel.

coûteux, une fois que nous avons en main les représentations floues des concepts source mise à disposition par l’algorithme de scoring hiérarchique.

7 Conclusion et futurs travaux

Quand la collaboration entre des ressources de connaissances hétérogènes est nécessaire, il est important de fournir des procédures qui rendent explicites aux utilisateurs les relations entre les différents termes de ces ressources. En tant que tentative pour résoudre ce problème, nous avons proposé une approche théorique floue pour construire une ontologie commune pour un ensemble d’ontologies source qui contient ces relations, ainsi que les degrés de vérité de ces relations. Cette ontologie floue peut être interrogée par différentes parties dans un cadre collaboratif.

Dans nos travaux futurs, nous allons étudier l’impact du choix de l’ontologie de référence sur la représentation floue de nos concepts et sur la qualité de l’ontologie commune floue conçue par la suite ainsi que les critères exacts auxquels cette ontologie de référence doit répondre. En outre, la démarche sera étendue avec des éléments de OWL 2, considérant des relations et des axiomes définies pour les instances qui ne sont pas couverts par la définition d’une ontologie utilisée dans ce travail.

Références

- [1] J.-I. Akahani, K. Hiramatsu, and T. Satoh. Approximate query reformulation based on hierarchical ontology mapping. In *SWFAT*, pages 43–46, 2003.
- [2] A. Bahri, R. Bouaziz, and F. Gargouri. Dealing with similarity relations in fuzzy ontologies. In *Fuzzy Systems Conference, 2007. FUZZ-IEEE 2007. IEEE Intl*, pages 1–6. IEEE, 2007.
- [3] F. Bobillo. *Managing vagueness in ontologies*. PhD thesis, PhD thesis, University of Granada, Spain, 2008.
- [4] P. Buche, J. Dibie-Barthélemy, and L. Ibanescu. Ontology mapping using fuzzy conceptual graphs and rules. In *ICCS Supplement*, pages 17–24, 2008.

Concept \mathcal{A} :	LSCM :truck vs.	LSCM :sports vs.	LM :computer vs.	LM :animal vs.
Concept \mathcal{B} :	LSCM :ground vehicle	LSCM :basketball	LM :elec. device	LM :bird
$\mathbf{is_a}(\mathcal{A}, \mathcal{B})$	1	0.007	1	0.004
$\mathbf{is_a}(\mathcal{B}, \mathcal{A})$	0.012	1	0.011	1
$\mathbf{is_a}^{\text{mean}}(\mathcal{A}, \mathcal{B})$	1	0.052	1	0.062
$\mathbf{is_a}^{\text{mean}}(\mathcal{B}, \mathcal{A})$	0.326	1	0.07	1
Base Sim.	0.848	0.959	0.915	0.390
Eucl. Sim.	0.835	0.908	0.854	0.350
SupMin Sim.	0.435	0.545	0.359	0.309
Jacc. Sim.	0.870	0.814	0.733	0.399
Cosine Sim.	0.974	0.994	0.975	0.551

Concept \mathcal{A} :	LM :gondola vs.	LSCM :group vs.	LSCM :truck vs.	LSCM :truck vs.
Concept \mathcal{B} :	LSCM :boat_ship	LM :audience	LM :vehicle	LM :conveyance
$\mathbf{is_a}(\mathcal{A}, \mathcal{B})$	0.016	0.006	0.022	0.022
$\mathbf{is_a}(\mathcal{B}, \mathcal{A})$	0.009	1	0.012	0.012
$\mathbf{is_a}^{\text{mean}}(\mathcal{A}, \mathcal{B})$	0.86	0.022	0.748	0.769
$\mathbf{is_a}^{\text{mean}}(\mathcal{B}, \mathcal{A})$	0.167	1	0.301	0.281
Base Sim.	0.72	0.78	0.58	0.58
Eucl. Sim.	0.66	0.71	0.40	0.38
SupMin Sim.	0.069	0.082	0.22	0.22
Jacc. Sim.	0.49	0.42	0.54	0.52
Cosine Sim.	0.69	0.82	0.66	0.67

TABLE 1 – Exemples de paires de concepts alignés venant d’une même ontologie (ci-dessus) et des ontologies différentes (ci-dessous).

- [5] S. Calegari and D. Ciucci. Fuzzy ontology, fuzzy description logics and fuzzy-owl. In *WILF*, pages 118–126, 2007.
- [6] S. Calegari and E. Sanchez. A fuzzy ontology-approach to improve semantic information retrieval. In *URSW*, 2007.
- [7] V. Cross and X. Yu. A fuzzy set framework for ontological similarity measures. In *WCCI 2010, FUZZ-IEEE 2010*, pages 1 – 8. IEEE Computer Society Press, 2010.
- [8] J. Euzenat and P. Shvaiko. *Ontology Matching*. Springer-Verlag, 1 edition, 2007.
- [9] A. Ferrara, D. Lorusso, G. Stamou, G. Stoilos, V. Tzouvaras, and T. Venetis. Resolution of conflicts among ontology mappings : a fuzzy approach. OM’08 at ISWC, 2008.
- [10] N. Fuhr, J. Kamps, M. Lalmas, and A. Trotman, editors. *Focused Access to XML Documents, 6th INEX 2007*, volume 4862 of *LNCS*. Springer, 2008.
- [11] A. Gal and P. Shvaiko. Advances in web semantics i. chapter Advances in Ontology Matching, pages 176–198. Springer-Verlag, Berlin, Heidelberg, 2009.
- [12] M. S. Lacher and G. Groh. Facilitating the exchange of explicit knowledge through ontology mappings. In *FLAIRS*, pages 305–309. AAAI Press, 2001.
- [13] G.A. Miller. WordNet : a lexical database for English. *Communications of the ACM*, 38(11) :39–41, 1995.
- [14] B.C. Russell, A. Torralba, K.P. Murphy, and W.T. Freeman. Labelme : A database and web-based tool for image annotation. *Intl Journal of Computer Vision*, 77(1), 2008.
- [15] E. Sanchez and T. Yamanoi. Fuzzy ontologies for the semantic web. *Flexible Query Answering Systems*, pages 691–699, 2006.
- [16] J.R. Smith and S.F. Chang. Large-scale concept ontology for multimedia. *IEEE Multimedia*, 13(3) :86 – 91, 2006.
- [17] H. Stuckenschmidt. Approximate information filtering on the semantic web. In M. Jarke, G. Lakemeyer, and J. Koehler, editors, *KI 2002*, volume 2479 of *LNCS*, pages 195–228. Springer Berlin / Heidelberg, 2002.
- [18] K. Todorov, P. Geibel, and K.-U. Kühnberger. Mining concept similarities for heterogeneous ontologies. In P. Perner, editor, *Advances in Data Mining. Applications and Theoretical Aspects*, volume 6171 of *LNCS*, pages 86–100. Springer Berlin / Heidelberg, 2010.
- [19] B. Xu, D. Kang, J. Lu, Y. Li, and J. Jiang. Mapping fuzzy concepts between fuzzy ontologies. In R. Khosla, R. J. Howlett, and L. C. Jain, editors, *KES*, volume 3683 of *LNCS*, pages 199–205. Springer Berlin / Heidelberg, 2005.
- [20] L.A. Zadeh. Fuzzy sets. *Information and Control*, 8(3) :338 – 353, 1965.