



HAL
open science

Intégration d'une ontologie de domaine dans un système de question-réponse

Farah Benamara, Mouna Kamel

► To cite this version:

Farah Benamara, Mouna Kamel. Intégration d'une ontologie de domaine dans un système de question-réponse. 18es Journées Francophones d'Ingénierie des Connaissances, Jul 2007, Grenoble, France. not specified. hal-00509865

HAL Id: hal-00509865

<https://hal.science/hal-00509865>

Submitted on 16 Aug 2010

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.

Intégration d'une ontologie de domaine dans un système question-réponse

Farah Benamara et Mouna Kamel

Institut de Recherche en Informatique de Toulouse,
{benamara, kamel}@irit.fr

Résumé : Dans cet article, nous proposons une méthode pour intégrer une ontologie au sein d'un système question-réponse (SQR) nommé WEBCOOP. Comme la plupart des SQR intégrant une ontologie de domaine, ce système se base sur une ontologie où chaque concept est décrit uniquement par ses propriétés intrinsèques. Nous proposons d'étendre ce système afin de prendre en compte une modélisation plus fine des concepts et d'étudier l'impact de ce type de modélisation sur les procédures de raisonnement nécessaires à la phase d'extraction de réponses. Nous montrons comment caractériser les différents types de propriétés qu'un concept peut posséder, tout en nuancant leur existence, puis, nous présentons une mesure de similarité qui prenne en compte des critères sémantiques tels que les propriétés structurelles et fonctionnelles.

Mots-clés : Systèmes question-réponse, modélisation des connaissances, relaxation de contraintes, mesure de similarité.

1 Introduction

Les systèmes question-réponse à domaines fermés (SQR-DF) permettent à des experts du domaine ainsi qu'à des non-spécialistes d'interroger des documents de domaines d'application spécialisés afin d'obtenir les réponses les plus pertinentes possibles. Ce type de système est un challenge supplémentaire pour les SQR actuels car ils utilisent des collections documentaires restreintes qui rendent les techniques classiques à base de redondance d'information inutilisables. De plus, les SQR-DF manipulent des ressources spécifiques dépendantes du domaine d'application telles que des lexiques ou des ontologies, contrairement aux SQR à domaine ouvert qui utilisent des ressources génériques telles que WordNet ou FrameNet. Les SQR-DF posent donc très clairement des problématiques d'ingénierie des connaissances que l'on peut situer sur deux plans : (1) quel est le bon niveau de granularité pour décrire les connaissances du domaine ? (2) comment ce mode de représentation peut-il être utilisé dans le processus de recherche de réponses ?

Dans ce contexte, différents travaux se sont intéressés à l'intégration d'ontologies dans un SQR, principalement pour deux objectifs:

- *la représentation de la question.* Les questions traitées sont des questions factoides où le focus (l'objet de la question) ainsi que ses contraintes sont des concepts de l'ontologie. Par exemple, dans la question '*je cherche un hôtel 3*

étoiles à Paris'', *hôtel* est un concept, *3 étoiles* et *Paris* sont respectivement des propriétés de catégorie et de localisation géographique du concept *hôtel*.

- *la recherche de réponses*, où l'ontologie est utilisée dans des procédures de raisonnement, comme par exemple, relaxer les concepts de la question qui causent son échec en utilisant diverses méthodes de relaxation de contraintes via un processus de généralisation/spécialisation [Bidault, 2002], réécrire la requête avec le vocabulaire de la base de connaissances [Bidault, 2002][Vargas & Motta, 04], généraliser une longue liste de réponses [Han & al, 1996]. D'autres travaux se sont intéressés à l'extraction de réponses à partir de sources d'information différentes en utilisant des méthodes de rapprochement d'ontologies [Atzeni & al., 2004]. D'une façon générale, ces procédures de raisonnement se basent sur le calcul d'une mesure de similarité entre les concepts de la requête avec ceux décrits dans l'ontologie. Ces mesures ne vont pas au-delà de la relation hiérarchisante « est-un » de l'ontologie.

Ce travail propose une méthode d'intégration d'une ontologie de domaine au sein de WEBCOOP [Benamara, 2004], un SQR coopératif sur le domaine du tourisme. Comme la plupart des SQR intégrant une ontologie de domaine, ce système se base sur une ontologie où chaque concept est décrit uniquement par ses propriétés intrinsèques. Nous proposons d'étendre ce système afin de prendre en compte une modélisation plus fine des concepts et d'étudier l'impact de ce type de modélisation sur les procédures de raisonnement nécessaires à la phase d'extraction de réponses.

L'article est organisé comme suit. Nous présentons, en section 2, le cadre du travail ainsi que les principales motivations de notre approche. Puis, nous donnons, en section 3, une représentation des connaissances relatives aux concepts de notre ontologie. La section 4 décrit comment la modélisation que nous proposons s'intègre dans le processus de relaxation de contraintes de WEBCOOP. Nous définissons alors une mesure de similarité qui a pour but de prendre en compte des critères sémantiques autres que la relation hiérarchisante 'est-un'. Nous terminons cet article par une évaluation de cette métrique de similarité.

2 Cadre du travail et motivations

Nous nous situons dans le cadre du système WEBCOOP [Benamara, 2004], un SQR sans dialogue et sans modèle de l'utilisateur qui va au-delà de l'extraction de portions de textes répondant directement à la question. L'un des principaux défis de ce type de système est la génération d'explications et de justifications en LN qui valident les réponses. Ce système intègre des procédures de raisonnement (détection d'un malentendu ou d'une fausse présupposition, relaxation de contraintes, réponses intentionnelles) ***qui opèrent sur une ontologie de domaine où chaque concept est décrit uniquement par ses propriétés intrinsèques*** ainsi que sur une collection de bases de connaissances décrivant des connaissances de sens commun.

Contrairement aux approches plus classiques qui consistent à supprimer dans la question le ou les concepts responsables de l'échec ou les remplacer par un de leurs

généralisants dans l'ontologie, l'approche adoptée consiste à prendre en compte des formes plus subtiles de relaxation comme remplacer les concepts échecs par leurs concepts frères les plus proches dans l'ontologie en termes de propriétés. Dans cet article, nous nous sommes intéressés à la relaxation sur critère ontologique qui consiste à rechercher dans l'ontologie de domaine le concept C' qui relaxera au mieux le concept C, cause de l'échec de la question grâce à une métrique de similarité.

Le mécanisme de relaxation examine une à une les sous-requêtes de la question à l'origine de l'échec. Notre stratégie consiste alors à présenter à l'utilisateur différentes solutions qui correspondent à différents axes de relaxation et à le laisser choisir celui qui répond le mieux à ses attentes. Un seul axe de relaxation est utilisé à la fois. Cette stratégie permet de corriger les malentendus ou les fausses présuppositions de la question et pallie l'absence de modèle usager dans le système, comme le montre l'exemple suivant :

Question : *je cherche un village vacances avec piscine*

Réponse : *Nous ne disposons pas de réponses à votre question. Nous vous proposons les solutions alternatives suivantes :*

- *Un village vacances proposant des structures aquatiques*
- *D'autres types d'hébergements touristiques avec piscine.*

Dans cet exemple, les concepts *village vacances* et *piscine* sont relaxés chacun à leur tour, les liens hypertextes (texte souligné) permettent d'accéder aux pages web décrivant les solutions proposées. De plus, le système explique à l'utilisateur les principales étapes du processus de relaxation en utilisant des méthodes de génération du langage naturel [Benamara, 2004].

L'objectif de notre travail est d'étendre ce mécanisme de relaxation afin d'être en mesure de proposer des explications de réponses plus pertinentes, comme celles décrites dans le second exemple suivant :

Question : *je cherche un village vacances avec piscine*

Réponse : *Nous ne disposons pas de réponses à votre question. Nous vous proposons les solutions alternatives suivantes :*

Réponse 1 : Un village vacances proposant des structures aquatiques

Réponse 2 : D'autres types d'hébergements touristiques avec piscine proche du village vacances en termes :

- * d'équipement : appart-hôtel, gîte, chalet*
- * d'animation : camping.*
- * de prestation : camping*

Dans ce second exemple, nous observons un niveau d'explication supplémentaire (cf réponse 2) où les concepts les plus similaires au concept *village vacances* sont présentés à l'utilisateur selon différents axes ontologiques. Ce choix a été guidé par l'étude d'un corpus d'environ 80 paires question-réponse tirées de forums et de sections FAQ de différents services web dans le domaine des hébergements touristiques. Ces questions et réponses sont produites par des humains et

communiquées par le biais du courrier électronique¹. Nous nous sommes focalisés sur des questions où le focus ainsi que ses contraintes sont des concepts de l'ontologie. Cette étude de corpus a permis de définir le bon niveau de granularité de représentation des connaissances dans notre domaine :

- Dans la plupart des cas les contraintes de la question sont des propriétés du focus. Les questions sont exprimées soit d'un point de vue structurel (*un hébergement avec sauna*), soit d'un point de vue fonctionnel (*un hébergement avec remise en forme*). La configuration qui en ressort est : (focus)(propriétés_focus)*. C'est pourquoi, il nous semble intéressant d'utiliser ces propriétés dans le processus d'explication, comme le montre le second exemple.
- Très peu de questions comportent des contraintes relatives aux propriétés du focus i.e : (focus)(propriétés_focus (propriétés_propriétés_focus))* , comme par exemple : *je cherche un hôtel avec une piscine d'au moins 25m de longueur*. C'est pourquoi, nous n'apportons pas d'explication supplémentaire (cf Réponse 1).

Pour mettre en œuvre un tel processus de relaxation ainsi que les mécanismes d'explication associés, nous allons enrichir notre ontologie de domaine par une description plus fine des concepts, en caractérisant *les différents types de propriétés* qu'un concept peut posséder, et *en pondérant leur existence*.

3 Modélisation des connaissances

La grande diversité du vocabulaire touristique nous a conduits à réutiliser puis à adapter, selon nos besoins, des hiérarchisations existantes trouvées sur le web². Notre objectif n'est pas d'avoir une ontologie regroupant tous les termes du domaine du tourisme mais d'aboutir à une hiérarchisation réutilisable, manipulant des concepts 'standards' communément admis par les professionnels du tourisme. Notre ontologie des hébergements touristiques comporte, pour le moment, environ 110 concepts et six relations.

Un extrait est présenté dans la figure 1. Dans cette figure, la hiérarchie a été définie en fonction de critères de différenciation relatifs à notre point de vue, tels que les hébergements avec ou sans structure d'accueil, chez le particulier ou non, etc. Au vu de cette classification, les concepts *appart-hôtel* et *meublé* d'une part, *gîte d'étape* et *ferme auberge* d'autre part (en gras dans la figure), paraissent éloignés. Pourtant, ces concepts peuvent être proches si l'on parle respectivement en terme d'équipement en électroménager et/ou d'infrastructure (ces hébergements proposent les mêmes commodités) ou en terme d'environnement (hébergements ruraux).

¹ Pour une description détaillée de la démarche d'analyse de corpus voir [Benamara, 2004]

²TourinFrance (www.tourinfrance.net) développé par le secrétariat d'Etat au Tourisme ; <http://www.tourisme-alsace.com>; http://www.vacancesuvertes.asso.fr/pages/info_vacances/dossier/lexique/

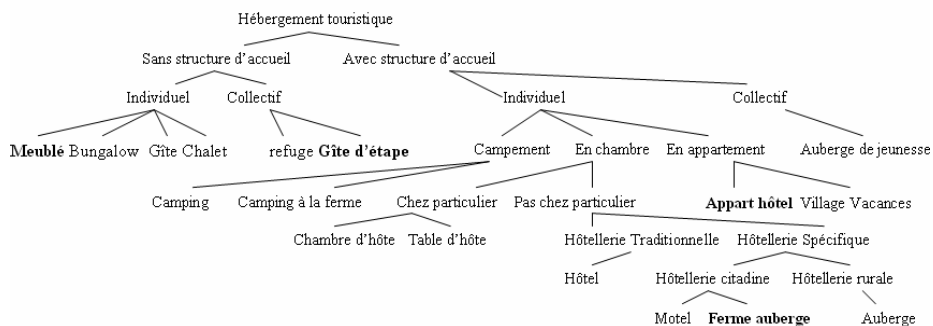


Figure 1. Extrait de l'ontologie des hébergements touristiques

Pour pouvoir rapprocher des concepts sur différents aspects (autres que taxonomiques), des critères dits structurels et/ou sémantiques au niveau des concepts devront être définis par le concepteur dans le but d'enrichir l'ontologie, fournissant ainsi pour nous des moyens de trouver des similitudes graduelles entre concepts et d'affiner le processus de relaxation de contraintes dans WEBCOOP.

3.1 Définition d'un concept

Un concept sera donc défini par *ses propriétés structurelles* qui décrivent un état et *ses propriétés fonctionnelles* qui décrivent les fonctions que doit remplir le concept, ce qui peut être réalisé sur lui ou avec lui (et qui sont inspirés de la structure Qualia [Pustejovsky, 1995]). Une propriété fonctionnelle que nous appelons aussi *rôle* peut être réalisée grâce à l'aspect structurel (l'existence de structures d'animation dans le *Village Vacances* permet la réalisation du rôle *Divertir*) et/ou à d'autres rôles (le rôle *Détendre* peut être déduit du rôle *Divertir*).

Par ailleurs, les experts en ingénierie des connaissances modélisent habituellement un concept d'ontologie en fonction des seules propriétés intrinsèques au concept [Breuker, 06] ; dès qu'une de ces propriétés est modifiée, le concept devient autre. La méthode de spécification semi-formelle OntoSpec [Kassel, 2002] offre la possibilité de différencier les propriétés essentielles des propriétés dites contingentes, éliminant l'aspect booléen lié à la présence d'une propriété. Cette possibilité nous a semblé très intéressante car dans diverses applications et notamment dans notre contexte, certaines propriétés peuvent être contingentes pour un concept, et essentielles pour un autre. Par exemple, la propriété *petit déjeuner compris* est essentielle pour le concept *chambre d'hôte*, contingente pour le concept *hôtel*. Si la plupart des hôtels fournissent le *petit déjeuner*, les concepts *chambre d'hôte* et *hôtel* peuvent être proches de ce point de vue.

Ainsi, pour chaque concept C, nous intégrerons dans notre modèle *les propriétés essentielles* qui sont les propriétés nécessaires et suffisantes pour spécifier C et *les propriétés contingentes* qui représentent des propriétés que peut posséder C. Chaque propriété (essentielle ou contingente) sera marquée d'une probabilité pondérant son existence (les propriétés essentielles auront une probabilité égale à 1).

Le rapprochement entre concepts pourra alors se faire sur un ensemble de propriétés jugées pertinentes, et sera évalué en fonction des valeurs des probabilités. Nous définissons alors un concept de l'ontologie par :

- le concept père
- les propriétés structurelles qui peuvent être essentielles ou contingentes
- Les propriétés fonctionnelles qui peuvent être aussi essentielles ou contingentes

3.2 Description de l'ontologie et représentation formelle

Les hébergements touristiques et les propriétés des hébergements sont représentés au sein de l'ontologie par des concepts. Le fait qu'une propriété soit associée à un concept de type hébergement est matérialisé par **une relation** entre ces deux concepts. Les rôles sont définis par l'expert du domaine, et chaque rôle est matérialisé par une relation entre 2 ou 3 concepts selon les arguments nécessaires à la réalisation du rôle. Nous donnons ci-dessous un extrait de notre ontologie :

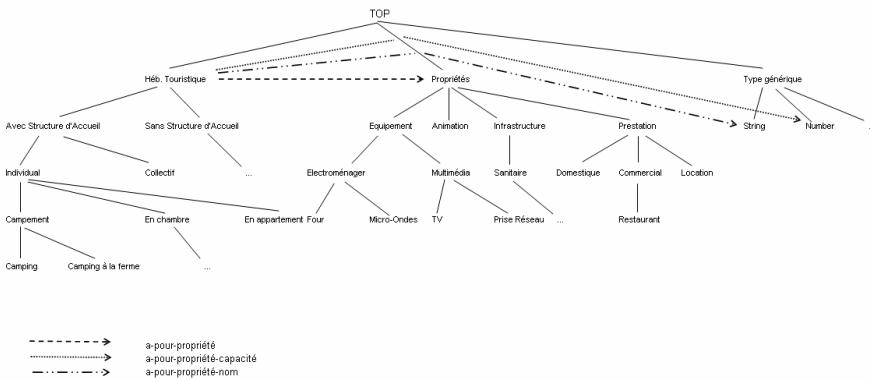


Figure 2. Extrait de notre ontologie de domaine

Comme les concepts issus du concept "Héb. Touristique" sont déjà proches du fait de leur appartenance à un domaine spécifique qui est celui des hébergements touristiques, nous avons opté pour associer l'ensemble des propriétés au concept le plus général (Héb. Touristique). Toute propriété pour tout hébergement se verra affecter d'une probabilité, ce qui permettra de comparer deux hébergements sur les valeurs des propriétés plutôt que sur leurs noms.

Formellement, nous définissons notre ontologie par le tuple $O = (C, Rel, Ro)$ où C est l'ensemble des concepts, Rel l'ensemble des relations tel que : $Rel = R_p \cup R_n$, et Ro est l'ensemble des rôles. Un concept est défini par la structure:

Concept { Nom: c tq $c \in C$,
 $Rel = \{ R_n = \{ \text{est-un}(c, c') / \text{"est-un"} \text{ est la relation hiérarchique et } c' \in C \},$
 $R_p = \{ (R(c, c'), p) / R \text{ est une relation, } c' \in C, \text{ et } p \in [0,1] \}$
 $Ro = \{ (Rl(c_1, \dots, c_n), p) / Rl \text{ est un rôle, } n = \text{arité du verbe identifiant le rôle avec } n \in \{1,2,3\},$
 $c_i \in C \text{ et } c \in \{c_1, \dots, c_n\}, p \in [0,1] \}$ }

On donne ci-dessous un exemple partiel de la définition du concept *village vacances*

Concept { nom : Village Vacances,
 Rel= { $R_h = \{ \text{est-un (Village Vacances, En Appartement)} \}$,
 $R_p = \{ \{ \text{a-pour-propriété-nom (Village Vacances, STRING), 1),$
 (a-pour-propriété-capacité (Village Vacances, NUMBER), 1),
 (a-pour-propriété-catégorie (Village Vacances, STRING), 1),
 (a-pour-propriété (Village Vacances, kitchenette), 0.7),
 (a-pour-propriété (Village Vacances, Sauna), 0.4), \}
 $R_o = \{ \text{divertir (Village Vacances, Client), 0.9),} \}$

4 Une proposition d'intégration d'une ontologie dans un SQR

La proximité entre concepts va être calculée en fonction des aspects structurels et/ou fonctionnels. La métrique proposée sera ajustable selon le degré de granularité souhaité. Du fait de la structuration hiérarchique des propriétés, le rapprochement entre concepts pourra se faire soit sur une propriété (*posséder un terrain de tennis*) soit sur une propriété plus générique (*posséder des équipements sportifs*).

4.1 Mesure de similarité : état de l'art

L'étude de l'évaluation de la similarité entre concepts d'ontologies a fait l'objet de nombreux travaux, et fournissent globalement deux familles de mesures : celles qui se basent essentiellement sur la fonction structurante de la taxonomie et concernent la similarité entre concepts d'une même ontologie, et celles qui font appel à des méthodes terminologiques, structurelles et/ou extensionnelles et qui concernent la similarité entre concepts d'ontologies différentes.

Pour évaluer la similarité entre concepts d'une même ontologie, [Rada & al., 1989] propose de mesurer la distance entre 2 concepts par le nombre d'arcs qui les sépare, mais cela suppose que les distances soient sémantiquement homogènes. [Wu, Palmer 1994] et [Zargayouna & Salotti, 2004] améliorent cette distance en calculant les distances par rapport à la racine (resp. aux feuilles). [Resnik, 1995] associe à chaque concept son contenu informationnel, afin que sa mesure ne dépende pas uniquement de la conception de l'ontologie. D'autres mesures sont définies en proposant des variantes ou des combinaisons des mesures de Rada et Resnik [Jiang & Conrath, 1997][Lin, 1997]. [Hirst & St Onge 1998] évaluent la distance entre 2 concepts en calculant un chemin guidé par les relations WordNet, en incluant les changements de direction générée par ces relations sur le chemin qui sépare les 2 concepts. La plupart de ces mesures ne fait pas intervenir l'aspect structurel des concepts, et ne permet donc pas d'ajuster la métrique en fonction de critères souhaités.

Les mesures de similarité entre concepts d'ontologies différentes mettent en œuvre des mécanismes pour pouvoir comparer deux concepts sur leurs noms, propriétés, relations, voisinages et extensions. Deux ensembles de propriétés, relations, voisinages ou extensions sont comparés en fonction de l'intersection et de

l'union de ces ensembles [Maedche & Staab, 2001][Rodriguez & Egenhofer, 2003]. Les relations peuvent aussi être comparées sur leur domaine et co-domaine [Ehrig & Sure, 2004], [Cullot & al, 2003], [Noy, 2004] (on parle alors de méthodes structurelles internes), ou encore sur leurs propriétés comme la symétrie, la transitivité, etc. [Ehrig & Sure, 2004][Navathe & Buneman, 1986] (ces méthodes sont dites structurelles externes). Les méthodes extensionnelles comparent les instances des concepts. Ces méthodes peuvent être élargies aux voisinages des concepts, en comparant un concept source aux concepts auxquels le concept cible est relié. On parle alors de méthodes globales, par opposition aux méthodes locales. Toutes ces mesures font appel à des méthodes terminologiques qui comparent les chaînes de caractères désignant les items, ou qui utilisent des connaissances linguistiques annexes (relations Wordnet, lexiques, thesaurus, etc.).

4.2 Métrique de similarité pour la relaxation de contraintes

Notre étude se situe dans le cadre d'une seule ontologie, et les concepts seront comparés sur l'ensemble de leurs propriétés et leurs rôles selon des méthodes structurelles réservées jusque là aux calculs de similarité entre concepts d'ontologies différentes. Par contre, la comparaison entre propriétés ou rôles se fera sur les probabilités associées, et non par sur des méthodes terminologiques comparant les chaînes de caractères les caractérisant.

Notre métrique prendra en charge l'aspect taxonomique de l'ontologie, l'aspect structurel via les propriétés, l'aspect sémantique via les rôles, tout en tenant compte de l'existence potentielle (probabilités) des propriétés et des rôles.

La mesure de similarité normalisée entre deux concepts C_1 et C_2 , notée $sim(C_1, C_2)$ qui répond à ces attentes, est définie par :

$$sim(C_1, C_2) = (w_1 \times sim_h(C_1, C_2)) + (w_2 \times sim_{prop}^\Phi(C_1, C_2)) + (w_3 \times sim_{role}^R(C_1, C_2)) \text{ avec } \sum_{i=1}^3 w_i = 1$$

où Φ (rep. R) est un ensemble de propriétés structurelles (resp. de rôles) sur lesquels la similarité sera évaluée. Les poids w_i permettent de pondérer les similarités, selon que l'on souhaite plus ou moins mettre l'accent sur l'aspect ontologique, structurel ou fonctionnel.

(1) sim_h est la similarité basée sur la relation hiérarchique et est évaluée par la mesure classique de distance (dite sémantique de [Wu & palmer, 1994]) :

$$sim_h(C_1, C_2) = \left(\frac{2 \times d(C_{sub})}{d(C_1) + d(C_2)} \right)$$

$d(C_i)$ est la distance entre la racine et C_i , et C_{sub} est le concept qui subsume C_1 et C_2

(2) sim_{prop}^Φ est la similarité calculée sur l'ensemble des propriétés appartenant à Φ . Soient C_1 et $C_2 \in C$, $p \in \Phi$. S'il existe $R \in R_p$ telle que $(R(C_1, p), p_1)$ et $(R(C_2, p), p_2)$, alors la distance entre deux concepts C_1 et C_2 pour la propriété p est donnée par :

$$dist_p(C_1, C_2) = \begin{cases} (p_1 - p_2)^2 & \text{si } p_1 \neq 0 \text{ et } p_2 \neq 0 \\ 0,75 & \text{si } p_1 = 0 \text{ et } p_2 = 0 \\ 1 & \text{si } p_1 = 0 \text{ ou (exclusif) } p_2 = 0 \end{cases}$$

La distance entre deux propriétés de concepts est maximisée (égale à 1) lorsque seul un des deux concepts possède la propriété. Différentes évaluations nous ont permis de fixer à 0.75 la distance entre 2 concepts lorsque les deux probabilités sont nulles (petite similarité). Enfin la distance lorsque les 2 probabilités sont différentes de 0 est calculée de façon classique, selon la distance euclidienne.

La distance sur un ensemble de critères est alors $dist_p(C_1, C_2) = \frac{\sum_{p \in \mathbb{P}} dist_p(C_1, C_2)}{|\mathbb{P}|}$

La similarité normalisée est évaluée par : $sim_{prop}^p(C_1, C_2) = 1 - dist_p(C_1, C_2)$

(3) $sim_{rôles}^R$ est la similarité calculée sur l'ensemble des rôles appartenant à R_r . Soient C_1 et $C_2 \in C$, $r \in R_r$, et tel que $r \in R_o$. Si le rôle r est associé à C_1 et C_2 alors la similarité entre ces deux concepts pour le rôle r est calculée de la même façon que pour les propriétés structurelles.

5 Implémentation et évaluation de la métrique de similarité

Dans le cadre des ontologies, les propriétés structurelles d'un concept sont représentées par des relations binaires liant deux concepts. Nous avons alors été confronté au fait que les langages du Web sémantique (RDFs, OWL qui sont aujourd'hui les langages de référence pour la description des ontologies) considèrent les propriétés comme des relations binaires, qui lient deux instances ou une instance à une valeur. Or dans notre cas, un degré de certitude est associé à chaque relation d'une part, et d'autre part, les rôles relient n concepts ($n \in \{1,2,3\}$). Les relations sont donc des relations n -aires et en OWL, ce type de relation ne peut être représenté que par des classes [Noy & Rector, 2004]. Les instances de cette classe correspondent alors aux instances de la relation.

L'évaluation de la métrique nécessite tout d'abord l'évaluation des probabilités associées aux propriétés et aux rôles. L'évaluation que nous reportons ici n'a été réalisée que sur la relaxation des contraintes relatives aux propriétés. Par la suite, nous envisageons d'élargir cette étude en prenant en compte l'aspect hiérarchique et les rôles décrits dans l'ontologie.

5.1 Evaluation des probabilités

Soit *Prob* la fonction qui calcule la probabilité d'existence d'une/des propriétés.

a) Probabilités des propriétés

1) Soient $R \in R_p, \{c, c'\} \in C$ tq $R(c, c')$ et il n'existe pas $c'' \in C$ tel que $R_h(c'', c')$ (c' n'a pas de fils). Alors : $\text{Prob}(R(c, c')) = p$ avec

$$p = \text{nombre d'instances de } c \text{ possédant } c' / \text{nombre d'instances de } c$$

2) Soient $R \in R_p, \{c, c'\} \in C$ tq $R(c, c')$ et $\exists c_i'' \in C$ tel que $R_h(c_i'', c')$, pour $i=1, \dots, n$. (c' possède n fils c_i'')

$$(1) \quad \text{Pr ob}_1(R(c, c')) = \sum_{i=1}^n \text{Pr ob}(R(c, c''_i)) - \sum_{i,j=1}^n \text{Pr ob}(R(c, c''_i), R(c, c''_j)) + \dots + (-1)^{n-1} \sum_{i_1, \dots, i_n} \text{Pr ob}(R(c, c''_{i_1}), \dots, R(c, c''_{i_n}))$$

où $\text{Prob}(R(c, c_{i_1}), \dots, R(c, c_{i_n})) = p_2$ avec

$$p_2 = \text{nombre d'instances de } c \text{ possédant } c_{i_1} \dots c_{i_n} / \text{nombre d'instances de } c$$

b) Probabilités associées aux rôles

Pour le moment, nous n'évaluons que les rôles qui sont uniquement liés à des propriétés structurelles du concept, et les probabilités de ces rôles seront évaluées en fonction des probabilités de leurs propriétés structurelles. Dans le cas où le rôle $r(c_1, \dots, c_n)$ d'un concept c_i (avec n compris entre 1 et 3, et $c_i \in \{c_1, \dots, c_n\}$) nécessite l'existence de deux propriétés structurelles associées aux concepts c_{p1} et c_{p2} :

Si $\exists R \in R_p$ tel que $(R(c_i, c_{p1}), p1)$ et $(R(c_i, c_{p2}), p2)$ alors $\text{Prob}(r(c_1, \dots, c_n)) = p1.p2$

5.2 Evaluation de la métrique

Le tableau 1 ci-dessous montre quelques résultats obtenus pour la mesure de similarité $\text{sim}_{\text{prop}}^p (w_1=0, w_2=1 \text{ et } w_3=0)$ dans le cas de différents hébergements. Une similarité maximale égale à 0.25 dénote une faible ressemblance. On note cependant que la similarité entre les concepts *camping* et *camping à la ferme* (2 concepts très proches ontologiquement) n'est pas maximale de plusieurs points de vue. Cette évaluation nous a permis, dans un premier temps, de valider notre approche sur les propriétés. Par la suite, nous envisageons d'élargir cette étude en prenant en compte l'aspect hiérarchique et les rôles décrits dans l'ontologie.

Hébergement	Propriété	Hébergements les plus similaires
Meublé	Electroménager	Village Vacances (0.75) , Appart Hôtel (0.75)
	Infrastructure	Gîte(0.68)
	Sanitaire	Appart hôtel (0.63), Bungalow (0.63), Chalet (0.63)
	Animation	Bungalow (0.25), Chalet (0.25), Gîte (0.25)
	Sport	Bungalow (0.25), Chalet (0.25), Gîte (0.25)
Camping	Commercial	Motel (0.75)
	Domestique	Appart hôtel (0.32)
	Location	Camping à la ferme (0.67)
	Restauration	Village Vacances (0.54)

Tableau 1. Evaluation de la métrique de similarité

Les propriétés pour lesquelles le degré de similarité est supérieur à 0.5 seront présentées dans les réponses (cf. Réponse 2 de l'exemple de la section 2).

6 Conclusion

Nous avons présenté une méthode pour l'intégration d'une ontologie au sein de WEBCOOP, un système question-réponse sur le domaine du tourisme, qui se base sur une ontologie où chaque concept est décrit par ses propriétés intrinsèques. Nous avons proposé d'étendre ce système afin de prendre en compte une modélisation plus fine des concepts et d'étudier l'impact de ce type de modélisation sur les procédures de raisonnement nécessaires à la phase d'extraction de réponses. Nous nous sommes focalisés sur les hébergements touristiques. Nos principales contributions sont les suivantes :

- La représentation des connaissances que nous adoptons s'inspire largement d'une étude de corpus de paires question-réponse où nous avons analysé la forme et le contenu des questions posées par les utilisateurs. Cette étude, couplée à l'analyse d'ontologies existantes sur le tourisme, a permis de décrire finement les concepts en termes de propriétés structurelles et fonctionnelles, essentielles et contingentes en nuancant leur existence via des probabilités.
- L'utilisation de cette ontologie dans le cadre de la relaxation de contraintes grâce à une mesure de similarité qui prend en compte ces différentes propriétés.
- La présentation des réponses en expliquant à l'usager en LN les principales étapes du processus de raisonnement qui reflètent la structuration de l'ontologie de domaine.

Dans le futur, nous envisageons : (1) d'évaluer l'adéquation linguistique et cognitive des explications produites face à des utilisateurs potentiels, (2) d'enrichir l'ontologie par des axiomes afin d'inférer de nouvelles propriétés fonctionnelles (un rôle peut être induit de propriétés et/ou de rôles). Et enfin (3), nous envisageons d'étendre cette approche à divers types d'application, en définissant finement à la fois les concepts et leurs propriétés (voire les propriétés de propriétés) afin de fournir aux utilisateurs, toujours dans le cadre des SQR, des niveaux d'explications de granularité différente.

Références

- ATZENI P & AL. (2004). Ontology-based question answering in a federation of university sites: the MOSES case study. *NLDB 2004*.
- BENAMARA F. (2004). Un système de Question Réponse coopératif sur le Web. *Thèse de Doctorat de l'Université Paul Sabatier*.
- BIDAULT A. (2002). Affinement de requêtes posées à un médiateur. *Thèse de Doctorat, Université d'Orsay*.
- BREUKER J. (2006). Ontology, epistemology and models, with examples from legal knowledge engineering. *SKIS – Course - May 2006*.
- CULLOT N. & JOUANOT F. & YETONGNON K. (2003). Une Méthode de Réconciliation Sémantique pour l'Extraction de Connaissances. *RSTI – EGC 2003*. p481-493

- EHRIG M. & SURE Y. (2004). Ontology Mapping – an Integrated Approach. *Proc. 1st ESWS, Hersounisous (GR)*. volume 3053 of lecture notes in Computer Science, p76-91
- HAN W & HUANG Y & CERCONI N & FU Y J. (1996). Intelligent Query Answering by Knowledge Discovery Techniques. *IEEE TRANS. ON KNOWLEDGE AND DATA ENGINEERING*, VOLUME 8.
- HIRST G. & ONGE D. (1998). Lexical chains as representations of context for the detection and correction of malapropisms. *WordNet : An Electronic Lexical Database*. Cambridge MA – The MIT Press
- KASSEL G. (2002). OntoSpec : une méthode de spécification semi-formelle d'ontologies. *Actes des 13^è journées francophones d'Ingénierie des Connaissances. IC'2002*. pages 75-87. France.
- JIAN J. & CONRATH D. (1997). Semantic Similarity based on Corpus Statistics and Lexical taxonomy. *Proceedings on International Conference on Research in Computational Linguistics*. Taiwan
- LIN D. (1997). Using Syntactic Dependency as Local Context to resolve word sense ambiguity. *Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics and 8th Conference of the European Chapter of the Association for Computational Linguistics*
- MAEDCHE A. & STAAB M. (2001). Ontology Learning for the Semantic Web. *University of Karlsruhe*. IEEE Intelligent Systems
- NAVATHE S. & BUNEMAN P. (1986). : Integrating User Views in Database Design. *Computer*, 19 (1) - p50-62
- NOY F-N. (2004). Tools for Mapping and Merging Ontologies. *Handbook on Ontologies 2004* – pp 365
- NOY F-N. & RECTOR A. (2004). Defining N-ary relations on the Semantic Web : Use with individuals. *W3C Working Draft*. World Wide Web Consortium : Boston, MA p. W3C Working Draft.
- PUSTEJOVSKY J. (1995). The generative lexicon. *MIT Press – Cambridge - Massachusetts*
- RADA R. & MILI H. & BICKNELL E. & BLETTNER M. (1989). Development and Application of a metric on Semantic Nets. *Man and cybernetics*, vol 19 pp 17-30, 1989
- RESNIK P. (1995). Using Information Content to Evaluate Semantic Similarity in a Taxonomy. *Proceedings of the 14th International Joint Conference on Artificial Intelligence*. Montreal
- RODRIGUEZ A. & EGENHOFER M. (2003). Determining Semantic Similarity Among Entities Classes From Different Ontologies. *IEEE Transactions on Knowledge and data Engineering* - 15 (2) : 442-456 . 2003
- WU Z. & PALMER M. (2003). Verb Semantics and Lexical Selection. *Proceedings of the 32nd Annual meetings of the Associations for computational Linguistics - Pages 133-138*
- ZARGAYOUNA H. & SALOTTI S. (2004). Mesure de similarité sémantique pour l'indexation de documents semi-structurés. *12^{ème} Atelier de Raisonnement à partir de Cas*.
- VARGAS-VERA, M. & MOTTA E. (2004). AQUA- Ontology-Based Question Answering System. In Monroy, Dr R. and Arroyo-Figueroa, G. and Sucar, L.E. and Sossa-Azuola, J.H., Eds. *Proceedings Third International Mexican Conference on Artificial Intelligence (MICAI-2004)*, Mexico City, Mexico.