

# Analyse sémantique spatio-temporelle pour les ontologies OWL-DL

Alina-Dia Miron, Jérôme Gensel, Marlène Villanova-Oliver  
Laboratoire d'Informatique de Grenoble,  
681 rue de la Passerelle BP 72, 38402 Saint Martin d'Hères Cedex  
prenom.nom@imag.fr

**Résumé.** L'*analyse sémantique* est un nouveau paradigme d'interrogation du Web Sémantique qui a pour objectif d'identifier les *associations sémantiques* reliant des individus décrits dans des ontologies OWL-DL. Pour déduire davantage d'*associations sémantiques* et augmenter la précision de l'analyse, l'information spatio-temporelle attachée aux ressources doit être prise en compte. A ces fins - et pour combler l'absence actuelle de raisonneurs spatio-temporel défini pour les ontologies RDF(S) et OWL-, nous proposons le système de représentation et d'interrogation d'ontologies spatio-temporelles ONTOAST, compatible avec le langage OWL-DL. Nous présentons les principes de base de l'algorithme de découverte d'*associations sémantiques* entre individus intégré dans ONTOAST. Cet algorithme utilise deux contextes, l'un spatial et l'autre temporel qui permettent d'affiner la recherche. Nous décrivons enfin l'approche mise en œuvre pour la déduction de *connexions spatiales* entre individus.

## 1 Introduction

L'une des conséquences de la croissance soutenue d'Internet est que les moteurs de recherche sont devenus des acteurs centraux dans l'infrastructure du Web (de Kunder, 2008). Les moteurs de recherche actuels retrouvent les documents du Web en fonction des correspondances syntaxiques qui existent entre leurs contenus textuels et des mots-clés donnés. Mais la précision des algorithmes de recherche est contestable étant donné l'important volume de données numériques disponibles sur le Web.

Le Web Sémantique (Berners-Lee *et al.*, 2001) vise à offrir des solutions pour accroître la performance et le rappel des moteurs de recherche, en annotant le contenu des ressources Web à l'aide de concepts ontologiques compréhensibles et exploitables par les machines. En associant une couche descriptive aux pages Web classiques, le Web Sémantique rend possible l'évolution des *données* vers des *connaissances* et marque le début d'une nouvelle étape dans l'exploitation d'Internet. Cette nouvelle étape nécessite le développement de nouveaux paradigmes d'interrogation, notamment l'*analyse sémantique* (Sheth *et al.*, 2002). Cette dernière a pour but l'identification automatique d'*associations sémantiques* (appelées aussi  *$\rho$ -paths*) reliant deux *individus*  $x$  et  $y$ , au sein d'un graphe RDF(S), en vue de répondre à des questions telles que : «*l'entité  $x$  est-elle reliée (même non directement) à l'entité  $y$ ?*». Les différents types d'*objets* sont reliés de façons complexes et souvent inattendues,

## Analyse sémantique spatio-temporelle pour OWL-DL

d'où l'intérêt de *l'analyse sémantique*, qui offre de nouvelles perspectives sur la découverte de connexions entre *individus* sans lien *a priori* apparent. Ce nouveau paradigme permet ainsi d'associer un contexte à chaque requête, pour mieux préciser l'intérêt de l'utilisateur, mais aussi pour éliminer le bruit (les résultats non pertinents). *L'analyse sémantique* a déjà été utilisée avec succès dans des domaines tels que la sécurité d'un territoire, la recherche de brevets biomédicaux, ou la détection de conflits d'intérêt...

Jusqu'ici, *l'analyse sémantique* s'est essentiellement concentrée sur la dimension thématique des métadonnées, en analysant, par exemple, les liens de collaboration entre deux personnes, membres d'une organisation. Néanmoins, étant donné la popularité croissante des applications spatiales sur le Web (Google Earth, Mappy, ViaMichelin, Geoportail, Virtual Earth 3D, etc.), nous pensons que la prise en compte des dimensions spatiale et temporelle des données lors de *l'analyse sémantique* est d'un enjeu crucial, adhérent en cela à la vision du futur Web Sémantique Géospatial proposée par Egenhofer (Egenhofer, 2002). Cette gestion contribuera à la fois à déduire davantage *d'associations sémantiques* et à rejeter celles présentant des incompatibilités avec le contexte spatio-temporel de la requête. Par exemple, partant d'une description ontologique de deux personnes qui ont vécu dans la même résidence pendant dix ans, une *analyse sémantique spatiale et temporelle* pourra déduire qu'il est très probable que les deux personnes se connaissent. La proximité et la simultanéité de résidence est ici automatiquement prise en compte par *l'analyse sémantique spatio-temporelle* pour suggérer qu'il pourrait y avoir un lien entre ces deux personnes.

Les recommandations à l'initiative du Web Sémantique, suffisamment génériques, permettent de répondre avec succès à la plupart des exigences d'expressivité imposées par les divers domaines d'application abordés. Néanmoins, en termes de traitement de l'information spatio-temporelle, on observe une absence de modélisation standardisée (Agarwal, 2005) et de mécanismes de raisonnement dédiés (Egenhofer, 2002 ; O'Dea *et al.*, 2005). En effet, le raisonnement géospatial impliquant des calculs mathématiques complexes, les environnements de modélisation ontologique et les raisonneurs associés aux langages à ontologies recommandés par le W3C (RDF (S), OWL...) se révèlent inadaptés pour le traitement de ce type de données. Dès lors, le traitement de l'information spatiale et temporelle ne peut se faire que par le biais d'un couplage faible entre ces langages à ontologies et des outils externes de traitement spatio-temporel. Une fois obtenus, les résultats de ces calculs doivent être intégrés dans l'espace ontologique, afin d'être mis à la disposition de différents raisonneurs pour répondre à des requêtes.

Dans (Miron *et al.*, 2007a), nous avons proposé un système appelé ONTOAST (pour ONTOlogies en Arom-ST), afin de combler l'absence actuelle de moteurs d'inférence spatiale et temporelle spécialisés, construits pour les ontologies OWL-DL. ONTOAST est à la fois un environnement de modélisation d'ontologies spatio-temporelles et un système d'inférence, capable de raisonner sur des connaissances spatiales, temporelles et thématiques. Compatible du point de vue de la représentation avec OWL-DL (Miron *et al.*, 2007b), il peut être utilisé dans le contexte du Web sémantique.

Dans le présent article, nous illustrons l'utilisation d'ONTOAST dans le processus de découverte *d'associations sémantiques* entre individus. Nous décrivons

l'architecture du *système d'analyse sémantique spatiale et temporelle* que nous proposons et qui utilise les capacités de raisonnement spatial et temporel d'ONTOAST, d'une part pour limiter l'espace de recherche, et, d'autre part pour construire de nouvelles *associations sémantiques*. Dans le premier cas, ONTOAST filtre les connaissances ontologiques pour ne retenir que celles qui sont compatibles avec le contexte spatial et temporel spécifié par l'utilisateur. Dans le deuxième cas, en utilisant les informations spatiales et temporelles décrivant des *individus* et des *propriétés dont la valeur est un individu*<sup>1</sup>, ONTOAST peut déduire des relations spatiales implicites. Ces relations seront ensuite utilisées pour la construction de nouvelles *associations sémantiques*.

L'article est organisé comme suit. Dans la section 2, la définition des *associations sémantiques* pour les structures de OWL-DL est présentée. La section 3 illustre l'architecture du *Système d'Analyse Sémantique Spatiale et Temporelle* que nous proposons. L'approche d'interrogation est détaillée dans la section 4 ainsi que les algorithmes sous-jacents. La section 5 conclut en donnant le bilan et les perspectives de ce travail.

## 2 Associations Sémantiques en OWL-DL

Le concept d'*association sémantique* (ou de  *$\rho$ -path*) a été introduit par (Sheth *et al.*, 2002) pour décrire les *relations indirectes* (impliquant un ensemble d'*individus* et de *relations* intermédiaires) entre deux *objets*  $x$  et  $y$  d'un graphe RDF,  $G$ . Il désigne une séquence  $x(P_1 \cdot P_2 \cdot \dots \cdot P_n)y$  tel qu'il existe un ensemble d'objets  $o_1, o_2, \dots, o_{n-1}$  appartenant à  $G$  et qui respectent :  $xP_1o_1 \wedge o_1P_2o_2 \wedge \dots \wedge o_{n-1}P_ny$ , où pour tout  $i \in [1, n]$ ,  $P_i$  est une *propriété dont la valeur est un individu* de  $G$ , dont le graphe contient la suite d'*objets*  $(o_{i-1}, o_i)$  avec  $o_0=x$  et  $o_n=y$ . Le choix d'utiliser RDF(S) comme langage de modélisation d'ontologies peut être considéré comme trop restrictif compte tenu de son expressivité limitée par rapport aux logiques de description ou aux langages de représentation de connaissances par objets. Ce constat a motivé nos travaux d'adaptation des techniques *d'analyse sémantique* aux ontologies OWL-DL (Miron *et al.*, 2008). Notre idée est, d'une part, de travailler sur des ontologies plus complexes qui exploitent le pouvoir d'expression d'OWL-DL, et, d'autre part, de réaliser des inférences étendues en explorant, par exemple, les axiomes d'OWL-DL ou encore l'alignement d'ontologies.

Pour toute ontologie OWL-DL  $\Omega$ , il est possible de construire un graphe orienté  $G=(V_G, E_G)$ , dont l'ensemble de sommets,  $V_G$ , contient les *individus* et les valeurs littérales définis dans la partie déclarative de  $\Omega$ . L'ensemble d'arêtes orientées  $E_G$ , correspond aux *instances des propriétés dont la valeur est un individu* (ou *tuples*), introduits dans la partie déclarative de  $\Omega$  (*ABox*) ou déduits à partir des axiomes de propriétés définies dans la partie terminologique de  $\Omega$  (*TBox*). Par exemple, le graphe illustré par la figure FIG.1, est construit à partir de l'ontologie OWL-DL présentée dans la figure FIG.2, en utilisant les assertions de la ABox (voir figure FIG.2), mais aussi l'axiome *InverseObjectProperties(worksFor hasEmployee)* pour inférer l'existence de l'arête *worksFor* reliant l'individu *pers3* à l'individu *UIP*. Dans ce

---

<sup>1</sup> Relations binaires entre individus, en anglais *object properties*.

## Analyse sémantique spatio-temporelle pour OWL-DL

contexte, deux individus  $x$  et  $y$  ( $x, y \in V_G$ ) sont sémantiquement associés si le graphe  $G$  contient au moins un chemin qui commence avec le sommet  $x$  et passe par un ensemble d'objets intermédiaires ( $o_i \in V_G, i \in [1, n-1]$ ) reliés par des tuples ( $e_i \in E_G, i \in [1, n]$ ) pour arriver au sommet  $y$ . Afin de décrire les  $\rho$ -paths de façon non ambiguë, dans le présent document, nous utilisons la notation suivante:

$$\rho\text{-path}(x, y) = x \xrightarrow{e_1} o_1 \rightarrow \dots \rightarrow o_{n-1} \xrightarrow{e_n} y, \quad o_i \in V_G, 0 \leq i \leq n, \quad x = o_0, y = o_n, \quad (1)$$

$$e_j \in E_G, 1 \leq j \leq n.$$

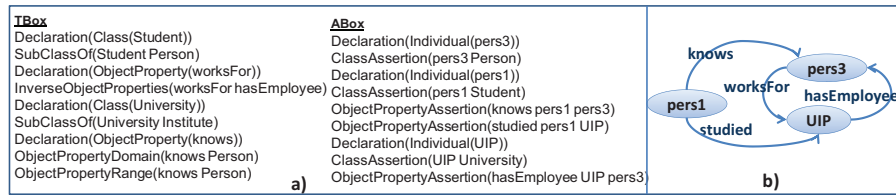


FIG. 1a) Exemple d'ontologie simple OWL-DL<sup>2</sup>. b) Graphe ontologique construit à partir de l'ontologie illustrée par a).

Pour illustrer nos propos, considérons l'ontologie simple présentée dans la figure FIG 1a) et le graphe ontologique correspondant (voir figure FIG.1b)). Le graphe ontologique (et donc l'ontologie) contient trois  $\rho$ -paths reliant les individus  $pers1$  et  $pers3$  :  $p$  est un  $p$ -path direct car il relie les deux individus en utilisant un seul tuple, et  $q$  et  $r$  sont des  $p$ -paths indirects car ils contiennent respectivement un et deux objets intermédiaires. Leur structure est détaillée ci-dessous.

$$p = pers_1 \xrightarrow{\text{studentOf}} pers_3,$$

$$q = pers_1 \xrightarrow{\text{studiesAt}} UIP \xrightarrow{\text{hasEmployee}} pers_3 \text{ et}$$

$$r = pers_1 \xrightarrow{\text{studentOf}} pers_2 \xrightarrow{\text{teachesAt}} UIP \xrightarrow{\text{hasEmployee}} pers_3.$$

Notre travail est axé sur les techniques d'inférence capables de déduire des relations spatiales et temporelles reliant les objets d'un graphe d'ontologie  $G$ , et qui peuvent conduire à la découverte de nouveaux  $p$ -paths. A ces fins, nous proposons l'utilisation d'un système de découverte d'associations sémantiques dont l'architecture est présentée dans la section suivante.

### 3 Le système d'analyse sémantique spatio-temporelle

Notre proposition s'appuie sur l'utilisation d'ONTOAST, un environnement de modélisation d'ontologies spatio-temporelles qui est également un système d'inférence et d'interrogation sémantique. ONTOAST est une extension du système AROM (Page *et al.*, 2001), compatible du point de vue de la représentation avec

<sup>2</sup> La syntaxe utilisée pour cet exemple est la syntaxe du style fonctionnel de OWL 2.

OWL-DL (Miron *et al.*, 2007a). Il intègre un modèle spatial prédéfini qui gère un ensemble de relations spatiales qualitatives. Celles-ci sont qualifiées de relations *topologiques*, de *distance*, et de *direction* et leurs sémantiques respectives sont définies dans (Miron *et al.*, 2007b). En s'appuyant sur l'utilisation du Langage de Modélisation Algébrique (LMA) d'AROM (Page *et al.*, 2001) et sur un ensemble de règles de déduction de relations spatiales qualitatives à partir d'autres relations qualitatives existantes et de données numériques (Miron *et al.*, 2007b), ONTOAST peut compléter les connaissances spatiales sur les objets modélisés et permet une formulation de requêtes plus proche par sa flexibilité du langage naturel. Pour réaliser des raisonnements analogues sur le temps, ONTOAST gère également l'ensemble des relations temporelles d'Allen (Allen, 1983) (*before*, *after*, *starts/started-by*, *finishes/finished-by*, *during/contains*, *equals*, *meets/met-by*, *overlaps/overlapped-by*). Le but est de répondre à des requêtes spatiales et temporelles à la volée, en combinant raisonnablement qualitatif et raisonnement quantitatif, sans avoir pré-calculé et stocké dans la base ontologique toutes les relations spatiales possibles entre les entités.

Nous décrivons ici l'utilisation d'ONTOAST dans le processus de découverte des associations sémantiques entre individus. Le *système d'analyse sémantique spatio-temporelle* que nous proposons, illustré dans la figure FIG. 2, comporte cinq modules principaux, organisés comme des outils d'exploitation distincts construits au-dessus du système ONTOAST. Afin d'exploiter les raisonnements fournis par ONTOAST, les connaissances ontologiques doivent être traduites vers le formalisme d'AROM et stockées dans un entrepôt de connaissances orientées-objets (étape 1, FIG.2). Pour interroger l'entrepôt, l'utilisateur spécifie sa requête (étape 2, FIG.2) dans la forme d'une paire d'individus ( $x, y$ ) en utilisant l'*Interface d'interrogation*, qui permet aussi la spécification des *Contextes Spatial* et *Temporel* attachés à la requête. Afin de réduire l'espace de recherche, le processus de découverte des  $\rho$ -paths (étape 3, FIG.2) débute par le filtrage des connaissances disponibles, en suivant les spécifications des *Contextes Spatial* et *Temporel*. Les connaissances filtrées sont exploitées par le module de *Découverte des P-paths*. Le *Raisonneur ONTOAST* est utilisé à la fois dans la phase de filtrage (pour déduire des caractéristiques géographiques des *individus* ainsi que la validité temporelle des *individus* et des *tuples*), et dans la phase de découverte de  $\rho$ -paths (pour déduire les connexions spatiales entre *individus*). Les  $\rho$ -paths découverts sont ensuite classés par le module de *Classification de résultats* (étape 4, FIG.2), en utilisant les spécifications de contexte et présentés à l'utilisateur par le module de *Visualisation des résultats* (étape 5, FIG.2).

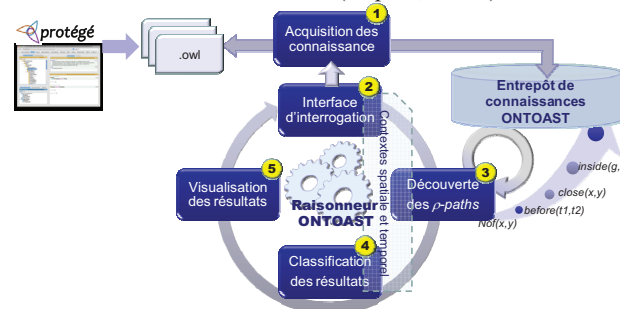


FIG. 2 - L'architecture du système d'analyse sémantique spatio-temporelle.

## Analyse sémantique spatio-temporelle pour OWL-DL

Lors de la déduction de nouvelles relations spatiales et temporelles (étape 3, FIG.2), celles-ci sont stockées dans l'entrepôt de connaissances et ensuite exportées au format *owl* afin d'être utilisées par les outils connexes d'annotation et/ou de raisonnement. Il est patent que le traitement de l'information spatiale et temporelle augmente le champ d'application de l'*analyse sémantique*, mais, en même temps, soulève de nouveaux défis de représentation et de raisonnement. Par exemple, le système de typage adopté par RDF(S) et OWL est assez limité et ne propose pas de support particulier pour des extensions spatiales. En conséquence, afin de modéliser des données spatiales dans OWL, des concepts dédiés, capables de supporter les types de données spatiales, doivent être définis et utilisés. Par exemple, un *polygone* (instance d'une classe *Polygone*) sera représenté par une liste d'objets *Points* ( $x, y$ ) connectés à un *Système de Coordonnées*. Plusieurs ontologies géospatiales modélisant des caractéristiques géométriques ont été proposées jusqu'à présent, mais aucune d'entre elles n'a été reconnue comme standard. Suivant les recommandations d'une étude récente (Ressler et Dean, 2007) qui compare 45 ontologies géospatiales et temporelles, nous avons choisi la version OWL de l'ontologie *GeoRSS-Simple* comme référence pour la modélisation des concepts spatiaux. *GeoRSS-Simple* est un format très léger pour la description des données spatiales que les développeurs et les utilisateurs peuvent rapidement et facilement intégrer dans leurs applications, ou bien utiliser pour créer des annotations sémantiques avec peu d'efforts. Il définit les géométries de base (*point, ligne, polygone...*) et couvre les besoins typiques pour l'encodage de lieux. Le module *d'acquisition de connaissances* est construit pour reconnaître les types de données spatiales modélisés en utilisant des concepts ontologiques décrits dans cette ontologie.

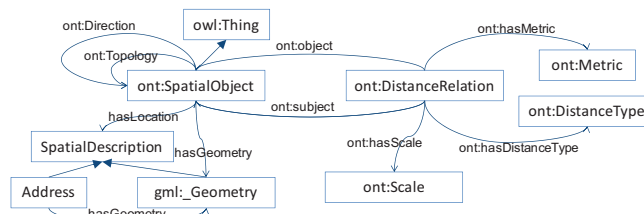


FIG. 3. La méta-ontologie utilisée pour modéliser des objets spatiaux. Elle intègre les ontologies *QualitativeSpatialRelations* (préfixée *ont*) et *GeoRSS* (préfixée *gml*).

Afin de représenter en OWL les relations spatiales qualitatives qui peuvent être traitées par le raisonneur d'ONTOAST, nous proposons l'utilisation de l'ontologie *QualitativeSpatialRelations*. Elle introduit deux types de relations spatiales, *Direction* et *Topology*, définies comme étant des *propriétés à valeur d'individus*. La relation de *Direction* dispose de neuf spécialisations qui permettent l'expression de directions cardinales (*Nof, Sof, Eof, Wof, Cof, NEof, NWof, SEof, SWof*)<sup>3</sup> existant entre les objets géographiques. Des spécialisations de la relation *Topology* ont été également définies: *Disjoint, Contains, Crosses, Touches, Equals, Intersects, Overlaps* and *Within*. Les

<sup>3</sup> *Nof* signifie "au nord de", *Cof* "au centre de", etc.

relations de *Distance*, quant à elles, sont plus difficiles à représenter car elles possèdent en général des attributs pour préciser, par exemple, le système métrique employé pour le calcul de la distance (distance euclidienne, de plus court chemin, en voiture, etc.), l'échelle ou le type de distance lorsque des régions de l'espace sont décrites (la distance moyenne des points frontaliers, la distance entre les centres de gravité, entre les centres administratifs...). Comme OWL-DL ne permet pas la définition d'attributs pour les propriétés à valeur d'individus, les relations de distance entre les objets géographiques sont réifiées comme des objets de la classe *DistanceRelation* (voir figure FIG. 3). Quatre spécialisations de *DistanceRelation* (*VeryFar*, *Far*, *Close*, and *VeryClose*) sont définies afin de permettre la spécification de distances absolues entre les *objets*. La figure FIG. 3 illustre la méta-ontologie OWL-DL que nous avons utilisée pour annoter les objets spatiaux. Nous avons considéré que les informations spatiales peuvent être attachées à des objets spatiaux (*SpatialObjects*) de trois façons: 1) à l'aide de la propriété *hasGeometry* qui désigne une géométrie pour l'objet géographique spécifié, 2) à l'aide de la propriété *hasLocation* qui peut faire référence à une adresse ou à une *Géométrie* concrète, 3) par le biais d'une relation qualitative spatiale.

Afin d'exprimer des données temporelles dans OWL, on peut exploiter ses types dédiés: *xsd:dateTime*, *xsd:date*, *xsd:gYearMonth*, *xsd:gMonthDay*, *xsd:gDay*, *xsd:gMonth*. Ces types de données peuvent être utilisés pour exprimer les caractéristiques temporelles des individus, mais pas pour exprimer la validité des *tuples* dans les frontières décidables d'OWL-DL. En OWL-Full, les propriétés à valeur d'individus peuvent être utilisées comme des classes, et donc leurs instances peuvent être décrites par des propriétés à valeur de données. Etant donné qu'ONTOAST n'est pas compatible avec OWL-Full, nous ne pouvons pas adopter cette solution. Une approche intéressante, centrée sur l'estampillage des triplets RDF par des instants ou des intervalles de temps, est présentée dans (Gutierrez *et al.*, 2005). Comme la recommandation actuelle pour le langage OWL ne prend pas en charge les estampilles temporelles, nous proposons de les simuler par des annotations spécifiques gérées par le module *d'acquisition de connaissances* d'ONTOAST. Trois exemples d'annotations temporelles sont illustrés dans la figure FIG 4. Les deux premières associent des intervalles temporels à deux *tuples* et la dernière estampille le *tuple* correspondant avec un instant de temps.

```
ObjectPropertyAssertion(Comment("temporalValidity(2005-10-24,T10:30:00 -)") studentOf pers1 pers4)
ObjectPropertyAssertion(Comment("temporalValidity(1993-01-10,2004-30-06)") hasEmployee UIP pers4)
ObjectPropertyAssertion(Comment("temporalValidity(2008-08-10)") colaboratesWith pers2 pers3)
```

FIG. 4 - Exemples d'annotations temporelles pour des *tuples* définis en OWL DL.

## 4 Approche d'interrogation

Les associations sémantiques existant entre deux individus peuvent être découvertes en utilisant des algorithmes *de recherche en profondeur d'abord* pour la découverte des chemins dans des graphes. Dans l'environnement ouvert du Web Sémantique, ces algorithmes peuvent renvoyer un grand nombre de résultats, pas forcément pertinents. Il est donc nécessaire de procéder à un filtrage en vue

## Analyse sémantique spatio-temporelle pour OWL-DL

d'améliorer la précision en tenant compte de certains critères, notamment le contexte spatial et le contexte temporel de la requête, comme présenté dans les sections suivantes. Pour illustrer notre approche, nous montrons dans la figure FIG 5 un exemple concret d'ontologie décrivant des relations de collaboration entre chercheurs. Pour ces connaissances explicites, nous examinons une requête qui recherche les liens potentiels existant entre les individus *pers2* et *pers4*. Cette requête a comme résultat immédiat les  $\rho$ -paths illustrés dans le TAB.1. Ces résultats peuvent être affinés ou bien étendus en utilisant les informations contextuelles attachées à requête, comme le montrent les sections suivantes.

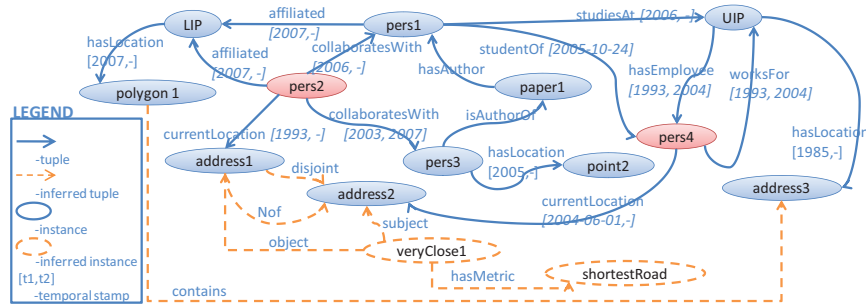


FIG. 5 - Exemple de graphe construit à partir d'une ontologie OWL-DL qui modélise un réseau social dans le domaine de la recherche scientifique.

	<i>collaboratesWith</i>	<i>studentOf</i>								
$p_1 = pers_2$	$\rightarrow$	$pers_1$	$\rightarrow$	$pers_4$						
	<i>collaboratesWith</i>	<i>studiesAt</i>	<i>hasEmployer</i>							
$p_2 = pers_2$	$\rightarrow$	$pers_1$	$\rightarrow$	$UIP$	$\rightarrow$	$pers_4$				
	<i>collaboratesWith</i>	<i>isAuthorOf</i>	<i>hasAuthor</i>	<i>studentOf</i>						
$p_3 = pers_2$	$\rightarrow$	$pers_3$	$\rightarrow$	$paper_1$	$\rightarrow$	$pers_1$	$\rightarrow$	$pers_4$		
	<i>collaboratesWith</i>	<i>isAuthorOf</i>	<i>hasAuthor</i>	<i>studiesAt</i>	<i>hasEmployer</i>					
$p_4 = pers_2$	$\rightarrow$	$pers_3$	$\rightarrow$	$paper_1$	$\rightarrow$	$pers_1$	$\rightarrow$	$UIP$	$\rightarrow$	$pers_4$

TAB 1 – Les  $\rho$ -paths reliant *pers2* et *pers4* dans l'ontologie de la figure FIG.5.

### 4.1 Le contexte temporel

Le contexte temporel  $C_{temp}$ , agit comme un filtre pour limiter la recherche aux individus qui satisfont une certaine relation de proximité temporelle par rapport à un intervalle de temps donné. Il est constitué d'un *intervalle* [*date\_début*, *date\_fin*] ou (*date*, *durée*), et d'une *relation temporelle topologique* qui doit être vérifiée par tous les *objets* et les *tuples* inclus dans le  $\rho$ -path. Lorsque l'intervalle temporel est décrit par une date de *début* et une *durée*, les attributs temporels décrivant les *tuples* ou les *objets* considérés seront comparés à l'intervalle [*date*, *date+durée*]. L'intervalle temporel implicite commence avec l'origine des temps et s'étend jusqu'à l'instant présent, mais il peut être adapté aux besoins de l'utilisateur. Toutes les relations temporelles qualitatives d'ONTOAST (voir section 3) peuvent être utilisées; la relation implicite est celle d'inclusion temporelle (*during*). Lors de l'analyse d'une *instance e* (*tuple* ou *objet*), candidate à l'intégration dans un  $\rho$ -path, le système



compare sa validité, exprimée par ses attributs temporels (pour les *objets*) ou par les estampilles temporelles (pour les *tuples*), au contexte temporel fourni,  $C_{temp}$ . Si  $e$  n'a pas d'attributs temporels, le système considère qu'elle existe depuis toujours et qu'elle est éternelle. Dans l'exemple illustré par la figure FIG 5, si  $C_{temp} = ([2005, 2008], overlaps)$ , les  $\rho$ -paths  $p_2$  et  $p_4$  sont éliminés de l'ensemble des résultats. En effet, la validité du *tuple* (*hasEmployee UIP pers4*), spécifiée par son estampille [1993, 2004], ne satisfait pas la relation *overlaps* avec l'intervalle  $C_{temp}$ . La construction des deux  $\rho$ -paths ne peut donc pas aboutir.

## 4.2 Le contexte spatial

Les inférences spatiales offertes par ONTOAST peuvent jouer le rôle de filtre afin d'éliminer (infirmer) les  $\rho$ -paths dont les éléments à dimension spatiale ne respectent pas une relation spatiale qualitative donnée vis-à-vis d'une région de l'espace. Pour cela, on définit  $C_{spatial}$ , le contexte spatial d'une requête comme une paire (*objet spatial, relationSpatialeQualitative*). De façon analogue au contexte temporel, lors de l'analyse d'un objet  $o$  (*objet intermédiaire*) candidat à l'intégration dans un  $\rho$ -path  $\alpha$ , on confronte sa *spatialité*, exprimée par sa géométrie ou inférée par le système, au contexte spatial  $C_{spatial}$ . Si  $o$  n'est pas en concordance avec  $C_{spatial}$ , il ne sera pas pris en compte pour le  $\rho$ -path  $\alpha$ . Les instances qui n'ont pas de dimension spatiale sont considérées comme étant cohérentes avec  $C_{spatial}$ . Dans l'exemple de la figure FIG. 5, on considère que  $address_1$  et  $address_2$  décrivent deux adresses à Paris, à une rue d'écart l'une de l'autre,  $address_3$  est située à Versailles,  $polygone_1$  décrit le contour géométrique de tous les sites administratif du Laboratoire d'Informatique de Paris et  $point_2$  représente des coordonnées géographiques à Orléans. Si l'utilisateur est uniquement intéressé par les connexions entre les chercheurs de la région Ile de France, il imposera comme filtre spatial  $C_{spatial} = (IleDeFrance, contient)$ .  $Pers_3$ , dont l'emplacement actuel (Orléans) ne satisfait pas la relation d'inclusion, et devient donc incompatible avec le contexte spécifié. En conséquence, la construction des  $\rho$ -paths passant par  $pers_3$  ( $p_3$  et  $p_4$  du tableau TAB 1) ne peut pas aboutir.

## 4.3 Inférences spatiales pour la construction de $\rho$ -paths

Les capacités de raisonnement spatial d'ONTOAST s'avèrent très utiles dans l'analyse sémantique, car elles permettent de construire des  $\rho$ -paths entre des objets apparemment sans lien thématique, mais qui ont une proximité spatiale et/ou temporelle intéressante<sup>4</sup>. L'inférence des  $\rho$ -paths spatiaux existants entre  $x$  et  $y$  suit les étapes suivantes :

1. Si  $x$  n'est pas un objet spatial ou bien si sa géométrie n'est pas connue, le système construit l'ensemble  $S_x$  des objets spatiaux reliés à  $x$  par des  $\rho$ -paths  $\alpha_1, \dots, \alpha_m$  d'une longueur maximale préétablie  $l_{Max} : |\alpha_i| \leq l_{Max} \quad \forall i \in 1..m$ . Tous les éléments de l'ensemble  $S_x$  doivent respecter le contexte spatial. Pour un objet spatial  $x$  dont on connaît la géométrie, l'ensemble  $S_x$  contient exclusivement l'objet lui-même.
2. On construit de la même façon l'ensemble  $S_y$ .

<sup>4</sup> Faute de place, dans cet article nous décrivons seulement les inférences spatiales. Les inférences temporelles suivent la même logique.

## Analyse sémantique spatio-temporelle pour OWL-DL

3. Entre les objets appartenant aux ensembles  $S_x \setminus S_y$  et  $S_y$ , le système énumère de façon exhaustive quelles relations spatiales qualitatives sont établies. Les relations inférées de cette façon (*relSpatiale*) sont ajoutées à l'ontologie et l'association sémantique obtenue :  $\alpha_i \rightarrow \text{spatial\_relation} \rightarrow \beta_j$  sera prise en compte comme résultat de l'analyse sémantique.

A l'aide de cet algorithme, de nouveaux  $\rho$ -paths peuvent être découverts. Par exemple, dans l'ontologie de la figure FIG. 5, après avoir construit les ensembles  $S_{pers_2} = (address_1, polygon_1)$  et  $S_{pers_4} = (address_2, address_3)$ , le raisonneur spatial d'ONTOAST infère les relations spatiales qualitatives établies entre les quatre paires:  $address_1$  et  $address_2$ ,  $address_1$  et  $address_3$ ,  $polygon_1$  et  $address_2$ ,  $polygon_1$  et  $address_3$ . Nous nous concentrons sur les inférences relatives à la première paire. Plusieurs services Web de géocodage existe à ce jour (Yahoo! Maps, Google Maps, MapPoint...), capables de transformer une adresse physique, par exemple 7, Place de la République, Paris, en une paire de coordonnées géographiques de type (*latitude*, *longitude*) correspondante (pour l'exemple considéré (48.8666170, 2.3644100)). De ce fait, obtenir une position géographique pour une adresse est relativement aisé. Les données spatiales quantitatives ainsi obtenues seront utilisées par le raisonneur spatial d'ONTOAST pour déduire, par le biais de calculs géométriques, les relations spatiales qualitatives existant entre les adresses. Sachant que  $address_1$  et  $address_2$  sont à une proximité immédiate et que  $address_2$  est au Nord par rapport à  $address_1$ , les tuples  $veryClose(address_1, address_2)$ ,  $Nof(address_1, address_2)$  et  $disjoints(address_1, address_2)$  seront ajoutés à l'ontologie. Ces ajouts de connaissances entraînent alors la découverte de trois nouveaux  $\rho$ -paths:

$$\begin{aligned}
 p_5 = pers_2 & \xrightarrow{\text{currentLocation}} address_1 \xrightarrow{\text{object}} veryClose_1 \xrightarrow{\text{subject}} address_2 \xrightarrow{\text{currentLocation}} pers_4; \\
 p_6 = pers_2 & \xrightarrow{\text{currentLocation}} address_1 \xrightarrow{Nof} address_2 \xrightarrow{\text{currentLocation}} pers_4; \\
 p_7 = pers_2 & \xrightarrow{\text{currentLocation}} address_1 \xrightarrow{disjoint} address_2 \xrightarrow{\text{currentLocation}} pers_4;
 \end{aligned}$$

## 5 Bilan et perspectives

Nous avons proposé ici une méthode d'analyse sémantique spatio-temporelle qui utilise le système ONTOAST (Miron *et al.*, 2007a), en tant que raisonneur spatio-temporel pour les connaissances du Web Sémantique. A l'aide d'ONTOAST, de nouvelles déductions peuvent être réalisées qui peuvent conduire à la découverte de nouvelles associations par lesquelles les individus sont connectés.

Une approche similaire, présentée par (Perry *et al.*, 2007), propose l'intégration d'ontologies RDF et de bases de données spatio-temporelles ORACLE, à travers l'utilisation d'ORACLE Semantic Data Store qui fournit les capacités de stocker, d'inférer et d'interroger des données sémantiques sous la forme de simples descriptions RDF ou d'ontologies basées sur RDFS. Or, cette approche, outre le fait qu'elle soit basée sur l'utilisation d'un outil propriétaire, ne permet pas de combiner les raisonnements qualitatif et quantitatif. De plus, pour interroger une base de données, il faut connaître à l'avance les relations qui existent entre les tables, et donc qui relie les individus. Traduit en analytique sémantique, cela signifie qu'il faut

prédisposer des  $\rho$ -paths qui seront ultérieurement filtrés en fonction des relations spatiales vérifiées par leurs éléments ou vis-à-vis d'une région de référence donnée.

En tant qu'objectif principal pour le futur, nous souhaitons mettre en œuvre un prototype du *système d'analyse sémantique spatio-temporelle* et tester nos algorithmes sur une base ontologique large, afin de quantifier leur pertinence et leur performance dans des scénarios réels de découverte *d'associations sémantiques*.

## Références

- Agarwal P. (2005). *Ontological Considerations in GIScience*. International Journal of Geographical Information Science. Vol 19, No.5.
- Allen, J. F. (1983). *Maintaining Knowledge about Temporal Intervals*; Communicaiton of the ACM. Vol 26, 832-843.
- Anyanwu K. et Sheth A. (2003).  *$\rho$ -Queries: Enabling Querying for Semantic Associations on the Semantic Web*. WWW2003, Budapest, Hungary.
- Berners-Lee T., Hendler J.A. et Lassila O. (2001). *The Semantic Web*. *Scientific American*. 284(5), 34-43.
- Egenhofer M.J. (2002). *Towards the Semantic Geospatial Web*. GIS'02.
- Gutierrez C., Hurtado C. et Vaisman A. (2005). *Temporal RDF*. Second European Semantic Web Conference. Heraklion, Crete, Greece.
- Halaschek C., Aleman-Meza B., Arpinar B.I. et Sheth A.P. (2004). *Discovering and Ranking Semantic Associations over a Large RDF Metabase*. Proceedings of the 30th VLBD Conference, Toronto.
- de Kunder, M. (2008). The size of the world wide web. [www.worldwidewebsite.com](http://www.worldwidewebsite.com)
- Miron A., Gensel J., Villanova-Oliver M., Martin H.: *Towards the Geo-spatial Querying of the Semantic Web with ONTOAST*. 7th International Symposium on Web and Wireless GIS (W2GIS 2007), Cardiff, UK (2007a).
- Miron A., Capponi C., Gensel J., Villanova-Oliver M., Ziébelin D., Genoud P.: *Rapprocher AROM de OWL...*, Langages et Modèles à Objets, in French, Toulouse, France (2007b).
- Miron A.D., Gensel J., Villanova-Oliver M.: *Thematic and Spatio-Temporal Semantic Analysis for OWL-DL*. Internal Report, LIG, Grenoble, France (2008).
- O'Dea D., Geoghegan S. et Ekins C. (2005). *Dealing with Geospatial Information in the Semantic Web*. Australian Ontology Workshop, Sydney, Australia.
- Page M., Gensel J., Capponi C., Bruley C., Genoud P., Ziébelin D., Bardou D., Dupierris V.: *A New Approach in Object-Based Knowledge Representation: the AROM System*. Lecture Notes in Artificial Intelligence 2070. 113-118. Budapest, Hungary, (2001)..
- Patel-Schneider P., Hayes P. et Horrocks I. (2004). *OWL Web Ontology Language Semantics and Abstract Syntax*. W3C Recommendation.

## Analyse sémantique spatio-temporelle pour OWL-DL

Perry M. et Sheth A. (2008). *A Framework to Support Spatial, Temporal and Thematic Analytics over Semantic Web Data*. Technical Report KNOESIS-TR-2008-01

Ressler J. et Dean M. (2007). *Geospatial Ontology Trade Study*. Ontology for the Intelligence Community (OIC-2007), Columbia, Maryland.

Sheth A., Arpinar B.I. et Kashyap V. (2002). *Relationships at the Heart of Semantic Web: Modeling, Discovering, and Exploiting Complex Semantic Relations*. Enhancing the Power of The Internet: Studies in Fuzziness and Soft Computing, Springer-Verlag.

### Summary

The *semantic analysis* is a new query paradigm for the Semantic Web knowledge, whose purpose is to identify complex associations between individuals described in OWL-DL ontologies. For increasing the accuracy of this *semantic analysis* and for improving the discovery results, spatial and temporal information attached to resources must be used. But for now, there is no spatio-temporal reasoner defined for RDF (S) and OWL ontologies. As a first response to this limitation, we propose the use of the ONTOAST system, a spatio-temporal ontology modeling and semantic query environment compatible with OWL-DL. We present a general algorithm for the discovery of *semantic associations* between individuals, which uses both a spatial and a temporal context in order to filter the search space. We also describe how spatial connexions that occur between individuals are inferred.