Graphes de Connaissances et Ontologie pour la Représentation de Données Immobilières Issues d'Annonces en Texte Libre

Lucie Cadorel, Andrea G. B. Tettamanzi, Fabien Gandon

Graphes de Connaissances et Ontologie pour la Representation de Données Immobilières Issues d'Annonces en Texte Libre

Lucie Cadorel[1,2,3], Andrea G. B. Tettamanzi[1,2], Fabien Gandon[1,2]

[1] Inria Sophia-Antipolis
[2] I3S, Université Nice Côte d'Azur, CNRS
[3] Septeo Proptech

lucie.cadorel@inria.fr, andrea.tettamanzi@univ-cotedazur.fr, fabien.gandon@inria.fr

Résumé

Nous décrivons nos premiers travaux sur la conception d'une ontologie et d'un graphe de connaissances pour la représentation des données immobilières issues d'annonces. Nous identifions plusieurs scénarios motivants à partir des données extraites afin de justifier notre modélisation. Puis nous proposons une modélisation pour représenter les données spatiales incertaines et leurs géométries dans un graphe de connaissances.

Mots-clés

Ontologie, Graphe de Connaissances, Texte, Immobilier, Données spatiales

Abstract

We describe our initial work on the design of an ontology and a knowledge graph for the representation of real estate data from advertisements. We identify several motivating scenarios from the extracted data to justify our modelling. Then we propose a model to represent uncertain spatial data and their geometries in a knowledge graph.

Keywords

Ontology, Knowlede Graph, Text, Real Estate domain, Spatial data

1 Introduction et Motivations

1.1 Contexte

Le secteur de l'immobilier joue un rôle important dans l'économie, au point que son poids dépasse celui de l'industrie et de l'agriculture réunies [2]. Les données immobilières sont de ce fait très étudiées par les professionnels du secteur, notamment pour estimer les prix et les tendances du marché. Ces données sont à la fois temporelles et spatiales, et proviennent de différentes sources (notaires, agents immobiliers, particulier, etc.). Plus particulièrement, les annonces immobilières constituent une source très abondante de données facilement accessibles, exhaustives et mises à jour. Les caractéristiques du bien et de son environnement sont en général décrites par l'annonceur, et peuvent être extraites à l'aide d'un modèle de langage [3, 4]. Cependant, ces données ne sont pas structurées, ce qui constitue un obstacle à leur exploitation, notamment pour effectuer du raisonnement. Si les ontologies et les graphes de connaissances peuvent aider à modéliser et représenter les informations issues des textes d'annonces, et faciliter leur interopérabilité, il en reste que certaines informations, notamment la description de l'environnement et de la localisation, sont parfois incertaines et floues (par exemple "proche du centre-ville"), et nécessitent une représentation appropriée.

1.2 Motivations et Questions de compétences

Afin de spécifier notre formalisation, l'ontologie attenante, et les applications futures, nous avons défini plusieurs cas d'usage. Pour cela, nous avons suivi la méthodologie "classique" des scénarios motivants et questions de compétences [1]. Nous avons d'abord identifié et dialogué avec des utilisateurs potentiels, notamment des professionnels de l'immobilier et des chercheurs en géographie. A partir de ce recueil, nous avons identifié des scénarios ainsi que les questions de compétences qui en découlent. Les tableaux 1, 2, 3 et 4 nomment (*Nom*), décrivent (*Desc.*), exemplifient (*ex 1, 2*) les scénarios et déduisent des questions de compétences (*QC*).

Nom	Recherche d'un bien immobilier
Desc.	Un acheteur recherche un bien immobilier dans une ville et un secteur particulier. Pour prendre sa décision, il a besoin de connaître les caractéristiques du bien, son prix et son environnement.
Ex. 1	Mathilde recherche un bien à acheter à Cannes, avec vue sur la mer et proche de la Croisette. Elle voudrait un 2 pièces pour moins de 300 000 euros.
Ex. 2	Paul déménage à Nice pour le travail. Il ne connaît pas la région et voudrait trouver un appartement proche des transports et des commerces, tout en étant au calme.
QC	- Quelles sont les caractéristiques du bien ? - Dans quelle ville se trouve le bien ? - Dans quel quartier se trouve le bien ? - À proximité de quels lieux se trouve le bien ? - Dans quel environnement (ex. sonore) est le bien ?

TABLE 1 – Scénario 1 : Recherche d'un bien immobilier

Nom	Étude du marché immobilier
Desc.	Un agent immobilier a besoin de comprendre le marché et les biens en vente/vendus pour positionner le bien qu'il a à vendre sur le marché.
Ex. 1	Denis a obtenu le mandat de vente d'une très belle maison à Antibes. Cependant, ce n'est pas le secteur sur lequel il a l'habitude de travailler. Il aimerait en savoir plus sur les prix de ce secteur et les biens vendus et/ou en vente similaires à ce bien avant de mettre en ligne son annonce.
Ex. 2	David est promoteur et cherche le meilleur quartier où construire son prochain immeuble. Il a besoin pour chaque quartier de savoir les prix au m^2 et les services (écoles, transports, etc.) s'y trouvant.
QC	- Quels sont les autres biens du même secteur ? - Sont-ils similaires à celui à vendre ? - Quel est le prix moyen ? - Quels sont les services souvent mentionnés dans ce secteur ? - Quels sont les autres lieux mentionnés dans ce secteur ? - Quels sont les volumes de ventes des quartiers ? - Quels sont les prix au m^2 des quartiers ?

TABLE 2 – Scénario 2 : Etude du marché immobilier

Nom	Analyse du territoire
Desc.	Les agents immobiliers ont une bonne connaissance du territoire et en parlent au travers des annonces immobilières. Ceci permet d'avoir de nouvelles connaissances (plus proche du réel) qui peuvent être analysées.
Ex. 1	Alicia est chercheuse en géographie et voudrait étudier la forme sociale des espaces urbains afin de comprendre quelle partie du territoire est plus adaptée à une population ou à une autre.
Ex. 2	Clément travaille dans le service d'urbanisation de la mairie de Nice et voudrait comprendre la politique de la ville en matière de transports (zone qui manque de transports) et d'accès aux services par la population pour ajuster ses recommandations.
QC	- Comment est perçu un quartier (recherché, résidentiel, etc.) par les agents immobiliers ? - Quels services sont mentionnés dans un secteur ? - Quels services ne sont pas mentionnés ? - Quels autres lieux sont mentionnés dans ce secteur ? - Est-ce que certains lieux sont souvent / toujours / jamais mentionnés ensemble ? - Dans quelle partie de la ville parle-t-on le plus du tramway ? - Est-ce que les annonces qui mentionnent les transports en commun sont situé loin de certains lieux centraux (centre-ville) ? - Quelles sont les zones dans lesquelles les déplacements piétons sont mentionnés ? - Est-ce qu'un lieu est inclus dans un autre lieu ?

TABLE 3 – Scénario 3 : Analyse du territoire

1.3 Contributions

Nous avons conçu une ontologie pour la représentation des données immobilières issues du texte des annonces et répondant aux scénarios et questions identifiés. Pour cela, nous proposons une représentation des données spatiales floues et incertaines dans un graphe de connaissances. Le reste de cet article est organisé de la manière suivante. Dans la section 2, nous présentons l'état de l'art et ses limites à la lumière des scénarios présentés. La section 3 détaille la théorie des ensembles flous appliquée aux objets spatiaux incertains. Enfin, nous présentons nos choix techniques pour la modélisation de l'ontologie dans la section 4.

Nom	Utilisation d'annotations textuelles
Desc.	Les entités extraites à partir du texte peuvent constituer un très grand jeu de données pour la reconnaissance d'entités nommées (géographiques). Néanmoins, il faudra ajouter un terme de confiance pour chaque entité extraite.
Ex. 1	Julien est doctorant en NLP et voudrait tester son nouveau modèle d'extraction d'entités nommées sur un jeu de données en français avec des catégories liées à la géographie.
Ex. 2	Fabrice souhaite intégrer dans son moteur de recherche sur les annonces les scores de confiance pour trier les résultats et faire remonter ceux pour lesquels la confiance est maximale.
QC	- Quelles sont les annonces avec des entités appartenants à la catégorie « Toponym » et ayant une confiance supérieure à 0.8 ? - Quelles sont les annonces dont les entités extraites ont toutes une confiance supérieur à 0.5 ?

TABLE 4 – Scénario 4 : Utilisation d'annotations textuelles

2 État de l'art

Dans cette section, nous discutons les ontologies et graphes de connaissances existants pour le domaine de l'immobilier et les données spatiales, ainsi que leurs limites à la lumière de nos scénarios. La construction d'une ontologie pour l'immobilier dépend du point de vue adopté et du type de données utilisées. Dans [10], les auteurs comparent plusieurs ontologies appliquées à l'immobilier selon plusieurs perspectives : le territoire et notamment le cadastre [11], les transactions [12] et la juridiction [13]. L'ontologie *proDataMarket* [14] regroupe les trois domaines précédents et permet d'étudier le marché immobilier au travers des parcelles et des transactions. Néanmoins, cette ontologie se base sur des données anciennes (par exemple *Demande de Valeur Foncière* [1] en France) et étudie seulement les parcelles. Ainsi, il n'est pas possible de rechercher des biens immobiliers récents et en vente selon leurs caractéristiques intérieures (nombre de pièces, étage, etc.) ou leur environnement. L'ontologie *NAREO* [15] s'intéresse plutôt à l'environnement d'un bien immobilier avec la description des services et aménités à proximité du quartier dans lequel le bien est localisé. L'application de cette ontologie est la recommandation de quartiers selon des critères de localisation et d'environnement. Cependant, les auteurs ne décrivent pas les caractéristiques du bien et utilisent des jeux de données officiels tels que *OpenStreetMap* et les données de l'*INSEE*. Ni la perception de l'agent immobilier dans la description d'un lieu de vie, ni les noms de lieux vernaculaires ne sont pris en compte. La dimension spatiale des données est primordiale en immobilier puisqu'un bien immobilier est localisé sur le territoire (par exemple sur une parcelle) et que sa localisation joue un rôle prépondérant dans la décision d'achat. De nombreux graphes de connaissances ont intégré des entités spatiales tels que *DBPedia* [19], *Yago2Geo* [18], *WorldKG* [17] ou *KnowWhereGraph* [16]. Les données de ces graphes proviennent principalement d'agences gouvernementales (*INSEE*, *IGN*) ou de projets participatifs comme *Wikipedia* ou *OpenStreetMap*. Cependant, dans notre étude nous avons des données

1. https ://app.dvf.etalab.gouv.fr/

qui ne sont pas toujours répertoriées dans ces sources, ce qui limite leur utilisation. Différentes approches ([20]) ont été développées pour extraire des lieux vernaculaires (i.e., locaux, non-officiels) sur le Web et ainsi enrichir les gazetteers, mais ne proposent pas une manière standard de représenter le concept de lieu. D'un point de vue ontologique, il existe plusieurs manières de décrire une entité spatiale. *GeoNames*[2] utilise les concepts *SKOS* pour décrire des classes haut-niveau. *GeoLinkedData*[3] définit trois ontologies selon le domaine d'utilisation (administratif, transport, hydrographie) en réutilisant des vocabulaires existants. Enfin, l'*IGN*[4] a développé sa propre ontologie en s'appuyant sur la *BDTOPO* pour décrire les entités topographiques et administratives du territoire (bâtiment, réseau routier, végétation, etc.). Les limites de ces ontologies sont qu'elles classent toutes les entités selon leur nature et leur topographie. Or, notre application suggère une représentation des entités spatiales selon leur perception et leur utilisation. Dans [21] et [22], les auteurs proposent une représentation du concept de lieu notamment en mettant l'accent sur la formalisation de la provenance des informations et leur date d'utilisation. Cependant, ces représentations restent limités pour les lieux cognitifs. Finalement, pour rassembler les communautés du Web Sémantique et de la Géographie, *GeoSPARQL*[5] a été développé pour représenter et requêter les données spatiales. Son ontologie, composée de trois classes haut-niveau (*SpatialObject*, *Feature* et *Geometry*), offre une grande flexibilité pour décrire des entités spatiales et leurs géométries selon le domaine d'application. Celle-ci se base sur un ensemble de standards du Simple Feature Access[6] qui définit (1) une architecture commune pour la géométrie et sa représentation en text (WKT) ainsi que (2) un extension spatial des fonctions SQL.
En résumé, le but de ce travail est de représenter à la fois les informations immobilières des annonces et les données spatiales incertaines en un seul graphe de connaissance.

3 Localisation imprécise des lieux

Les scénarios et questions précédentes expriment le besoin de représenter les limites des lieux extraits dans les annonces immobilières, et plus particulièrement les limites des lieux vernaculaires (i.e., propres à la région étudiée) et non-officiels. En effet, la localisation et l'environnement du bien sont en général décrits dans les annonces immobilières mais au travers du regard de l'agent immobilier. Ainsi, les limites des lieux mentionnés sont celles perçues par l'agent immobilier et peuvent être différentes des limites administratives ou bien exagérées dans le but de vendre le bien [6]. Ces limites ne peuvent donc pas être simplement représentées par un point ou un polygone et demandent une représentation intégrant cette imprécision. Pour cela, nous avons décidé d'utiliser la théorie des ensembles flous.
Dans la théorie des ensembles flous [5], un sous-ensemble flou A d'un ensemble E est caractérisé par une application appelée fonction d'appartenance et notée μ_A. Celle-ci donne le degré d'appartenance à l'ensemble flou A, pour chaque élément x de E. Le degré d'appartenance est généralement dans l'intervalle [0,1]. On dit que si $\mu_A(x) = 1$, alors x appartient totalement à A tandis que si $\mu_A(x) = 0$, alors x n'appartient pas du tout à A.
Nous appliquons cette théorie aux lieux extraits pour capturer une approximation de la localisation en calculant le degré d'appartenance de chaque point de l'espace. De plus, il est possible d'obtenir des limites nettes (par exemple pour projeter la localisation sur une carte) en utilisant les α-coupes. Une α-coupe d'un ensemble flou A notée $\tilde{A}_\alpha$ est un sous-ensemble net dont chaque élément a un degré d'appartenance supérieur ou égal à α :

$$\tilde{A}_\alpha = \{x \in A; \mu_{\tilde{A}}(x) \geq \alpha\}.$$

Le noyau et le support sont des α-coupes particulières pour lesquelles α est égal respectivement à 1 et 0 :

$$\begin{aligned} noy(A) &= \{x \in A; \mu_A(x) = 1\}, \\ supp(A) &= \{x \in A; \mu_A(x) > 0\}. \end{aligned}$$

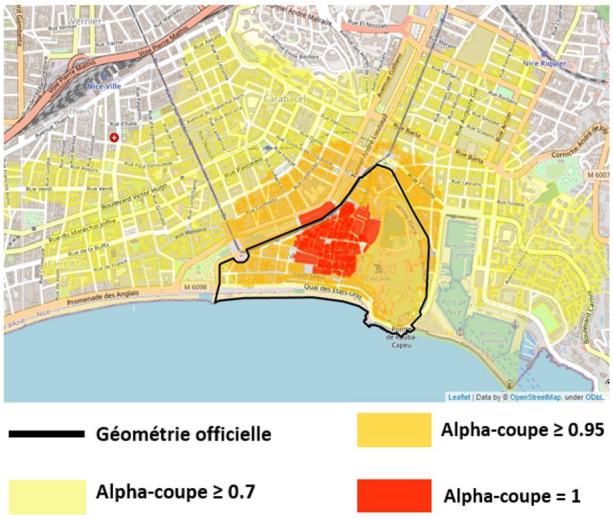

FIGURE 1 – Exemple de notre représentation incertaine VS les limites officielles du Vieux Nice

Dans l'étude des annonces immobilières, nous avons extrait un grand nombre de lieux vernaculaires qui sont mentionnés dans des annonces géolocalisées précisément (i.e., la latitude et la longitude sont sur ou proche du bâtiment). Nous avons donc pu estimer les limites d'un lieu en appliquant d'abord une méthode d'estimation par noyau gaussien (KDE) sur les coordonnées des annonces précisément géolocalisées et évoquant ce lieu. Cette méthode retourne une densité de probabilité sur l'espace étudié qui peut être facilement transformée en fonction d'appartenance d'un ensemble flou. La figure 1 montre un exemple de représentation incertaine du quartier du Vieux Nice calculée à

2. http ://geonames.org/ontology/ontology_v3.0.rdf
3. http ://geo.linkeddata.es
4. http ://data.ign.fr/def/topo/20190212.htm
5. http ://www.opengeospatial.org/standards/geosparql
6. https://www.ogc.org/standard/sfa/

partir des annonces immobilières et sa géométrie officielle. Pour cet exemple, nous avons représenté la géométrie incertaine à partir de 3 alpha-coupes projetées sur les parcelles de Nice. En effet, les biens immobiliers se trouvent sur des parcelles donc nous donnons un degrés d'appartenance à un lieu pour chaque parcelle. Nous pouvons remarquer que notre noyau ($\alpha = 1$) se situe dans la partie du quartier où les habitations sont très anciennes tandis que le reste du quartier a un degré d'appartenance un peu plus faible. Enfin, nous pouvons remarquer que le quartier du Vieux Nice a une grande influence puisqu'il est cité au-delà de ses limites officielles.

4 Modélisation ontologique

Dans cette section, nous décrivons et discutons la modélisation choisie pour définir les entités et les relations de notre ontologie. La figure 2 montre un exemple de notre modélisation appliquée à une annonce immobilière. Nous utilisons le préfixe `sure:` dans le reste de la section pour faire référence à l'ontologie *SURE*[7] (Spatial Uncertainty and Real Estate) que nous avons développée et publiée selon les standards et bonnes pratiques des données liées. Cette ontologie contient des classes et propriétés relatives au domaine d'étude et décrites ci-dessous, ainsi que des classes générées à partir des textes des annonces immobilières.

4.1 Représentation du bien immobilier

Les deux premiers scénarios (1,2) nous permettent d'identifier les entités et relations liées au bien immobilier. Nous avons notamment besoin de modéliser les termes suivants :

1. Le bien et son type ;
2. Les caractéristiques du bien (prix, étage, surface, nombre de pièces, calme, rénové, etc.) ;
3. La localisation du bien : la ville, les coordonnées et les lieux mentionnés dans l'annonce avec les relations spatiales.

Nous avons utilisé les vocabulaires *GeoSPARQL* et *schema.org* pour représenter le bien immobilier. La classe *Accommodation* de *schema.org* nous permet de modéliser le type du bien (*Apartment*, *House*, etc.) et certaines caractéristiques du bien grâce aux propriétés préalablement définies (*numberOfRooms*, *floorLevel*, etc.). D'autre part, le bien immobilier est aussi un objet spatial puisqu'il peut avoir des coordonnées ou des relations spatiales avec des lieux. Nous avons donc choisi de créer une sous-classe *RealEstate* de la classe *Feature* de *GeoSPARQL*. Ceci nous permet de créer une géométrie si nous connaissons sa position ou d'utiliser les propriétés de *GeoSPARQL* pour le localiser dans la ville et dans les lieux extraits (*geo:sfWithin*).

4.2 Représentation des lieux et localisations

Le troisième scénario (tableau 3) décrit les besoins relatifs à la représentation des lieux géographiques. Néanmoins, pour modéliser ces lieux, nous devons d'abord définir ce qu'est un lieu. Dans [7], les auteurs décrivent quatre manières de parler d'un lieu :

- *Place-Names* : la manière la plus simple de parler d'un lieu est d'utiliser son nom propre (Nice, Promenade des Anglais) ;
- *Place-Like Count Nouns* : utilisation d'un nom commun pour lequel un objet peut être localisé "dans" (ville, quartier, environnement, etc.) ;
- *Locative Property Phrases* : composition d'un nom propre ou commun avec une relation spatiale ("proche de la promenade des Anglais", "non loin de la gare", etc.) ;
- *Definite Descriptions* : description d'un objet à partir d'un autre objet et d'une relation spatiale ("la rue derrière la gare", "l'école à côté de la place Masséna").

D'autres travaux ([8], [9]) définissent seulement deux catégories pour parler d'un lieu : lieu absolu et lieu relatif. Le lieu absolu s'apparente à la catégorie *Place-Names* tandis que le lieu relatif est composé d'un lieu absolu et d'une relation spatiale.

Dans notre approche, nous avons décidé de créer un classe générale *Place* et une sous-classe *RelativePlace* qui correspond aux lieux composés d'une relation spatiale autre que "dans". La classe *Place* est une sous-classe de *geo:Feature*. La classe *RelativePlace* a deux propriétés supplémentaires permettant de définir le type de relation spatiale et l'objet de la relation spatiale (*hasSpatialRelation*, *hasAnchor*).

Les instances de *Place* sont principalement les lieux définis avec des noms propres (Place Masséna, Nice, etc.). Néanmoins, les noms communs qui s'apparentent à des lieux et dans lesquels le bien peut être localisé (centre-ville, zone piétonne, rue, etc.) sont aussi des instances de cette classe. Pour choisir si un nom commun peut être vu comme un lieu ou non, nous avons créé deux classes, *Amenity* et *LocativeArea*, qui sont aussi des sous-classes de *geo:Feature*. *LocativeArea* regroupe les noms communs qui localisent le bien. *Amenity* regroupe les entités qui se trouvent à un certaine proximité du bien et qui lui donnent de la valeur (gare, école, port, plage, etc.). Enfin, les lieux composés d'un nom propre ou d'un nom commun (*LocativeArea* ou *Amenity*) et d'une relation spatiale sont des instances de la classe *RelativePlace*. Nous avons généré automatiquement des sous-classes de *LocativeArea* et *Amenity* à partir du texte des annonces immobilières (*Quartier*, *Gare*, *Rue*, etc.). A ce stade, aucun traitement n'a été appliqué a posteriori, à l'exception d'une heuristique requérant au moins deux instances pour qu'une classe soit conservée i.e. deux annonces mentionnant la classe. Une perspective sera de travailler sur l'amélioration de la qualité de cette extraction.

Enfin, nous représentons les géométries de ces lieux à l'aide de la théorie des ensembles flous présentée dans la section 3. Nous avons choisi de créer une classe *AlphaCut* qui est une sous-classe de *geo:Geometry*. Cette classe a la propriété *hasAlpha* pour définir le degré d'appartenance correspondant. Enfin, *GeoSPARQL* permet d'associer une collection de géométries à un même objet ce qui nous permet d'associer plusieurs *AlphaCut* à un même lieu pour re-

7. `http://ns.inria.fr/sure#`

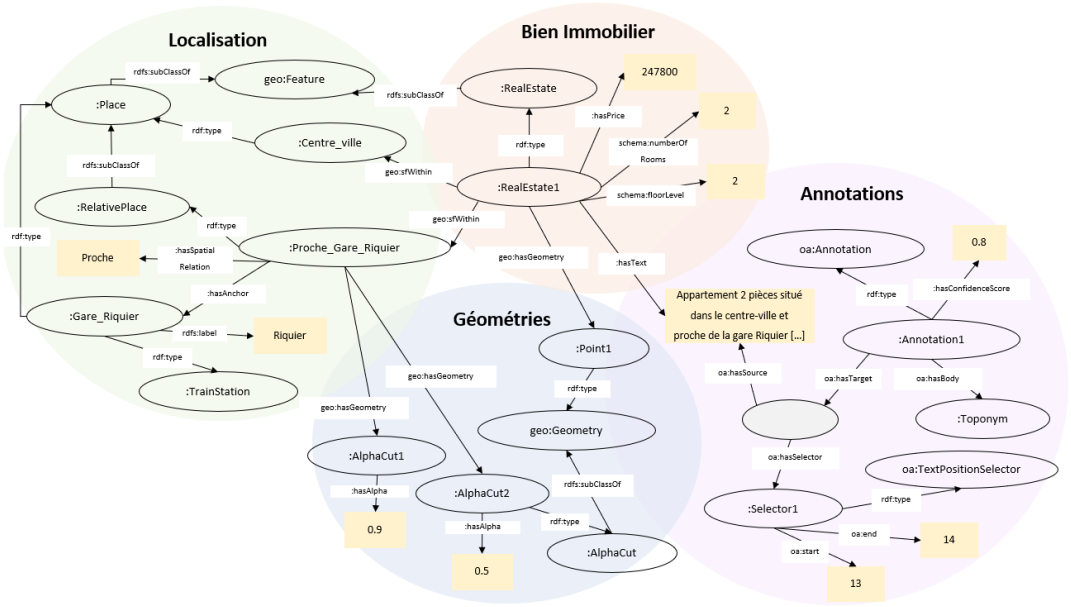

FIGURE 2 – Exemple d'un graphe RDF représentant les informations issues d'une annonce immobilière

présenter de manière aussi fiable que possible sa frontière floue. Le listing 1 donne un extrait de cette formalisation.

```
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns
    #> .
@prefix geo: <http://www.opengis.net/ont/geosparql#> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix owl: <http://www.w3.org/2002/07/owl#> .
@prefix xsd: <http://www.w3.org/2001/XMLSchema#> .
@prefix : <http://ns.inria.fr/sure#> .

############# Classes et Proprietes
:TrainStation rdfs:subClassOf geo:Feature.
:Place rdfs:subClassOf geo:Feature.
:RelativePlace rdfs:subClassOf :Place.
:AlphaCut rdfs:subClassOf geo:Geometry.

:hasAlpha a rdfs:Property ;
   rdfs:domain :AlphaCut ;
   rdfs:range xsd:double .

:hasAnchor a rdfs:Property ;
   rdfs:domain :RelativePlace ;
   rdfs:range geo:Feature.

:hasSpatialRelation a rdfs:Property ;
   rdfs:domain :RelativePlace ;
   rdfs:range xsd:string.

############# Instances
:Gare_Riquier a :TrainStation, :Place;
  rdfs:label "riquier"@fr.

:Proche_Gare_Riquier a :RelativePlace;
  :hasAnchor :Gare_Riquier;
  :hasSpatialRelation "proche";
  geo:hasGeometry :AlphaCut1, :AlphaCut2.

:AlphaCut1 a :AlphaCut ;:hasAlpha "0.5"^^xsd:double;
    geo:asWKT "MULTIPOLYGON (((43.6957 7.280889,..., 43.
     69578 7.280882)))"^^geo:wktLiteral.

:AlphaCut2 a :AlphaCut ;:hasAlpha "0.9"^^xsd:double;
    geo:asWKT "MULTIPOLYGON (((43.695 7.2808,..., 43.
     6955 7.280892)))"^^geo:wktLiteral.
```

LISTING 1: Exemple de la syntaxe RDF d'un lieu incertain et de ses limites floues.

4.3 Représentation des textes annotés

Le dernier scénario (tableau 4) a un objectif qui ne touche pas à l'immobilier. Nous proposons d'utiliser les annotations textuelles produites par le modèle de reconnaissance d'entités dans le texte comme un jeu de données réutilisable pour mener d'autres recherches en traitement du langage naturel, notamment sur l'extraction d'entités nommées. Le vocabulaire *Web Annotation Data Model* permet d'annoter les entités retrouvées dans un texte. Nous avons donc fait le choix d'utiliser ce vocabulaire. Néanmoins, les annotations ne sont pas certaines puisqu'elles proviennent des prédictions du modèle de reconnaissance d'entités. Ainsi, nous avons ajouté le score de confiance donné par le modèle à l'annotation à l'aide d'une nouvelle propriété *hasConfidenceScore*.

5 Conclusion et Perspectives

La représentation des données immobilières issues de l'extraction d'information des annonces présente plusieurs enjeux que nous avons décrits dans cet article. Nous avons proposé une modélisation ontologique et justifié nos choix à l'aide de nos scénarios motivants. Ainsi, nous avons modélisé le bien immobilier, ses caractéristiques et sa localisation. Nous avons montré que la localisation est en général incertaine et nécessite une représentation particulière. Nous avons proposé d'utiliser la théorie des ensembles flous et d'intégrer les alpha-coupes dans notre ontologie. Enfin, nous avons ajouté les annotations textuelles issues d'un modèle de reconnaissance d'entités nommées afin de créer un jeu de données réutilisable pour mener d'autres recherches en traitement du langage naturel. La prochaine étape de ce travail est donc le peuplement de l'ontologie et la création du graphe de connaissances à partir des annonces immobilières localisées dans les Alpes-Maritimes

dans un premier temps, et dans toute la France dans un second temps. Nous nous attacherons aussi à évaluer notre modélisation. Finalement, le graphe de connaissances produit pourra permettre de retrouver des biens immobiliers similaires et créer, à termes, un système de recommandation.

Références

[1] Uschold, Mike, and Michael Gruninger. "Ontologies : Principles, methods and applications." The knowledge engineering review 11.2 (1996) : 93-136.

[2] Bosvieux, Jean. "L'immobilier, poids lourd de l'économie", Constructif, vol. 49, no. 1, 2018, pp. 10-14.

[3] Bekoulis, Giannis, Deleu, Johannes, Demeester, Thomas, and Develder, Chris. 2017. "Reconstructing the house from the ad : Structured prediction on real estate classifieds." In Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics : Volume 2, Short Papers, pages 274–279, Valencia, Spain.

[4] Lucie Cadorel, Alicia Blanchi, and Andrea G. B. Tettamanzi. 2021. "Geospatial Knowledge in Housing Advertisements : Capturing and Extracting Spatial Information from Text." In Proceedings of the 11th on Knowledge Capture Conference (K-CAP '21). ACM, USA, 41–48.

[5] L.A. Zadeh. 1965. "Fuzzy sets." Information and Control 8, 3 (1965), 338–353.

[6] Grant McKenzie and Yingjie Hu. 2017. "The "Nearby" Exaggeration in Real Es- tate (Position Paper)." In Proceedings of the Cognitive Scales of Spatial Information Workshop (CoSSI 2017) (L'Aquila, Italy), Werner Kuhn, Dan Montello, Scott Fre- undschuh, Crystal Bae, Thomas Harvey, Sara Lafia, and Daniel Phillips (Eds.). 4–8.

[7] Bennett, B., Agarwal, P. (2007). Semantic Categories Underlying the Meaning of 'Place'. In : Winter, S., Duckham, M., Kulik, L., Kuipers, B. (eds) Spatial Information Theory. COSIT 2007. LNCS, vol 4736. Springer.

[8] Lesbegueries, Julien, Christian Sallaberry and Mauro Gaio. "Associating spatial patterns to text-units for summarizing geographic information." Workshop on Geographic Information Retrieval (2006).

[9] Syed, M. A., Arsevska, E., Roche, M., and Teisseire, M. : GeoXTag : Relative Spatial Information Extraction and Tagging of Unstructured Text, AGILE GIScience Ser., 3, 16,

[10] Ling Shi and Dumitru Roman. 2018. Ontologies for the Real Property Domain. In Proceedings of the 8th International Conference on Web Intelligence, Mining and Semantics (WIMS '18). Association for Computing Machinery, New York, NY, USA, Article 14, 1–8.

[11] D Sladić, M Govedarica, D Pržulj, A Radulović and D Jovanović (2013) Ontology for real estate cadastre, Survey Review, 45 :332, 357-371, DOI : 10.1179/1752270613Y.0000000042

[12] Erik Stubkjaer. 2017. The ontology and modelling of real estate transactions. Routledge

[13] Jesper M Paasch. 2005. Legal Cadastral Domain Model : An Object-orientated Approach. Nordic journal of surveying and real estate research 2, 1 (2005), 117–136.

[14] Ling Shi, Nikolay Nikolov, Dina Sukhobok, Tatiana Tarasova, and Dumitru Roman. 2017. The proDataMarket Ontology for Publishing and Integrating Crossdomain Real Property Data. journal Territorio Italia. Land Administration, Cadastre and Real Estate 2 (2017)

[15] Wissame Laddada, Fabien Duchateau, Franck Favetta, and Ludovic Moncla. 2020. Ontology-Based Approach for Neighborhood and Real Estate Recommendations. In Proceedings of the 4th ACM SIGSPATIAL Workshop on Location-Based Recommendations, Geosocial Networks, and Geoadvertising (LocalRec'20). ACM, USA, Article 4, 1–10.

[16] Krzysztof Janowicz, Pascal Hitzler, Wenwen Li, Dean Rehberger, Mark Schild- hauer, Rui Zhu, Cogan Shimizu, Colby K Fisher, Ling Cai, Gengchen Mai, et al. 2022. Know, Know Where, KnowWhereGraph : A densely connected, cross-domain knowledge graph and geo-enrichment service stack for applications in environ- mental intelligence. AI Magazine 43, 1 (2022), 30–39.

[17] Dsouza, Alishiba and Tempelmeier, Nicolas and Yu, Ran and Gottschalk, Simon and Demidova, Elena. WorldKG : A World-Scale Geographic Knowledge Graph. 30th ACM International Conference on Information and Knowledge Management (CIKM), 2021.

[18] Nikolaos Karalis, Georgios M. Mandilaras, and Manolis Koubarakis. 2019. Extending the YAGO2 Knowledge Graph with Precise Geospatial Knowledge. In Proc. of the ISWC 2019, Part II (Lecture Notes in Computer Science, Vol. 11779). Springer, 181–197

[19] Sö ren Auer, Christian Bizer, Georgi Kobilarov, Jens Lehmann, Richard Cyganiak, and Zachary G. Ives. 2007. DBpedia : A Nucleus for a Web of Open Data. In Proc. of the ISWC 2007 (Lecture Notes in Computer Science, Vol. 4825). Springer, 722–735

[20] Christopher B. Jones, Ross S. Purves, Paul Clough, Hideo Joho : Modelling Vague Places with Knowledge from the Web. International Journal of Geographic Information Science 22(10), 1045 – 1065 (2008)

[21] Karl Grossner and Ruth Mostern. "Linked places in world historical gazetteer." Proceedings of the 5th ACM SIGSPATIAL International Workshop on Geospatial Humanities. 2021

[22] Andrea Ballatore. "Prolegomena for an ontology of place". Advancing geographic information science, 91-103. 2016