| | | |
|---|---|---|
| Auteur(s) | : | C. Trojahn, H. Arenas, C. Comparot, N. Aussenac-Gilles |
| Source **(SOCIETE)** | : | IRIT |
| **REFERENCE** | : | Livrable 4.5.1 |
| Version | : | 1 |
| Date du document | : | 14/02/2017 |
| Nombre de pages | : | 9 |
| Propriétaire | | |

# Livrable 4.5.1 : Inventaire des ressources et des approches d'alignement

# Deliverable 4.5.1: Inventory of resources and approaches for ontology matching

© Copyright 2016, Atos S.E. Tous droits réservés. Toute reproduction en tout ou partie de ce document est interdite sauf autorisation préalable expresse et écrite de l'auteur. Pour toute question ou remarque se rapportant à ce document, veuillez contacter Atos, 05 34 36 37 38.

# Table des Matières

# Liste des modifications

| Version | Date | Description | Auteur(s) |
|---|---|---|---|
| 1 | 14/02/2017 | Description of the main approaches and resources for ontology and instance matching | C. Trojahn, H. Arenas, C. Comparot, N. Aussenac-Gilles |
| | | | |
| | | | |

# Résumé/Abstract

Ce document constitue le livrable 4.5.1: Inventaire des ressources et des approches d'alignement. L'alignement d'ontologies consiste à déterminer des correspondances entre les éléments de différentes ontologies. Nous nous intéressons ici à la problématique d'alignement d'ontologies et des données géo-spatiales. Nous présentons un état de l'art succinct des approches et des ressources impliquant ce type de données.

This document stands for to Deliverable 4.5.1: Inventory of Resources and Approaches for Ontology Matching. Ontology matching consists in determining a set of semantic correspondences between the elements of different ontologies (in the broad term). Here, we are interested in the problem of matching ontologies and data with a geo-spatial dimension. We present a brief survey of the state of the art on the topic.

# 1. Introduction

Ontology (and instance) matching [Euzenat et al., 2013] consists in determining the correspondences between the elements of different ontologies (including their instances).

Several matching approaches have been proposed in the literature. These approaches can be classified according to the characteristics present in the ontologies (labels, structure, instances, semantics), or according to the techniques used (e.g., statistics, combinatorics, semantics, linguistics, learning, or data analysis).

In order to systematically evaluate the different matching techniques, the Ontology Alignment Evaluation Initiative (OAEI) [Achichi et. al. 2016] have been carried out for many years. As part of these campaigns, several systems involving both schema and instance matching have been presented. An inventory of them can be found in [Bergman, 2014].

Here, we are interested in the problem of matching geo-spatial ontologies and data.

# 2. Brief overview on matching approaches and resources with geo-spatial features

As far as the matching of ontologies is concerned in GeoSciences, the problem arises in particular at the level of the data (instance matching, entity matching or entity resolution). This task is performed either by applying terminological and lexical similarities for entities [Raimond and Mustire, 2008; Karam et al., 2010; Morana et al., 2014], by exploiting spatial information [Safra et al., 2010], by applying machine learning algorithms [Sehgal et al., 2006] or by computing coherence and consistency checking at schema level [Du et al. 2012].

The matching approach in [Morana et al., 2014] consists in computing the similarity between the attributes of entities, using different strategies according to the kind of attribute. For instance, the coordinates of two entities are compared according to the Euclidean distance, while the Levenhstein measure is applied between the names (or titles) of entities (phone number or web site, for instance). The manual alignment of the type's hierarchies is also used for computing the similarity value between two types. The authors also present GeoBench, a tool for assisting users in the discovery and integration of corresponding spatial entities.

In [Sehgal et al., 2006], while location names are compared using Jaccard measure, coordinate matching is defined as the inverse of the coordinate distance, and location type similarity is compared using co-occurrence probabilities (if two locations are duplicates they are likely to have similar location types). The authors then integrate spatial and non-spatial features and apply different machine learning algorithms (logistic regression, neural network, and support vector machines) for learning a classifier that combines both kinds of features and tunes their weights.

While those works focus on pairswise matching and lexical comparison of attributes, [Berri et al., 2005] propose a set of algorithms for location-based matching of three or more sources, exploiting the spatial relations between objects. The matching is based on the following assumptions : locations of objects are recorded as points and more complex forms of recording locations (e.g, polygons) can be approximated by points (e.g., by computing the center of mass); and distinct objects represent distinct real-world entities. The similarity between locations is then computed using different join algorithms, such as the one-side nearest-neighbor, mutually nearest join and normalized-weighs methods.

In [Du et al., 2013], the spatial relations are also exploited. The authors propose a method for establishing \`sameAs' and \`partOf' relations between geospatial instances from different ontologies. The method is based on mereological partOf in geometry and similarity of labels. If for two instances have \`partOf' relations in both directions, a \`sameAs' can be established between them. This work has been extended with other relations involving geometries (\`bufferedEqual' and \`bufferedPartOf') and the implementation tof he MatchMaps system [Du et al., 2015].

While those approaches mostly focus on data level, [Du et al., 2012; Du et al., 2013a] propose a semi-automatic method to match geospatial ontologies (GeoMap), based on coherence and consistency checking in description logic and domain experts' knowledge. They consider in fact both geospatial concepts (Tbox) and facts about geospatial individuals (Abox) with geometry and location information. They define a set of disjointness axioms to improve the quality of the alignment. The matching process has two main steps: generating assumptions and calculating a consistent and coherent assumption set using a reasoner. In case of inconsistencies and incoherences, the user expert is able to correct them.

In [Cruz et al. 2005], the structural level of geospatial ontologies is exploited in the matching process and a tool helping users in the edition of alignments. A different approach is adopted in [Bharambe et al., 2012], focusing on the resolution of uncertainty in geospatial information. They propose a hybrid approach that integrates string matching and information theory based approaches.

A challenge in the field is to have benchmarks to evaluate the different aspects of entity matching with geospatial entities. In [Delgado et al. 2013] the evaluation of ontology matching techniques is performed, in the task of performing an automatic integration of geospatial information modeled from different viewpoints. To that, an evaluation methodology was designed, and it was applied to the discovery of relationships between CityGML and ontologies coming from the building information modeling and Geospatial Semantic Web domains.

Recently, [Cheatham et al., 2016] proposes to use the GeoLink data repository as an instance matching benchmark with a particular focus on coreference resolution. The GeoLink project brings together seven datasets related to geoscience research (BCO-DMO, DataONE, IEDA, IODP, LTER, MBLWHOI Library, and R2R). Several correspondences between instances are expressed as owl:sameAs and skos:closeMatch links. The sameAs links were manually generated by the data providers, while the closeMatch links were generated by an automated coreference resolution system. The authors highlight three different classes within the GeoLink schema that pose opportunities for evaluating and challenging coreference resolution systems: Person, Cruise, and Organization.

A similar aim has been addressed by [Berjawi et al. 2015], which propose PABench, a benchmark for spatial entity matching that includes a taxonomy of observed differences, inconsistencies and errors between different location-based service (LBS) providers. A taxonomy that characterizes differences, heterogeneities and errors between LBS providers at four levels (schema, terminology, spatial and availability) has been proposed.

However, as stated by [Fundulaki et al. 2016], inking spatial resources requires techniques that differ from the classical mostly string-based approaches. In particular, considering the topology of the spatial resources and the topological relations between them is of central importance to systems driven by spatial data. The authors propose a benchmark for geo-spatial link discovery tasks. This benchmark relies on widely accepted Linked Data datasets such as GeoNames, LinkedGeoData, and Dbpedia. The matching tasks focus on the different types of spatial object representations and provide different data transformation levels. In these transformations, objects may keep their representation, they may change their geometry, type or attributes, merge with other objects, or can completely disappear. This is a scenario that stems from the heterogeneous datasets (in structure and semantics) used to describe geo-spatial entities.

# 3. Conclusions

This deliverable has briefly introduced the state-of-the-art on matching geospatial ontologies. Different approaches to the problem have been proposed, from terminological approaches evaluating values of attributes to spatial relations exploiting the geometries of locations.

In our work, we have defined a set of vocabularies that are used to represent satellite images records enriched with contextual data (weather, etc.). The integration of data from different sources is then performed by linking the entities according to their coordinates (latitude and longitude).

As future work, we aim at matching the integrated knowledge to external sources in the Linked Open Data, by combining the approaches presented in this deliverable. We are as well particularly interested in exploiting the spatial relations between geometries from satellite images, in order to establish topological relations between images.

# 4. References

Achichi, M.; Cheatham, M.; Dragisic, Z.; Euzenat, J.; Faria, D.; Ferrara, A.; Flouris, G.; Fundulaki, I.; Harrow,I.; Ivanova, V.; Jiménez-Ruiz,e.; Kuss, E.; Lambrix, P.; Leopold, H;. Li, H.; Meilicke, C.; Montanelli, S.; Pesquita, C.; Saveta,T.; Shvaiko, P.; Splendiani, A.; Stuckenschmidt, H.; Todorov, K.; Trojahn, C.; Zamazal O. Results of the Ontology Alignment Evaluation Initiative 2016. OM@ISWC 2016: 73-129.

Beeri, C.; Doytsher, Y.; Kanza, Y.; Safra, E.; Sagiv, E. Finding corresponding objects when integrating several geo-spatial datasets. Proceedings of the 13th annual ACM international workshop on Geographic information systems (GIS '05). ACM, New York, NY, USA, 87-96.

Bergman, M. 50 Ontology Mapping and Alignment Tools. http://www.mkbergman.com/1769/50-ontology-mapping-and-alignment-tools/, 2014 (available in 14/02/2017).

Berjawi, B.; Duchateau, F.; Favetta, F.; Miquel, M.; Laurin, R. PABench: Designing a Taxonomy and Implementing a Benchmark for Spatial Entity Matching. Proc. of the 7th Int. Conf. on Advanced Geographic Information Systems, Applications, and Services, Feb 2015, Lisbon, Portugal.7-16, 2015.

Cheatham, M. ; Amini, R. ; Patel, C. Matching instances in GeoLink. OM@ISWC 2016: 237-238.

Cruz, I.; Sunna, W. ;Ayloo, K. Concept Level Matching of Geospatial Ontologies. Proceedings of the GIS Planet Second Conference and Exhibition on Geographic Information, 2005.

Delgado, F.; . M. Martínez-González; Finat, J. An evaluation of ontology matching techniques on geospatial ontologies. International Journal of Geographical Information Science. 27:12, 2279-2301, 2013.

Du, H.; Alechina, N.; Jackson, M.; Hart, G. Matching geospatial ontologies. OM@ISWC 2102:250-251.

Du, H.; Alechina, N.; Jackson, M.; Hart, G. Matching geospatial instances. OM@ISWC 2013:239-240.

Du, H.; Alechina, N.; Jackson, M.; Hart, G. Matching Formal and Informal Geospatial Ontologies." Geographic Information Science at the Heart of Europe, Springer International Publishing, 155-171, 2013a.

Du, H.; Alechina, N.; Jackson, M.; Hart, G. A tool for matching crowd-sourced and authoritative geospatial data. Inter. Conference on Military Communications and Information Systems, 1-8, 2015.

Euzenat, J. and Shvaiko, P. (2013). Ontology matching. Springer-Verlag, Heidelberg (DE). 2nd edition.

Karam, R.; Favetta, F;. Kilany, R., Laurini, R. Integration of similar location based services proposed by several providers. Networked Digital Technologies, 2010, pp. 136–144.

Morana, A., Morel, T.; Berjawi, B.; and Duchateau F.: GeoBench: a geospatial integration tool for building a spatial entity matching benchmark. SIGSPATIAL/GIS 2014: 533-536

Safra, E.; Kanza, Y.; Sagiv, Y.; Beeri, C.; Doytsher, Y. Location-based algorithms for finding sets of corresponding objects over several geo-spatial data sets. International Journal of Geographical Information Science, vol. 24, no. 1, 2010, pp. 69–106.

Raimond, A. and Mustire, S. Data matching - a matter of belief. International Symposium on Spatial Data Handling (SDH), 2008, pp. 501–519.

Sehgal, V.; Getoor, L.; Viechnicki, P. Entity resolution in geospatial data integration. ACM International Symposium on Geographic Information Systems, 2006, pp. 83–90.