# Ontology alignment in the urban domain

Sylvie Calabretto

# Ontology Alignment in the Urban Domain

**Sylvie Calabretto**

## 4.1 Introduction

Concepts in the domain of Urban Civil Engineering are often categorized and described using ontologies. Such ontologies may be designed independently by domain experts who have a minimal communication or no communication between them. As a result, similar concepts may be described differently and their categorization may result in heterogeneous ontologies.

More and more ontology-based urban systems are being built in different countries. However, most of the language-processing oriented ontologies that have been built so far have English or another language as basis. Since there is a growing need for multilingual ontologies, it is natural to ask for multilingual ontology alignment and viewpoint confrontation.

In this chapter, we first introduce several justifications for heterogeneity and give illustrations on urban problems. We then give some definitions on ontology matching and alignment, and some elements on ontology alignment approaches. Then we propose an overview of ontology alignment in urban or GIS domain and of viewpoint confrontation systems. Finally, we present the Hyppodamos tool as a solution for multilingual ontology alignment.

## 4.2 Heterogeneity in Urban Problems

Heterogeneity does not lie solely in the differences between goals of the applications according to which they have been designed or in the expression formalisms in which they have been encoded. They have been different classifications to types

S. Calabretto (✉)
Institut National des Sciences Appliquées de Lyon (INSA-Lyon), France
e-mail: sylvie.calabretto@insa-lyon.fr

of heterogeneity: syntactic heterogeneity, terminological heterogeneity, conceptual heterogeneity and semiotic heterogeneity. Usually, several types of heterogeneity occur together. The terminological heterogeneity occurs due to the variations in names when referring to the same entities in different ontologies.

### *4.2.1 Syntactic Heterogeneity*

Syntactic heterogeneity occurs when two ontologies are not expressed in the same ontology language. This happens when two ontologies are modelled by using different knowledge representation formalisms, for instance OWL and F-logic. A solution to this heterogeneity consists in establishing equivalences between constructs of different languages. But this is not always possible. For instance if a language is more expressive than another one, not all F-logic expressions can be translated to OWL.

### *4.2.2 Terminological Heterogeneity*

The terminological heterogeneity occurs due to the variations in names when referring to the same entities in different ontologies. This can be caused by the use of different natural languages, e.g. Paper vs. Articulo, different technical sublanguages, e.g. Paper vs. Memo, or the use of synonyms, e.g., Paper vs. Article. The Fig. 4.1 is an example in the urban domain. It is based on the glossary of urban, regional and environmental planning terms established in 2004 by Calderon and Ventura (Fig. 4.2).

### *4.2.3 Conceptual Heterogeneity*

Conceptual heterogeneity stands for the differences in modelling the same domain of interest. This type of heterogeneity is also called semantic heterogeneity in Euzenat (2001) or logical mismatch in Klein (2001). This can happen because of the use of different axioms for defining concepts or because of the use of totally different concepts (geometry axiomatised with points as primitive objects or with spheres as primitive objects). Benerecetti et al. (2001) identifies three different reasons for conceptual heterogeneity to hold: difference in coverage, difference in granularity and difference in perspective.

### *4.2.4 Semiotic Heterogeneity*

Semiotic heterogeneity is also called pragmatic heterogeneity in Bouquet et al. (2004). This heterogeneity is concerned with how entities are interpreted by people.

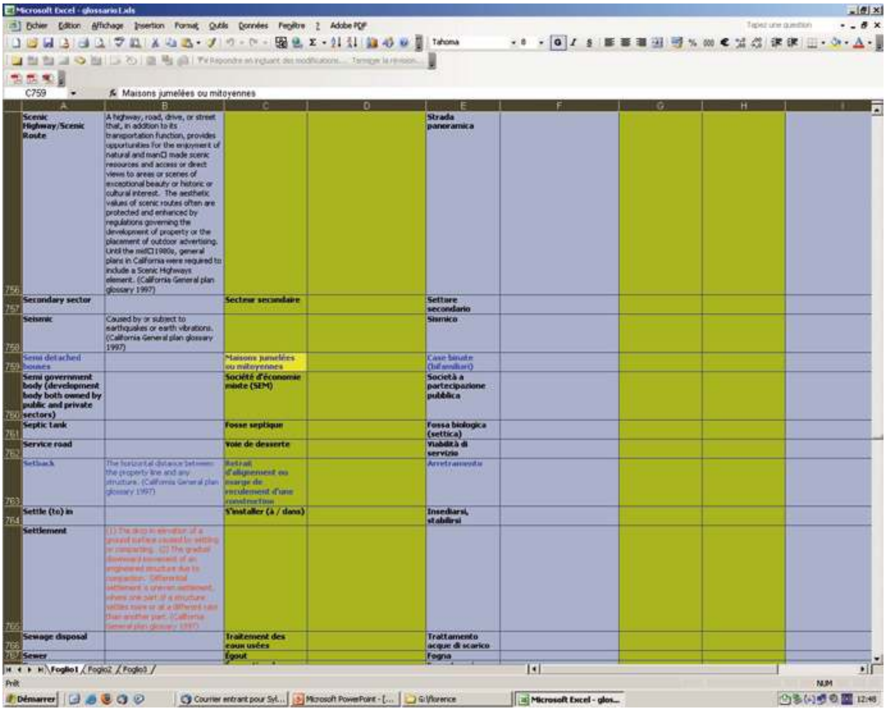


**Fig. 4.1** Glossary of urban planning terms in english, french and italian languages

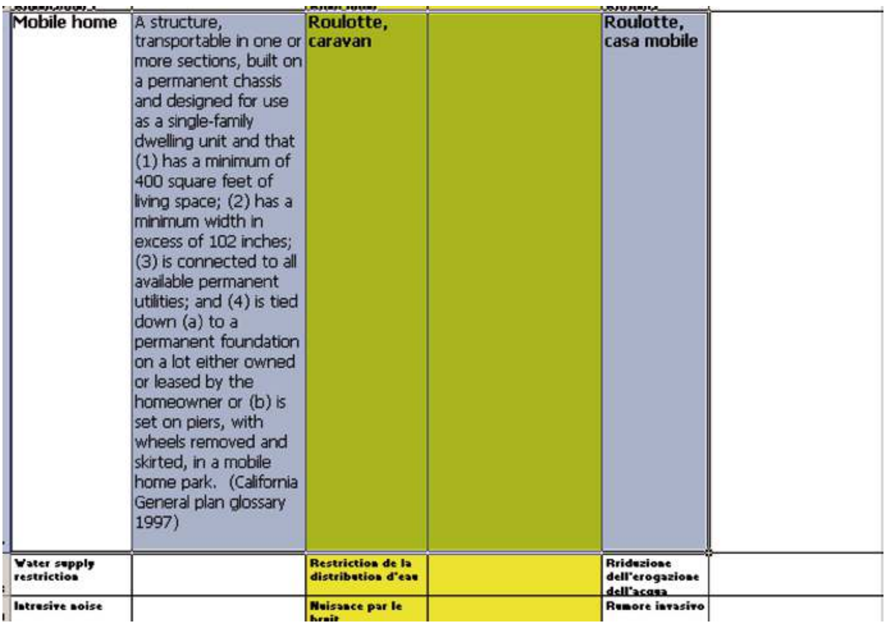


**Fig. 4.2** Glossary with focus on mobile home term (caravane in french and casa mobile in italian)

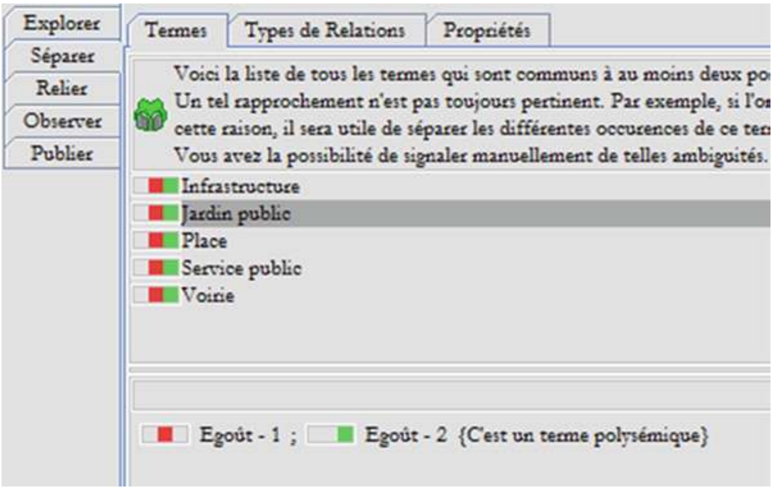


**Fig. 4.3** Example of semiotic heterogeneity

In the example of Fig. 4.3, the term "Egoût" is interpreted differently by the Expert 1 and the Expert 2. This type of heterogeneity is very difficult to detect and solve by a computer.

### *4.2.5 Terminology in Ontology Alignment*

Ontology matching aims at finding correspondences between semantically related entities of different ontologies. These correspondences may stand for equivalence as well as other relations, such as consequence, subsumption, or disjointness, between ontology entities. Ontology entities, in turn, usually denote the named entities of ontologies, such as classes, properties or individuals. Ontology matching results, called alignments, can thus express with various degrees of precision the relations between the ontologies under consideration (Euzenat and Shvaiko 2007).

Ontology alignment is a set of correspondences between two or more (in case of multiple matching) ontologies (by analogy with molecular sequence alignment). The alignment is the output of the matching process.

Alignments can be used for various tasks, such as ontology merging, query answering, data translation or for browsing the semantic web.

Ontology merging is the creation of a new ontology from two, possibly overlapping, source ontologies. The initial ontologies remain unaltered. The merged ontology is supposed to contain the knowledge of the initial ontologies, e.g., consequences of each ontology are consequences of the merge. This concept is closely related to that of schema integration in databases.

### 4.2.6 Ontology Alignment Approaches

The ontology alignment problem can be expressed as: in How to find the relationships that hold between the entities represented in different taxonomies?

We can identify two approaches (Nogueras-Iso et al. 2006) for the ontology construction: manual and automated ontology construction.

In the manual approach, we use the matching of terms (names and acronyms) between the different taxonomies. We can consider three categories of matches:

- Exact match
- Partial match: one concept is broader or narrower No match
- Provisional match: taxonomy errors (homonyms) imply erroneous matches

The automated approaches are used because manual mappings are time consuming and because some mappings may not be successful (content creators have not assigned the correct feature type). Two main approaches have been discussed in the literature: one which exploits the abstract data (nodes) represented by its names (lexical methods) and another which exploits the relationships (edges) between the various classes that form the structure of the ontology, (structural methods). Consequently, some of these techniques attempt to compare text strings that describe the entities in the ontologies (terminology-based ontology alignment) while others calculate the similarity measures between entities taking into account the structure of their corresponding ontologies (structural ontology alignment).

### 4.2.7 Overview of Ontology Alignment Tools

The state of the art of ontology alignment methodologies was recently surveyed by Euzenat and Shvaiko (2007). Previously, Rahm and Bernstein surveyed schema matching in databases (Rahm and Bernstein 2001).

In this section, we cover ontology and alignment tools even if most of them do not focus specifically on the urban domain. A notable exception is offered by Fonseca et al. (2002). They introduce an ontology-driven geographic information system (ODGIS), which is used to drive the creation of ontologies that will enable the integration of geospatial data.

Chimaera (McGuinness et al. 2000) is a software tool developed by the KSL group at Stanford, which provides tools for merging ontologies and checking the correctness of ontologies. Chimaera is web-based. Its graphical user interface supports a set of commands accessible via spring-loaded menus as well as drag and drop editing. The interface displays the knowledge base being edited and allows for users to check an automated merging procedure by highlighting the classes that require the user's attention. The authors of Chimeara consider the task of merging two ontologies to be one of combining two or more ontologies that may use different vocabularies and may have overlapping content. The major two tasks are to (1) to coalesce two semantically identical terms from different ontologies so that they are referred to by the same name in the resulting ontology, and (2) to identify terms that should be related by subsumption, disjointness, or instance relationships and provide support for introducing those relationships.

COMA++ (Aumueller et al. 2005) is a schema and ontology mapping tool, which is in many ways similar to our own mapping tool. However, both tools have been developed independently. COMA++ supports an iterative and automatic matching of ontology components and multiple matching algorithms. COMA++ supports multiple ontology and schema formats such as OWL, XSD, and XML.

The MAFRA toolkit is a mapping framework for distributed ontologies which adopts an open architecture in which concept mappings are realized through semantic bridges. A semantic bridge is a module that transforms source ontology instances into target ontology instances manually defined. The MAFRA toolkit supports a graphical user interface that provides domain experts with functionalities that are needed for the specification of semantic bridges. In the MAFRA toolkit, the ontologies are represented as graphs and in particular cases as trees using the Touch Graph library (http://www.touchgraph.com).

Falcon-AO (Jian et al. 2005) is an automatic ontology alignment tool that uses linguistic and graph matching techniques. It attempts to align ontologies using linguistic similarity between two entities relying on their names, labels, comments and other descriptive information. Falcon-AO relies on a graph matcher, which measures the structural similarity between the graphs that represent the ontologies.

Clio (Hernández et al. 2001) is a graphical tool used for the semi-automatically mapping of relational and XML schemas. In contrast, our mapping tool is mainly intended to match ontologies and therefore supports the mapping of XML and OWL/RDFS ontologies represented in XML, RDFS, OWL, or N3 (Berners-Lee et al. 2005). Using Clio, the user loads a source schema and a target schema and establishes connections between objects in both schemas graphically. Such connections are referred to as value correspondences, which express how one or more objects in the source schema are transformed into a target value. Clio has a mapping engine that incrementally produces database (SQL) queries that realize the mappings implied by the correspondences. The AgreementMaker generates a document that shows the mappings between concepts and can be used in a variety of ways, including in generating database queries.

MapOnto (An et al. 2005), which is inspired by Clio, is a research prototype for mapping between a database schema and an ontology as well as between two different database schemas. MapOnto works in an interactive and semi-automatic manner, taking input from the user for creating simple attribute-to-attribute correspondence and allowing the user to select a set of logical formulas that can be used to establish correspondences between related attributes. These logical formulas are generated by the tool using knowledge embedded in the ontologies. These logical formulas are ordered to suggest to the user the most reasonable mapping between the two models. MapOnto supports a graphical interface, which is based on Protégé (Gennari et al. 2003). Unlike our tool, the correspondences between attributes are not represented by lines in the interface, but as logical formulas displayed in a separate pane.

Cruz et al. (2007) have proposed an integration framework, in geospatial domain, to facilitate access to the information that is contained in distributed and heterogeneous databases (Cruz et al. 2002). Their approach relies on the alignment of ontologies. When such mappings have been established, we say that the two ontologies are aligned or matched. They consider two different architectures: a centralized architecture and a peer-to-peer (P2P) architecture. In the former architecture, there is a global ontology. Each distributed ontology is aligned with the global ontology. As a consequence, a query expressed in terms of the concepts of the global ontology can be translated into a query to one of the distributed or local databases using the mappings that are established during the alignment process. In the latter architecture, a query to a source peer can be translated into a query to a target peer, provided that the ontologies of the two peers have been aligned. Whichever the architecture, querying can be easily extended to new databases, and therefore to new regions.

Nogueras-Iso et al. (2000) use URBISOC, a thesaurus focused on Spanish terminology for Town Planning, developed by the CINDOC/CSIC institute (Centre for Scientific Information and Documentation / Spanish National Research Council). The proposition is made through the IDEZar Project (collaboration agreement signed in March 2004 between the City Council and the University of Zaragoza). The Objective of the project is the development of a local SDI for Zaragoza, to facilitate, increase and coordinate the use of spatial data by the Council and to develop applications for the citizens and to provide them with access to public sector information. Formal Concept Analysis (FCA) approach is used (it enables the extraction of a hierarchy of concepts from the feature instances contained in the source repositories) and seems to be more flexible: it allows dynamic building of the ontology (at least, a draft), it doesn't need to define the concepts, it just need to observe the data that exists. They have also created a domain specific ontology that facilitates the interoperability (synchronization, update and merge) of the separate repositories.

#### 4.2.7.1 Overview of Viewpoints Confrontation Systems

For the purpose of confrontation, we defined the notion of opinion-viewpoint as opposed to the notion of viewpoint which is an emerging paradigm in Computer Science and especially in Information Systems Design (that is, a view angle on an entity). An opinion-viewpoint is a dynamic, non-consensual theory which is expressed on a domain for the purpose of sharing. It can be easily found in Sciences in the process of theory elaboration, and, to give an immediate example, this paper for instance is an opinion-viewpoint.Very few existing systems include confrontation of viewpoints in their functionalities. Indeed, allowing confrontation of viewpoints implies that the notion of viewpoint is well defined. Viewpoints-based systems, such as Bénel et al. (2001, 2002, 2006), Porphyry (2004) and Zaher et al. (2006), allow some form of confrontation. Porphyry especially includes a graph filtering system that shows, when several viewpoints are considered, which of them contain a given resource (Fig. 4.4).

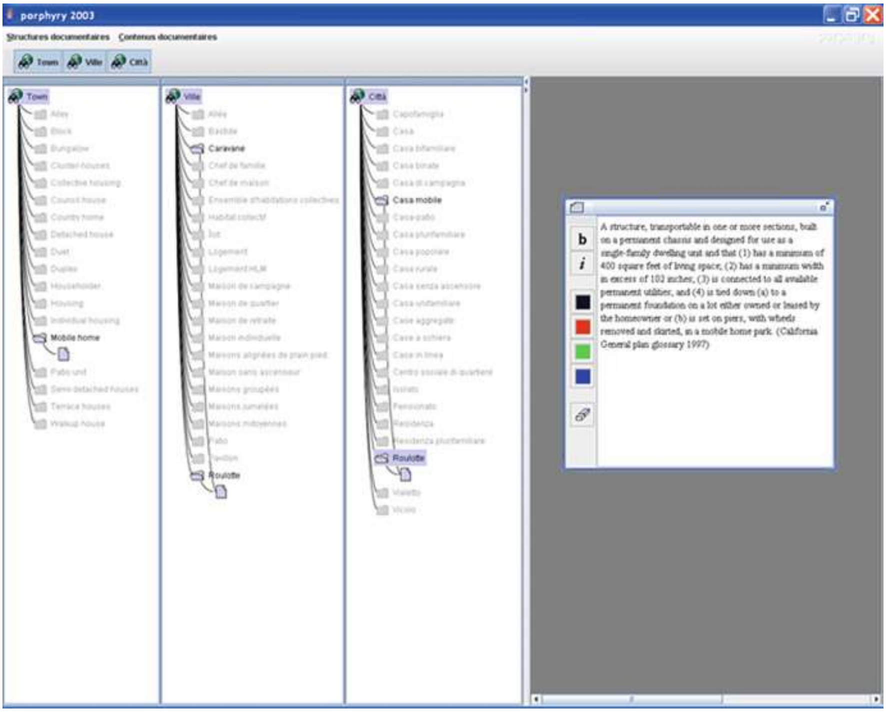


**Fig. 4.4** In this first test, the terms are "in flat" (no hyponyms or meronymes). The viewpoints are the three languages. The shared documentary entities are the definitions in one or another language [NB: Look at the terms appearing as synonyms in a same language]

The reason why confrontation is not included in the current approaches of collective work is that the same software is generally used by small communities that do not necessarily wish to share their research work (Ribière and Dieng 2002). Confrontation can only be envisaged either as an inner functionality that works within a single community, or as a general tool that works only on published work.

Generating inventiveness through confrontation is a difficult task in a digital environment. There are three issues that we had to deal with. First of all: confrontation by digital computation. At this level of expertise, when even terminology can vary from an expert to another, any algorithm is overtaken by the complexity of the semantics that is involved in the process. It is important to limit the bias introduced by inaccurate matching algorithms. Therefore, we planned the environment as being used by the experts themselves, and the algorithms as being mere tools to test on the subject of study_validating or refusing their result. The second issue comes right from the solution of the first one: if the experts are supposed to control the environment, it is important to build it such as they can use it without being very proficient with the computer. We have thus kept the GUI as simple as possible, limiting the number of options and merging all algorithms into three options: exhaustive search,

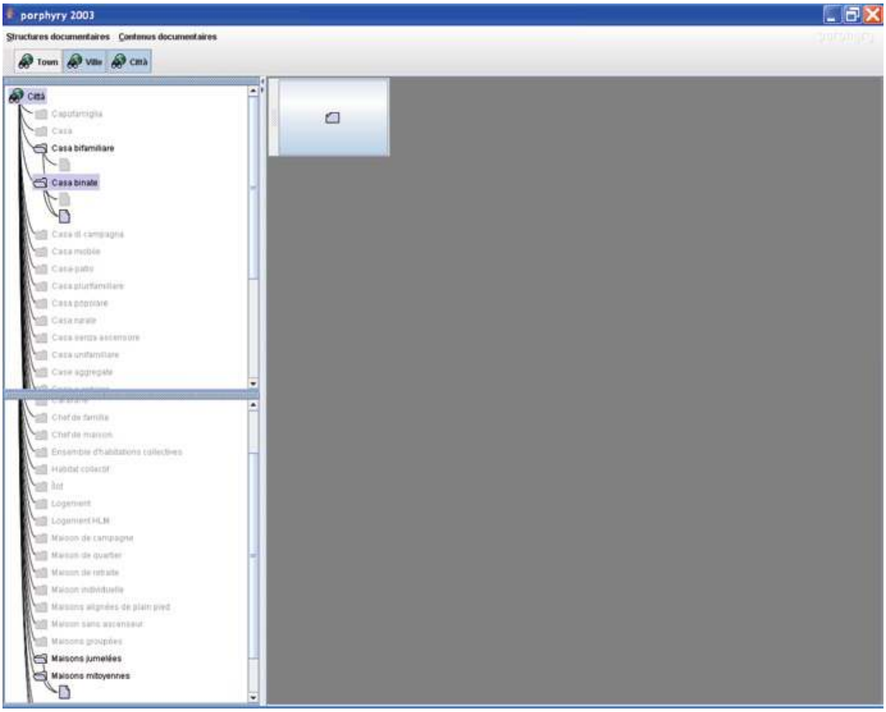


**Fig. 4.5** Revealing of a little bit complex overlapping of terms and definitions

quick search and immediate search (since the algorithms merely propose their result, *search* is the most adapted term for what they do from the expert's viewpoint).

The third issue is at the level of data and representation. Digital processing, especially when it comes to matching things, implies that some graph structure will be used. Expressing a urban viewpoint as a graph can also induce a bias. We do not express viewpoints ourselves (at the moment), so we use the solutions taken by whatever source we have for them. For instance, Porphyry uses a representation that puts little constraint in the formalism, arguing that when experts in humanities are involved, interpretation is more important that formalness.

In the following, we show how we have used Porphyry for modelling the glossary of urban, regional and environmental planning terms established in 2004 by Calderon and Ventura (Figs. 4.5 and 4.6).

#### 4.2.7.2 The Hyppodamos Tool

The goal of the Hyppodamos environment (Gesche et al. 2006, 2007; Gesche 2008; Berdier et al. 2008) is to allow an expert to confront and to align several ontologies on compatible subjects. We do not limit ourselves to a single formalism, thus we

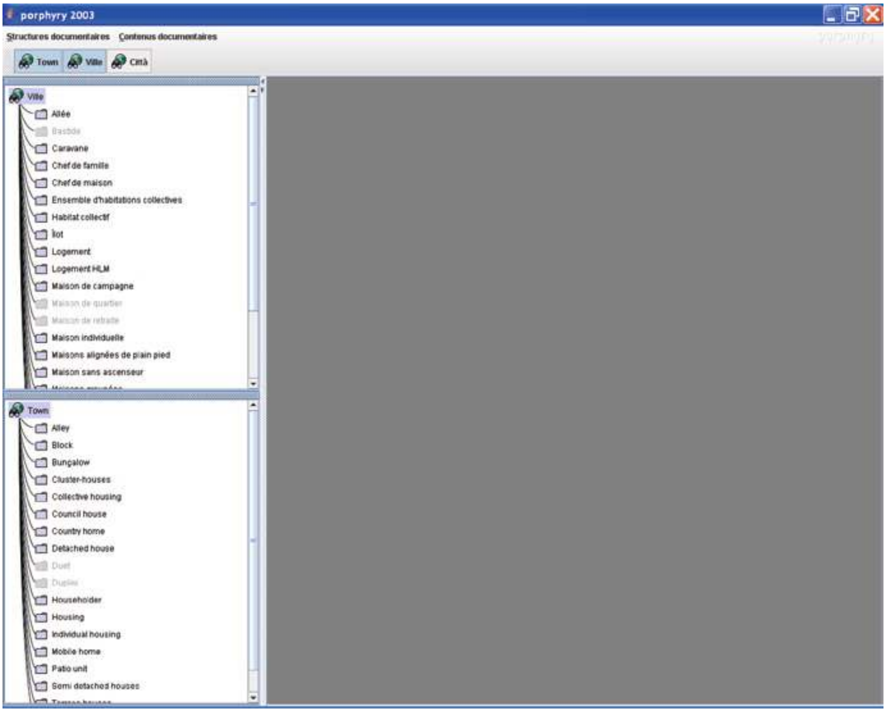


**Fig. 4.6** Revealing of untranslatable terms from french towards english and conversely

created a generic formalism to allow the import from the formalisms used by the software that already allow the expression of viewpoints (Fig. 4.7) (Towntology for instance). We are not limited to a single media either (for now, we use text and images that are the main means of knowledge representation), nor to a single language (however, the expert actually doing the confrontation must still be able to understand what he is working on).

The environment itself is organized according to the computer-aided paradigm: it provides a place where the viewpoints can be imported (a virtual desk), and a set of tools that allow either an automated or a manual processing on them. These tools can be used at any moment, in any state of the viewpoints (Fig. 4.8). Viewpoints are thus never overwritten, instead a save file is issued linking to them.

There are five actions that we expect an expert to use while confronting. Three of them have been included in the environment, and the remaining two have to do with building the graphs (which we assume is already done). They are:

- Observation: the main, even though almost passive activity of the expert is to watch the viewpoints, study them and observe the effects of the other actions on them.

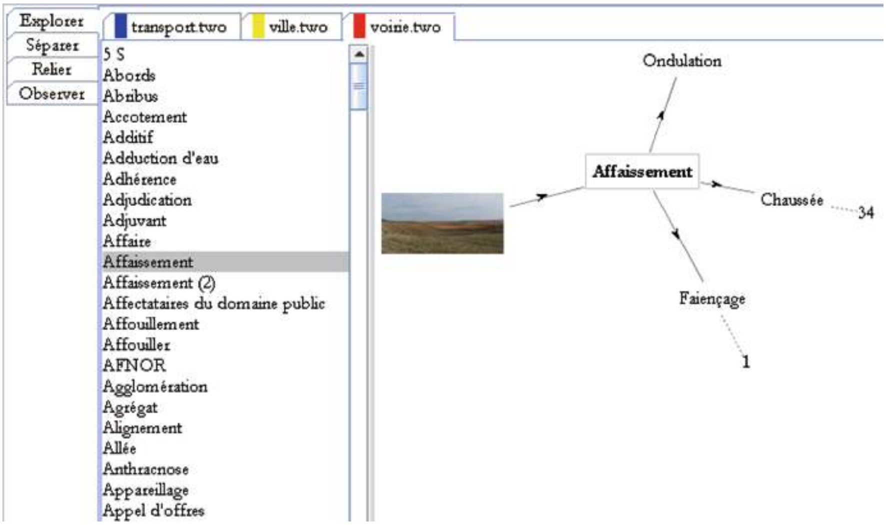


**Fig. 4.7** A viewpoint in urban domain

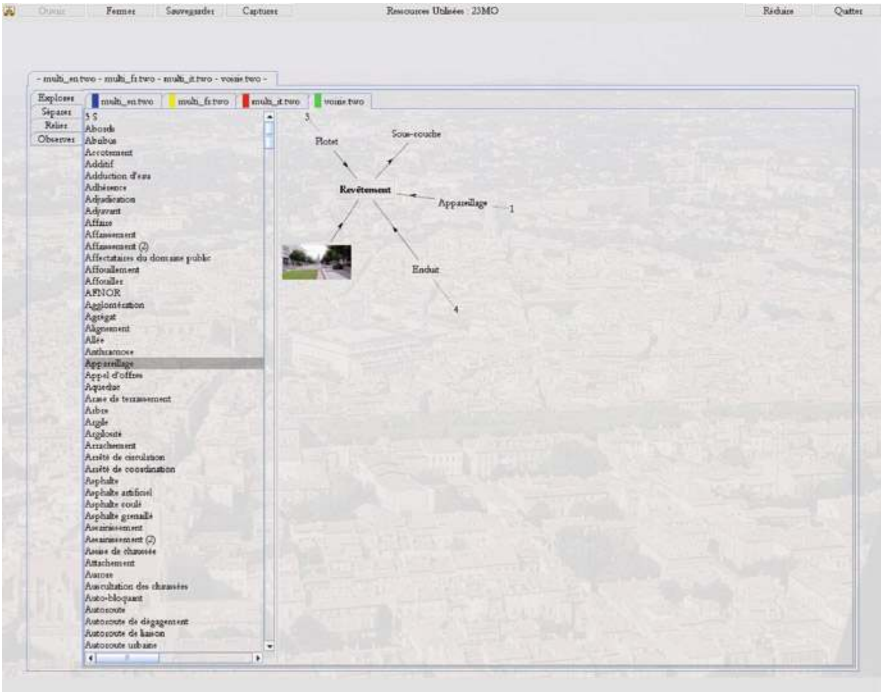


**Fig. 4.8** The Hippodamos tool

- Extraction and Organization: the actions involved in graph building. Interesting patterns have to be extracted from the raw viewpoints (for example a digital paper, or the mind of the expert) and they must be organized within a graph structure.
- Connection and Dissociation between patterns of different viewpoints.

Ontology matching only involves a single action, connection. Its aim is to find matches between patterns of matched ontologies, in order to allow interoperability most of the time. The algorithms we took from this domain have the same goal, finding any relation between the viewpoints. However, since we deal with an expert/machine partnership where the expert holds the power of decision, this task had to be split in two. Indeed:

1. When identical names are used for different meanings in several points of view, they must be dissociated (it was, for instance, the case of Thebes, a name of many cities in the antique world).
2. Whenever one of the matching algorithms points out that some terms could be connected, and it is not the case, it is also useful to explicitly dissociate them. These dissociations are not only correcting some error of a matching algorithm. Most of the time, they carry just as much sense as most connections. Among the experts using Porphyry especially, terminology can be as much a stake as diverging interpretations on a given subject.

### *4.2.8 Conclusion: Open Problems and Research Challenges*

As a conclusion we can mention some directions in which research on ontology alignment should evolve.

- Foundations: Available model-theoretic semantics are sufficiently similar, so they could eventually converge. Recent work on categorical characterisation of ontology matching raised some questions about the statement in categorical terms of expressive alignments, which go beyond equivalence. Therefore, interesting and useful work could be pursued in this direction (Euzenat and Shvaiko 2007).
- Representing alignments: In the long term, we expect progress on the framework for integrating different alignment systems. Infrastructures are now still missing. Such an infrastructure should match ontologies and process the alignment on specified data. Therefore, alignment formats and metadata become crucial. Furthermore, graphical alignment editors are needed. They should be easy to use for ordinary users.
- Explaining alignments: there are only a few matching systems that provide a justification of their results. Explanation is a challenge for ontology alignment as well as user interfaces.
- Processing alignments: Currently, many systems are rather monolithic and provide ontology alignment at once. In the future, we hope to see more modularisation and also more alignment processors to be developed.

# References

An, Y., Borgida, A., Mylopoulos, J.: Inferring complex semantic mappings between relational tables and ontologies from simple correspondences. In: OTM Confederated International Conferences (Part II). Lecture Notes in Computer Science, vol. 3761, pp. 1152–1169. Springer, Berlin (2005)

Aumueller, D., Do, H.H., Massmann, S., Rahm, E.: Schema and ontology matching with COMA++. In: ACM SIGMOD International Conference on Management of Data, pp. 906–908. ACM Press, New York (2005)

Bénel, A., Egyed-Zsigmond, E., Prié, Y., Calabretto, S., Mille, A.: Truth in the digital library: from ontological to hermeneutical systems. In: Proceedings of the Fifth European Conference on Research and Advanced Technology for Digital Libraries, Darmstadt, September 4–9, 2001. Lecture Notes in Computer Science, Vol 2163, pp. 366–377. Springer-Verlag, Berlin (2001)

Bénel, A., Calabretto, S., Iacovella, A., Pinon, J.M.: Porphyry 2001: semantics for scholarly publications retrieval. In: Proceedings of the Thirteenth International Symposium on Methodologies for Intelligent Systems, Lyon, June 26–29, 2002. Lecture Notes in Artificial Intelligence, Vol, 2366, pp. 351–361. Springer-Verlag, Berlin (2002)

Bénel, A., Iacovella, A., Calabretto, S.: Porphyry and Steatite: software layers for sense makers in humantities. In: Proceedings of the Workshop on Indexing and Knowledge in Human Sciences, Nantes, June 26–28, 2006, pp. 72–75 (2006)

Benerecetti, P., et al.: On the dimensions of context dependence: partially, approximation and perspective. In Proc. 3$^{rd}$ International and Interdisciplinarity Conference on Modeling and Using Context (CONTEXT), vol 2116 of LNCS, pp. 59–72, Dundee (UK) (2001)

Berdier, C., Calabretto, S., Caplat, G., Gesche, S.: Managing heterogeneity in urban ontologies. In: 3rd Workshop of COST Action C21: Construction of Multilingual Ontologies for Urban Civil Engineering Projects, Octobre 2008, Zaragoza, Espagne (2008)

Berners-Lee, T., Connolly, D., Prud'homeaux, E., Scharf, Y.: Experience with N3 rules. In: W3C Workshop on Rule Languages for Interoperability, Washington, D.C (2005)

Bouquet, P. et al.: D2.2.1 Specification of a common framework for characterizing alignment, knowledge Web (FP6-507482), http://knowledgeweb.semanticweb.org/semanticportal/deliverablesD2.2.1v1.pdf (2004)

Cruz, I.F., Rajendran, A., Sunna, W., Wiegand, N.: Handling semantic heterogeneities agreements. In: Tenth International ACM GIS Symposium, pp. 168–174. ACM, New York (2002)

Cruz, I.F., Sunna, W., Makar, N., Bathala, S.: A visual tool for ontology alignment to enable geospatial interoperability. J. Vis. Lang.Comput. **18**(3), 230–254 (2007)

Euzenat, J.: Towards a principled approach to semantic interoperability. In: Proc. IJCAI Workshop Ontologies and Information Sharing, Seatle (WA US), pp. 19–25, (2001)

Euzenat, J., Shvaiko, P.: Ontology Matching. Springer, Heidelberg (2007)

Fonseca, F.T., Egenhofer, M.J., Davis, C.A., Câmara, G.: Semantic granularity in ontology-driven geographic information systems. Ann. Math. Artif. Intell. **36**(1–2), 121–151 (2002)

Gennari, J.H., Musen, M.A., Fergerson, R.W., Grosso, W.E., Crubézy, M., Eriksson, H., Noy, N.F., Tu, S.W.: The evolution of protégé: an environment for knowledge-based systems development. Int. J. Hum. Comput. Stud. **58**(1), 89–123 (2003)

Gesche, S.: Exploitation de l'hétérogénéité entre points de vue-opinion. In: Actes du XXVème congrès INFORSID, INFORSID'2008, Fontainebleau, France, 27–30 mai 2008, pp. 101–116 (2008)

Gesche, S., Calabretto, S., Caplat, G.: Un modèle pour la Confrontation d'opinions numérisées sous Porphyry. In: Colloque International sur le Document Electronique, CIDE'2006, Fribourg, Suisse, 18–22 septembre 2006, pp. 253–267 (2006)

Gesche, S., Caplat, G., Calabretto, S.: Managing difference of opinion in semantic structures. In: International Workshop On Semantically Aware Document Processing and Indexing held in cooperation with ACM SIGWEB, SADPI'07, Montpellier, France, may 21–22, 2007, pp. 79–86 (2007)

Hernández, M.A., Miller, R.J., Haas, L.M.: Clio: a semi-automatic tool for schema mapping. In: ACM SIGMOD International Conference on Management of Data, p. 607 (2001)
Jian, N., Hu, W., Cheng, G., Qu, Y., Falcon, A.O.: Aligning ontologies with Falcon. In: K-CAP 2005 Workshop on Integrating Ontologies, CEUR Workshop Proceedings 156, Banff, Canada (2005)
Klein, M.: Combining and relating ontologies: an analysis of problems and solutions. In: Proc. IJCAI Workshop Ontologies and Information Sharing, Seatle (WA US)(2001)
McGuinness, D.L., Fikes, R., Rice, J., Wilder, S.: An environment for merging and testing large ontologies. In: Seventh International Conference on Principles of Knowledge Representation and Reasoning (KR 2000), Colorado, pp. 483–493 (2000)
Nogueras-Iso, J., López, F.J., Lacasta, J., Zarazaga-Soria, F.J., Muro-Medrano, P.R.: Building an address gazetteer on top of an urban network ontology. 1st Workshop of COST Action C21: "Ontologies for Urban Development: Interfacing Urban Information Systems", Geneva, 6–7 November 2006 (2006)
Porphyry.: Porphyry Project. http://www.porphyry.org (2004). Accessed 4 Feb 2010
Rahm, E., Bernstein, P.A.: A survey of approaches to automatic schema matching. VLDB J. **10** (4), 334–350 (2001)
Ribière, M., Dieng, R.: A viewpoint model for cooperative building of an ontology. 10th International Conference in Conceptual Structures. LNAI, vol. 2393, pp. 220–234. Springer, Heidelberg (2002)
Zaher, H., Cahier, J.P., Zacklad, M.: The Agoræ / Hypertopic approach. In: Harzallah, M., Charlet, J., Aussenac-Gilles, N. (eds.) Proceedings of Indexing and Knowledge in Human Sciences, Nantes (2006)