Jacques Teller
Anne-Françoise Cutting-Decelle
Roland Billen (eds.)

Future of Urban Ontologies

Proceedings of the Final Conference of the COST Action C21 – Towntology: Urban Ontologies for an Improved Communication in Urban Development Projects
Liège, 9&10 March 2009.

Jacques TELLER
LEMA - Universite de Liege
1 chemin des Chevreuils B52
4000 Liege
Belgium
Jacques.Teller@ulg.ac.be

Anne-Françoise CUTTING-DECELLE
Ecole Centrale de Lille
Laboratoire de Génie Industriel (LGIL)
Cité Scientifique, BP 48
59651 VILLENEUVE D'ASCQ CEDEX
France
afcd@ec-lille.fr

Roland BILLEN
Unité de géomatique
Bât. B5 Géométrologie et topographie
allée du 6 Août 17
4000 Liège 1
Belgique
rbillen@ulg.ac.be

Foreword

COST- the acronym for European COoperation in Science and Technology - is the oldest and widest European intergovernmental network for cooperation in research. Established by the Ministerial Conference in November 1971, COST is presently used by the scientific communities of 35 European countries to cooperate in common research projects supported by national funds.
The funds provided by COST – less than 1% of the total value of the projects – support the COST cooperation networks (COST Actions) through which, with EUR 30 million per year, more than 30 000 European scientists are involved in research having a total value which exceeds EUR 2 billion per year. This is the financial worth of the European added value which COST achieves.
A "bottom up approach" (the initiative of launching a COST Action comes from the European scientists themselves), "à la carte participation" (only countries interested in the Action participate), "equality of access" (participation is open also to the scientific communities of countries not belonging to the European Union) and "flexible structure" (easy implementation and light management of the research initiatives) are the main characteristics of COST.
As precursor of advanced multidisciplinary research COST has a very important role for the realisation of the European Research Area (ERA) anticipating and complementing the activities of the Framework Programmes, constituting a "bridge" towards the scientific communities of emerging countries, increasing the mobility of researchers across Europe and fostering the establishment of "Networks of Excellence" in many key scientific domains such as: Biomedicine and Molecular Biosciences; Food and Agriculture; Forests, their Products and Services; Materials, Physical and Nanosciences; Chemistry and Molecular Sciences and Technologies; Earth System Science and Environmental Management; Information and Communication Technologies; Transport and Urban Development; Individuals, Societies, Cultures and Health. It covers basic and more applied research and also addresses issues of pre-normative nature or of societal importance.

Website: http://www.cost.esf.org

This publication is supported by COST
© COST Office, 2008

No permission to reproduce or use the contents of this book by any means is necessary, other than in the case of images, diagrams or other material from other copyright holders. In such cases, permission of the copyright holders is required. This book may be cited as: COST Action C21 – Future of Urban Ontologies.
Neither the COST Office nor any person acting on its behalf is responsible for the use which might be made of the information contained in this publication. The COST Office is not responsible for the external websites referred to in this publication.

Table of content

Adaptative urban ontologies

User Evaluation of a Software Interface for Communication in Urban Ontologies

Hyowon Ban[1] and Ola Ahlqvist[1]

[1]Department of Geography, The Ohio State University,
1036 Derby Hall, 154 N Oval Mall
Columbus, OH 43210, U.S.A.
{ban.11, ahlqvist.1}@osu.edu

Abstract. The study aims to evaluate a software interface that is developed for interactive representation and negotiation of spatial concepts of urban ontology terms such as periurbanisation to help communication and knowledge sharing between users. We conduct two types of user-evaluation; closed interviews focusing on general aspects of the interface, and in-depth interviews to deal with boundary issues of urban environment between stakeholders. From the result, we find that the software interface could be useful for urban ontology research and land use, with support of different ontologies of different actors for a common concept.

Keywords: user evaluation, software interface, communication, periurban

1 Introduction

Recently diverse actors including government, stakeholders, and citizens are participating in urban development for better understanding of local social context and environment (Nolmark 2007). Information that different actors would bring into urban planning and design activities could include a wide spectrum from local practical knowledge to general urban theory. The term 'ontology' means a formal, explicit specification of a common understanding of some domain that can be communicated across people and computers (Studer et al. 1998). As public participation in urban decision-making processes are increasing, communication, negotiation, and argumentation between the actors become increasingly important (Teller 2007).

Genesereth and Nilsson (1987) had argued that one of the main goals of applications for ontologies in information sciences is the ability of knowledge sharing and reuse. There have been efforts to improve communication and knowledge-sharing of urban ontologies using developed tools. Some examples tools aim to support representation of conceptual/spatial information of urban ontologies and knowledge in ontology-based models (Ban and Ahlqvist 2007b; Cataldo and Rinaldi 2007). Other tools simulate multiple urban-design scenarios to formally represent both implicit and explicit knowledge of ontologies (Caneparo et al. 2007). In addition, a tool for Geo

Semantic Web community was developed to support annotation of common interest between multiple users and browsing of information (Marcheggiani et al. 2007).

However, it would be hard to tell how such tools would be useful for dealing with issues of urban ontology in practice unless it is evaluated. For example, Metral et al. (2007) argue that a tool should provide relevant information for each actor that is easily understood according to different user profiles so that they could support "actor-specific ontologies", stressing the importance of evaluation of the tool and its interface interface with different urban actors. One reason for this need is the different "paradigms of visualization" that could vary in their usefulness for specific actors, such as people in different age groups (Marcheggiani et al. 2007). In addition, feedbacks from the diverse actors could be helpful to reveal current limitations of a tool and provide better ideas for its improvement.

This paper aims to evaluate an existing tool for communication and knowledge sharing between different actors in urban planning/design activities by interviewing actors from different user groups—i.e., policy makers, researchers, and residents. Another goal of this study is to see how different types of user-evaluation methods would contribute to evaluate the tool, ultimately to understand how urban ontologies in practice work in the domain of urban development.

Our research on urban ontologies has gone through three research phases. In the first phase, we revealed spatial implications of uncertainty in urban-ontologies terms such as periurbanisation using a fuzzy-set approach and geovisualisation techniques (Ban and Ahlqvist 2007a). The second phase of our study suggested a web-based software interface 'pinu' (program for identification and negotiation of uncertain concept definitions) developed for interactive visualization, communication, and negotiation of the periurban concept with empirical spatial data (Ban and Ahlqvist 2007b). In this paper we report on our findings from the third phase user evaluations of pinu's interface conducted through pilot interviews followed by in-depth interviews using a semi-structured format.

2 Theoretic Background

2.1 Uncertainty of Concepts in Urban Ontology between Actors

Because actors in urban management and policy making are from diverse social groups, have different roles, and come from individual backgrounds, there are potential issues of communication or knowledge sharing. Some of the issues may originate from uncertain aspects of urban ontologies. For instance, the concept of periurban area is sometimes represented by other terms such as "urban sprawl", "exurban area", or "urban-rural fringe" (Weaver and Lawton 2001). However, even when dealing with a same terminology for a concept, there may be multiple definitions of the concept (Ban and Ahlqvist 2007a; 2007b). Due to the reasons mentioned above, the inherent uncertainties in urban ontologies may lead to difficulties in communication between the actors and this could generate inefficiency and confusion in urban management. Especially, in European countries there are important issues such as multilinguality that may affect development of e.g. a European Spatial Data Infrastructure (Nowak et al. 2005). It is therefore essential to have a clear and transparent information exchange between the actors in urban decision-making. Some problems related to these or other types of uncertainty issues could be identified and possibly resolved by evaluating the tools with participation of the actors in practice.

2.2 Participatory GIS (PGIS) Approach

PGIS aims to promote public participation in policy-making processes by engaging individual citizens, nongovernmental organizations, grassroots groups, and community-based organizations using GIS, such as a spatial decision support system or a web-based municipal GIS (Sieber 2006). GIS's ability to explore spatial data and its visual representations is important in PGIS approaches because some grassroots organizations may use GIS more for "cartographic spatial narratives" than spatial analysis (Elwood 2006).

As PGIS are increasingly being adopted to urban research, qualitative PGIS brings quantitative GIS-methods and qualitative data together to interpret different forms of spatial knowledge from diverse social groups. However, even a PGIS approach may be criticized because the way used ontologies might include particular socio-political information and embed institutional factors in the spatial data and this in turn might generate uneven access to the GIS (Elwood 2006). When developing software for PGIS such as pinu, evaluating the effectiveness is therefore fundamental to a successful integration of societal goals with local context and objectivity (Sieber 2006; Haklay and Tobon 2003).

3 User Evaluation Design

Having an existing prototype application, 'pinu' (Ban and Ahlqvist 2007b), we developed a pilot-type user testing of pinu in order to see whether its interface supports the intended concept comparison/creation user activities, and to find existing limitations in pinu that could be further developed. The pilot interviews were focused on evaluating overall interface design and provide feedback for potential extensions of the software functionality. After this first round of interviews and revisions we conducted additional in-depth interviews for the purpose of evaluating the revised pinu interface illustrated in Figure 2. These in-depth interviews were focused not only on the interface design and usefulness, but also emphasized how each actor recognized periurban areas and how they felt about issues or conflicts related to the periurban boundary.

3.1 Designing the Pilot Interviews

The subjects of the pilot interviews consisted of seven graduate students in the Department of Geography, the Ohio State University, U.S.A. The students were recruited to represent different specializations in geography and to have varying computer/online-map literacy.

In the survey, we showed a cartoon developed for the interviews to inform the interviewees about an example of periurban boundary issues in practice. Then we gave each interviewee a brief presentation about the purpose of the study and basic concepts necessary for the interview. This was followed by a demonstration of the tool with an example scenario (Figure 1). Lastly, after the participants had experienced pinu, we asked the interview questions. The questions include evaluation about the readability, visual design, and usability of components in the interface (see examples in Table 1) that they had looked at and manipulated. The questions were formulated as closed questions in which interviewees are provided with alternative answers such as Yes/No (Parfitt 1997), however the participants could also elaborate freely on the questions during the interview.

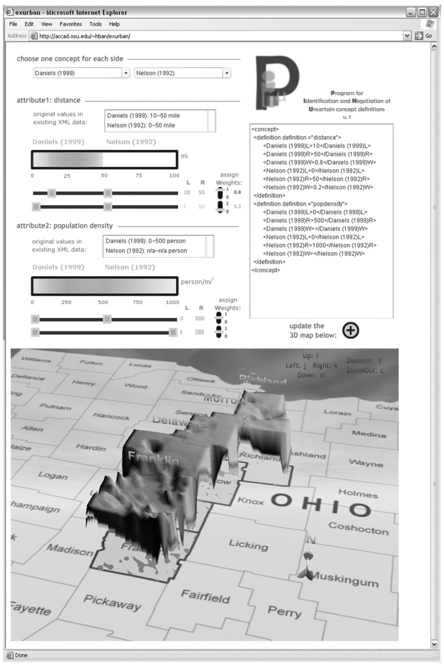


Figure 1. pinu used in the pilot-interviews (revised version of Ban and Ahlqvist 2007b, Fig 1)

Tab. 1. Example of closed questions designed for the pilot interviews

Are the sliders that define the weights of the attributes easy to use to assign the weights? ( ) No Yes a. b. c. d. e. f. Not sure Is pinu easy to use and understand the uncertain concept of exurban boundary? ( ) No Yes a. b. c. d. e. f. Not sure If not, could you tell me the reason?

3.2 Designing the In-depth Interviews

After the pilot evaluation, the pinu interface went through some revisions based on the user feedback (Figure 2). The in-depth interviews were then designed to focus not only on the tool's interface design and usefulness, but also emphasize how each actor recognize periurban areas and what they feel about issues or conflicts related to the periurban boundary. We conducted the in-depth interviews with five actors consisting of policy maker, resident, and researcher in Delaware County, Ohio, U.S.A. The interviews consisted of both closed and open-ended questions, allowing participants to express their opinion on matters that related to the use of this software (see examples in Table 2). The open-ended questions do not force respondents into particular answers (Parfitt 1997). The questions were designed to gain insights about software such as 1) what the actors find from the software by using it and 2) how they think that the software could be useful to deal with periurban boundary issues in urban development and management. During the interviews, the same procedures of the pilot interviews were used. This time, all responses were voice-recorded under consent of each interviewee when the open-ended questions were asked. On the other hand, in the pilot interviews we recorded all responses onto answering forms that were developed for the interviews.

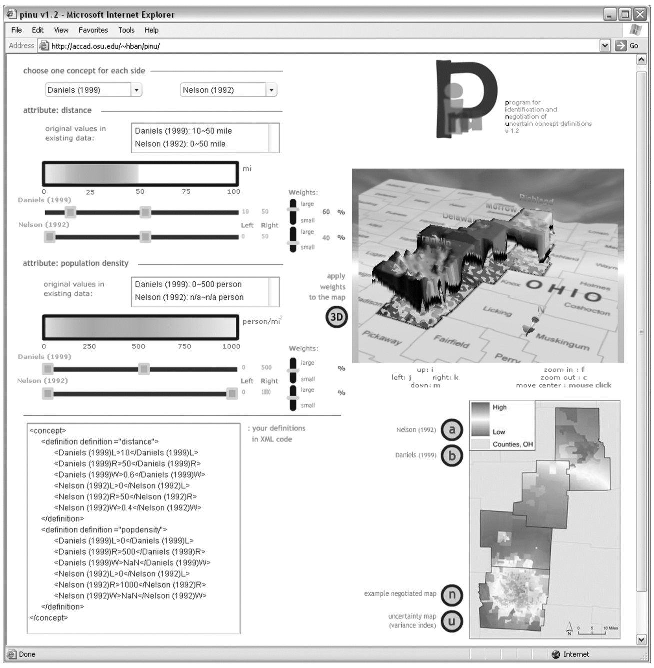


Figure 2. pinu's interface used in the in-depth interviews (revised version of Figure 1)

In Figure2, the interface is revised from previous version of pinu (Figure 1) based on the responses from the pilot interviews. For example, design of some components was relocated and/or resized, the 3D map had classified color schemes, and 2D maps were included for the user's reference and convenience of use.

Tab. 2. Example of open questions designed for the in-depth interviews

How would you describe an exurban area? Do you know or have you heard about any conflicts associated with the presence of these different types of places—such as areas more urban or more rural—in Delaware County? Do you think something like pinu would be useful for land-use policy making? Why or why not? (follow-up question if positive answer) How do you think you could use a tool like pinu in your work related to land-use policy making?

4 Results

4.1 Pilot Interviews with General Actors

As general remarks on the interface design, the participants recommended changes for the size, position, colors, and transparencies of the rectangular bars, sliders, and texts for better readability and manipulation. The interviewees also pointed out a need for multiple color schemes on the 3D object in the map (Figure 1), more landmark objects in the study area, two types of views of the spatial data in 2D and 3D. The participants also expressed an interest to explore and negotiate multiple periurban definitions to build their own understanding with the empirical data.

4.2 In-depth Interviews with Focused Actors

The in-depth interviews provided much feedback on the revised interface and their thinking about the periurban concept. All participants seemed to agree that the concept is uncertain and difficult to define. For instance, actors in the resident group were not aware of the term ‘periurban area’ or its exact meaning. Due to this limitation of knowledge, they tried to deal with the periurban boundary issues based on their experiences and knowledge of their home environment.

Asked about what they recognize from visualization of uncertain periurban boundary provided by pinu, all five interviewees responded that they realized the uncertainty. For questions about pinu’s usefulness for communication and knowledge sharing, most of them responded positively and mentioned that they would use a tool like pinu. Related to this, all of them suggested they would prefer if the tool could include more concepts with definitions since additional definitions may provide a better sense of the periurban boundary through empirical examples.

For enhancement of the interface, most of the interviewees suggested changes in some part of the interface such as size of the sliders and texts would for easier use. Specifically, the resident-interviewees suggested to develop another tool that could support representation and negotiation for planning and design of their home environment with other actors such as their neighbors. The researcher-interviewees recommended to expand the representation of periurbanisation by including more dimensions such as temporal change of periurban boundary. The policy maker responded that pinu’s visualization and negotiation functionalities would be useful for land-use management with multiple stakeholders. The interviewee also suggested the potential to use pinu to deal with other issues of urban development that local agencies and actors need to deal with.

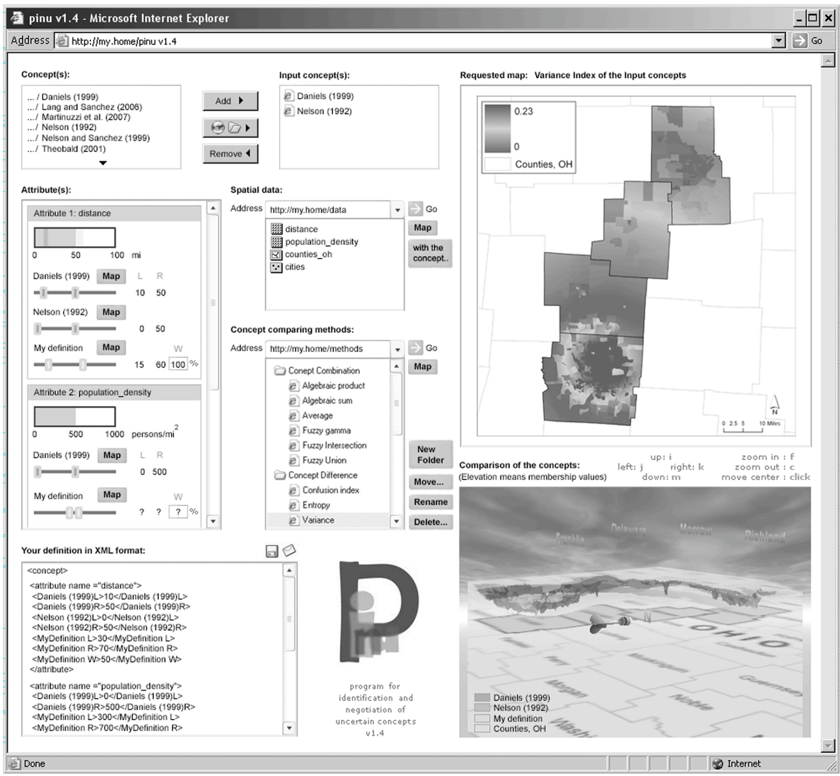


Figure 3.Example of future version of pinu's interface (Ban and Ahlqvist, in press)

The responses from in-depth interviews also indicated that even actors in the same social group may have issues of communication regarding urban ontologies. For instance, both of the two resident interviewees responded that people do not move to periurban areas only for their interests such as financial advantage they may earn in the periurban areas. At the same time, one of the resident interviewees used some characteristics of periurban area to describe suburban area.

Based on the responses from the pilot interviews, the interface revised in Figure 3 supports access to remote concept-definition resources. The re-developed interface thus includes more interactivity, enhanced usability, and interoperability. In addition, the user could create a new concept definition and visualize that together with existing definitions and empirical spatial data using the interface in Figure 3.

5 Concluding Discussion

This paper provided user-evaluations of a software interface developed for representation, communication, and negotiation of urban ontologies. A series of revisions of the software-interface based on the evaluations was shown in Figures 1, 2, and 3. This study also demonstrated the following results: first, the evaluated software interface, pinu, is potentially useful for participation, communication, knowledge sharing, and conflict resolution between actors in the domain of urban planning. According to the results of the evaluation, interactive-visualization functionality along with the user's manipulation was one of the strengths of the software interface. In future work the tool could be extended to include broader functionality such as Parallel Coordinate Plots (Inselberg 1985) that are useful for visualization of relationship(s) between records in the data. Figure 3 suggests another possibility of future developments of pinu's interface that supports original-concept creation by users, multiple choices of concept and data, and complementary use of 2-dimensional and 3-dimensional geovisualization (Ban and Ahlqvist, in press). To deal with qualitative data related to urban environment, analysis of multiple sources of data such as images of a certain place or movies showing narratives by actors for such a place would be useful for richer local information (Kwan and Ding 2008). Other functionalities such as metadata management by the users in Semantic Wiki could be considered (Krötzsch et al. 2005; Auer et al. 2006; Hepp et al. 2006).

In terms of methodology, we found that the two types of interviews conducted in this study were useful for specific questions being asked. For instance, the pilot interviews with closed questions were useful for measuring qualitative responses using multiple, but continuous choices (Table 1). When the interviewees had to answer some qualitative questions—such as "Is pinu easy to use and understand the uncertain concept of exurban boundary?"—they responded that the multiple choices showing the corresponding, continuous degrees between Yes and No in a continuous grayscale between dark gray and light gray provided them some guidelines to select an answer. On the other hand, the open-ended questions in the in-depth interviews were helpful to capture elastic flow of thinking and responding of the interviewees that was not frequently following simple linear order of answering for the questions. For example, some interviewees provided answers for a question asked as well as questions that will be asked later. Or, some following questions that were unplanned before the interview shed light on some responses for other questions.

For even more insight, more actors from various social groups such as land developers from private sectors in urban planning could be interviewed, and diverse focus-group meetings could be conducted for the evaluation of the tool. Clearly, user evaluations are an important instrument to provide for better design and our evaluations, albeit small and partial, provided many informative answers and helped improve our work.

References

1. Auer, S., Dietzold, S., Riechert, T.: The Semantic Web - ISWC 2006, 5th International Semantic Web Conference, ISWC (2006) http://www.informatik.uni-leipzig.de/~auer/publication/ontowiki.pdf
2. Ban, H., Ahlqvist, O.: Representing and Negotiating Uncertain Geospatial Concepts - Where are the exurban areas?. Computers, Environment and Urban Systems (in press), doi:10.1016/j.compenvurbsys.2008.10.001
3. Ban, H., Ahlqvist, O.: Visualizing the Uncertainty of Urban Ontology Terms. Studies in Computational Intelligence: Ontologies for Urban Development. 61 (2007a) 85-94
4. Ban, H., Ahlqvist, O.: Web-based Interactive Visualization of Uncertain Concepts in Urban Ontologies. In: Proceedings of the 2nd COST Action C21 conference, "Urban Ontologies for an improved communication in urban civil engineering projects", Castello del Valentino, Torino, Italy (2007b) 48-59
5. Caneparo, L., Collo, M., di Giannantonio, D., Lombardo, V., Montuori, A., Pensa, S.: Generating Urban Morphologies from Ontologies. In: Proceedings of the 2nd COST Action C21 conference, "Urban Ontologies for an improved communication in urban civil engineering projects", Castello del Valentino, Torino, Italy (2007) 162-177
6. Cataldo, A., Rinaldi, A.: Using an Ontology-based Model for Knowledge Representation in Rural Landscape. In: Proceedings of the 2nd COST Action C21 conference, "Urban Ontologies for an improved communication in urban civil engineering projects", Castello del Valentino, Torino, Italy (2007) 125-139
7. Elwood, S.: Critical issues in participating GIS: Deconstruction, reconstructions, and new research directions. Transactions in GIS. 10(5) (2006) 693-708
8. Genesereth M.R., Nilsson N.J.: Logical foundations of artificial intelligence. M.Kaufmann Pub., Los Altos (CA) (1987)
9. Haklay, M., Tobon, C.: Usability evaluation and PPGIS: towards a user-centred design approach. International Journal of Geographical Information Science. 17(6) (2003) 577–592
10. Hepp, M., Bachlechner, D., Siorpaes, K.: OntoWiki: community-driven ontology engineering and ontology usage based on Wikis. International Symposium on Wikis. In: Proceedings of the 2006 international symposium on Wikis, Odense, Denmark (2006) 143 – 144
11. Inselberg, A.: The plane with parallel coordinates. Visual Comput. 1 (1985) 69–97
12. Krötzsch, M., Vr, D., Völkel, M.: Wikipedia and the Semantic Web - The Missing Links. Proceedings of Wikimania 2005 (2005) http://www.aifb.uni-karlsruhe.de/WBS/mak/pub/wikimania.pdf
13. Kwan, M.-P., Ding, G.: Geo-Narrative: Extending Geographic Information Systems for Narrative Analysis in Qualitative and Mixed-Method Research. The Professional Geographer, 60(4) (2008) 443-465
14. Marcheggiani, E., Nucci, M., Tummarello, G., Christian, M. Geo Semantic Web Communities for Rational Use of Landscape Resources. Paper presented at the 2nd Workshop of COST Action C21 Towntology: Urban Ontologies for an Improved Communication in Urban Civil Engineering Projects, Ontologies for Urban Development: Conceptual Models for Practitioners. Turin, Italy (2007) 100-113
15. Métral, C.,Falquet, G., Vonlanthen, M.: An Ontology-based Model for Urban Planning Communication. Studies in Computational Intelligence (SCI). 61 (2007) 61–72
16. Nolmark, H.: COST Action C20 Urban Knowledge Arena: Cross-boundary Knowledge and Know-how on Complex Urban Problems. Studies in Computational Intelligence (SCI). 61, 15–25 (2007)

17. Nowak, J., Peedell, S., Nogueras-Iso, J.: Issues Of Multilinguality In Creating A European SDI – The Perspective For Spatial Data Interoperability. In: Proceedings of the 11th EC GI & GIS Workshop, ESDI Setting the Framework. Alghero, Sardinia, Italia (2005)
18. Parfitt, J.: Questionnaire design and sampling. In: Flowerdew, R.; Martin, D., eds. Methods in human geography: a guide for students doing a research project. Addison Wesley Longman Limited, Essex, England. (1997) 76–109
19. Seale, C. F.: Computer assisted analysis of qualitative data. In Handbook of interview research, edited by J. F. Gubrium and J. A. Holstein, 651–70. Thousand Oaks, CA: Sage (2002)
20. Sieber, R.: Public Participation Geographic Information Systems: A literature review and framework. Annals of the Association of American Geographers. 96(3) (2006) 491–507
21. Studer, R., Benjamins, V. R., Fensel, D.: Knowledge Engineering: Principles and Methods. In: *Data and Knowledge Engineering*. 25(1-2) (1998) 161-197
22. Teller, J. Ontologies for an Improved Communication in Urban Development Projects, Studies in Computational Intelligence (SCI). 61 (2007), 1–14
23. Weaver, D., Lawton, L.: Resident Perceptions in the Urban–Rural Fringe. Annals of Tourism Research. 28(2) (2001), 439-458

Updating Plans: A Historiography of Decisions over Time

Nikhil Kaza[1], Donovan Finn[2], Lewis D. Hopkins[2]

[1] University of Maryland at College Park
nkaza@umd.edu

[2] University of Illinois at Urbana Champaign
{dfinn, ldhopkin}@illinois.edu

Abstract. Plans provide information about how multiple decisions are structured over time, and what the intentions of a particular actor are. As and when these decisions get made, or not made, some parts of plans become irrelevant, while some other new relationships are discovered and considered. Recognising this provides a useful way to interpret the changes in plans by tracking the decisions and vice versa. In this paper, we illustrate the complexities of an ontology of urban systems which are needed to ensure the currency of plans, so that they could be effectively used in urban decision making. Especially, when actors are numerous, jurisdictions overlap, actions are interdependent and interests are unstable, this framework enables us to think about plans in a complex and changing urban environment and make them so that they remain useful.

1 Many Plans over Time and of Many Actors

Plans are useful because they consider interdependent sets of decisions ahead of time. These decisions taken at various scales and various times by multiple actors shape the aggregate pattern of cities both geographically and socially (1). Decisions about infrastructure investments or regulatory changes in urban settings are made by groups of actors who are partially cognisant of the intentions of others both within and outside the group. Thus, some of these decisions are informed by multiple plans of multiple actors. However, in making these decisions over time, the plans that inform them also are modified, thus fundamentally altering the relationships between current decisions and those that come after them. Furthermore, the decisions that are made represent a choice among the alternatives delineated within the plans, thus rendering some portions of the plan obsolete. Viewing plans as information that is useful in making decisions demands that the information within them be current and relevant. In this paper, we explore ontological approaches to keeping track of decision histories of various actors over time thus simultaneously modifying and updating plans that inform them.

The paper uses ongoing examination of planning, plan making and use of plans in Champaign, IL to illustrate how tracking and accessing the changing information in plans over time and among actors can help understand how urban development decisions are made. Specifically, we will use plans relating to downtown revitalisation, campus planning and community safety to illustrate how plans can and should consider the inherent interdependence of relationships in decisions and policy making within the urban landscape. In doing so, we look for information in plans that are situated *backward* and *forward* in time and *outward* in space and function with respect to other plans and those of other actors, by using both stylised and real cases. We extend previous work to provide illustrations of how an ontology of decisions which accounts for relationships among decisions, actions, and intentions helps us in maintaining the currency of plans.

In subsequent sections, we elaborate on the contingency relationships of decisions within a plan (typified by strategy) and interdependence relationships (typified by design). We then derive the temporal and other orderings of decision chains from these strategies and designs. By observing the decisions that are made and their situatedness in these plans, we archive some portions of the plan material as irrelevant to future decision making. This allows us to explore further relationships between decisions and their outcomes, which have not been envisaged in the previous versions of the plans. We then also touch upon situations in which a subsequent plan completely overrides a previous plans due to fundamental changes in the perceived importance of the relationships and other values which are inherently political. In doing so, we provide yet another justification for developing an ontology for plans in urban situations that is cognisant of the complexities of multiplicities of actors, intentions, goals, decisions and interactions.

2 Decisions and Plans

Plans provide guidance to decisions that are to be taken in different circumstances (2). Different actors have many plans, which change over time. Each plan considers interdependencies and contingencies among decisions by different actors at the particular time point and provide useful indicators about commitment. Decisions themselves are types of actions (3), made by an intentional actor prior to action. In this paper, we conflate actions and public decisions.

One of the key claims of (4) is that information in plans is organised according to relationships between sets of decisions and actions. Two such kinds of relationships are particularly important to us: Strategy and Design. Strategy deals with uncertainty of actions and outcomes, whereas designs deal with interdependence and complementarity between actions. Such structuring of relationships between decisions helps us understand if and how plans are being used when the decisions get made. In subsequent sections, we elaborate on these relationships and use them to identify how decisions that are based on them could be used to update plans.

2.1 Strategy

The most complicated and useful exercise of planning is to recognise uncertainty of outcomes of actions and plan strategically with respect to goals and criteria. Thus in a particular plan, strategies specify to sufficient depth (figure 1), possible outcomes as well as possible actions in response to those outcomes(2). If two actions (or sets of them) are considered substitutable with respect to a particular outcome then the resolution of which action to pursue has been left for a future time. These actions are in response to an uncertain situation which characterises the decision situation.

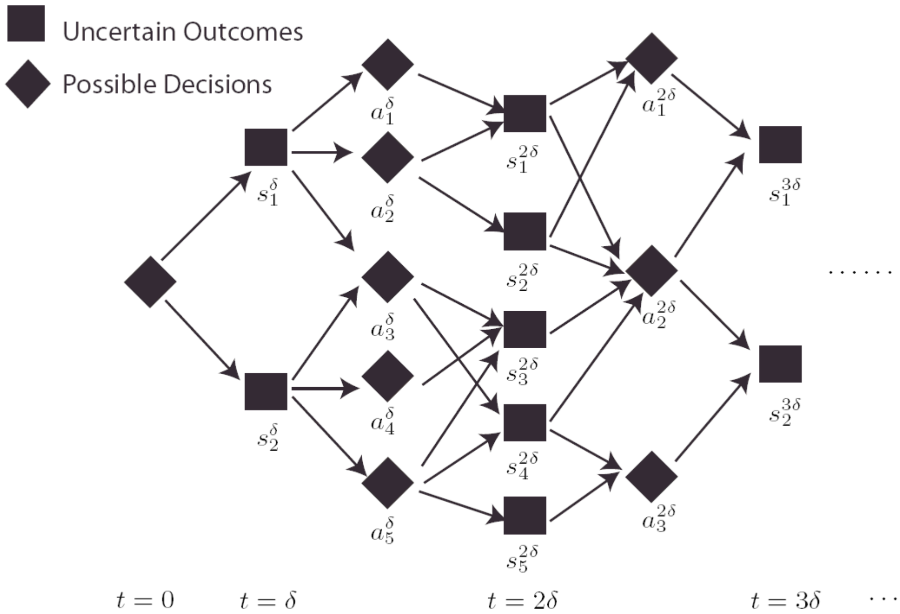


Fig. 1. Relationships between decisions in a Strategy

The strategy could thus be represented as a directed graph delineated by events.The representation in figure 1 uses time as opposed to events. Passage of time is one kind of event. The occurrence of any state could be an event(5). In case there is only one such outcome, such as taking action, there strategy merely represents contingency relationship between two actions. Thus, contingency is a crucial relationship made explicit by the strategy. When only one such contingency relationship is delineated between two actions, (4) call it Policy.

2.2 Design

Fundamentally, design is a tightly worked out set of actions which sit in relation to other actions and when taken together achieve a desirable result. Design can thus be thought of, as an intentional action set, whose member actions are related to each other and the make up is deliberate, and are to be taken in concert to bring about a particular state of the world. Such relationships that are of particular interest in urban planning are spatial relationships such as `adjacency` and `distance` , functional relationships such as `connectivity` , actor-asset relationships such as `ownership` or other actor-action relationships such as `responsibility`.

A familiar example of a design is the design of a building. It can be viewed as an outcome, where the constituent parts fit together coherently. A key point, however, is the design could be that of a building or more elaborately could be that of actions that bring into being the constituent parts (see figure 3). In other words, a construction management plan of a building is also a design, as are the architect's conceptual relationship diagrams, or the detailed construction diagrams. On the other hand, we are also interested in designs which draw from a social science perspective. Excellent examples include the hierarchical structure of an organisation (e.g. firm). A division of labour in a group working on a project, coherently specified, is a design. Lest it be taken that designs are necessarily static in nature, that need not be so. Any process can be viewed as a state or an outcome and thus specified relationships between outcomes could be relationships between processes.

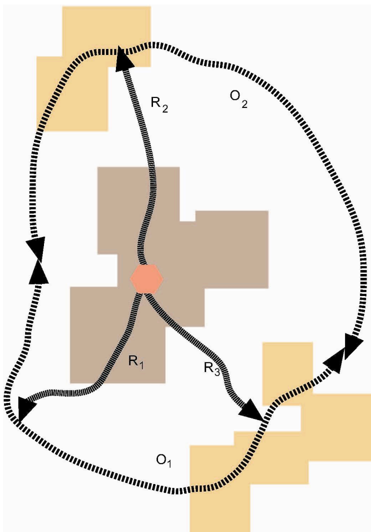


Fig. 2. Infrastructure investments as a Design

The transportation improvements plan in figure 2 could be modelled as a Design. In this case, the three radial links would be considered together because they would only be effective in strengthening the core if all the links were built. And the two ring road links would be considered together because they would only be effective in improving peripheral access if both were built. The response or anticipation of developers would then consider the construction or anticipated completion of combinations of links rather than individual links. This is a very simple instance of design relationship.

A design, however would be that the combinations of R_1 and R_2 has to be done in conjunction with the O_1 and O_2

```
ActionSet1(O1,O2,O3),
ActionSet2 (R1,R2)
NetworkConnect(O1,O2,O3)
...
NetworkConnect(R1,R2)
Precedes (ActionSet1, ActionSet2)
Connect(ActionSet1, ActionSet2)
```

Broadly construed, we are concerned with the design relationships which are spatial, functional, temporal and mereological in nature. Spatial relationships include distance or qualitative spatial relationships such as front and back (6). Functional relationships are the actor- asset relationships such as ownership, or asset-asset relationships such as connectivity (7). Mereological relationships such as parthood or membership and subset relationships are treated naively without references to the topological issues raised by (8).

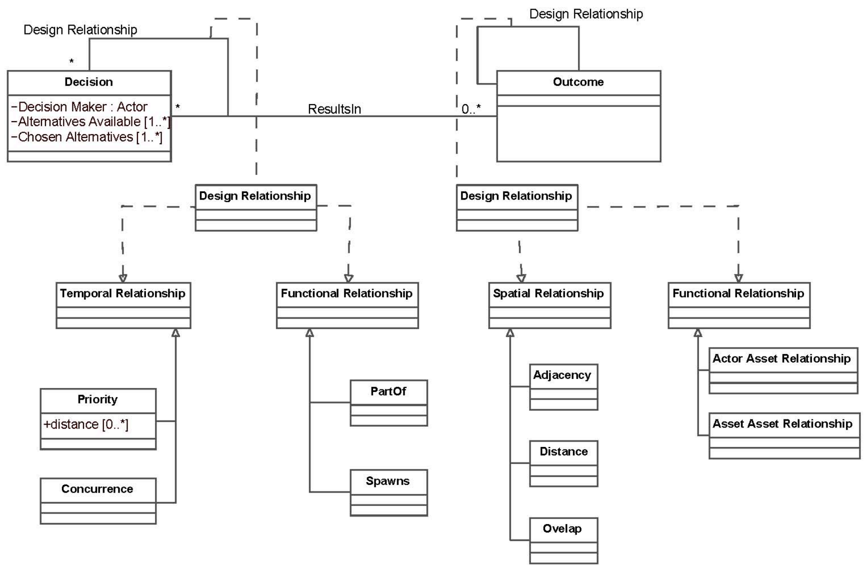


Fig. 3. Incomplete description of Design relationships

Figure 3 specifies an incomplete list of design relationships. Though contingency could also be a design relationship, since it is explicitly dealt within strategy we do not represent it in this diagram. This, however, points out to an important idea. Plans do not specify a pure design or a pure strategy. In fact, an action which is a part of a design could very well be contingent upon another or substitutable with another. Or the action that is specified as a response to an uncertain outcome in a strategy is a composite action which on close inspection turn out to be a design. For example, if building the a ring road and expanding the transit system are two possible reactions to an increase in traffic volume in the city, each of these seemingly monolithic actions are truly designs.

3 Priority & Contingency

Following the characterisations of (9) and (3), decisions are events not necessarily rooted in time. Just as we need to divorce the notions of Assets and Actors from location, we also need to divorce events from inherent underlying framework of time. In this paper, we are limit our scope to decisions that are instantaneous in nature, that is they do not occupy an interval in the characterisation of (10). However, the decisions may produce effects that unfold and endure over time. For example, a decision to `build (Road)` brings an `asset` into existence, which endures over its lifespan or till another action deliberately brings about its destruction.

(11) and (5) distinguish two different modes of representing event relationships. One is SNAPshots of states arranged on the temporal axis, and the other which is primarily focussed on the processes (existence, modification etc.) that occur in a SPAN of time. Both kinds of representation are useful and in fact reasoning for urban systems should consider both modes of representation without too much emphasis on the rigorous and exact translations between the two. Activities, such as shopping, travel, residing etc. are processes and the level of activity — volume of sales transactions, traffic count on a link etc.— are also Snapshots of states.

Since plans are made for contingent futures, the occurrence (or non- occurrence)of a particular future is an event the plan is supposed to address, irrespective of its location on the temporal axis. The temporal location of the event is useful only for the purposes of discerning relationships to other events. Thus, an event set can compose of temporal relationships such as `before` , `lag` , and `temporal adjacency` as primitives, without inferring them from location on the temporal scale.

Furthermore, these relationships are ephemeral and particularistic. For example, a plan of city government may suppose event *A* is prior to *B* (building the ring road first and expanding the connectors to the center later in the figure 2) and choose to plan for such a future (by scheduling the capital improvements plan and budgets accordingly). However, the suburban development at the fringes may suppose the opposite precedence relationship – it may be more useful to get to the employment center in the city first than to connect to other fringe development. Especially when the issue is who gets to act on which subset of actions in the design, plans of multiple agents may presume incongruent event relationships. Furthermore, different plans of the same agent may prescribe different courses of actions for event sets, which in different futures may be related differently.

To update plans using histories of decisions and actions, we consider two relationships which are closely tied to each other but nevertheless offer sharp distinction on reasoning about updating plans. They are contingency and priority.

An action is prior to another temporally. If action *A* occurs before or after *B*, then they share a temporal relationship. Temporal priority is a binary relationships which arranges actions on a temporal scale. The crucial difference is that we have to observe the occurrence of both *A* and *B* to make judgements about such arrangements. Functional priority implies temporal priority, while not the other way round.

Functional priority is contingency. An action *A* is functionally prior if it is necessary before the occurrence of *B*. In other words, if we decide to do *B*, we have implicitly decided to bring about *A* or at least observe that *A* has occurred. If a soil conservation group recognises the contingency of a corridor study and its recommendations as necessary for the building of an interchange *I* and if it believes that building *I* is inimical to its goals, then it could reason that lobbying for strong representation of its concerns during the study process would produce an unfavourable recommendation, and thereby not bring about *I*. This course of action when it is made explicit in their negotiations with the County, which favours the interchange, could enact regulations that conserve the farmland to assuage the concerns of the conservation group.

4 Decision Histories & Changing Plans

Partial Orders thus frame the reasoning for decisions within a plan on the temporal scale. Keeping in mind the adequatist, fallibilist and particularlist model of reasoning and representation (to use the words of (11)), partial orders are sufficient for the purposes of this paper. For example for actions *A*,*B* and *C*, `precedes(A, B)` and `precedes(C, B)` taken together make no claim about the precedence or any other relationship of *A* and *C*. If these two relationships are apparent in a plan, an action *X* is contingent on occurrence of *B* is also contingent on occurrence of *A* and *C*. When decisions are set in the indeterminate future, they are set typically without any reference to the underlying temporal structure but are ordered by the relationships among them.

However, when decisions are taken/observed they are situated in a well-ordered temporal axis. That is, each decision is time stamped. This allows us to reason with the actions and compare them to their situation, set in plans. If a plan *A* specifies `Contingent (C, D)`, we can infer the temporal relationship `precedes (C, D)`. However, if we observe C_{t_1} and D_{t_1}, thus deducing the temporal relationship `concurrent (C, D)`, then the plan is not useful in guiding us with the reasoning about intended actions and thus need to be updated.

On observing a sequence of decisions by an actor, we can match them to the plans of the actor. In particular, if any of these actions are specified in a strategy, then we can trace the contingency relationships by traversing the directed graph backward to see if any of the previous decisions are identified by this strategy. Furthermore, even when the decisions are not observed, but the outcomes are, then we can speculate on the actions that were taken that led to the outcome[1] . The subgraph that is rooted at the current decision node traversed backward is useful only for archival purposes and does not contribute to any future decision making. Thus, the strategy can be updated by archiving the subgraph.

If any of these decisions match a decision node of a strategy, we should look for decisions/actions that are forward from that node on to consider the effect of the current decision on the future decisions. That is the subgraph of the strategy that is rooted at the current decision node and traversed forward is most useful for determining the contingent decisions. The other uncertain outcomes and decisions that are not a part of either of the two subgraphs are still useful to maintain as a separate strategy for they may be useful in other decision making when such state of the world occurs.

Thus a particular strategy would undergo transformation into multiple strategies, some useful for their historical relevance and others useful for decision making. In particular, the two strategies that are still yet useful could be further elaborated to consider other uncertain outcomes that were not yet identified in the previous version of the strategy. This process maintains the currency of the strategies and therefore that of plans.

If any of the subset of decisions/actions match a part of design, then we can recognise that all the other actions that are part of the design when taken follow the 'design relationships' within the design. That is for assets, O_1 and O_2 `Network Connect(`O_1,O_2`)` , specifies a design relationship. Even when O_1 occurs without O_2, we can deduce that, if the design is followed, if and when O_2 occurs, O_1 and O_2 are related through a `connectivity` relationship. If, however when O_2 occurs and connectivity is not satisfied then the value of the design lies in identifying how such change affects all the other design relationships, in particular the connectivity relationships, between R_1, R_2 and R_3. Thus, the design is modified to suit the circumstances.

On the other hand, when the plan does get updated, due to a change in circumstances, then previously known contingency and priority relationships may be discarded and new ones are specified. When plans are changed in this fashion, they are changed either due to change in composition of actors who make the decisions, change in direction of policies and strategies or different modes of evaluation of the contingencies and interdependencies. Such changes, though not anticipated from a logical system point of view, once made, provide useful information about the relationships between decisions. In this case, the new plan supercedes the old plan by keeping some actions, discarding some, and changing some relationships between them. A case in point is the example below.

The 1974 City of Champaign Comprehensive Plan (12) addresses the city's core downtown area on less than three pages. Despite noting that Market Place Mall is set to open on the city edge, the authors of the plan emphasize only that the imminent

[1] Outcomes are determined by a particular state of the world or a measurement of an attribute of the state. For example, different levels of Vehicle Miles Travelled (VMT) on a highway network could be the various uncertain outcomes.

creation of a covered pedestrian mall on Neil Street is an “obtainable and reasonable solution to the problem of maintaining a strong tax base in the Central Business District (CBD) and preserving it as a viable part of the community’s cultural life," (p. 87). The only other downtown-oriented actions suggested in the plan are to ‘curtail’(p. 87) certain commercial uses on the edge of downtown and to replace a medium density residential district west of downtown with higher density uses to “buffer low density residential uses from the Central Business District," (p. 84).

By the next Comprehensive Plan six years later the attitude toward the downtown has shifted (13, 14). No longer just a brief mention, retention of existing users and attraction of new ones has become a more important aspect of the plan. The downtown is mentioned throughout the plan, under topic headings such as ‘Growth’ (p. 1-2), and ‘Commercial Development’ (p. 1-4), in which the issue of facilitating a more vibrant downtown through active government intervention is a primary focus, resulting in objectives such as, “Encourage downtown redevelopment and allow flexibility in building codes, but without jeopardizing life safety," (p. 1-5), and “Direct a substantial percentage of new development to areas inside the existing City limits, particularly to the downtown core", (p. 1-2).

By 1992, preservation and revitalization of downtown had emerged as a prominent enough issue to warrant creation of the 1992 Downtown Comprehensive Development Plan (15), followed by the more detailed Downtown Plan in 2006 (16). Both argue for the focus of substantial city efforts on the limited geographic area of the Downtown, in order to create a space that will benefit the city as a whole. As circumstances and stakeholders change over time, plans change, too. However, certain elements of old plans may have gained traction and persist in reality even as they are omitted from newer plans. The “stickiness” of these schemes has implications for urban development and there is utility in tracking them through time.

If a strategy *A* is present in a plan P_{t_1} and *B* is present in that of plan P_{t_2} with explicit provision that P_{t_2} supercedes P_{t_1} in informing decision making. However, *A* is only superceded by *B* only if the same decision node occurs in both *A* and *B*. If the subgraph that contains all the decision nodes forward from the all common decision in *A* be called *T*, then the directed *AT* is still useful information for other contingencies and uncertainties that are not considered in *B*. Thus, some portions of previous plan are still relevant for decision making.

In the case of design, however, inferring these changes in relationships is a more onerous task. One design may be alternative to another as a whole even if they do not share any constituent action items. Such alternatives are decided based on capability or location constraints or substitutability with regards to intentions (17). In such cases, the subsequent plan which explicitly supercedes the original plan, completely replaces the design. Further elaborations are left for future work.

5 Multiple Plans

Now that we can update single plans as and when decisions get made, we must consider the idea of multiple plans informing a single decision situation. Different actors have concurrent plans, some of which are interdependent. For the purposes of brevity, we consider here only those plans where explicit relationships between plans are acknowledged in plans which are still ‘valid’. In such cases, the actor responsible for the decision is an attribute of the decision situation and thus the reasoning is no different from earlier sections. If one plan specifies a decision is subsequent to decisions of other actors, and if such other actors identify other contingencies between those actions, then we can begin to reason about further interdependencies between decisions of one actor and actions of another. What becomes interesting, however, is that this provides another justification of the ontology as a standard representational device to encode the plans of all these actors so that such connections can be made.

The University of Illinois Campus Area Transportation Plan (CATS) published in 1999 (18) argues for enhancing multi-modal transportation access in and around the campus. Some tactics, however, extend well beyond the spatial extent of the campus. On page 74, the plan suggests widening Springfield Avenue where it connects the southern tip of downtown to the northwest portion of campus between Neil and Wright Streets. "In order to widen portions of Springfield Avenue," the plan notes, “it would be necessary to replace the existing viaduct at the railroad tracks to accommodate additional travel lanes and semi trucks. The City of Champaign and the University should work with [Illinois Department of Transportation] regarding the replacement of this viaduct." To increase multi- modal access within the campus, the University’s plan is contingent on the plans and actions of other actors well outside the campus; this is not an unusual situation.

Actor `University` has a plan which identifies `Widen Road` as decision contingent on `Replace Viaduct`. However, the actor who is responsible, or has the capability to replace the viaduct is the group of actors consisting of the `City`and the, `Dept. of Transportation`. If either actor balks at replacing the viaduct, then the University cannot expand the road. This has a cascading effect on other plans of the university which were contingent on the expansion of the road, such as if and where to build a parking structure or transit transfer station. Thus, the decision by the Dept. of Transportation, indirectly changes the plans of the University.

Likewise, actors may make plans which directly attach their own future actions to existing or planned actions of other actors. The Champaign Police Department, for instance published a Five Year Community Safety Plan in 1995 (19). The first goal in the plan was to address the issue of “Total Crime" in the community; the first strategy to address that goal was, “To comprehensively address infrastructure and service needs in threatened areas," (p. 28). The first action under this goal is to “Combine efforts between police and planning departments in the on-going evaluation of neighborhoods in which to begin to develop specific improvement plans," (p. 28), and to “Integrate neighborhood needs into the multi-year strategy for infrastructure improvement included in the [Capital Improvements Program] and public works maintenance programs throughout the city" (p. 28). By integrating land use and service provision within a policing and safety plan, the Police Department illustrates the utility of being able to track the plans and decision of other actors when making a plan; the police department is relying on the actions of the planning department as a way to combine these two currently unconnected activities in order to serve their own needs.

The indeterminacy of relationships of all actions in plans, provides useful opportunities to expand on existing plans. It is imperative to acknowledge that, to consider and encode 'all' possible contingencies and interdependencies between various sets and subsets of actions is unreasonable expectation of plans and planners. The plans specify only relationships between specific combination of actions that are of particular interest to the decision maker. As and when some alternatives and decisions recede from the decision situations, they provide opportunities to add or reconsider other actions and decisions and how they might relate to decisions that are being considered and yet to be taken.

6 Conclusion

We have illustrated some ways in which keeping track of the decision histories of various actors over time can help inform decisions within urban development situations. These arguments provide a rationale for development of an ontology of plans that recognizes the complexity of decision- making in towns and cities. Subsequent stages of this work must overcome the obstacles to development of such an ontology when multiple plans do not acknowledge one another's existence, yet are nevertheless contingent and/or interdependent. This situation is common, yet requires a system for unmasking these hidden connections if the system is to be useful 17. Nevertheless, maintaining the currency of plans is an important application of the ontology of urban systems, so that these different plans over time can be used in decision making by various agencies.

References

1. Bourne, L.S.: Urban structure and land use decisions. Annals of the Association of American Geographers **66**(4) (1976) 531–535
2. Friend, J., Hickling, A.: Planning under Pressure: The Strategic Choice Approach. Butterworth-Heinemann, Oxford, UK (1997)
3. Kaza, N., Hopkins, L.D.: Ontology for land development decisions and plans. In Teller, J., Lee, J., Roussey, C., eds.: Ontologies for Urban Development. Studies in Computational Intelligence. Springer-Verlag, Berlin (2007) 47–59
4. Hopkins, L.D., Kaza, N., Pallathucheril, V.G.: Representing urban development plans and regulations as data: A planning data model. Environment & Planning B : Planning and Design **32**(4) (July 2005) 597–615
5. Worboys, M.F.: Event-oriented approaches to geographic phenomena. International Journal of Geogrpahic Information Science **19**(1) (2005) 1–28
6. Freksa, C.: Using orientation information for qualitative spatial reasoning. In Frank, A.U., Campanari, I., Formentini, U., eds.: From Space to Territory: Theories and Methods of Spatio-Temporal Reasoning in Geographic Space, London, UK, Springer-Verlag (1992) 162–178
7. Kaza, N.: Towards a data model for urban planning: Ontological constructs for representing regulations and guidelines with geography. Master's thesis, University of Illinois at Urbana Champaign (2004)

8. Casati, R., Varzi, A.C.: Parts and Places: The Structures of Spatial Representation. MIT press, Cambridge, MA (1999)
9. Worboys, M.F., Hornsby, K.: From objects to events. GEM, the geospatial event model. In Egenhofer, M., Freksa, C., Miller, H., eds.: Third International Conference on GIScience 2004. Lecture Notes in Computer Science 3234. Springer-Verlag, Berlin (2004) 327–344
10. Allen, J.F., Ferguson, G.: Actions and events in temporal logic. In Stock, O., ed.: Spatial and Temporal Resoning. Kluwer Academic, Boston, MA (1997) 205–243
11. Grenon, P., Smith, B.: SNAP and SPAN: Towards dynamic spatial ontology. Spatial Cognition & Computation **4**(1) (2004) 69–104
12. Champaign County Regional Planning Commission: Champaign comprehensive plan, City of Champaign, Illinois. Champaign County Regional Planning Commission, Urbana, IL. (1974)
13. City of Champaign: Land use element of the Champaign comprehensive plan. City of Champaign, Champaign, IL. (1980)
14. Champaign County Regional Planning Commission: City of Champaign 3 year community development comprehensive plan. The Commission, Urbana, IL. (1979)
15. Urbanics, BRW: Downtown Area Comprehensive Development Plan. City of Champaign, Champaign, IL. (1992)
16. City of Champaign: 2006 Downtown Plan. Champaign, City of, Champaign, IL (2006)
17. Kaza, N.: Reasoning with plans: Inference of semantic relationships among plans about urban development. PhD thesis, University of Illinois at Urbana Champaign, Champaign, IL (2008)
18. University of Illinois: Urbana Campus Area Transportation Plan. University of Illinois, Urbana, IL (1999)
19. Champaign Police Department: Five Year Community Safety Plan. Champaign Police Department, Champaign, IL (1995)

Exploring Informal and Adaptive Ontologies for Communicating Urban Time-Depth in Strategic Planning

Stephen Dobson

Department of Landscape
University of Sheffield
arq06sd@shef.ac.uk

Abstract. An appreciation of the historic landscape in the urban domain introduces temporal processes which require dynamic rather than static models. A flexible, informal and practice-based ontology provides a means of theorising and describing space in a way that can be easily redefined and continually questioned. Traditional cartographic methods, whether paper-based or digital, will only ever provide static output. Geographical Information Systems (GIS) provide the most flexible form of mapping through the querying of underlying data, however, the result is a larger collection of static models rather than something truly dynamic.

The research examples illustrated here are based on participatory Action Research with Sheffield City Council in developing a landscape character ontology to communicate urban time-depth for the Green and Open Space strategy. GIS is used simply to explore and test informal and partial ontologies and is primarily a means of building understanding and facilitating stakeholder communication.

Keywords. Heritage, Historic Landscape Character, Urban Planning, Urban Design, Place, Identity, Green Space, GIS, Archaeology

1. Introduction

> "The main objective of the COST Transport and Urban Development Action C21 is to increase the knowledge and promote the use of ontologies in the domain of urban development, in the view of facilitating the communications between information systems, stakeholders and urban specialists at a European level." [1]

Multi-stakeholder spatial planning in the post-modern urban domain brings an increasing number of perspectives together in balancing complex models of urbanism, with the ultimate aim of addressing its multiplicity of functions; economic, social, aesthetic, cultural, educational, ecological etc. The number of components in this list continues to increase as our vision changes and their subcomponents multiply the more closely we examine them.

The objectives outlined by Teller require an understanding, communication and more importantly the ability to combine these complex components and their interrelationships. However, many stakeholders along with the general public (and even some specialists) are unaccustomed to deconstructing their realm of interest and modelling the ontological structure that defines its conceptual framework. In this case: "…conceptualizations are often tacit or implicit in the urban development domain and efforts to formalize these conceptualizations are generally viewed as 'oversimplifications' by experts that are struggling to defend their scientific and technical legitimacy." [1]

In a multi-stakeholder planning context, communication is then often based on assumed and domain-specific 'world views' where incompatibilities and conflict will inevitably arise. Whilst this may result from differing object/relationship *values* (ie. the perceived importance of certain environmental or social concerns) it can also occur through simple inconsistencies between the models that each stakeholder uses to understand the landscape – especially if these are communicated in natural language and without due period of reflexivity. The actual structure of these perspectives may not have previously been formally explored or articulated by the stakeholder and so wider frameworks that rationalise multiple concerns become difficult to achieve. This is particularly the case when exploring highly conceptual characteristics of urban space such as place, identity, attachment and character and, more significantly, how these might correspond to, and influence, the decision-making surrounding more 'traditional' urban concerns such as morphology, architecture and transportation networks.

> "Ontologies have also an important role to play in revealing the logical structure of existing conceptualizations. 'Conceptualizations are often tacit. They are often not thematized in a systematic way. But tools can be developed to specify and to clarify the concepts involved and to establish their logical structure, and thus to render explicit the underlying taxonomy'" [1, 2]

Place attachment, community identity, character and local significance are central to our appreciation of the urban landscape although *formal* ontologies of these realms may be highly subjective and difficult to agree upon. However, by rapidly visualising them using semi-complete, adaptive ontologies, multi-stakeholder communication may be progressed and at least aid the development of ideas relating to terms, values and their overarching concepts. Because of the rapid nature of such models they are flexible enough to be used in meetings, email communication and internal documentation and can be transformed quickly whilst communication progresses. Depending upon the domain in question, a final formal ontology may not be desirable or required however the early processes of definition - specification, knowledge acquisition and conceptualisation [3] - may provide a level of reflexivity that can help stakeholder's engage with the inherent complexities. These adaptive ontologies, used primarily throughout the process of research and communication, may however, become extremely useful to the formal ontologist at a later stage.

The specific examples featured in this paper examine of the role practitioner-researcher in helping to facilitate the use of Historic Landscape Characterisation (HLC) data in the formation of a pilot Green and Open Space Strategy (GOSS) for Sheffield's East Area Development Framework (EADF) region. The EADF GOSS [4] aimed to adopt a broad landscape perspective and so solely traditional heritage assessments (site-based and monument focused) would not be appropriate. Whilst HLC attempts to help stakeholders appreciate the wider historic landscape [5], there lacks critical assessment of how this might be achieved in practical urban planning.

2. The Urban Landscape and Time-depth

The extent to which the evidence of the past is interwoven with the present is a surprisingly ubiquitous urban concern. Character, place and community identity in the urban domain rely not only on residents having knowledge of how things might have changed but also their ability to see signs of this change in their surroundings. Understanding and communicating this in a strategy document was a key challenge for the research project and could only be achieved through critical deconstruction of key concepts that underpin the HLC dataset. These were composed as queries which were then applied to the GIS to provide outputs which visually tested whether less obvious historic 'assets' in the landscape could be highlighted through such attribute and relationship modelling.

> "…the analysis of 'morphological process' should not be restricted to the evolution of the built environment but also encompass the evolution in the way a given urban feature may be conceptualised over time." [1]

As a result of an iterative process of questioning and reflection on how urban character, time-depth and the importance of evidence relates to concepts of place and community, the multiple GIS queries and map outputs were 'by products' of the ontological discussions that they illustrated. The outcome therefore was a more considered understanding of how data relates to each other and was achieved in collaboration with the non-technical stakeholder to increase awareness of the domain in question. This was particularly important when addressing the depth of time in the urban landscape where emphasis was placed on historic *processes* rather than specifically its built form. In this case, evidence of historic land uses in the contemporary surroundings was considered more appropriate in assessing the character of the city over time and at multiple scales than a consideration of a small number of heritage 'points of interest' which are often without context and lack local significance.

> "In comparing the formal determinism of modernist urban planning and the more recent rise of neo-traditional 'New Urbanism', the cultural

> geographer David Harvey has written that both projects fail because of their presumption that spatial order can control history and process…His point is that projection of new possibilities for future urbanisms must derive less from an understanding of form and more from an understanding of process – how things work in space *and* time."
>
> "In conceptualizing a more organic, fluid urbanism, ecology itself becomes an extremely useful lens through which to analyze and project alternative urban futures…the complexity of interaction between elements within ecological systems is such that linear, mechanistic models prove to be markedly inadequate to describe them. Rather, the discipline of ecology suggests that individual agents acting across a broad field of operation produce incremental and cumulative effects that continually evolve the shape of an environment over time. Thus, dynamic relationships and agencies of process become highlighted in ecological thinking, accounting for a particular spatial form as merely a provisional state of matter, on its way to becoming something else." [6]

Girot, underlines this need for temporally dynamic models of urban process and highlights that the full complexity of any urban infrastructure can only be appreciated by recognising not only multiple scales of space but also time. From the historic processes at work in created these multi-layered places, to our very means of movement (transportation) and is affects on perceiving them.

> "How far from reality can the landscape design tools that we work with be? The gradual withdrawal from landscape as a place to landscape as a piece of paper or computer screen must be questioned, not only in terms of its conceptual shortcomings but also in terms of the very landscapes that result." [7]

To achieve this level of dynamism and fluidity the ontology process should be accessible to all stakeholders and informal enough to facilitate multiple reconceptualisations. Using Uschold and Gruninger's definitions of ontology structures we explore a role for informal and partial ontologies in "…facilitating the communications between information systems, stakeholders and urban specialists…" [1]

- Highly informal: the ontology is expressed loosely in natural language
- Semi-informal: to ontology is expressed in a restricted and structured form of natural language, greatly increasing clarity by reducing ambiguity
- Semi-formal: the ontology is expressed in an artificial formally defined language
- Rigorously formal: the ontology is meticulously defined with formal semantics, theorems and proofs. [8]

3. Example: Building and Rebuilding Relationships and Models

Our historic landscape resource is then made up of; 1) areas of obvious age and historic value (time-depth) 2) areas where the past may simply be ‘readable’ in the present. The ‘readability’ and stories of the past may remain where reuse after reuse has not managed, for whatever reason, to completely erase the previous patterns of use that may still be evident in the shape of a road, a line of trees, pockets of woodland, remnant urban hedgerows, the lines of walls and fencing or the very size and orientation of land units. This depth, authenticity and underlying coherence adds character, embodies community experience (individual or collective) and is often so interwoven with the present that it is difficult to perceive one without the other.

Figure 1 illustrates the results from combining two queries which are used to identify the extent of pre-1945 historic resources citywide. The [FROM] attribute (which assigns a start date to all current land use character areas) is used to classify earlier periods than 1945 - with land use age being represented by increasingly darker shades. Areas developed since 1945 are blanked out in white as these are not required by the question. However, some post-1945 developments may still contain readable (historically legible) elements and so by including these in the resource map, wherever [HISTORICLE] = ‘significant’ or ‘partial’, it is possible to present both types of resource - either of a particular intrinsic age (in this case pre-1945) or those areas that contain some readable ingredients of a pre-1945 past.

Figure 2 uses the same methodology to present an alternative definition of ‘historic’ which uses 1750 to define a pre-industrial, historic landscape resource. The implications of working either solely with time-depth or both time-depth and legibility on actual sites is explored in the following case study.

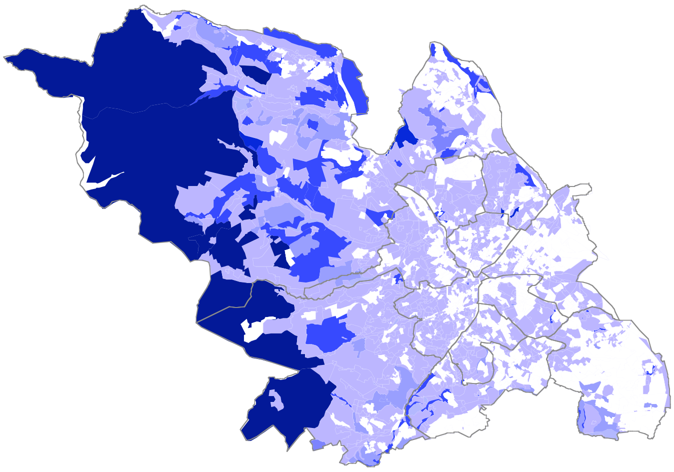


Fig. 1: Pre-1945 wider historic landscape resource

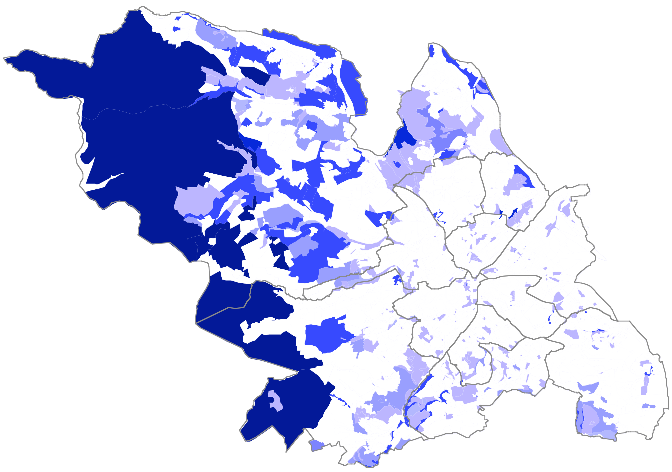


Fig. 2: Pre-1750 wider historic landscape resource

4. Case Study: Eccleshall Woods Surrounding Area

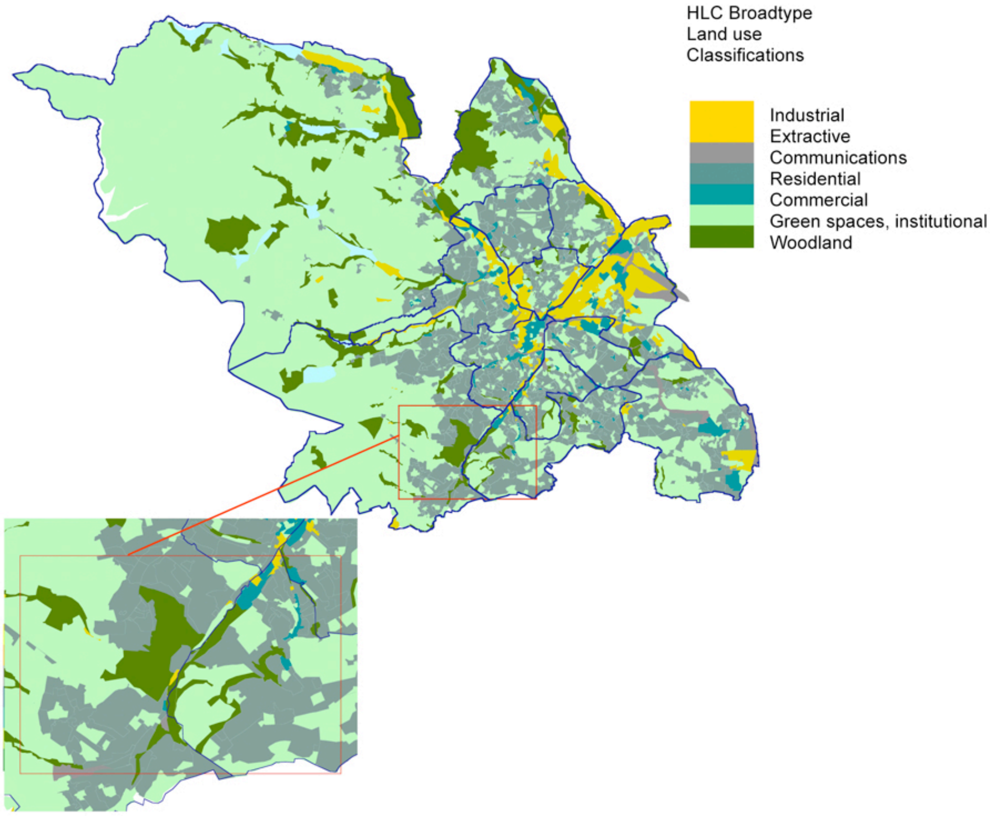


Fig. 3: Case study area

Temporally, the Eccleshall Woods area has great diversity as it lies in the urban fringe and also contains all [BROADTYPE] land use classifications.

Figure 4 presents the extent of the pre-Industrial landscape existing in today's case study area based solely on the [FROM] date of land use. It is created by filtering out, in the query, any areas which have been extensively developed since 1750 and is designed to highlight the more obviously historic parts of the landscape. We assume that these areas will be rich with history and as such have intrinsic value and significance. By comparing this plot with the land use map of the area in figure 5 we can appreciate that, apart from the historic residential areas of Dore and Greenhill and the Industrial works of Abbeydale and Whirlow Wheel, the predominant areas of age are green spaces and woodland.

However, following discussions with the Parks and Countryside stakeholder, it was noted that one of the city's most significant historic parks was not present in the original conceptualisation. The reason for this was that the original definition of an 'historic landscape' was purely based on age and did not include the 'readability' of the past in more recently developed areas and so the ontological model used to define the query had to be altered.

By including polygons that, whilst containing a [FROM] date later than 1750, have legibility of a pre-1750 earlier land-use, a number of new additions are included in the total historic resource map. These are presented in figures 6 and 7 with the additional green spaces labelled. One of the more notable additions is Graves Park (Figure 7: HSY2201, HSY2199).

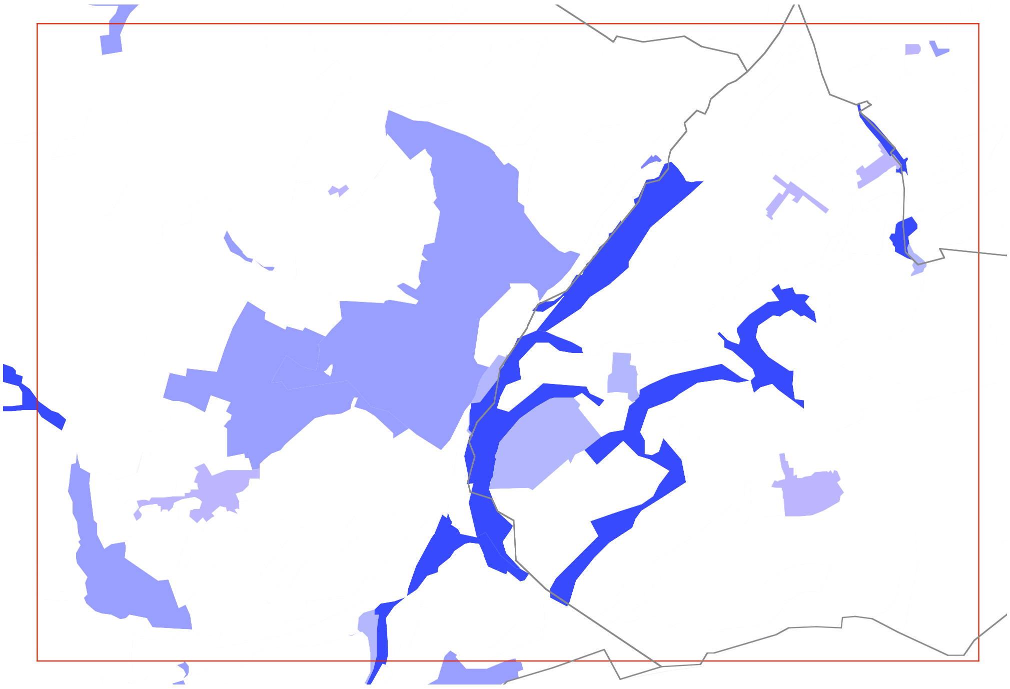


Fig. 4: Pre Industrial resource based solely on time-depth

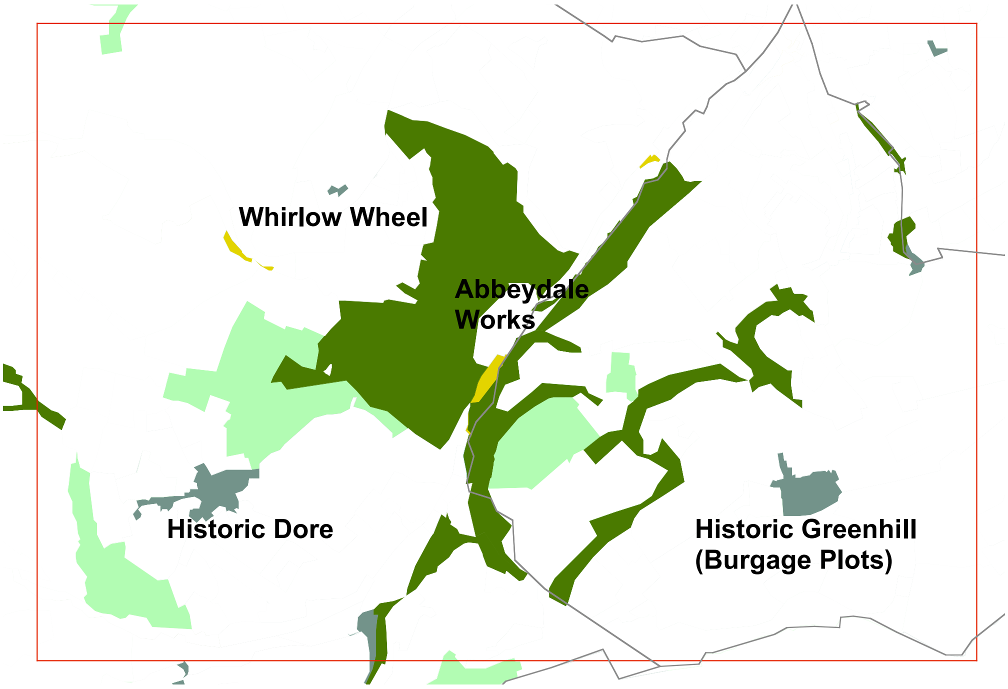


Fig. 5: Pre Industrial time-depth is predominantly evident in green spaces and woodland with exceptions such as: Historic Dore, Abbeydale Works, Whirlow Waterwheel and Historic Greenhill

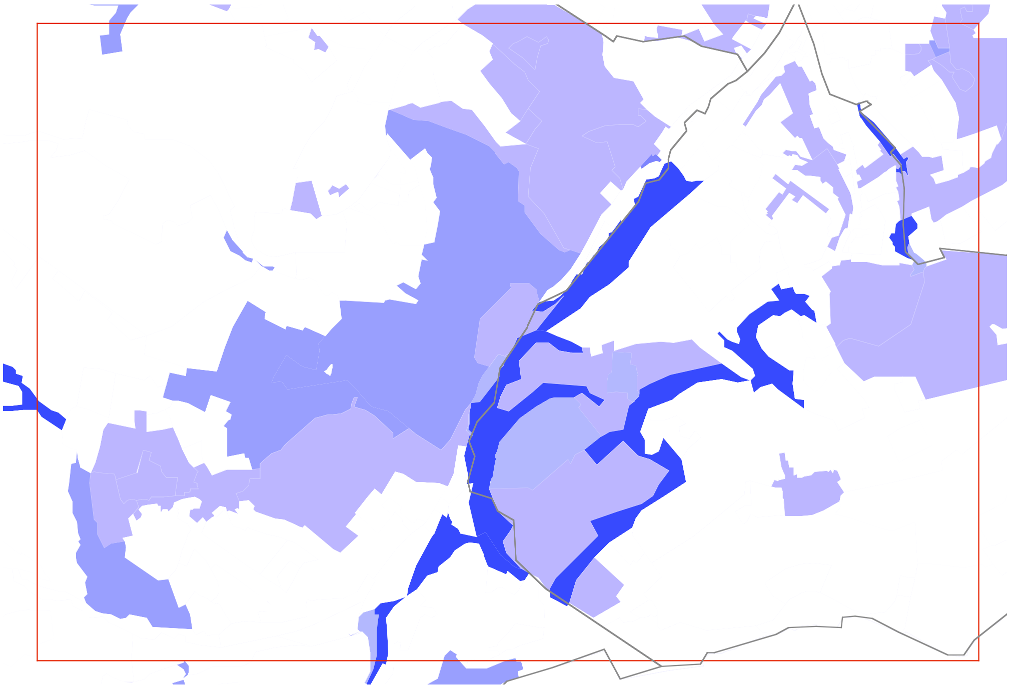


Fig. 6: Pre Industrial resource based on time-depth but including more recent areas with legibility of earlier land use

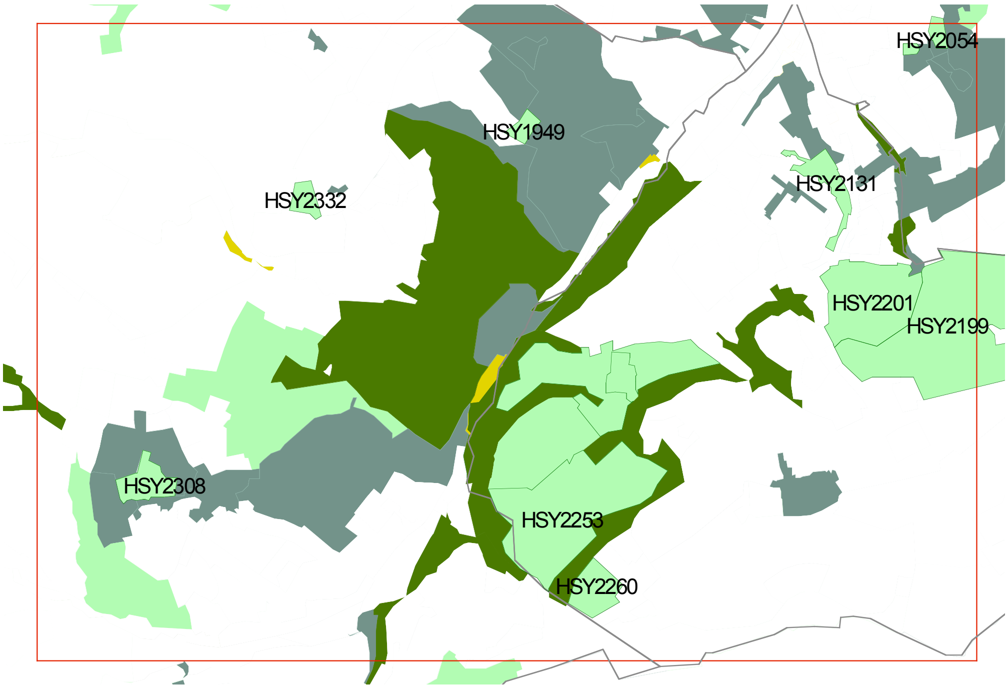


Fig. 7: By widening the scope of historic landscapes to include legibility many other green spaces emerge as having historic significance, including HSY2201/HSY2199 – 'Graves Park'.

HLCUID	**HLCTYPE**	**FROM**	**LEGIBILITY**	**PREVIOUS1B**	**PREVIOUS1H**	**FROM**
HSY1949	School	1972	Significant	Enclosed Land	Piecemeal Enclosure	1066
HSY2054	Allotments	1924	Partial	Enclosed Land	Piecemeal Enclosure	1066
HSY2060	Allotments	1952		Enclosed Land	Piecemeal Enclosure	1066
HSY2131	Public Park	1967	Partial	Enclosed Land	Assarts	1066
HSY2199	Public Park	1935	Significant	Ornamental, Parkland	Private Parkland	1750
HSY2201	Public Park	1931	Partial	Enclosed Land	Piecemeal Enclosure	1066
HSY2253	Golf Course	1900	Partial	Ornamental, Parkland	Private Parkland	1750
HSY2260	Sports Ground	1982	Partial	Enclosed Land	Piecemeal Enclosure	1066
HSY2308	Playing Fields	1950	Significant	Enclosed Land	Piecemeal Enclosure	1066
HSY2332	Private Parkland	1851	Significant	Enclosed Land	Piecemeal Enclosure	1066

5. Graves Park

Graves Park is the most notable addition to our areas of historic significance map. The park as it exists today was developed in the early 1930's however its origins were much older through its use as a deer park and as piecemeal enclosure agricultural land. If our historic landscape ontology equates value to landscape age then such spaces can become absent from a landscape assessment of the area. However, by exploring the field of historic legibility and how this relates to the concept of an historic landscape we are able to appreciate a wider range of urban spaces. This example demonstrates the importance of recognising historic legibility as a key part of landscape character since historic features and an essentially historic space can be appropriated in modern times. From the perspective of urban development it is then essential to establish which legible elements exist in a space so that creative changes may occur without compromising our ability to 'read' the landscape story. Particularly for a heritage park it is essential that management, planning and development may continue without eroding the very clues that form the historic 'feel' of a now modern place and are not eroded by uninformed incremental change. The preservation of legibility is therefore an important part of achieving levels of multi-functionality in historic spaces. The refined ontology was then used to define GIS queries for areas of very little known historic significance in an attempt to help local residents 'connect' to their place and value their surroundings. This process was facilitated by using the resulting maps to help define walking routes for a national urban walking initiative - 'Get Walking, Keep Walking'.

6. Summary

By testing the final model with walking routes we are able examine its ability to work dynamically as Girot suggests. The HLC provides landscape areas in a temporal sense and by documenting this landscape by process (successions of land uses) rather than purely built form we address the context of site rather than the site specific. Historic fragments and individual sites can obviously be related to their broader HLC context however their presence becomes part of an interconnected network of legibility and time-depth spaces within land-uses and so are not treated in isolation. This is particularly necessary at the landscape scale when considering multi-functionality, across multiple stakeholders through multiple time periods. No landscape is more subject to this than the urban landscape where complexity of space is matched only by the multiplicity of its use, our perception of it and the ways in which we attempt to conceptualise and describe this perception.

The examples here were visualised through GIS but the resulting knowledge acquisition regarding the practical relationships between data attributes means that the stakeholder is empowered and not 'bound' to a small number of static cartographic maps created for them by the GIS expert. In this case, a definition of the historic landscape ontology as required for GOSS purposes differs from that which had initially been defined by the archaeologist or historic environment professional and so it is an increased appreciation of the ontology behind the query that is of ultimate benefit to the stakeholder.

References

1. Teller, J. 2007 'Ontologies for an Improved Communication in Urban Development Projects' *Studies in Computational Intelligence* 61, 1-14 (2007)
2. Smith, B 2003 'Ontology' in L. Floridi (ed) *Blackwell Guide to the Philosophy of Computing and Information* Blackwell, Oxford 155-166
3. Roussey, C. 2005 *Technical Report No 1: Guidelines to Build Ontologies: A bibliographic study* COST Action 21 'Urban Ontologies for an improved communication in urban civil engineering projects'
4. Sheffield City Council (SCC) 2008 *East Sheffield Green and Open Spaces Strategy* SCC *http://www.sheffield.gov.uk/eastgoss*
5. Dobson, S 2008 'Exploring Ontologies of Historic Landscape Characterisation: Towards an approach for recognising the impact of incremental change to historic legibility in urban areas' in J. Teller; C. Tweed; G. Rabino (Eds.) *Conceptual Models for Urban Practitioners* Società Editrice Esculapio, Bologna, December, 2008, (121-132)
6. Corner, J. 2006 'Terra Fluxus' in. C. Waldheim (ed) *The Landscape Urbanism Reader* Princeton Architectural Press, New York
7. Girot, C. 2006 'Vision in Motion: Representing landscape in time' in. C. Waldheim (ed) *The Landscape Urbanism Reader* Princeton Architectural Press, New York
8. Uschold, M. and Grininger, M 1996 'Ontologies: Principles, methods and applications' in *Knowledge Engineering Review* Vol11:2, 93-136 (1996)

Knowledge Management for More Sustainable Water Systems

Stephen Mounce[1], Christopher Brewster[2], Richard Ashley[1], and Louise Hurley[1]

[1] Pennine Water Group, Department of Civil and Structural Engineering, University of Sheffield, S1 3JD, UK
{s.r.mounce, r.m.ashley, l.hurley,}@sheffield.ac.uk

[2] Department of Computer Science, University of Sheffield, S1 3JD, UK.
c.brewster@dcs.shef.ac.uk

Abstract. The management and sharing of complex data, information and knowledge is a fundamental and growing concern in the Water and other Industries for a variety of reasons. For example, risks and uncertainties associated with climate, and other changes require knowledge to prepare for a range of future scenarios and potential extreme events. Formal ways in which knowledge can be established and managed can help deliver efficiencies on acquisition, structuring and filtering to provide only the essential aspects of the knowledge really needed. Ontologies are a key technology for this knowledge management. The construction of ontologies is a considerable overhead on any knowledge management programme. Hence current computer science research is investigating generating ontologies automatically from documents using text mining and natural language techniques. As an example of this, results from application of the Text2Onto tool to stakeholder documents for a project on sustainable water cycle management in new developments are presented. It is concluded that by adopting ontological representations sooner, rather than later in an analytical process, decision makers will be able to make better use of highly knowledgeable systems containing automated services to ensure that sustainability considerations are included.

Keywords: Water Cycle Management, Ontologies, Text Mining.

1 Introduction

1.1 The Project

The inclusion of careful consideration of water management within the urban planning process is increasingly vital [1]. Reducible and irreducible risks and uncertainties associated with climate, demographic and behavioural changes require preparedness for a range of future scenarios and potential extreme events; to do this requires the sharing and management of complex information. Communication, knowledge sharing and a common vocabulary across professional domains and between stakeholders is central to more sustainable infrastructure management. New IT methodologies can help structure complex information and facilitate mutual understanding.

The UK Water Cycle management for New Developments (WaND) project (2003-2007) aimed to support the delivery of integrated, sustainable water management for new developments by provision of tools and guidelines for project design, implementation and management. The WaND research consortium comprised work packages researching technical, planning, financial and social science aspects of water management. The tools and procedures developed will assist in co-ordinated decision making across professional boundaries that takes into account risks and uncertainties. The WaND project has provided a case study for application of ontology learning techniques.

1.2 The Water Industry and Ontologies

Ontologies encode knowledge in a domain (a domain being a specific subject area such as the water industry) and also knowledge that spans domains - so called upper level ontologies. Ontologies include computer-usable definitions of basic concepts in the domain and the relationships among them and are increasingly valued because of the ever-increasing need for knowledge interchange. Currently, the Web is an incredibly large information source. Yet despite this, the main burden in information access, extraction and interpretation is left to the human user. Tim Berners-Lee coined the vision of a Semantic Web [2] that provides more automated services based on machine processable data and heuristics (methods or strategies, often informal 'rules of thumb' for problem solving) that make use of these meta data. He sees the Web as evolving into a universal medium for data, information, and knowledge exchange.

The Water Industry and related fields are an appropriate area for the application of Ontologies, since the management of complex data, information and knowledge is a fundamental and growing concern. For example, new sensor systems are providing increasing amounts of data. Sensors may be designed by different companies and often downloaded, processed (if at all) and stored on different computer systems. This results in a body of data from different sources that require intelligent algorithms to turn the data into useful information and knowledge for decision makers. Analysis of hydraulic data from water distribution systems for leak detection is one area requiring more sophisticated automated analysis than by simple manual inspection [3].

A variety of tools are available and under constant development for synthesising data into useful information and knowledge to advise decision makers. These tools can include handbooks, hydraulic models, GIS databases etc. Some tools and approaches which make particular use of ontological techniques are now described. The FLUMAGIS project is looking at interdisciplinary development of methods and data processing tools in support of the planning and management of river basins, with an emphasis on implementing the European Water Framework Directive (WFD). Software components enable planners and planning affected citizens to investigate, debate and evaluate planning measures in a co-operative process. The software tool being developed combines an object ontology (developed in Protégé [4]) with an inference machine with Causal Network (for reasoning about cause and effect) and Petri Nets (for modelling decision processes).

OntoWEDSS (Ontology-based Wastewater Environmental Decision Support System) is a DSS for wastewater management which augments rule-based reasoning and case-based reasoning with a domain ontology. It makes use of the generic WAWO (Waste Water Ontology) ontology and is implemented in LISP [5]. Schwering and Hart [6] look at a case study for the Semantic Translation of the WFD and a Topographic Database with a particular emphasis on semantic translation and portability of Ontologies with the Ontologies in question being the WFD ontology and the topographic ontology.

HarmoniQuA (Harmonising Quality Assurance in model based catchment and river basin management) forms part of the CATCHMOD cluster of European projects; supporting the implementation of the WFD. The HarmoniQuA project uses ontological knowledge engineering techniques to structure the knowledge and make it easily accessible. From this, software-based modelling Knowledge Base and support tools have been developed [7]. The knowledge Base for the tools is built with ontological knowledge engineering techniques and is implemented with Protégé-2000.

The European Project CD4WC (Cost effective Development of Urban Wastewater System for WFD compliance) has explored using an ontological knowledge base as a dissemination tool (a software program consisting of an OWL database and user interface) for stakeholders [8].

The joint Defra/EA project (FD2323) investigated the delivery of the improvement of data and knowledge management within flood and coastal erosion risk management (FCERM). One output of this project has been to describe and document the existence of data, information and knowledge applied to the full range of FCERM decisions, and the current roles and responsibilities for collection, management and use. Detailed ontologies in the form of maps have been assembled to represent the domain and particularly the structure of the relevant stakeholders' organisations [9].

Ordnance Survey is developing knowledge modelling methods to systematically construct conceptual and logical ontologies for flood risk management, while considering the requirements for interoperability between geographical and risk analysis information. Ordnance Survey is also collaborating with Oxford Brookes University to develop a freshwater ecology ontology to be used to test for interaction with their topographic ontology [10].

2 Methods

2.1 Constructing Ontologies

A domain expert defines classes to represent concepts in a domain, with slots (or roles) to represent properties and relationships between the concepts. An ontology, together with a set of instances, constitutes a knowledge base. An instance has properties with particular values. A class can have subclasses (is-a relation) to represent more specific concepts. Similarly, if some concepts share common properties, they may be generalized by creating a superclass, which holds the common properties. As a simple example, a sub-class of the surface water body class is river. The River Thames is an instance of river. The superclass of the surface water body class is the water body class. However, there is no single unique ontology for any domain [11].

A knowledge base consists of an ontology and a set of instances. Protégé and similar editors can be used for the hand construction of taxonomies (a subject-based classification that arranges the terms into a hierarchy). Alternatively, rather than being human-labour oriented, knowledge acquisition can be machine-aided. Automatic knowledge acquisition as a concept is at least as old as AI itself. It is particularly attractive due to the fact that extracting knowledge from a domain expert (via e.g. interviews or questionnaires) is an arduous, labour intensive task (the 'knowledge acquisition bottleneck'). Therefore excavating knowledge from human artefacts (whether minds or texts) is extremely costly [12]. Data driven knowledge acquisition is particularly desirable as text is massively available on the Web.

2.2 Semi-automatic Creation of Ontologies

Current research by computer scientists is exploring how a set of techniques (based on Information Extraction, Information Retrieval (IR) and Natural Language Processing (NLP)) can be applied to texts for automated Knowledge Acquisition since very complex domains are often extensively described by collections of text documents. These techniques include semi-automated construction of taxonomies of concepts and subsequent discovery of relations among concepts. Such technology is expected to be very important given that much knowledge in industry is kept in informal natural language repositories. Text mining is the process of deriving high quality information from text. Most of the work on text mining combines statistical analysis with various levels of linguistic analysis.

The area has its origins in techniques from the 1960s and 70s for Knowledge Acquisition for AI (e.g. Semantic Network extraction). Work on Thesaurus Extraction for Information Retrieval addressing the extraction of keywords, thesauri and controlled vocabularies has also been a major influence. More recent development for Lexical Knowledge extraction in NLP led to systems being developed for Extraction of Semantic Lexicons from corpora (a corpora is a large and structured set of texts) e.g CRYSTAL [13]. Current work focuses on a series of increasingly complex processes for ontology learning from texts referred to as the Ontology Learning Layer Cake [14]. Figure 1 illustrates this and the key stages are briefly covered here. Term extraction is the lowest level of the ontology learning process. Terms express more or less complex semantic units. Term extraction can consist of a number of techniques that determine most relevant phrases as terms:

- Statistical analysis (such as the comparison of frequencies between domain and general corpora – e.g. Constructed wetland will be specific to the Water Management domain, while River will be less specific to the Water Management domain). Scores are used such as MI (Mutual Information) used in co-occurrence analysis and TFIDF (Term Weighting which creates a normalised word frequency based on the number of times a word appears in a target document compared with a set of documents).
- Linguistic methods – several layers of analysis including tokenisation, Part-of-Speech and semantic tagging, morphological analysis, extraction of patterns (Adjective-Noun, Noun-Noun etc.) and ignoring names (ICE, IBM etc.) and certain adjectives.

The next step is to identify terms that share (some) semantics, i.e., potentially refer to the same concept (synonyms). As an example, SUDS (sustainable drainage systems) – BMPs (Best Management Practices) used in UK and US practice respectively. These may be multilingual or cross-cultural. Classification (using existing class systems such as WordNet [15]) and clustering (according to similar distributions) are used for this phase. Terms are then candidates for concepts. A concept can be described formally by, for example, a set of instances that the definition of this concept describes. Named-entity recognition and information extraction can then be used to extract instances for a concept from text (e.g. for the River class, extract the River Thames etc.)

Taxonomy extraction produces a taxonomy backbone (is-a relations). Techniques such as lexico-syntactic patterns, distributional similarity and hierarchical clustering (e.g. using Formal Concept Analysis (FCA)) are used. The linguistic paradigms are based on Harris' distributional hypothesis [16] that words that occur in the same contexts tend to have similar meanings: "You shall know a word by the company it keeps" [17]. Relation extraction from text, other than the is-a relation discussed above, has been addressed primarily within the biomedical field due to the ready availability of large numbers of texts. These relations include Part-Of and Attributes (the X of Y). The final level of the Layer Cake is Rules and Axioms i.e. generating logical rules between concepts, but this is an area in its infancy.

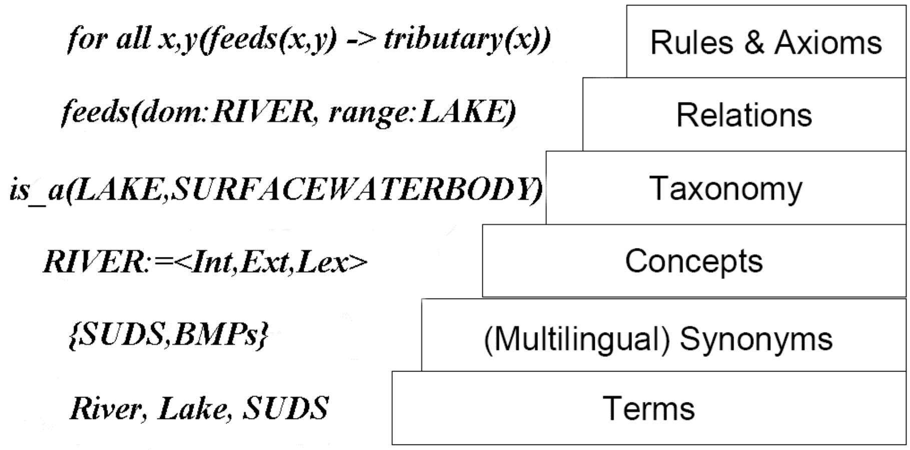


Fig. 1. The Ontology Layer Cake

2.3 Text2Onto

A tool/API called Text2Onto [18] which implements some of the techniques described in section 2.2 was identified as suitable for application to water domain corpora. Text2Onto combines machine learning approaches with basic linguistic processing and incorporates the general purpose ontology WordNet and the GATE (General Architecture for Text Engineering) [19] suite of tools created by the Sheffield University NLP group. Text2Onto introduces two new innovations for ontology learning. Firstly, Probabilistic Ontology Models (POMs) represent the learned knowledge at a meta-level in the form of instantiated modelling primitives which are based on Gruber's Frame Ontology [20] and include concepts (CLASS), the inheritance of classes (SUB_CLASS) etc. By being independent from any particular language, ontology writers can be used to translate to particular representation formats. The POMs also attach a probability to the results learned from the system allowing ranking according to certainty or only showing results above a certain confidence threshold. Secondly, the system monitors for changes in the corpus (documented information) and only recalculates based on those changes, thus avoiding processing the whole corpus every time it changes and enabling the user to trace the evolution of the ontology.

The aim of using software such as Text2Onto is to (semi) automatically generate an ontology. Why do we wish to create such an ontology? Firstly, it makes domain assumptions explicit, providing a community reference point for applications and allows a sharing of a consistent understanding of what information means. Knowledge management is one of the main areas for ontology use. Creating a shared vocabulary (concepts, relations, axioms) of the various actors in a Knowledge management information system allows information to be shared in a more precise way. A by-product of this is knowledge conservation, particularly useful in terms of organisational memory: most big companies lose track of their internal and external data, information, and capabilities. Another consequence is it makes it easier to understand and update legacy data. Ontological software can generate systems that can interface to the organizations' legacy systems and use them as data repositories.

In Text2Onto the user specifies a corpus (a collection of text, html or PDF documents) and starts the graphical workflow editor. The editor allows the selection of algorithms (including combinations) for different ontology learning tasks. The corpus is first pre-processed by a natural language processing component (tokenisation and sentence splitting etc.) before being passed to the algorithm controller. Each algorithm starts by detecting changes in the corpus and updating the reference store accordingly. Finally, it returns a set of requests for POM changes to its caller. After the process of ontology extraction is completed, the POM is presented to the user. Generally, user interaction (e.g. adding or removing concepts, instances or relations) will still be needed for transforming the POM into a high-quality ontology. Finally, the POM can be translated to a suitable representational format (e.g. OWL or RDF) with an ontology writer. It can then be used in any number of applications that need to process the content of information instead of just presenting information to humans including semantic search engines, for filtering and in ontology reasoning products.

3 Application in the WaND project

The part of WaND outlined here aimed to provide support to a range of stakeholders in understanding and implementing water systems that were more sustainable through a flexible framework accessible through the 'portal' - either a website or a stand-alone CD. The stakeholders included: water service providers; local authority planners; housing developers and householders. The portal comprises many hundreds of sources of information and software tools and steering a pathway through these aimed specifically at each stakeholder's individual needs was the primary aim of the ontological investigations.

Text2Onto extracts various types of ontology elements from a given corpus of text documents. The developers (Institute AIFB, University of Karlsruhe) advised the authors on the initial configuration and have been interested in how useful the tool is for application in the water field, as most of the existing experimentation has been based on corpora with technical documents or scientific papers taken from the computer science or knowledge management domains. Several corpora of texts were assembled in the WaND case study. The objective was to achieve the representation of domain knowledge from stakeholder texts (e.g. as regards decision making resources, legislation, organisational memory etc.) in a structured form that would assist a wide range of stakeholders. The corpora were processed with a set of algorithms for each modelling primitive (task). Several algorithms can combine their POM changes in order to obtain a more reliable probability for each primitive. Different pre-defined strategies specify the way individual probabilities are combined and in general a strategy to average the results has been adopted. Corpora 1 consists of WaND project work package summaries concerning techniques and options for more sustainable water management in new developments. Figure 2 shows the top ten concepts based on relevance score for the corpus with the most relevant concept being 'development'. Figure 3 provides the Subclass-of relations by highest relevance, after minor pruning. So for example, 'water supply' is a sub-class of 'water management'.

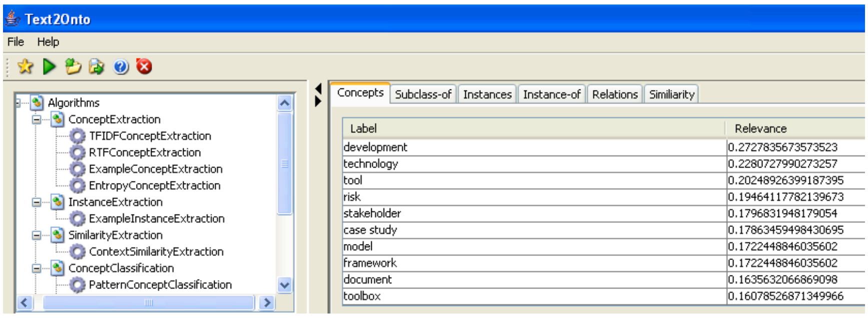


Fig. 2. The ten concepts by relevance for Corpora 1

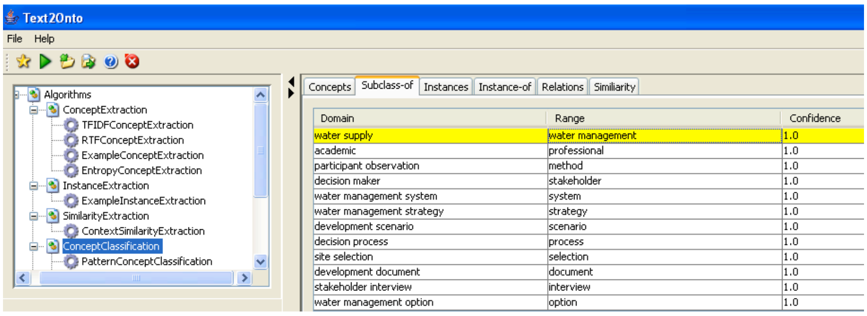


Fig. 3. Subclass-Of by relevance for Corpora 1

Corpora 2 consists of a set of stakeholder texts in this area and included documentation for local authority planners, governmental sustainable development strategy reports, the comprehensive 2006 House of Lords report on water management and interview transcripts on the topic with stakeholders (almost 1000 pages of text in total). The most relevant concept was again 'development'. Figure 4 shows some extracted Instance-Ofs with confidence 1.0. So that 'Breeam' is an instance of 'accreditation scheme' (www.breeam.org).

The extracted primitives for large corpora are, in practice, numerous and some results spurious. Editing is generally required to produce a high-quality ontology. A larger study will need to utilise expert human annotators to evaluate ontologies. Text2Onto was evaluated for the 'knowledge management' domain by five annotators and an average taken of their view on the correctness of the primitives.

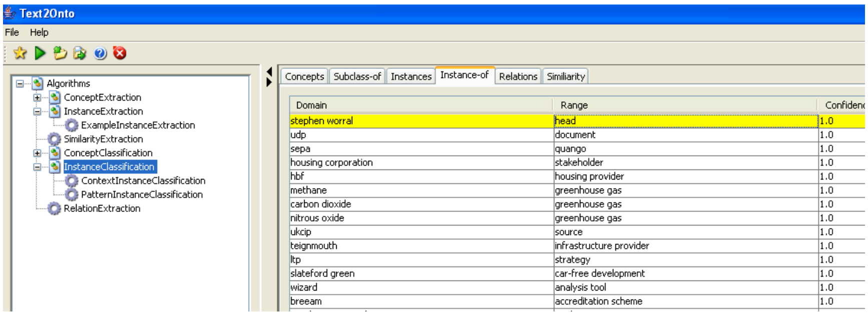


Fig. 4. Instance-of examples for Corpora 2

In practice, ontologies generated in this way were combined with hand constructed taxonomies to form one aspect of a ‘knowledge base’ which was used to create the multimedia portal for the WaND project. This portal provided stakeholder decision-making support and incorporates a rudimentary Expert System that illustrates the alternative routes/options open to various stakeholders faced with different stages in the lifecycle of a new housing development, with a focus on more sustainable water management. By traversing a decision tree comprising questions and organized answers, the stakeholder is helped to proceed in the most sustainable way by being provided with the relevant information, guidance and access to computational and various decision support tools [21]. One way that these decision-making processes of stakeholders can be represented formally is in OWL-DL with a formal reasoner (similar to the biomedical domain).

4 Conclusions

An ontology of a particular domain is not a goal in itself. Developing an ontology is akin to defining a set of data (usually expressed in a logic-based language) and their structure for other programs to use. Ontologies can be seen as meta data that explicitly represent semantics of data in a machine processable way. Problem-solving methods, domain-independent applications such as intelligent search engines and software agents (a piece of software that acts for a user or other program) use ontologies and knowledge bases built from them as data.

In the future, the Semantic Web, constructed from ontologies, will weave together a net linking very large parts of human knowledge and complement it with machine processability. Various automated services will support the human user in achieving goals via accessing and providing information present in a machine-understandable form. This process will ultimately lead to a highly knowledgeable system with various specialized reasoning services that may support human endeavors in nearly all aspects of our daily life. By adopting ontological representations sooner, rather than later, engineers will be able to make full use of these exciting facilities to guide and support their search for knowledge and the way in which this knowledge is used and presented to clients and the wider community.

The work presented here has shown potential for addressing knowledge engineering needs in the sustainable water management domain. It can be seen as providing some first steps in the ultimate aim of developing a comprehensive ontology for this field. This would facilitate greater communication, knowledge sharing and understanding in working towards the complex and challenging aim of sustainable water management.

References

1. House of Lords: Water management Volume 1: report. Science and Technology Committee 8th report of session 2005-6. HL Paper 191-I. The Stationery Office. ISBN 0 10 400871 7 (2006)
2. Berners-Lee, T., Fischetti, M.: Weaving the Web. HarperSanFrancisco, chapter 12. ISBN 9780062515872 (1999)
3. Mounce S.R., Khan A., Wood A.S., Day A.J., Widdop P.D., Machell J.: Sensor-fusion of hydraulic data for burst detection and location in a treated water distribution system, International Journal of Information Fusion, vol. 4, no.3, pp. 217-229 (2003)
4. Noy, F.N., Sintek, M., Decker, S., Crubézy, M., Fergerson, R.W., and Musen, M.A.: Creating Semantic Web contents with Protégé-2000. IEEE Intelligent Systems, vol. 16, no. 2, pp. 60–71 (2001)
5. Ceccaroni, L., Cortés, U. and Sànchez-Marrè, M.: OntoWEDSS: Augmenting Environmental Decision-Support Systems with Ontologies. Environmental Modelling & Software, vol. 19, no. 9, pp. 785-797 (2004)
6. Schwering, A. and Hart, G.: A Case Study for Semantic Translation of the Water Framework Directive and a Topographic Database, Crete University Press. 7th Conference on Geographic Information Science (AGILE), Heraklion, Greece. Seite(n) pp. 503-510 (2004)
7. Scholten, H., Refsgaard, J.C. and Kassahun, A.: Structuring multidisciplinary knowledge for model based water management: the HarmoniQuA approach. Proceedings iEMSs 2004 (2004)
8. Koegst, T., Tranckner, J., Blumensaat, F., Eichhorn, J. and Mayer-Eichberger, V.: On the use of an ontology for the identification of Degrees of Freedom in Urban Wastewater Systems. Proceedings of the 4th International Conference on Water Sensitive Urban Design, Melbourne, Australia (2006)
9. Robinson A D C., Ogunyoye F., Guthrie G., Williams A., Morris J., Romanowicz A., Ashley R M., Cashman A., Blanksby J R.: Improving Data and knowledge Management for Effective Integrated Flood and Coastal Erosion Risk Management. Volume 1. Development of Ontology. Defra/Environment Agency R&D Technical Report FD2323/TR1 (2006)
10. Annoni, A., Bernard, L., Douglas, J., Greenwood, J., Laiz, I., Lloyd, M., Sabeur, Z., Sassen, A.M., Serrano, J.J.: Orchestra: open architecture and spatial data infrastructure for risk management, Proceedings of the International Conference on Risk and Emergency Management, Hannover, Germany, 8-9 June 2005 (2005)
11. Noy, N.F. and McGuiness D.L.: Ontology Development 101: A Guide to Creating Your First Ontology by Noy, N.F. and McGuinness, D.L. SMI technical report SMI-2001-0880, Stanford University (2001)
12. Brewster, C., Iria, J., Ciravegna, F. and Wilks, Y.: The Ontology: Chimaera or Pegasus. Machine Learning for the Semantic Web Dagstuhl Seminar 05071, Dagstuhl, DE (2005)

13. Soderland, W., Fisher, D., Aseltine, J. and Lehnert, W.: CRYSTAL: Inducing a Conceptual Dictionary. In Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI), pp. 1314-1319 (1995)
14. Buitelaar, P., Cimiano, P. and Magnini, B.: Ontology Learning from Text: Methods, Evaluation and Applications, Volume 123 Frontiers in Artificial Intelligence and Applications, IOS Press (2005)
15. WordNet, http://www.cogsci.princeton.edu/~wn
16. Harris, Z.: Distributional structure. Word, vol. 10, no. 23, pp. 146-162 (1954)
17. Firth, J. R.: A synopsis of linguistic theory. In J.R. Firth et al. Studies in Linguistic Analysis. Special volume of the Philological Society. Oxford: Blackwell (1957)
18. Cimiano, P. and Voelker, J.: Text2onto - a framework for ontology learning and data-driven change discovery. In Proceedings of the 10th International Conference on Applications of Natural Language to Information Systems (NLDB'2005) (2005)
19. GATE, http://www.gate.ac.uk.
20. Gruber, T. R.: A Translation Approach to Portable Ontology Specifications, Knowledge Acquisition, 5(2), pp. 1999-2220 (1993)
21. Hurley, L. Mounce, S.R., Ashley, R., Butler, D. and Memon, F.: Flexible Process Frameworks for the Evaluation of Relative Sustainability in Water Management Decisions. International Journal of Technology, Knowledge and Society. Vol. 2, Issue 7, pp. 57-67 (2007)

Future Perspectives in Ontologies for Urban Regeneration

Francesco Rotondo

Department of Architecture and Town Planning, University of Bari Polytechnic – via Orabona 4, 70126, Bari, Italy Corresponding author.
Tel.: +39 347 5957508; fax: +39 80 5963348.
f.rotondo@poliba.it

Abstract. This paper, starting from previous experiences in the construction of an urban regeneration ontology (OUR, Rotondo, 2008), search for future perspectives in OUR development. An initial discussion on the aspirations of urban regeneration and of the theoretical background and possible definitions of an ontology is followed by a description of useful results obtained in prior experiences in which a prototype of an ontology for urban regeneration has been developed. Starting from these experiences it has in turn led to the clarification of particular user needs and requirements. Interesting results and suggestions of possible user needs and requirements are discussed in the paragraph "User needs and requirements for ontology specification". Future perspectives in ontologies for urban regeneration are examined in the final paragraph.

Keywords: Urban Regeneration; Ontology; User needs

1. Why It Could Be Useful Developing an Ontology for Urban Regeneration (OUR)

The European Union (EU) Community Strategic Guidelines 2007-2013 place particular emphasis on the specific needs of certain zones, such as urban and rural areas. The guidelines encourage an "integrated approach" towards cohesion policy, not only stimulating growth and creating jobs, but also pursuing certain social and environmental objectives.

Furthermore, the European Parliament, in its report on the urban dimension within the context of enlargement[2], welcomed the incorporation of sustainable urban development in cohesion policy.

The success of the URBAN Community Initiative[3] is in no small measure due to such an integrated approach. URBAN has targeted social and economic cohesion in

[2] Report on the urban dimension in the context of enlargement, rapporteur: Jean Marie Beaupuy, A6(2005) 0272 on 21.9.2005.

[3] The Community Initiative URBAN II (2000-2006): Communication from the Commission to the Member States of 28 April 2000 laying down guidelines for a Community initiative

parallel, thus removing barriers to employability and investment whilst at the same time focusing on social and environmental goals. The mobilisation of a broad range of partners with different skills has underpinned this approach. Forthcoming EU urban regeneration policies attempt to consolidate these successes with new policy initiatives directed towards the regeneration of deprived urban areas, with the JESSICA[4] policy a case in point. In this context, ontologies could play a significant role in developing and managing these new policies thereby strengthening integration, sharing ideas and increasing knowledge of problems specific to urban regeneration. Furthermore they could, in the context of the European Union, present a multilingual tool capable of demonstrating concepts, shared definitions and the relationships between them.

At present, institutions dedicated to the management of regeneration policies at all levels, whether EU, regional or municipal, often demonstrate real difficulty in terms of interpreting the language used by an architect, a planner, an ecologist or an economist due to discipline-specific terminology. Urban regeneration may therefore mean different things in different disciplines. Ontologies could be a useful tool in ordering, integrating and making transparent a range of possible meanings associated with a policy. It would appear to be particularly useful within the European context, where the coherence of different actions, even within the same field, is often difficult to establish whether at a European or municipal level.

2. Significant Elements of Urban Regeneration and its Multiple Dimension

Urban regeneration is an integrated urban policy approach mixing multiple dimensions: economic, social, cultural, spatial and environmental. New urban planning and design methods replaced rationalist architectural codes and conventions by locating some key points which, when seen alongside the *Leipzig Charter on sustainable European cities*[5], echo the aspirations of urban regeneration policies and strategies. in response to the Leipzig Charter, the same questions where developed during the Slovenian Presidency of the Council of the European Union in 2008, as demonstrated by the Ljubljana Declaration on Urban Regeneration & Climate Change[6].

This, may be expressed, for example, in the upgrading of the physical environment and encouraging sustainable urban transport, the strengthening of the local economy and labour markets, or in the promotion of proactive education and training policies for children and young people in deprived urban areas.

---

concerning economic and social regeneration of cities and of neighbourhoods in crisis in order to promote sustainable urban development (URBAN II), C(2000) 1100 of 28.04.2000.

[4] Joint European Support for Sustainable Investment in City Areas

[5] the *Leipzig Charter* is a document by the Ministers responsible for urban development policy of the EU member states, edited in its final draft version on 2 May 2007, available on line at: http://europa.eu.int/comm/regional_policy/index_en.htm.

[6] Ljubljana, 17 June 2008

The recommendations of the Charter summarize strategies put forward by the New Urbanism movement (Dutton, 2000) or those already declared in the *New Charter of Athens* (2003) by the European Council of Town Planners (2003) and embrace:

The creation of high quality public spaces and their reinforcement within city structure;

The improvement of energy efficiency in buildings and the modernization of infrastructure networks favouring a compact city form;

The use of greenery as a bio-infrastructure to enhance the sustainability of cities;

The encouragement of mixed-use neighbourhoods, buildings and blocks (shops, offices, apartments, and homes on the same site), with a wide diversity in demographic make-up in terms of age, income level, culture, and race;

The promotion of well-conceived social housing policies, with suitable and affordable housing;

The participation in urban policies which lead to a better level of education and training contributing to achieve their ambitions and to ensure equal opportunities on a long-term basis.

All of these elements are of crucial importance to deprived urban neighbourhoods not only in reducing inequalities but also in preventing social exclusion and improving the physical environment. Indeed, new EU initiatives, JASPERS[7] and JESSICA, and several particular measures of the European Regional Development Funds (ERDF) will support, from 2007 to 2013, exactly such policies within urban regeneration.

The multi-dimensional nature of urban regeneration processes encouraged by the European Union is therefore evident when seen in the context of the concrete objectives of urban regeneration itself and the support of specific European policies which target those objectives. To a region implicated in such European policy making it is therefore of primary importance to manage the multi-dimensional nature of the problem, by drawing upon different skills and competencies and sharing the same words and objects whether physical, economic or social (EC, 2006). Ontologies could be a potential way of organizing this complex and multifaceted task, as we attempt to outline in the following paragraphs.

3. A Possible Ontology for Urban Regeneration (OUR)

According to Gruber (1993), an ontology is an explicit, formal and shared conceptualization of a particular domain. The conceptualization process represents the attribution of unambiguous meanings to terms defining knowledge in that precise domain (domain ontology). Guarino (1998) defines an ontology as a set of logical axioms designed to account for the intended meaning of a vocabulary.

A domain ontology for urban regeneration is therefore expected to express the viewpoints and satisfy the informational needs of multiple stakeholders and interest groups, including, yet by no means limited to, town planners, environmental agencies, municipalities, police departments, owners and sellers of real estate, third sector

[7] Joint European Resources for Micro to Medium Enterprises

associations. These actors use different jargons and pursue different, occasionally conflicting tasks, even if they manage similar or related domains. Reports of the experience of ontology development in many fields of application[8], underline that different jargons and informational needs are hard to accommodate in a consensual ontology. In the case of OUR, it is not strictly necessary to arrive at a unique definition of a term, if, as may be the case, the obtaining of a common definition proves impossible. Rather, it is sufficient that all agents involved in the same or similar activities are at least informed and have the possibility of knowing others definitions. For example, in an "Objective 1" region such as Apulia in Italy, it is necessary that the Environmental Management Department, the Regional Planning Office, the Transport Bureau and the Economic Planning Bureau are aware of other definitions thus avoiding a scenario in which each of them promotes different and possibly incoherent funding policies on the same urban regeneration objective. Such a situation is not simply theoretical. Consider, for example, the different funding for the construction of cycle paths within the previous phase of cohesion policies (2000-2006), which may be considered one of the objectives of urban regeneration. Promoted by the Transport Bureau, this particular objective may take on a more functional image, useful in increasing the possibilities for movement in an urban context. This clearly does not correspond with an altogether different definition of a cycle path associated with leisure and nature, which could be built with natural materials such as compact sand and bordered by green hedges, perhaps intended as tourist routes through the countryside. Indeed, such a vision including all of the above elements was conceived during the same policy phase by the Environmental Management Office. In the same period the Regional Planning Office promoted urban regeneration processes in which it funded the same objective, in this case encouraging an alternative method for commuters to reach the work place, schools etc. With this in mind, an ontology for urban regeneration related to regional geographical information systems has the potential to begin forming a common platform for spatial policies. It would also perhaps avoid a repeat of such ineffective communication and a lack of knowledge, especially within an organization of any considerable size (as for example regional institutions). Indeed, this has already been highlighted by various experiments carried out by the Italian Ministry of Innovation through its centre for Informatics in Public Administration (CNIPA in the Italian acronym[9]), in a project entitled "ARIANNA"[10]. According to Barone and Di Pietro (2005), ontologies could introduce administrative simplification and efficiency, better control of public expenditure and the simplification of policy management. Moreover, a domain ontology could demonstrate all of the different definitions of a given policy and its corresponding objective, in a semantic integration of a certain number of databases, and with the software Towntology[11], it also offer the possibility of using images to represent concepts, thereby creating a visual dimension which can potentially offer a different pathway towards the same objective.

---

[8] Examples of ontologies are available at http://protege.stanford.edu/download/ontologies.html

[9] www.cnipa.it

[10] available at: http://arianna.diviana.net/

[11] release 2.0 produced by the LIRIS, Institut National des Sciences Appliquées de Lyon.

4. Future Perspectives in Ontologies for Urban Regeneration: the User Point of View

After these first, yet significant, experiences (Rotondo, 2008; Berdier, C., Roussey, C. (2007) in collecting impressions from participants it is possible to outline some user requirements:

- OUR could perhaps be of most use if used as an integrated tool within usual policies and policy making, rather than as an exceptional instrument;
- In order for OUR to be effective it has to be available on the web or at least on the intranet of the public office or institution involved;
- A unique multitask interface could be developed within the "Towntology" software with which the user would simultaneously be able to locate the list of terms (possibly with a multilingual description), their definitions, their relationships and any associated imagery where applicable.

As it seems it has been done in the release n°2 of the software Towntology represented in the following image.

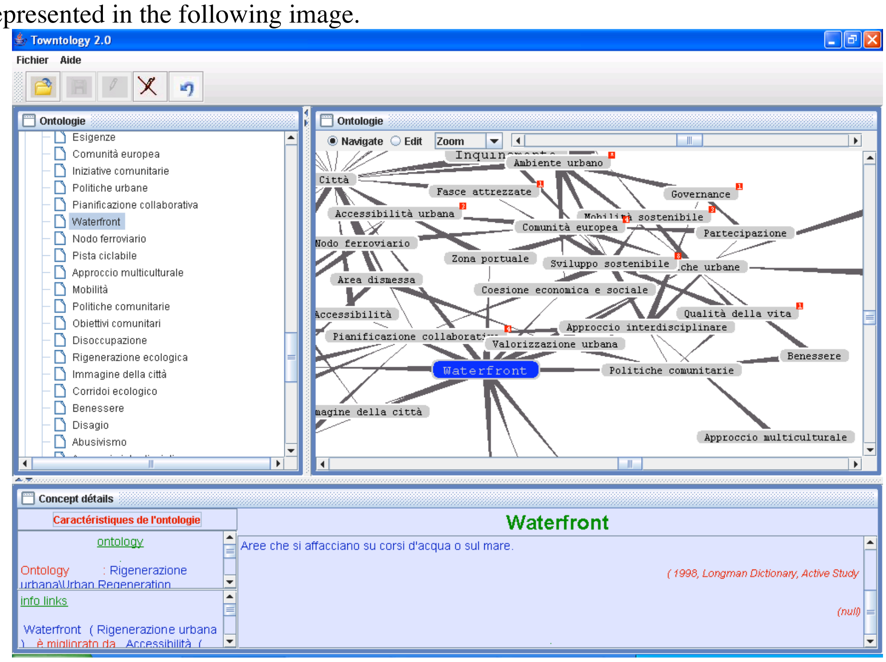


Fig. 1. The new interface of towntology software.

The availability of multiple on-line ontologies takes on a particular importance, especially when considering EU policies given that:

a) Public organizations are predominantly divided into a range of departments with a high level of specialization yet a low level of integration. If ontologies were to become an integrated tool which could be applied to even standard policy, or better still if applied on a GIS, public organizations could potentially arrive at a greater integration of policy content.

b) Ideally, OUR would be available on-line or at least on the intranet of the organization in question, as its value is determined by the possibility of being utilized by anyone involved in urban regeneration regardless of their physical working location**.** In this way the glossary will grow and every definition could be discussed and eventually shared in a unified way.

c) EU cohesion policies are frequently multifaceted and complex, often with various possible implementation choices, deriving in part from the particular characteristics of the nation in which it is applied.

In the case therefore of single large-scale organizations, ontologies could lead towards a better cohesion in the way that different member states may apply the same EU policy.

Although yet to be sufficiently experimented and diffused, considering the initial results of first experiences, it may be possible to assume that OUR is potentially a tool which could foster improved communication between stakeholders. Possible future directions for research in the field of ontologies for urban regeneration with reference to EU policies could be a compared evaluation between ontologies as seen within different languages and cultures as, for example, with a French urban renewal ontology, as has been developed by the EDU Laboratory in Lyon (Berdier, Roussey, 2007), alongside another in English thereby making an ontology available in the official language of EU. A step beyond this would perhaps be the conception of a more extensive experiment involving EU offices, in which regeneration policies are developed and managed. Te possibility of using OUR in practice is strictly related to the wider diffusion of ontologies within public administration routine. From the first definition by Gruber (1993) of an ontology in the sphere of Artificial Intelligence, only within the last few years have we seen some experiences. The greater the increase in the availability of data sets, the more an ontology lends itself to being a useful instrument in providing clear definitions and corresponding relationships within a specific domain.

References

1. Barone, A., & Di Pietro, P. (2005). "Ontologia. Elemento Chiave per la Costruzione della Società della Conoscenza". *IterLegis 5-6*, 9-15.
2. Berdier, C., Roussey, C. (2007). "Urban Ontologies: the Towntology Prototype towards Case Studies". In J. Teller, J. Lee & C. Roussey (Eds.), *Ontologies for Urban Development* (pp. 143-155). Berlin: Springer.
3. European Council of Town Planners. (2003). *The New Charter of Athens 2003*. Firenze: Alinea.
4. Dutton, J. A. (2000). *New American Urbanism: Re-forming the Suburban Metropolis*. Milano: Skira.
5. European Commision, Commission Staff Working Document (2006). *State Aid Control and Regeneration of Deprived Urban Areas. Vademecum*. Document available on-line at: http://europa.eu.int/comm/regional_policy/index_en.htm.
6. Gruber, T. R. (1993). "A Translation Approach to Portable Ontologies". In *Knowledge Acquisition*, 5(2), 199-220.
7. Guarino, N. (1998). "Formal Ontology in Information Systems". In N. Guarino (Ed.), *Proceedings of FOIS'98, Trento, Italy, 6-8 June 1998* (pp. 3-15). Amsterdam: IOS Press.

8. Rotondo F. (2008), "Ontologies for urban regeneration: opportunities and weakness for their development in cohesion policies for cities", in Teller, J., Tweed, C., Rabino G. (Eds.), *Conceptual Models for Urban Practitioners*. Società Editrice Esculapio, Bologna, ISBN 88-7488-208-4, pp. 149-158

Ontologies for the Classification of Urban Characteristics: Opportunities for Urban Designers

Simon Lannon and Orly Linovski

Welsh School of Architecture, Cardiff University
King Edward VII Avenue
Cardiff, UK CF10 3NB
linovskio@cardiff.ac.uk, lannon@cardiff.ac.uk

Abstract. Urban designers and morphologists are often hampered by a lack of easily obtainable information about the city fabric. Data often has to be collected manually in a time and labour-intensive process. This paper examines the opportunities for integrating ontological systems into the research process for urban designers and researchers. Using the case study of urban energy efficiency modeling, the possible benefits and limitations are discussed. Through the creation of a system that focuses on the 'building blocks' of city form, useful information could be easily collected for urban designers and other researchers. It is believed there are significant instances where an ontological approach could aid in the study and design of cities.

Key words: urban design, morphology, design research, urban modeling, design indicators

Research into the built form of cities is an integral part of the decision-making process for urban planners and designers. This research often lays the foundation for future design work as well as influencing policy decisions. Despite the critical importance of these types of studies, there is little guidance or direction as to how such studies should be conducted. Indeed, while there is a significant history of different approaches to urban design research, including morphological studies and space syntax, there are few that can accommodate a more global view of city form. While scholars such as Whitehand and Carr [1999, 231] bemoan the lack of attention paid to the built form of British cities, a fair amount of this can be attributed to the lack of easily obtainable information. This paper will examine the opportunities for applying an ontological system of classification to aid in the design research process and the benefits associated with this type of research method. Through an examination of an energy efficiency case study, it becomes clear that there are significant advantages to incorporating an ontological system into the urban design process and research.

1. Design Research Studies

For professionals working in the built environment field, the background research process (referred to as the 'design research stage' here) informs a variety of later-stage decisions. Historically, researchers have dealt with the need for information in several ways [Aspinall and Whitehand 1980; Kohler and Hassler 2002; Whitehand and Carr 2001]. Because of the idiosyncratic nature of these approaches, transparency may be sacrificed in favour of methods that are considered as more technically rigorous. The following brief discussion of typo-morphology studies illustrates some of the shortcomings associated with traditional research methods.

1.1 Typo-morphology Studies

Urban morphologists try to gain an understanding of the city, and the role of humans within it, through analysis of its form. As complex social and economic forces change the urban landscape, it is believed that the study of the resulting built form can offer new insight into these processes. Morphological studies often offer the most visual and direct method of understanding the urban fabric. From Nolli's first figure ground studies of the courtyards and cathedrals of Rome to more recent experiments with space syntax, visual expressions of city space are often the most resonant.

The process of anayzing information about the built environment has traditionally been a time-intensive exercise due to the complexity in obtaining raw data. Moudon-Vernez explains how morphologists approach city form: "… they analyze a city's evolution from its formative years to its subsequent transformations, identifying and dissecting its various components. [It is viewed as] the accumulation and the integration of many individual and small group actions, themselves governed by cultural traditions and shaped by social and economic forces over time" [1997, pg.3]. From this definition of urban morphology, it becomes clear that the information required for this type of study is significant. The emphasis on not just the current built form but on how it has *changed* over time means that very specific types of information are required. For example, Whitehand and Carr's [2001] in-depth study of post-war change in British suburbs was based on field surveys, extensive interviews and detailed examination of building permits, city plans and survey maps. Other similar studies have required equally as intensive data collection.

2. Ontologies

The use of ontological systems, and geographical information sciences more broadly, provide a rich source for the practice of urban planning and design. However, it is often felt that there is a mismatch between the needs of users and the design of systems. As Kuhn explains: "information system ontologies, as they are understood and designed today, contain a view of the world that has less to do with human activities than with existing data holdings" [2001, pg. 613]. There is a valid argument for more thought to be given to the needs of end users at the initial planning stages of geographical systems. An important consideration is the different information needs of built environment professionals. The needs of a transport engineer are quite different from those of an urban designer or planner. Guarino's distinction between an application ontology from a generic knowledge base emphasizes the community of users. While a generic knowledge base is considered to describe facts that are always true for a community of users, ontological definitions can shift depending on the group of users. This is what informs the desire to accommodate the needs of urban design specialists when designing an ontological system.

3. Opportunities for Integration

3.1 Energy Efficiency Modelling Case Study

Energy efficiency research often requires specific information about the built environment that is difficult to obtain. For example, the approximate age of a building can provide much of the information needed for energy modeling but is currently only available through costly and time-intensive manual surveying. Modeling software has been developed to quantify energy emissions but requires this detailed information of the built environment and age of buildings (Jones et al. 2000).

Research has been undertaken evaluating the use of GIS to collect required age data from land survey maps. While the collection of certain types of information (such as the size of buildings and the distance between buildings), can be automated and obtained from digital maps it is unclear how to relate this information to eras of development. First iterations of this model involved relying on the 'expert' classification of built form characteristics into development periods. While this technique is relatively reliable, it is idiosyncratic in nature and difficult to replicate.

The process for determining when a building or community was constructed has a strong foundation in the methods of urban morphologists. Similar to the work done by these scholars, it is rooted in the belief that social and economic factors informed and shaped the built environment. Where there have been major societal shifts, likewise there are changes in the way buildings are constructed. While these changes can be brought on by legislation or other government-led initiatives, it is often fundamental changes in societal values that are reflected by changes in government policy. The starkest of these changes is the Tudor-Walters report which espoused the reaction against the dense terrace housing of the pre-war period [Barrett and Phillips 1987; Colquhoun 1999]. This report, published in 1918, advocated for lower densities, minimum distances between homes and larger living spaces, was in response to growing opinion that pre-war housing was 'unhygienic' and 'unhealthy'. Yet, government suggestions (the Tudor-Walters report had not been mandated into law) were insufficient to promote a wide-spread change in domestic building without accompanying economic stimulus. It was when the cul-de-sac became more cost-effective than the standard pre-war terrace, that the massive shift in domestic building began [Barrett and Phillips, 1987]. This example of the changes that occurred during the inter-war period demonstrates the sort of information that can be gleaned from a careful analysis of the urban fabric. It is the process of extricating and processing this data that provides opportunities for the integration of ontological systems.

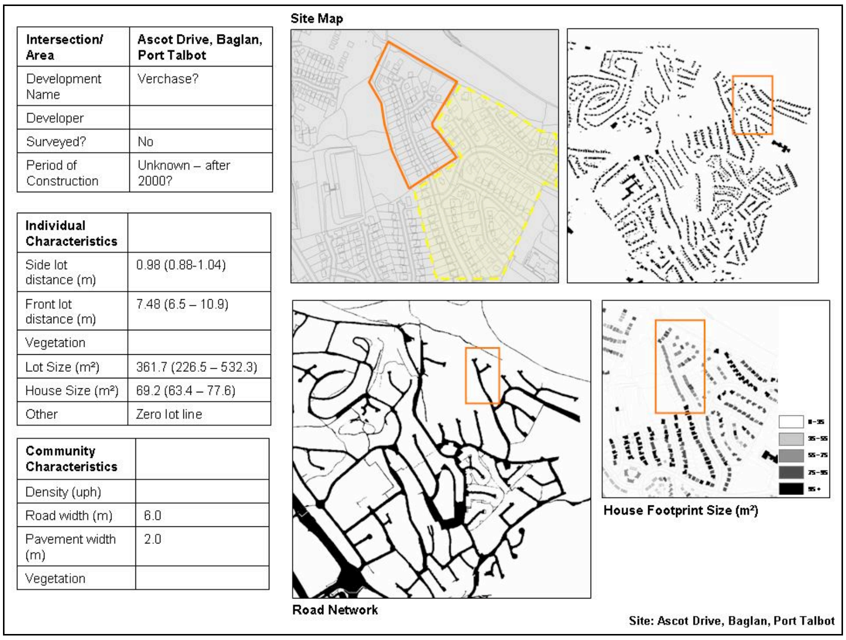
3.2 Methodology

Two study areas in Wales, UK were chosen for in-depth analysis of the built form characteristics that could be utilized by urban designers and morphologists. As the intention was to provide information for energy efficiency modeling, it was necessary to identify characteristics that would assist in determining the age of buildings and communities. The age of buildings is strongly linked to changes in building materials and construction methods, making this an important data source for modelling energy use. For example, homes built before World War I have solid wall construction while those built after 1919 generally have cavity walls. This makes a considerable difference in the energy profile of the building. Therefore, knowing the age of buildings provides a useful proxy for determining the energy efficiency of buildings on a large-scale.

3.3 Creating Ontologies for Urban Designers

While some of the changes that occurred in domestic building have been well-documented, mostly in the pre- and inter-war periods [for example: Edwards, 1981; Oliver, 1981; Whitehand and Carr, 2001; Harris and Buzzelli, 2005], other periods have not been studied as intensively. Therefore, it was necessary to identify the built form factors that could be indicative of built form changes in other periods. The next section will describe the method for identifying these factors, and how ontological systems may be integrated into this process to aid the work of urban designers and researchers.

Based on the existing literature related to domestic architecture and urban design, a series of case studies were undertaken in selected areas where the approximate age of buildings was known. Around 20 different indicators were measured (such as the average distance between buildings and average plot size) and compared between ages to identify the most important characteristics [see Figure 1]. The analysis of these factors gave some direction for the creation of ontologies for urban designers and morphologists. Secondly, a more global analysis of these factors was undertaken to gain a better understanding of how indicators interact. For example, Figure 2 visually displays the distances between building and road in the Wales case study area. In this type of analysis it becomes clearer that this factor is an important indicator of construction date; there is a distinct pattern of similar distances being grouped by building age.



Intersection/ Area	Ascot Drive, Baglan, Port Talbot
Development Name	Verchase?
Developer	
Surveyed?	No
Period of Construction	Unknown – after 2000?

Individual Characteristics	
Side lot distance (m)	0.98 (0.88-1.04)
Front lot distance (m)	7.48 (6.5 – 10.9)
Vegetation	
Lot Size (m²)	361.7 (226.5 – 532.3)
House Size (m²)	69.2 (63.4 – 77.6)
Other	Zero lot line

Community Characteristics	
Density (uph)	
Road width (m)	6.0
Pavement width (m)	2.0
Vegetation	

Fig. 2. Example of a case study worksheet for different development periods. Factors analyzed include lot size, road width and distance between buildings.

Fig. 3. Distance from building to road (meters)

There are several conclusions that can be drawn from the built form analysis. It is hoped that these ‘built form lessons’ can provide greater guidance for the structuring of future urban ontologies.

Firstly, often the relationship between buildings is more important than the buildings themselves. Urban design is commonly defined as the design of the space between buildings. It is this space that is of most importance for designers and morphologists. An ontological system that is able to accommodate information about the complex relationships at different levels, for example neighbours to block level to community level, would be of great use to those working with the built environment. In our example of development period modeling, one of the strongest predictors of age is home many homes a terrace house is linked to. In this case, it the relationship between one house and many others that is of importance. The concept of nearness, or more specifically spatiality, is integral to the understanding required by designers.

Secondly, the interaction between different types of the built environment is also of importance. In this sense, one must understand the different levels of the built environment and how they interact with each other. In the case study, it was shown that the relationship between buildings and the streets they abut are important indicators of different development periods. As shown in Figure 3, buildings constructed prior to WWI have a clear relationship with the street with almost all homes at a 90° angle to the road. In contrast, buildings constructed in recent times have a much less clear relationship with the street. The differences in the manner of constructing communities can be seen as an expression of changing tastes and economic circumstances. This sort of information contained within the built environment can provide a strong foundation for the work of designers.

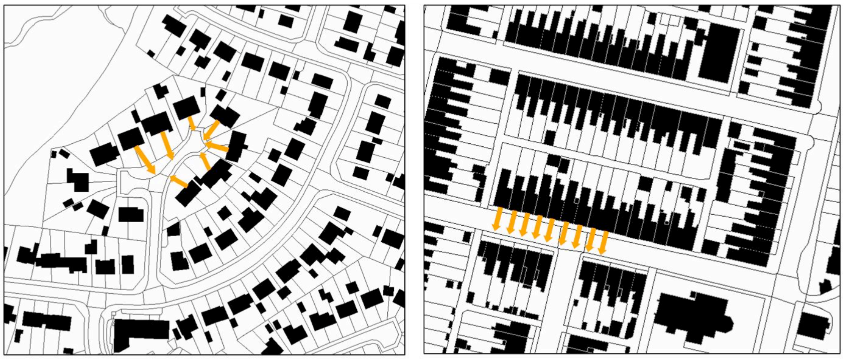


Fig. 4. Relationship between building and road in a contemporary development *(left)* and pre-1919 housing *(right)*

The last indicator, and perhaps the most difficult for future ontological systems to incorporate, is an ability to describe and accommodate change over time. As mentioned previously, morphologists are primarily concerned with how the urban fabric changes over time and how this can be interpreted. With the increasing digitization of historic maps (for example, surveys from as early as the 1800s are available for many parts of the UK), this provides an excellent opportunity to incorporate this data into ontological systems. A simple demonstration of this is seen below in Figure 4. In this example, the historic map from 1919 is overlaid with the current built form. The integration of historic data provides a significant opportunity for allowing a more nuanced view of city form.

Fig. 5. Historic survey map from 1919 overlaid with current Ordnance Survey map

4. Limitations

While the opportunities for integrating ontological systems into improved tools for urban designers and researchers are substantial, there are also limitations that should be addressed. Data quality often varies significantly between countries and even cities. While the case study presented used data that is available for all of the UK, this may not be the case in other areas. Secondly, careful attention should be paid to the eventual end user of any ontological system. This case study was specifically tailored for urban designers and those studying urban morphology and as such may be considered quite narrow. Analysis of user's needs can ensure that ontological systems are of use to urban professionals.

Conclusions

It is believed that a more robust method for identifying built form characteristics is possible using an ontological approach. Through the integration of information that addresses the space between buildings, the relationship between different urban forms and historical data, urban ontological systems can provide a viable tool for professionals in the built environment field. While urban researchers have often struggled with laborious processes for acquiring and analyzing built form data, the careful construction of ontological systems can assist in these processes.

References

1. Aspinall, P. J. and Whitehand, J. W. R. 1980. Building Plans - a Major Source for Urban Studies. *Area* 12(3), pp. 199-203.
2. Barrett, H. and Phillips, J. 1987. Suburban Style: The British Home, 1840-1960. London: McDonald & Co.
3. Colquhoun, I. 1999. RIBA Book of 20th Century British Housing. Oxford: Butterworth-Heinemann.
4. Edwards, A. M. 1981. The Design of Suburbia: A Critical Study in Environmental History. London Pembridge Press.
5. Fonseca, F. T. et al. 2000. Ontologies and Knowledge sharing in urban GIS. *Computers, Environment and Urban Systems* 24(3), pp. 251-272.
6. Guarino, N. 1998 Formal ontology and information systems. In FOIS'98 (Trent, Italy: IOS Press), http://www.loa-cnr.it/Papers/FOIS98.pdf
7. Jones, P. et al. 2000. Planning for a sustainable city: an energy and environmental prediction model. *Journal of Environmental Planning and Management* 43(6), pp. 855-872.
8. Kohler, N. and Hassler, U. 2002. The building stock as a research object. *Building Research & Information* 30(4), pp. 226-236.
9. Kuhn, W. 2001. Ontologies in support of activities in geographical space. *Int. J. Geographical Information Science*. 15(7), pp. 613-631
10. Smith, B. and Mark, D. M. eds. 1998. *Ontology and Geographic Kinds*. International Symposium on Spatial Data Handling. Vancouver. International Geographic Union.
11. Vernez Moudon, A, 1997. Urban morphology as an emerging interdisciplinary field. *Urban Morphology*. 1, pp. 3-10
12. Whitehand, J. W. R. and Carr, C. M. H. 1999. Morphological periods, planning and reality: the case of England's inter-war suburbs. Urban History 26(2), pp. 230 - 248.
13. Whitehand, J. W. R. and Carr, C. M. H. 2001. *Twentieth-Century Suburb: A Morphological Approach*. London: Routledge

Ontology-Based Models for Improving the Interoperability of 3D Urban Information

Claudine METRAL[1], Roland BILLEN[2], Anne-Françoise CUTTING-DECELLE[3], Muriel VAN RUYMBEKE[4]

[1]Centre Universitaire d'Informatique - Pôle en Sciences de l'Environnement, University of Geneva, 7 route de Drize, 1227 Carouge, CH
claudine.metral@unige.ch

[2]Unité de Géomatique, University of Liège, 17 Allée du 6-Août, 4000 Liège, B
rbillen@ulg.ac.be

[3]Ecole Centrale de Lille / Laboratoire de Génie Industriel de Lille, Cité Scientifique, 59651 Villeneuve d'Ascq, F
afcd@ec-lille.fr

[4] European Centre of Archaeometry, University of Liège, 17 Allée du 6-Août, 4000 Liège, B
mvanruymbeke@ulg.ac.be

Abstract. 3D geodata are more and more available as well as realtime visualization possibilities with free three-dimensional viewers such as Google Earth. This implies a growing demand of 3D city models, which are 3D representations at the scale of the city. Despite their intended wide range of applications, such models cannot be used for many urban tasks as they cannot represent the urban information associated with these tasks. On the contrary, ontologies have proven their capacity and usability in the representation of information and knowledge of various domains. In this paper we will present, on the basis of case studies, how ontologies can overstep the semantic limitation of 3D city models and how ontology-based models can be interconnected thus increasing the interoperability of urban information.

Keywords. 3D city models, knowledge, ontologies, interoperability, information

1. Introduction

A 3D city model is a digital mock-up containing the 3D representation of the geometric elements of a city, such as buildings, terrain, streets or vegetation. An increasing number of cities and companies are building 3D city models all around the world. The intended applications are wide (3D cadastre, disaster management, mobile telecommunication, vehicle and pedestrian navigation, tourism, etc.), the main application being urban planning. If the first 3D city models were centered on the geometrical aspects, there is now a trend towards models including semantic and topological aspects. CityGML, the newly standard for 3D city models, emphasizes such aspects.

We argue that CityGML is insufficient for representing the semantics of urban information and thus that 3D city models based on CityGML are insufficient for being used in many urban tasks. For example urban projects involve many actors ranging from urban planners to inhabitants, and many tools ranging from plans (traditional tools for specialists) to 3D representations (more suited to the general public). So using 3D city models is a good way for communicating urban projects to inhabitants. Let us take the example of transportation issues. The transportation feature of City GML, that consists in infrastructure aspects (such as roads) and in more advanced aspects (such as TransportationComplex associated to a function and a usage) cannot represent many transportation or mobility issues such as soft mobility aspects. Archaeology is another case where the package of CityGML features does not offer the appropriate tools to incorporate the whole knowledge generated by data.

These semantic gaps can be filled in by defining and using urban ontologies for representing the different types of urban information. These ontologies can be connected to CityGML (represented itself as an ontology) in order to benefit from the city objects and attributes defined in CityGML, particularly the geometry and appearance of these objects. By defining models based on these urban ontologies, we obtain not only semantically enriched 3D city models that can be used for various urban applications but also interconnected models, thus contributing to the interoperability of urban information.

In this paper we (1) briefly describe the semantics of CityGML, (2) describe, on the basis of case studies, which urban information is necessary to explicitly and formally define ontologies as well as the resulting models and their applications, (3) explain how and why these experiments could be generalised to improve the interoperability of 3D urban information.

2. CityGML

In August 2008 the OGC (Open Geospatial Consortium Inc.) defined an OpenGIS standard for 3D city models named CityGML (City Geography Markup Language). CityGML defines the classes and relations for the most relevant topographic objects in cities and regional models with respect to their geometrical, topological, semantic and appearance properties [1]. CityGML also differentiates five Levels of Detail (LOD) ranging from LOD0 to LOD4. As city objects become more detailed with increasing LOD, their geometry and their thematic is differentiated. A building, for example, is represented in LOD1 as a block with flat roof while having differentiated roof structures (such as overhangs or antennas) and thematically differentiated surfaces (representing walls, roofs, etc.) in LOD2 and in higher LODs. In LOD0, transportation complexes are modeled by center lines, thus establishing a linear network. In LOD1 and in higher LODs, a TransportationComplex provides a surface geometry describing the actual shape of the object. In LOD2 and in higher LODs, it is further subdivided thematically into TrafficArea (representing the areas used for the traffic of cars, trains, public transport, airplanes, bicycles or pedestrians) and AuxiliaryTrafficArea (associated to grass for example). The figure 1 below shows the transportation model of CityGML, as defined in UML (Unified Modeling Language) [2].

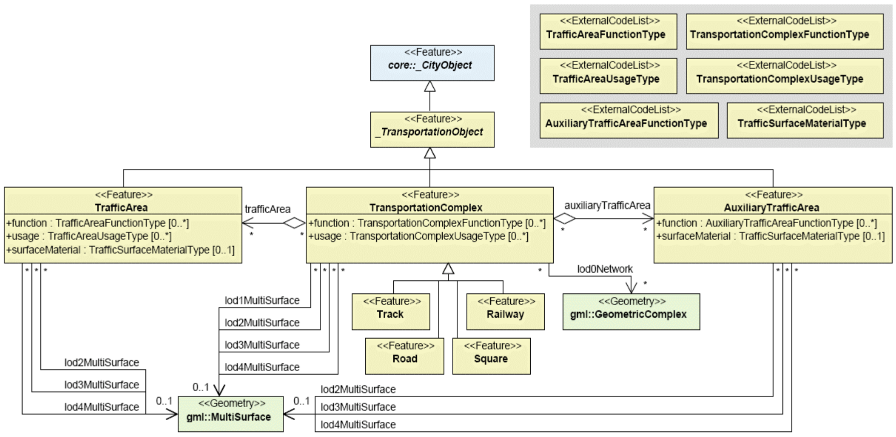


Fig. 1 : UML class diagram of the transportation model in CityGML

In CityGML, the city objects (relief, buildings, water bodies, vegetation, city furniture) or the city thematics (transportation, land use, etc.) are defined by classes related to geometric primitives (such as polygons, lines, points) but also to non-geometric attributes (such as function, usage, height, material, address). In fact, there exists an ontology (very simple) behind CityGML. For example, the UML diagram of the transportation model of CityGML can be translated into the OWL language [3] with the ontology editor Protégé [4]. The UML classes and relations of CityGML can be directly translated into OWL classes and properties. The attributes can be either translated into datatype properties or object properties. The cardinality restrictions can be represented by formulas in descriptive logic.

3. An Ontology-based Model for the Communication of Urban Projects

Urban projects involve many actors (urban planners, politicians, inhabitants, etc.) and many tools, such as plans, legal texts or 3D representations closer to our vision of the real world. The use of 3D city models for the management of communication in urban projects is thus tempting. However CityGML is not sufficient to represent such projects since the concepts handled are essentially based on physical objects. More abstract concepts such as Right_of_way are missing, as shown on Figure 2. It would be interesting to associate a geometric form to a Right_of_way and to display it in the 3D scene associated to the geometric objects (such as parcels) to which it is associated. Furthermore, the class TransportationComplex, although associated to a function and a usage and with subclasses TrafficArea and AuxiliaryTrafficArea, is not sufficient for many transportation or mobility aspects such as soft mobility aspects.

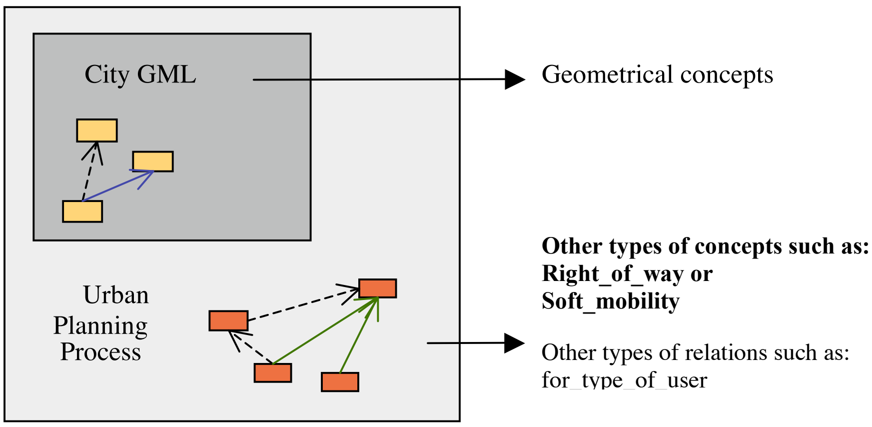


Fig. 2 : Examples of the semantics missing in CityGML

For those reasons we decided to define:

- an Ontology of Urban Planning Process named OUPP, including in particular the semantic aspects identified previously in order to be able to represent the information of urban projects
- semantic links between CityGML and OUPP in order to use the geometrical representations of the objects that exist in CityGML

3.1 OUPP

In this paper we describe the part of OUPP related to *soft mobility* aspects. This part of the OUPP ontology has been defined with the aim of providing an urban actor (an inhabitant for example) with an integrated view of the various aspects related to soft mobility, in order to promote this way of travelling [5]. The legal aspects (which are important to urban planners or politicians) are not described in this paper in order to focus on some other aspects such as the duration of travelling for a type of user (as these aspects seem to be an important issue to many potential users). We have also decided to define a general enough ontology to represent soft mobility in different places but sufficiently fitted to Geneva to be directly used and tested as a communication tool in this city. That is why the promenades through parks (public or even private if a right of way has been negotiated), have been represented as they are described in legal texts. If soft mobility concerns all the ways of transportation muscularly propelled, the major urban realizations are for pedestrians or cyclists. The figure 3 below shows (as a graph) the ontology that we have defined for representing soft mobility aspects within OUPP.

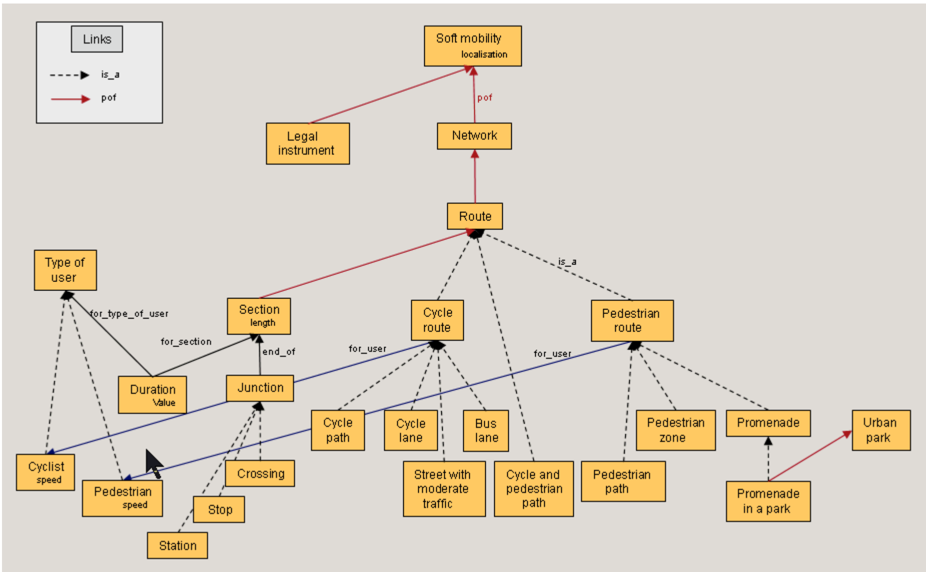


Fig. 3 : Part of the ontology of soft mobility

We then coded the ontology using the OWL language with the editor Protégé (see the Figure 4). We also created a knowledge base of soft mobility in Geneva. We worked on various documents (legal, associative, etc.) available on web sites and on data coming from the Information System of Geneva (*Système d'information du Territoire Genevois*, SITG) [6]. More precisely, we defined:

- instances, such as Promenade_des_parcs which is an instance of Route
- semantic annotation links between, on one hand, these data and documents and, on the other hand, concepts and instances of OUPP, see the Figure 5.

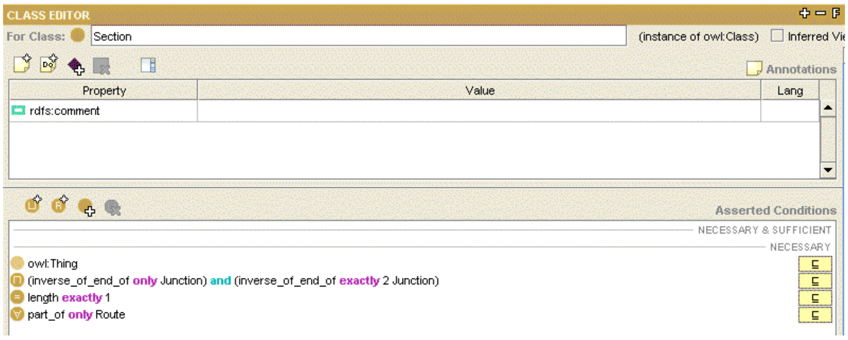


Fig. 4 : Part of the ontology of soft mobility defined in OWL with Protégé

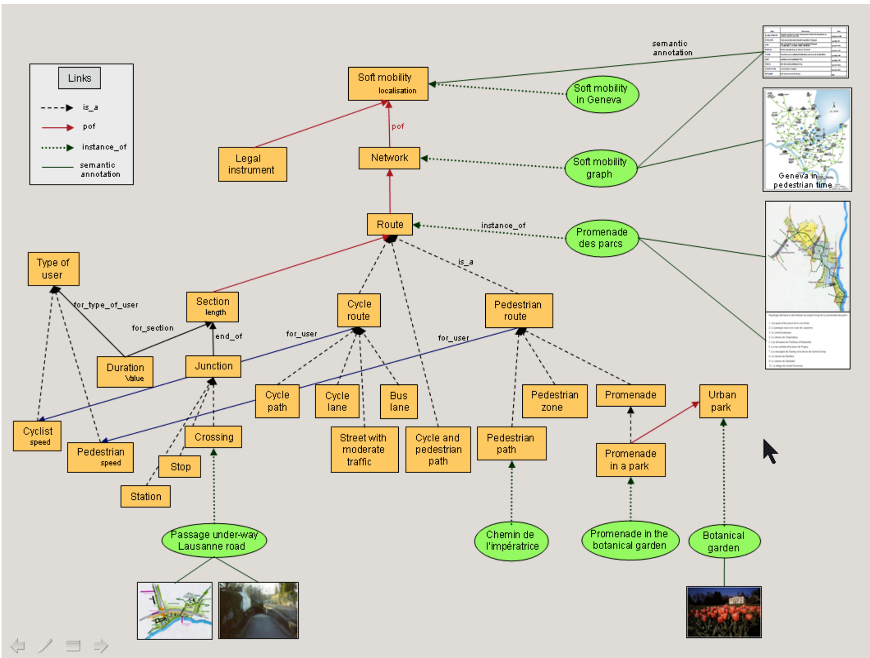


Fig. 5 : Knowledge base of soft mobility in Geneva

As illustrated in the figure 6 below, we performed the *semantic integration* of data and documents of different types, previously disseminated on various sites.

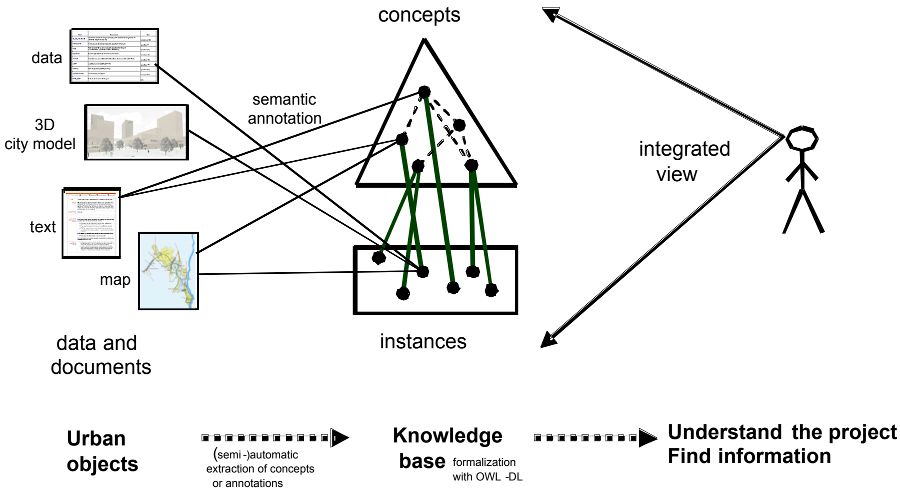


Fig. 6 : Semantic integration of the urban information

With the knowledge base of soft mobility we can compute the duration of a particular route (Promenade_des_parcs for example) for a type of user (Pedestrian). As Promenade_des_parcs is an instance of a Route, it is part of a Network which is in this case the Soft_mobility_graph of Geneva. With the SITG, we have the Sections that form the Promenade_des_parcs with their length. A Duration can then be associated to each Section for a Type_of_user, in our case a Pedestrian.The Duration value is computed from the speed of a Pedestrian x the length of the Section. The values of the different sections of Promenade_des_parcs can then be added to obtain the duration value for a pedestrian travelling through this promenade.

3.2 Interconnection between CityGML and OUPP

If we look at CityGML we see that the concept of TrafficAreas:

- provides elements which are important in terms of traffic usage like car driving lanes, pedestrian zones or cycle lanes
- has a function including crosswalk, green spaces, footpath, cyclepath, combined footpath/cyclepath
- enables a usage including pedestrian, bicycle, horse.

TrafficArea and Section can be connected by a subsumption relation: Section $\sqsubseteq$ TrafficArea, as shown on Figure 7.

Through the use of the geometry and appearance that are associated to objects in CityGML we can now present the urban information (that has been previously integrated) to the user within a 3D scene. This visualization can be adapted to the profile and the centers of interest of the user [7].

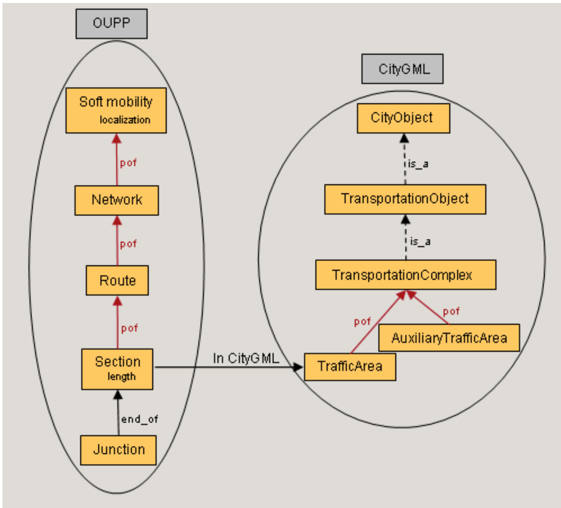


Fig. 7 : Interconnection between OUPP and CityGML

3.3 Interconnection between OUPP and OTN

As a possible candidate of our investigation, we also identified an Ontology of Transportation Networks (OTN), defined within the framework of the REWERSE project [8]. OTN describes various transportation aspects but none related to soft mobility. As it is, OTN seems to provide a complementary approach to our ontology of soft mobility and can be used to extend this ontology to other transportation issues such as public transportation for example. The Figure 8 below shows an excerpt of the OTN ontology represented in OWL with Protégé.

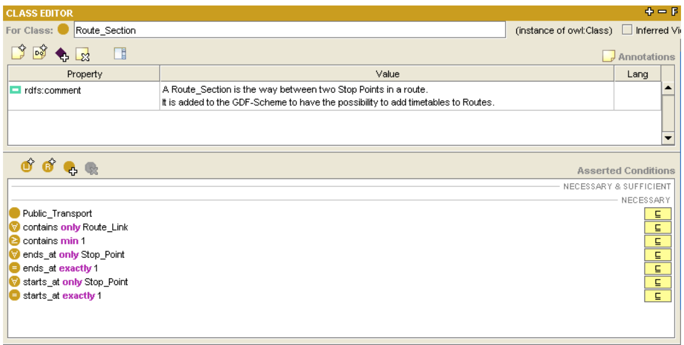


Fig. 8 : Excerpt of the OTN ontology defined in OWL with Protégé

OTN differentiates a Start_Point and a Stop_Point for a Route_Section, in order to be able to represent a travelling direction.. Even if we did not perform such differentiation for soft mobility, we have similar conceptual structures on both sides: Routes containing Route_Sections ended by Stop_Points in OTN, Routes composed of Sections ended by Junctions in OUPP, as shown on Figure 9. The main difference is that OUPP is related to soft mobility issues while OTN represents public transport.

In order to represent different types of urban transport, it is possible to generalize OUPP by defining Junctions, Sections, Routes and Networks as general concepts and to associate those concepts to an extra-attribute indicating, at the instance level, whether we are considering soft mobility, public transport or other means of transport or not. When OUPP and OTN are interconnected, it is possible to define routes partly by foot and partly with public transportation systems.

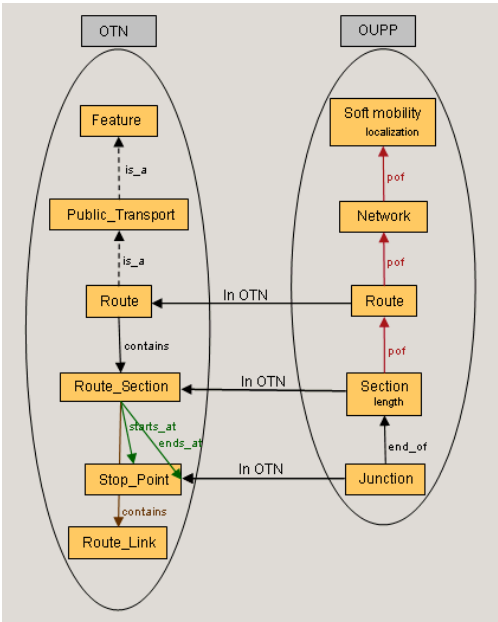


Fig. 9 : Interconnection between OUPP and OTN

3.4 Interconnection between OUPP, OTN and CityGML

With OUPP, OTN and CityGML interconnected (see the Figure 10), it is possible to visualize the information and knowledge contained in the ontologies within 3D city models based on the standard CityGML. This is possible because TrafficAreas is associated to MultiSurface geometries.

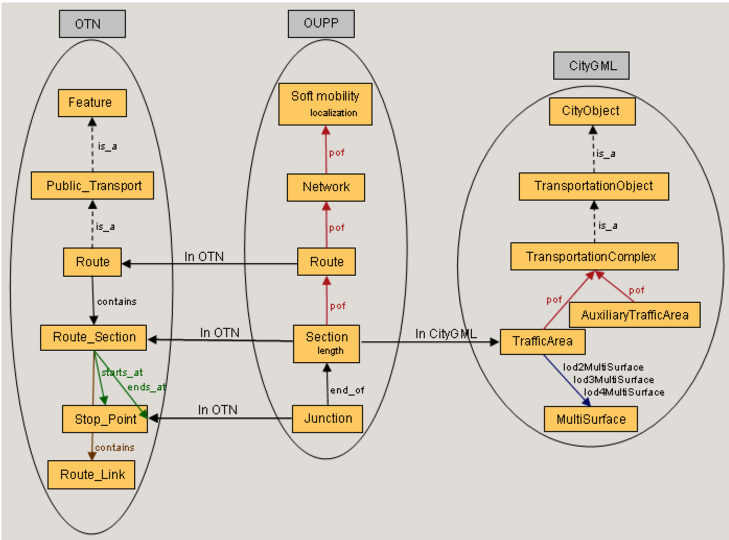


Fig. 10 : Interconnection between OUPP, OTN and CityGML

4. An Ontology-based Model for Archaeological Purposes

Archaeological remains are important components of our current cities. It is therefore sensible to consider them in an urban modelling context. However, their integration in such a virtual system is far from being straightforward. Indeed, archaeological and more generally cultural heritage domains present some specific issues one of these being the uncertainty; knowledge is almost always unsupported (notably concerning shape, function or chronology) and information is linked to structures with a very often incompletely known geometry. Moreover, those archaeological structures are nowadays hidden or replaced by other, more recent. Sometimes they have disappeared. Besides, archaeological data are known for their complex semantics. For example, a building is likely to have host several levels of function (main, minor, symbolic...) and sometimes different functions connected to a same level. This non-permanence of knowledge gives rise to interpretation changes [9]. For this reason, we believe that such information should be handled in a flexible and evolving way by combining different levels of archaeological ontologies and connecting them to urban ontologies or models.

This approach has been adopted in the case of the Structure II Sub C of Calakmul (Yucatan, Mexico) (see Figure 11). It is the oldest building of a very important Maya city, founded during the Late Pre-classical period (between - 300 and + 250) and left at the beginning of Post-Classic (at the end of the 1st millenary). This building consists in two superimposed platforms, on which a space decorated with red stuccos rests. A monumental staircase, surrounded by macaroons, goes to a large vaulted room, the only example of barrel vault known for all Maya architecture.

Lastly, a summit crest probably tops it off. For the current specialists of Maya world, this building giving off monumentality and verticality is interpreted like a temple - mountain. The macaroons feature the supernatural powers, undoubtedly of the gods. The vaulted room symbolizes the *Xibalba* (the infra-world). The crest, an openwork, is designed to create sets of shades and light. In the history of Calakmul, this building certainly played the following role: symbol of the holy mountain evoked in the myth of Maya creation, the mountain from where came the "original" twins, the building pointed out the sacred origin of "royal" lineage. It was the place where the *ahaw* (the "king") came in contact with the heavenly forces, the place inside of which, thanks to men supports, the sun could be regenerated and reappears as each morning the rays proved it which crossed the crest [10].

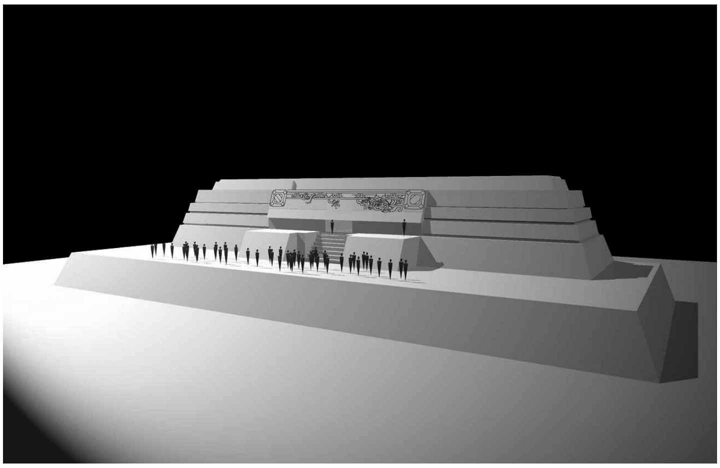


Fig. 11 : A vision, without crest, of the Structure II Sub C of Calakmul
© Massimo Stefani - animasalva@tin.it [11]

To perform the integration of this information and to be able, when needed, to update it easily, it is necessary to consider three successive ontological models at least: a generic archaeological ontology, an ontology dedicated to the Maya world and, finally, an ontology specifically dedicated to the city of Calakmul. In the following, we present an excerpt of what could be such a succession of ontologies : The first ontologies describing archaeological "universal concepts", the second one gathering classes of terms and concepts developed by archaeologists, historians and anthropologists to describe the preclassical Maya civilization and finally the third one dealing with specific objects and situations linking the Calakmul site. To illustrate our approach, we consider a schematic semantic net involving a very small number of concepts, see Figure 12. In addition to the link between archaeological concepts, we wish to use existing models, such as CityGML, to deal with structural and geometrical issues. In that case, building descriptions of CityGML seem to be well suited to the representation of the geometry and the structural nature of some archaeological items.

An extension of CityGML could be envisaged for some domain applications, but archaeological domain is so far away from the original CityGML considerations that such extensions seem impossible. Therefore, the use of ontologies seems to be the only sensible solution to integrate archaeological knowledge and information in a broader urban model.

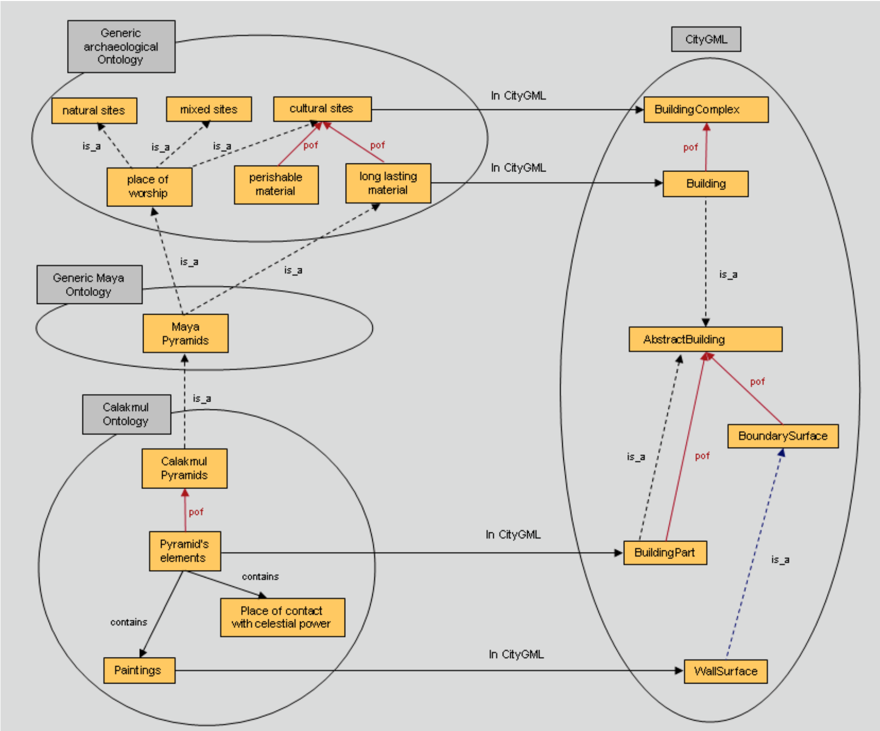


Fig. 12 : Schematic semantic net of the different ontologies

5. Towards Generic Models based on Urban Ontologies

5.1 Specificities, Problems, Stakes for the Future

Based on our experience, it becomes clear that developing a unique universal urban model is impossible; urban reality perception is diverse and multiple. On the other hand, developing isolated models for each application is unconceivable. The solution we have adopted for the OUPP project and the CALAKMULL project can be generalised; interconnecting models by means of ontologies in order to gradually build a strongly interconnected set of models representing different perceptions of urban environment. Such an interconnected set of models could be built around the CityGML model, as proposed on the Figure 13.

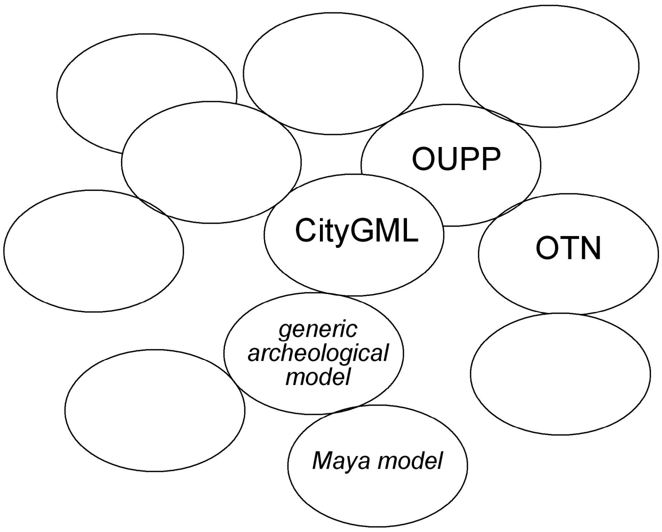


Fig. 13 : A set of models around CityGML

However, to ensure a strong interconnection between those models, rules must be adopted. Models should conform to existing standards when available and applicable, for example ISO or OGC geometric features. Beyond comprehensive feature data catalogues, it is crucial to have detailed semantic descriptions of modelled objects to avoid semantic heterogeneity issues. Each database model ontology must be provided together with, if possible, links to existing application and domain ontologies. The more central (or connected) in the set a model, the more strongly described it must be. In our example, if elements of the Maya model change, this change does not have any impact on the other connected models. However, a change of some CityGML object definitions could have a huge impact on model interconnections.

Building such a set of interconnected models allows for processing specific queries: for example, combining soft mobility and public transportation for routing purposes or exploring buildings that had worship function in Maya's culture (see Figure 14).

CityGML seems to be a good candidate as a central model dealing with urban fabric and geometry. It however shows some conceptual drawbacks which should be overcome. As mentioned above, models must be strictly described to allow for a high level of interaction. Ideally, they should be associated with a meta-model enabling a clear and strict model definition. It is not to date the case of CityGML. A solution would be to strengthen the ontological bases of CityGML as proposed by Billen *et al.* [12]. In this position paper, the authors show that some CityGML's objects could be retrieved from the associated hierarchical meta-model.

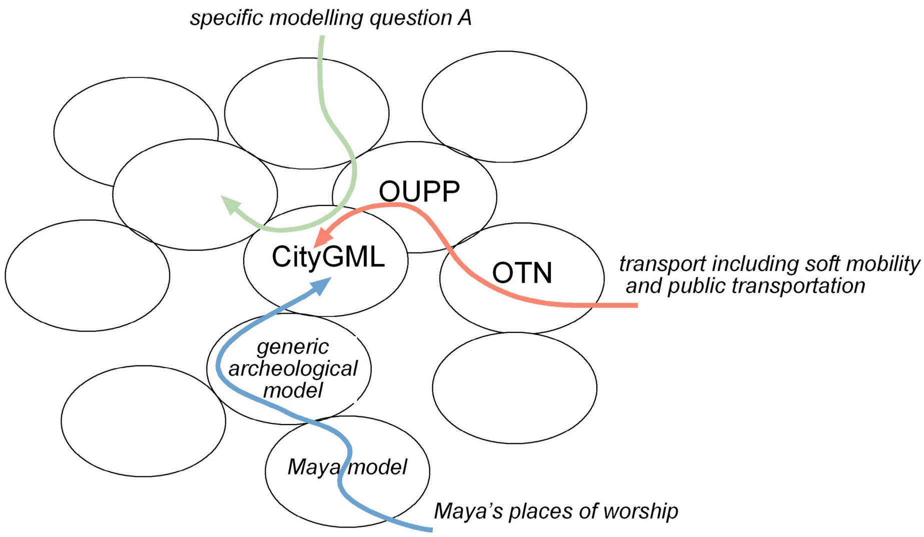


Fig. 14 : Towards an interconnectet set of models around CityGML

6. Issues and Perspectives

In this paper, after a short presentation of the semantics of CityGML, we have described some case studies integrating urban information, more particularly the kind of urban information necessary to explicitly and formally define ontologies as well as the resulting models and their applications. We have also tried to explain how and why these experiments could be generalised to improve the interoperability of 3D urban information.

Work is currently on-going in this domain, the authors of the paper are working together to build up a representation of the different kinds of urban information that is common and ontology-based. For this development, they need to rely on a formal and explicit representation of the knowledge embedded in the models.

Another important feature of the models developed is to enable the interoperability of the urban information on the basis of those common representations [13].

All this work can be considered as a first step towards the development of generic models, representing different points of view. It also provides a first stage towards the development of patterns of models that can be tailored to specific needs, thus enabling their use and their adaptation to various specific applications.

References

1. *OpenGIS City Geography Markup Language (CityGML) Encoding Standard*. Document reference number: OGC 08-007r1, Editors: Gerhard Gröger, Thomas H. Kolbe, Angela Czerwinski, Claus Nagel, 2008, www.opengeospatial.org/standards/citygml.
2. *Unified Modeling Language Specification*, UML version 1.4.2, Document formal/05-04-01, http://www.omg.org/docs/formal/05-04-01.pdf
3. *OWL Web Ontology Language Overview*, W3C Recommendation 10 February 2004, http://www.w3.org/TR/owl-features/
4. *The Protégé Ontology Editor and Knowledge Acquisition System*, available on the web site http://protege.stanford.edu/
5. *La mobilité douce : un concept d'avenir*. In Feuille d'avis officiel de la République et canton de Genève, 2003, http://www.ge.ch/Fao/2003/doc/20031001.pdf
6. *Système d'Information du Territoire Genevois*, http://etat.geneve.ch/sitg/accueil.html
7. Métral C., Falquet G. and Vonlanthen M. An Ontology-based System for Urban Planning Communication. In *Studies in Computational Intelligence (SCI)*, 61, 61-72, Springer-Verlag, Berlin Heidelberg, 2007.
8. Lorenz B., Ohlbach H. J. and Yang L. *Ontology of Transportation Networks*. REWERSE-DEL-2005-A1-D4, Editor: Bernhard Lorenz, 2005, http://rewerse.net/deliverables/m18/a1-d4.pdf
9. *Definition of the CIDOC Conceptual Reference Model*, Produced by the ICOM/CIDOC Documentation Standards Group, continued by the CIDOC CRM Special Interest Group. Version 4.3, November 2008, Editors: Crofts, N., Doerr, M., Gill, T., Stead, S., Stiff, M., http://cidoc.ics.forth.gr
10. Calakmul Subcb, http://www.archeografica.org
11. Rodrigo Campero O, *La Gran Plaza de Calakmul*. In *Arqueología Mexicana*, 42, 22 - 27, 2000.
12. Billen R., Laplanche F., Zlatanova S., Emgard L. *Vers la création d'un méta-modèle générique de l'information spatiale 3D urbaine*. In *Revue XYZ – association française de topographie*, 114(1), 37 - 42, 2008.
13. Grangel R., Métral C., Cutting-Decelle A.F., Bourey J.P. and Young R.I.M. Ontology based Communications through Model Driven Tools : Feasibility of the MDA Approach in Urban Engineering Projects. In *Studies in Computational Intelligence (SCI)*, 61, 181-196, Springer-Verlag, Berlin Heidelberg, 2007, 211 p., ISBN 978-3-540-71975-5

Enabling Geolocating via Ontologies[12]

Aneta J. Florczyk [1], Francisco Javier López-Pellicer [1], Rodolfo Rioja [2], Javier Nogueras-Iso [1], Francisco Javier Zarazaga-Soria [1]

[1] Computer Science and Systems Engineering Department
University of Zaragoza
C/ María de Luna, 1, 50.018, Zaragoza, Spain
{florczyk,fjlopez,jnog,javy}@unizar.es
[2] GeoSpatiumLab
C/ Carlos Marx, 4, local der., 50.015 Zaragoza, Spain

Abstract. This paper presents an architecture for compound geocoding Web services built on diverse Web services of geographic information, especially the gazetteer and geocoding services. The proposed architecture uses ontologies as a key element to source selection and data integration. This approach intends to satisfy the user needs for an adaptive geocoding service of general purpose, which might be a core component of geolocating service architecture.

Keywords: Geocoding, Ontology, Geolocating, Service Architecture, Web Service.

1 Introduction

Public authorities charged with the responsibility of dealing with the maintenance of public infrastructure require urban management systems able to manage large amount of resources. Especially in case of emergency situations (e.g. floods, fires), an effective management of resources included in the emergency plan involves a centralized and coordinated management system. These systems often require support from Geographic Information Systems for geocoding street addresses. Thus, there is a need for up-to-date data of high quality (e.g. coverage at the national level, high precision, reliability). However, assurance of data quality is a hard task. The decentralized public administration is responsible for data maintenance, and the spatial information changes continuously, especially in the urban area (e.g. cadastral data).

---

[12] This work has been partially founded by the Spanish Government through the projects "España Virtual" (ref. CENIT 2008-1030) and TIN200765341, and by the Government of Aragon through the project PI075/08. The work of Aneta J. Florczyk has been partially supported by an FPU Grant from the Spanish Ministry of Science and Innovation (ref. AP2007-03275). The work of Rodolfo Rioja has been partially supported by the Spanish Ministry of Science and Technology (Torres Quevedo program ref. PTQ06-2_0790).

In this paper, geocoding means *the act of turning descriptive locational data such as a postal address or a named place into an absolute geographic reference* [1]. This includes relative description of location as input data [2], and any geographic description as the output, i.e. a point, a polygon or three-dimensional geospatial entity [3].

The most prevalent way of providing geocoding functionality is a geocoding service. Nowadays there are lots of geocoding services with diverse characteristics determined by the quality of geo-service, the provided data and the terms of service (ToS). Today there is no problem in using online geocoding providers but to find and to choose a proper provider.

The main differences among geocoding Web services are caused by the provided data, i.e. the type of content (e.g. point of interest, address). The quality of geo-service is influenced by factors that depend on the typical QoS requirements (e.g. response time, reliability) and the quality of the spatial data.

The geocoding services also differ in their ToS. In general, regardless of service origin, which might be private sector (e.g. Google, ViaMichelin or Yahoo), public sector or volunteer communities (e.g. GeoNames[13], Megalithic Portal[14], Geograph British Isles[15], OpenStreetMap[16]), the services may be divided into three groups due to their ToS: paid access services, free of charge services with restricted use, and services of free use.

The private sector offers the ad-hoc designed paid services, which guarantee the quality of data and service. Free services offered by the public sector or open communities provide less quality than the dedicated ones. Usually, the largest providers offer free access to their address geocoding services with lower quality and some use restrictions. Their ToS restrict the presentation (e.g. the license requires use of the supplier's visualization APIs), prohibit the reuse of data, and have influence on the quality of applications based on that service, i.e. establishing limits, such as rate limit or the maximum number of requests per day.

Finally, we should also consider that new types of geocoding service applications such as support of mobile application, demand supplementary characteristics. Location-based services require the support of geocoding services for tracking of user location and the reverse geocoding at the operation system level (e.g. the Android[17] or GeoClue[18] projects). The availability and capabilities of these services have to be adjusted to the requirements of mobile devices (battery life, cellular network, access to the Web, or GPS availability).

The choice of service is determined by the use case. Free geocoding Web services are appropriate for "geotagging" (i.e. the process of adding the geocoded information to any kind of media) the local news or incidents (e.g. water supply shortage, planned roadwork) because such information does not require high quality geocoding services or spatial data. On the other hand, the systems on which depend public health [4],

[13] http://www.geonames.org/
[14] http://www.megalithic.co.uk/
[15] http://www.geograph.org.uk/
[16] http://www.openstreetmap.org/
[17] http://code.google.com/android/
[18] http://www.freedesktop.org/wiki/Software/GeoClue

public security [5] or environmental services [6] require high quality of service and data. For example, quickness and efficacy of fire-fighters depend on the information they possess such as the characteristics of the building in fire (e.g. number of floors, shape, location of entrances and the accessibility, nearby buildings) or the localization of fire hydrants.

The vast heterogeneity of geocoding services and the specific features of geographic data set up the open problem of provider selection. There are many works in the context of the service discovery and selection. Some of proposals need prior service evaluation (e.g. rating agency [7] or user [8] pre-evaluation), but most of the works in this area use typical QoS features [9 – 11] (e.g. end-to-end delay, overall cost, service reliability, availability). Recently, researchers show interest in services of geographic information [12, 13] but they include only basic concepts (e.g. coverage) and do not exploit specific characteristics of geographic data in discovery and selection processes (e.g. reasoning based on coverage, quality of geographic objects).

Our proposal for compound geocoding architecture attempts to be a first step in the research on the problem of the geocoding service selection via geo-ontologies. This approach uses a framework that allows building hybrid solutions composed of different services which provide geographic information (e.g. geocoding, gazetteer or cadastral service). Each solution applies administrative unit ontology for (i) provider selection and (ii) data integration. This approach may increase the flexibility and adaptability of applications. In case of the public services, it provides access to different services, national and local, in a transparent manner and ensures the use of updated data.

Additionally, integration of multiple geocoding sources and/or services is one of the open issues of research in field of geolocating, which will be presented in the section 2. Our approach aims to develop the fundaments for designing core components of geolocating services.

The paper is organized as follows. Section 2 presents issues that appear in the context of geolocating. The next section reviews the characteristics of geocoding services. Section 4 presents the architecture of the compound geocoder. The following section shows the state in geocoding of urban management applications in Spain, and presents the implementation of the application. Finally, there are presented conclusions and the description of the future work.

2 Geolocating within Urban Systems

Existing geocoding services are generally limited to assign a geographic coordinate to an absolute location such as a street address. However, urban management systems need to geocode location description in a more flexible way. For example, it is common that a citizen who calls an emergency centre does not know the address where he/she is and the descriptive information that provides is ambiguous or even confusing (e.g. “100 meters of a memorial statue in the park, which is situated by Colon street”). In such a situation the user needs to be “geolocated”.

Hutchinson and Veenendaal [2] define "geolocating" as the evolution of the geocoding process that permits to assign valid geographic codes to freeform textual descriptions of locations. Conventional geocoding services that work with absolute locations will not be able to determine the coordinates of the place of incident. The best solution is a geolocating service, which might interpret correctly the relative location (e.g. distance, direction), with landmarks (e.g. park, railway station) and features (e.g. coffee shop). Hutchinson and Veenendaal [2] enumerate the following elements that contribute to the concept of geolocating: query syntax flexibility, user interaction, user context consciousness and complex site representation. It is important to stress that the authors argue for a new methodology, however, they do not include details of how one would implement it.

The principal design goals of the architecture proposed in this paper are flexibility and extension facility. Due to the fact that the geolocating process demands spatial information of wide range of types, the compound approach seems to be suitable for the architecture model of a geolocating system.

3 Characteristics of Geocoding Service

The proposed architecture allows the service selection according to the use case requirements. Therefore, the proper description of each source is vital for the behaviour of the whole system. Fig.1 presents the properties that are used in the strategy for provider evaluation, which is essential for the service selection and the decision making process.

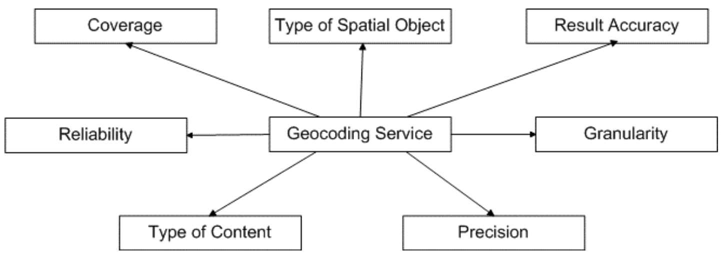


Fig. 6. Characteristics of geocoding service.

The main features of each geocoding service are the spatial *coverage*, the *type of content* and the *type of spatial object*. The first two of them are always given by the provider or are indicated by the name of the service. The first one defines the area in which offered data are situated and, usually, this area corresponds with the jurisdictional geographic object, the concept defined in administrative unit ontology. The second feature, the type of content, strictly depends on the georeferenced types of features. The last one indicates the list of provided types of spatial object, such as point, polygon or 3D entity. For example, the cadastral service of Spain (Servicio de Catastro de España), as the name of the service indicates, has the coverage of Spain and offers coordinates of parcels. Google Maps has world coverage and its type of content is street address geocoded via point, which is provided by the service description.

Another feature is the *result accuracy* and should not be misinterpreted as “data accuracy”, the term commonly used in literature to describe spatial data accuracy. Result accuracy is defined for each source and is derived from the analysis of the source data model and the search data model. It indicates the level of overlapping of the search and data models. Let us define the search data model for the address search in Spain. According to the general approach [14], it will contain at least province, municipality, zip code, street name and number. The data model of the cadastral service of Spain maintains this minimum (the accuracy is 'number' or simply 'address'), but the data model of the geographic information service does not have streets and the search accuracy will be 'municipality'.

From the analysis of spatial data, it is possible to obtain two additional indicators (of range 0 - 1): **reliability** and **precision**. The reliability indicates the capacity of representation of elements of physical world in the content. The service that offers all elements of real world has a reliability value equal to '1'. The indicator of precision informs about the average *positional error* [15] of the whole dataset. It is important to note that this indicator may be influenced by difference between the provided spatial object and the search one. For example, using cadastral data for the address geocoding, there will be a decrease in spatial data precision.

Usually, the reliability of a street address service varies in function of the area relevance, for example a new suburb might be even omitted. Such a feature might be indicated by an additional indicator, called **granularity**, which respects the level of detail.

4 Architecture Overview

The compound geocoding architecture uses different geographic information services such as gazetteer services, geocoding services and cadastral services. Each source has to be properly described as the descriptors values are the clues for source selection, which determines the adequate functionality of the system. The result accuracy is estimated via comparison of the search model and the data model of each source. This requires a well defined model of searched information, possibly with an administrative unit ontology [16]. Also, the data model of each provider has to be described with the help of the administrative unit ontology that permits the mapping of data models.

The main elements of the compound geocoding architecture (see Fig. 2) are an input data processor component, a core component and a mediator component. The input data processor performs the pre-processing of input data. The steps in this phase of geocoding are common techniques among geocoders [2]: cleaning, parsing and standardizing.

The core element is responsible for the whole process of source selection and data evaluation. The rules used by the source selector component and the data evaluator component are implemented in the decision maker component. The rules apply the search criteria that are defined in the same terms as the source characteristics. The search criteria are provided by the application context and/or the user requirements. The first evaluation of the results is done in the mediator and then the data evaluator performs posterior evaluation according to the predefined criteria of search and the characteristics of the content (e.g. data precision, content type).

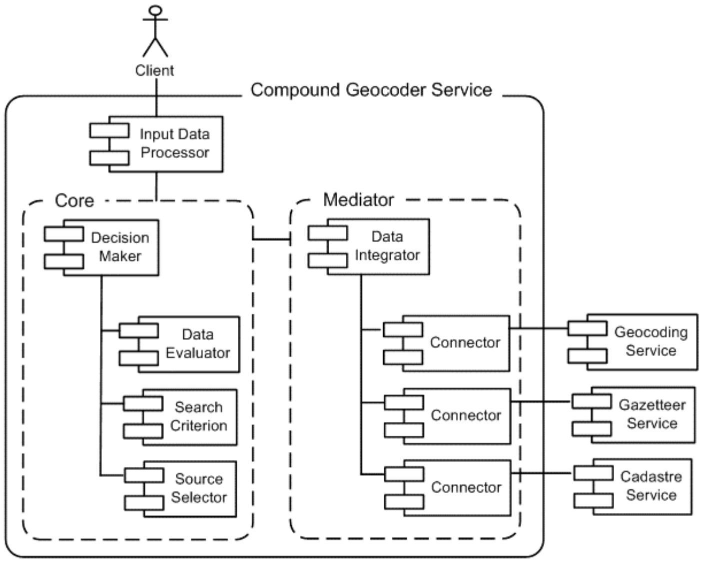


Fig. 7. The main components of the compound geocoding architecture.

The mediator component consists of pluggable service connectors and a data integration component. The main advantage of the connectors is the abstraction from communication protocol, invocation styles or interfaces used. The data integration component is responsible for data harmonisation which consists of data mapping and coordinate transformation if necessary.

This architecture allows geocoding different types of named places and due to complementing data from one source with data from the others; it improves the reliability and the data precision. It also gives the user more freedom in deciding the search strategy. Due to the access to several services, the search strategy may select the best response of the entire system, the best answer for each source or the best answers from a chosen source. In addition, as the details of implementation are hidden in the mediator, this allows incorporating any type of georeferenced data.

5 Address Geocoding in Urban Management of Spain

In Spain, in public sector, there are several alternative services in the field of geocoding of street addresses, such as cadastral services, CartoCiudad[19], Geopista[20],

[19] http://www.cartociudad.es/ignnomenclator/
[20] http://www.geopista.com/

or services provided by local authorities (e.g. IDEZar Street Gazetteer Service[21] of the Zaragoza municipality, IDEZarSG).

The national cadastral service is offered by the *Dirección General del Catastro*[22], DGC (an administrative registry which belongs to the Ministry of the Treasury). The service provides updated data and spatial representation of parcels of high quality. However, the spatial data precision is lower for address search (i.e. parcel centroid), and the data access is uncomfortable and might result confusing (definition of the province and then the municipality).

The alternative providers are Geopista or CartoCiudad. The first one uses data that comes from the DGC. Although the second one offers services with data coming from the harmonisation and integration of digital data produced by several official suppliers, the main source of CartoCiudad is again DGC. As both of them use stand-alone data bases, in a short time there will be problems of data reliability. Moreover, they do not provide any update procedures.

The other alternative might be the services of street data provided by local administration, such as the IDEZarSG. These services are characterized mainly by high precision. However, the level of granularity differs depending on the area (e.g. urban centre, town) and there are areas without coverage, such as new highways or suburban areas.

In Spain, the geocoding services in the public sector are not adequate for urban management systems, for there is lack of update procedures, reliability, and precision or even simplicity of use (step strategy). One of the solutions to this problem might be the data integration from different sources via compound service architecture.

Table 1. Characteristics of the selected services in the context of address search in municipality of Zaragoza. The values are obtained via series of provider tests according to the description presented in section 3.

	Coverage	Content Type	Result Accuracy	Reliability	Precision
IDEZarSG Service	Municipality of Zaragoza	Street Data	Number	0.98	1
GoogleMaps Service	World	Address	Number	0.96	0.99
Cadastral Service of Spain	Spain	Parcel	Number	1	0.95
IGN Concise Gazetteer Service	Spain	Geographic Features	Locality	1	not applied

The implementation of a compound service for address street geocoding uses the following geocoding services: the IDEZarSG Service, the GoogleMaps Service, the Cadastral Service of Spain, and the Concise Gazetteer Service of the National Geographic Institute (*Instituto Geográfico Nacional*, IGN). The characteristics of each one are described in the table 1.

[21] http://idezar.unizar.es/SRW/client.html
[22] http://www.catastro.meh.es/

The first step is definition of a search model. The street address model is described by the administrative unit ontology that permits assignation of specific legal roles to administrative unit entities, in this case, the address role. In Spain, the address is defined via province, municipality, zip code and address line, i.e. the compound of street name, street type, number with additional information if necessary, e.g. floor and door. As the zip code is not supported by any of the plug-in sources, the data disambiguation requires extending the official address model with locality and suburb fields.

The domain data model and the source data models are defined to satisfy the address search model. The administrative unit ontology is used to define each model that helps in the mapping between them as Fig.3 and Fig.4 show.

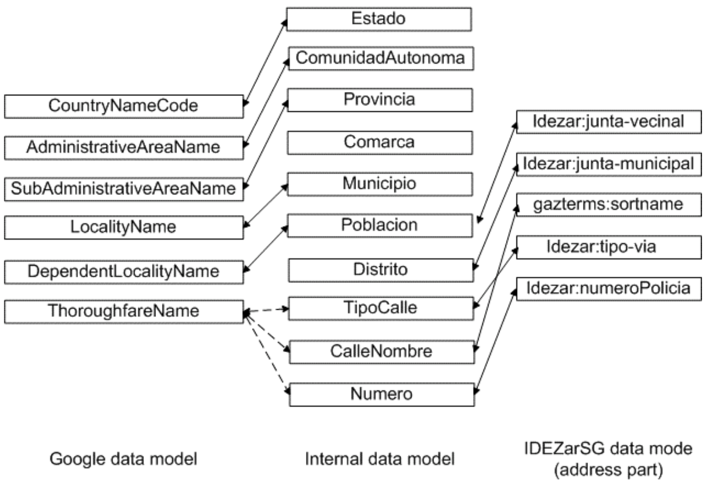


Fig. 8. Data model mapping between the domain model, the Google data model and the IDEZarSRW

The proposed architecture uses several geographic information suppliers to obtain the highest service quality. The features of the compound service allow eliminating the problems indicated in the previous section, such as the need for data update, manual step strategy of search, or omissions in coverage or granularity.

Currently, an instance of this service is used as the geocoding service of the official web of the Zaragoza municipality council[23]. The main data source is the IDEZarSG service but the result data is complemented with the data from other sources, which improves the reliability of the service.

[23] http://idezar.zaragoza.es/callejero/index2.jsp

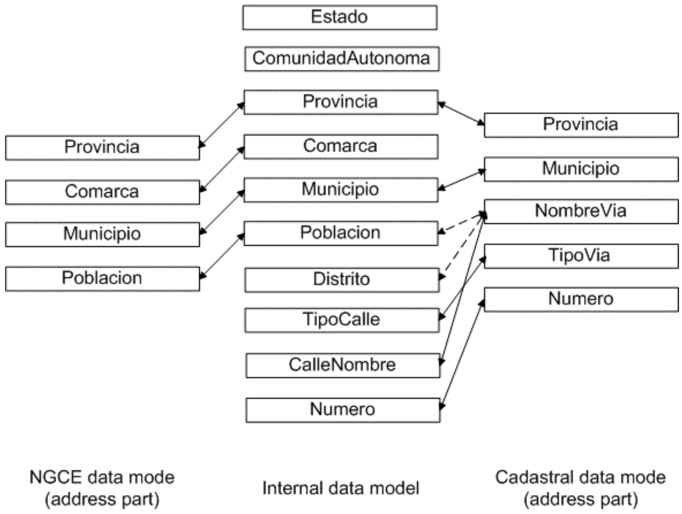


Fig. 9. Data model mapping between the domain model, the address part of the MNE data model and the cadastral service data model

6 Conclusions and Future Work

The compound geocoding architecture ensures the improvement of geocoding results (e.g. reliability, spatial data quality) thanks to the use of different geographic information suppliers. The use of multiple sources involves the development of ontology for describing the geocoding sources. Moreover, the use of ontologies yields an advanced architecture in terms of extensibility, flexibility and adaptability.

The framework for geocoding service selection permits to develop a methodology to geocode diverse categories of data types (e.g. spatial features, points of interest), which is an essential functionality of a geolocating service. In this context, there is a strong demand for a generic search model that involves the formalization of search model. Additionally, knowledge integration system adds new issues to the gazetteer concept. Usually, the spatial data of gazetteers is obtained from geocoding processes and comes from diverse data sources. Nevertheless, data models of gazetteers do not offer information about spatial data accuracy and its origins.

The principal disadvantage of this approach is the need of implementation of a pluggable connector for each source. Although the implementation permits adapting to changes, the effort of connector creation is significant.

As a future work, we will study the current techniques in Web service interoperability that permits automatic discovery and use of the geographic data providers. An effort will be made to create the formal definition of the ontology for geo-service description that includes the features enumerated in this paper. As the next step, we will apply the ontology reasoning for improvement of the service selection.

References

1. Goldberg, D.W., Wilson, J.P., Knoblock, C.A: From Text to Geographic Coordinates: The Current State of Geocoding. URISA Journal. 19(1), pp. 33--46 (2007)
2. Hutchinson, M., Veenendall, B.: Towards Using Intelligence to Move from Geocoding to Geolocating. In: Annual URISA GIS in Addressing Conf., Austin, TX. (2005)
3. Beal, J.R.: Contextual Geolocation, A Specialized Application for Improving Indoor Location Awareness in Wireless Local Area Networks. In: 36th Annual Midwest Instruction and Computing Symposium, Duluth, Minnesota. (2003)
4. Bonner, M.R., Han, D., Nie, J., Rogerson, P., Vena, J.E., Freudenheim, J.L.: Positional Accuracy of Geocoded Addresses in Epidemiologic Research. Annals of Epidemiology. 17, pp. 464--470 (2003)
5. Ratcliffe, J.H.: Geocoding crime and a first estimate of a minimum acceptable hit rate. Int. Journal of Geographical Information Science. 18, pp. 61--72 (2004)
6. Ratcliffe, J.H.: On the accuracy of TIGER type geocoded address data in relation to cadastral and census areal units. IJGIS. 15, pp. 473--485 (2001)
7. Sriharee, N.: Semantic Web Services Discovery Using Ontology-Based Rating Model. In: WI '06. IEEE Computer Society, pp. 608--616 (2006)
8. Manikrao, U.S., Prabhakar, T.: Dynamic Selection of Web Services with Recommendation System. In: NWESP '05. IEEE Computer Society, 117 (2005)
9. Yu, T., Jay Lin, K.: Service Selection Algorithms for Composing Complex Services with Multiple QoS Constraints. In: 3rd Int. Conf. on Service Oriented Computing, pp. 130--143 (2005)
10. Wang, X., Vitvar, T., Kerrigan, M., Toma, I.: A QoS-aware Selection Model for Semantic Web Services. In: ICSOC, Springer, pp. 390--401 (2006)
11. Tsesmetzis, D.T., Roussaki, I.G., Papaioannou, I.V., Anagnostou, M.E.: QoS Awareness Support in Web-Service Semantics. In: Int. Conf. on Internet and Web Applications and Services. (2006)
12. Fallahi, G.R., Mesgari, M.S., Rajabifard, A., Frank, A.U.: A Methodology Based on Ontology for Geo-Service Discovery. World Applied Sciences Journal 3(2) pp. 300--311 (2008)
13. Lan, G., Huang, Q.: Ontology-based Method for Geospatial Web Services Discovery. In: Int. Conf. on Intelligent Systems and Knowledge Engineering. (2007)
14. Walker, R.: A General Approach to Addressing. In: ISO Workshop on Address Standards: Considering the issues related to an international address standard. (2008)
15. Christen, P., Churches, T., Willmore, A.: A Probabilistic Geocoding System Based on a National Address File. In: Australasian Data Mining Conference, Cairns. (2004)
16. López-Pellicer, F.J., Florczyk, A.J., Lacasta, J., Zarazaga-Soria, F.J., Muro-Medrano, P.R.: Administrative Units, an Ontological Perspective. In: ER Workshops. (2008) To appear in: LNG&C

Combining Conceptual and Ontological Models for Representing Spatio-temporal Data and Semantic Evolution in GIS

Chamseddine Zaki[1], Myriam Servières[1], Guillaume Moreau[1]

École Central de Nantes / CERMA / IRSTV
1 rue de la Noë,
44321 Nantes, France
{Chamseddine.Zaki, Myriam.Servieres, Guillaume.Moreau}@ec-nantes.fr

Abstract. In this paper, we present a new approach as a spatio-temporal data model for the geographic information systems (GIS). We base our approach on the MADS conceptual model that we will be combine to a corresponding new generated ontology model using a set of transformation rules. We chose the MADS model for its interesting features like multi-presentation, multi-perception and uniform handling of spatiotemporal data, and ontologies because they allow adding new concepts and handling concept changes. Thus, this approach allows the tracking of historical changes of spatio-temporal objects and concepts over time. We also provide an example of an urban development plan in order to explain our approach.

Keywords: GIS, Conceptual models, Spatio-temporal Data, Ontology, History Tracking, Concept Evolution.

1 Introduction

Human beings know that the world around them changes both in space and in time. Men started developing Geographic Information Systems (GIS) the moment they felt the need to represent on maps the various elements from which their environment is constituted, and with the development of information technology, the will to save and manage this information by geographic databases started growing and appeared very essential and effective. On the one hand, current users of geographic information, who seem to be more and more interested by the notion of historicity of the geographic information, increasingly see their willingness to integrate more the temporal dimension in their data sets in order to enrich their spatial analysis by an evolutionary study of the considered phenomena. And on the other hand, this development of information also created a growing need for understanding the available geographic information. By allowing the representation of semantics of information, ontologies, that enable the specification of knowledge approved by a community of people, seem to be the promising answer to this objective.

In the following sections, we will first present the spatio-temporal models used in GIS, then give an overview of ontologies. Then, we will propose an approach based on the combination of both the MADS model and some geographical ontologies. We will first describe those concepts and show how to combine them in a single model which is illustrated by a real case study.

2 Related Work

2.1 Spatio-temporal Models for GIS

Simultaneous management of spatial and temporal dimensions (spatio-temporal) must be ensured within the geographic databases used in a GIS to make sure that time elements stored in these databases reflect, as closely as possible, the actual development of the entities, and in consequence many spatio-temporal data models have been proposed. Pelekis [1] enumerates ten distinct spatio-temporal models. Among them, the Snapshot Model [2] is the earliest and one of the simplest. This model uses time-stamping layers that enable the incorporation of temporal information into the spatial data model.

Many other models were presented like The Space-Time Composite (STC) [3], the Simple Time Stamping (STS) [4] and the Event-Oriented Spatio Temporal Data Model (ESTDM) [5]. STC creates a base map and then enhances it by applying changes through time. This model supports many types of spatio-temporal queries but still has difficulties in facilitating queries about spatio-temporal relationships. In ESTDM, instead of storing snapshots of instances the model stores changes in relation to previous state starting from a base map, and observations of an event are shown in a temporal sequence by grouping time-stamped layers. Every event is composed of a list of components indicating where, when and what changes have occurred. Even though, this model finds it difficult to search the unchanged parts.

The Three Domain Model [6] is an innovative model based on ESTDM and separately represents semantics, space and time. This model is able to handle both movements and changes, which is another improvement over existing models that only handle one of them at once. By handling spatial, temporal and semantic information separately, the model gains the capability of handling simple and complex temporal and spatio-temporal queries.

The Object Based approach added many advantages in the spatiotemporal modeling such as the effective handling of temporal data and the uniform treatment of temporal and spatial data. Worboys is the first one to propose object orientation for spatio-temporal modeling [7]. He proposed an object-oriented spatio-temporal model that improves the space-time composite model by using persistent object identifier that can resolve a new set of space-time composite without overlaying them. In his latest works Worboys provides a computational model for defining and detecting spatiotemporal events. He explains how this given model can be used in order to represent continuous change, regularity and connectedness in dynamic regions [8].

The object relationship approach added the description of the process which caused the changes in space and time as a relationship type between spatio-temporal objects. This description provides a better description for spatio-temporal events. Under this approach we find the MADS (Modeling of Application Data with Spatio-temporal features) [9], a conceptual spatio-temporal data model with multi-perception and multi-representation features (A detailed description of MADS is given in section 3.1).

Few of the proposed spatio-temporal models allow keeping track of feature history. They have mainly focused on the spatial changes of a feature.

Another model named feature-based temporal model (FTBM) represents space and time changes independently [10]. In his model that allows instances history tracking, Choi benefits from the features of both the object oriented spatio-temporal model and the three domain model and then adds an explicit temporal relationship structure in order to enhance querying.

2.2 Ontologies and GIS

Ontology in information science describe a particular way to understand a part of the world [11], it is defined by [12] as an explicit specification of a conceptualization.

The Ontology Web Language OWL provides the means for defining structured web ontologies. It is a XML (Extensible Mark-up Language) dialect based on the RDF (Resource Description Framework) syntax and is certified by the World Wide Web consortium.

Ontologies help in sharing common understanding of the structure of information among people, enable analysis and reuse of domain knowledge. They are mainly used to share knowledge and are an important step toward GIS interoperability [13].

Unlike databases, where the schema definition is completed before instances are created, ontologies allow some flexibility in their definition and their evolution, for example the specification of a concept can be changed anytime, irrespectively of the fact that the ontology holds instances of the concept or not.

There are many types of ontologies including taxonomic, descriptive and geographic ontologies. Many environments have been proposed to model, query and manage ontologies but these are not mainly dedicated to the modeling of geographic information. An important challenge is the evolution of descriptive ontologies to ontologies that consider and manage geographic information, the so called geographical ontologies.

Geographic ontologies cover three main domains [14]:

1- Space ontologies specifically dedicated to the description of concepts that characterize the area as the point, line, etc.
2- Ontologies of geographical domains (i.e. ontology modeling concepts of electric networks, etc).
3- Spatial (spatio-temporal) ontologies whose concepts are located in space.

For those last ontologies, a time component is required in order to model geographic information, because geographical applications often handle spatio-temporal data. Since 1998 some studies have handled space [15], whereas those handling time have started in 2001 with [16]. Later on, other studies combined space and time together: [17] uses the labeling technique, and [18] distinguishes spatio-temporal objects and spatio-temporal processes.

The work of Smith falls in this direction. He considers two types of entities: “continuants” which are entities that exist at any instant of time and preserve their identity through time no matter the changes that might happen to them and, “occurents” which are entities that “occur in time and unfold themselves through a period of time”. Based on this differentiation, and these two different ways of existing in time, Smith distinguishes two basic types of ontologies that they call SNAP (for continuants) and SPAN (for occurents) [19]. Smith’s work provides a frame in which it is possible to formulate relations between different spatial and spatio-temporal ontologies in order to solve problems pertaining to dynamic and geographical ontologies. Other researchers interested in ontologies and their role for GIS are Kathleen Stewart and Yuan. The former is more into modeling geospatial semantics, including event modeling for dynamic GIS, moving objects research and reasoning about spatio-temporal phenomena over multiple granularities while the latter studies the continuous phenomena and distinguishes between static and dynamic objects in his studies [20].

On the other side, many interoperability architectures using mappings from databases to ontologies started to exist since available data are mainly stored in databases and due to the important role that ontologies play in resolving semantic heterogeneity by providing a shared comprehension of a given domain. Under this topic we distinguish between two types of approaches. The first one creates a new ontology model from the existing database and then populates the generated ontology with the contents of the database, while the second approach maps databases to already existing ontologies [21].

3 Our Approach: MADS and Ontologies

The objective of our study is to propose a spatio-temporal database model that allows the evolution of the model and insures the multi-presentation of geographic objects and history tracking of concept evolution as well as instance changing. For these reasons, our research work is mainly concerned in two research axes: “mapping databases to ontologies” and “automatically handling changes within ontologies” which are both still relatively new areas. We believe that a spatiotemporal ontology study can significantly enrich representation and understanding of geographical spatial information, so our proposed model maps the MADS conceptual model used for its interesting features to a new ontology in which concept changes will be allowed without dropping the initial MADS model.

We apply our approach in the Urban Development Planning field where many projects already exist like for example Towntology [22] that builds an ontology about urban planning in which the terminology used by French urban actors is well clarified and organized.

3.1 Modeling of Application Data with Spatiotemporal Features (MADS)

MADS [9] is a conceptual spatio-temporal data model with multi-perception and multi-representation features, enriched by the object-oriented structure that has multiple advantages like inheritance, effective manipulation of temporal data, and uniform handling of spatio-temporal data. MADS handles its structural, spatial, temporal, and multirepresentation modeling dimensions in a way that intentionally makes them orthogonal. In MADS the multiplicity of level and perspectives is handled at both the schema and instance levels which provides a comprehensive approach for multi-level and multi-perspective modeling. Moreover, the spatial and temporal features are both included into the conceptual model.

The definition of a relationship in MADS is done in two phases. First, a link is set between the object instances using an association or a multi-association. Then, the link is given its specific semantics when needed. These semantics can be: aggregation, generation, transition, topological, synchronization and many others that deal with spatial, temporal, and multirepresentation features.

MADS supports specialisation/generalisation links or “is-a” links between relationship types and object types. The description of spatial and temporal features is allowed with either a discrete or a continuous view. A spatio-temporal object type in MADS is an object type that either has both a spatial and a temporal extent, separately, or its spatial extent changes over time. The same definition applies for a spatio-temporal relationship type that holds spatial and temporal information relevant to the relationship. Concerning attributes, both spatial (geometry) and temporal (lifecycle) attributes are present. In addition MADS presents space varying attributes and time varying attributes in order to support the continuous view. MADS explicitly supports multiple perceptions for the same database which is a very important feature for us since geographical applications have strong requirements in terms of multiple representations [23]. In order to organize multiple representations[24], two complementary techniques are provided. The first one consists in building a single object type with several representations and the stamps of the properties of the type provides the perception to which a representation belongs. The second technique is to define the same object type two times, each one bearing the corresponding perception stamp. Then both objects definitions are linked using a relationship type holding a specific inter-representation semantics [23].

Despite all the great and important features of MADS, it still works with the closed work assumption while, for our approach, we need to be able to add and change concepts and relationships and explicitly keep a history of their evolution through time. Since we, as other researchers [11], believe that better spatio-temporal data representation for information systems and including geographic ones can be reached by the use of ontologies, and in order to fulfill the shortcomings of MADS, our approach consists on creating an ontological model from the original MADS model and combining them afterwards, as explained in the section below.

[24] An object or relationship type is a multi-representation if at least one of its characteristics has at least two different representations

3.2 Proposal

Our approach is based on the assumption that conceptual models are the representation of the user's perception of a certain domain of interest and that they cannot be updated regarding the progress and evolution of information technology. Therefore, they can be considered as an interesting starting point for a more powerful model where these evolutions can be applied. In our aim to improve conceptual modeling, we choose the MADS conceptual model to be the starting point of our approach for his great capabilities and important features discussed before. Then we use the first approach of mapping from databases to ontologies that consists on creating a new ontology model from the existing database and populating the generated ontology with the contents of the database. This new ontology is combined to the existing MADS model. Mappings are defined as correspondences between original database components (table, relationship ...) and ontological components (concept, property ...). We enrich our generated ontology by taking into consideration the semantics of the relationships defined between database components (topological, generation …). We also propose adding temporal attributes to the definition of all ontological components (Classes, Properties …).

Now that we have the ontological structure, we can add new concepts to the model, thus concept evolution is no longer impossible. Ontologies gives us the ability to add new concepts which will be treated identically to primitive concepts by the ontology, thus these new concepts can be instantiated. In order to track instances history we propose a better method instead of MADS time-stamped attributes. Our method is implemented in the ontology and takes into consideration the new instances of the defined concepts (MADS doesn't). This method consists on defining new classes[25] for the three domains (Space, Time and Theme) in addition to a Concept class, then relating all these classes with their relative object through an explicit object property. We illustrate this approach by an example based on the real Urban Development Planning (UDP) of the city of Nantes – France. In this planning, we distinguish between five main categories of territories: "Artificial Territories", "Agricultural Territories", "Forest and semi natural environment", "Wetlands" and "Water Surface". "Artificial Territories" are composed of "Urban Zones", "Industrial and commercial areas", "Mines", "Artificial nonagricultural Green spaces" and "ROW water treatment". "Urban Zones" can be "Continuous urban fabric", "Discontinuous urban fabric" or " Specialized urban areas". In the "Continuous urban fabric" we have the "Downtown Type" where we are the "Dense Core" and the "Suburbs". We limit our example to the "Dense Core".

In this limited example:

1. A "Dense Core" is composed of "Structure Types".
2. Structure Types are "Building", "Road" and "Land".
3. Buildings are either public or private.
4. Many lands may be joined together, or a land can be split in many smaller lands.

[25] These classes are implemented intentionally in the ontology in order to study the saved information (and the explicit relationships) that might can be re-organized and make use of the accumulated metadata using classification and Data Mining techniques.

5. Roads have borders with lands.

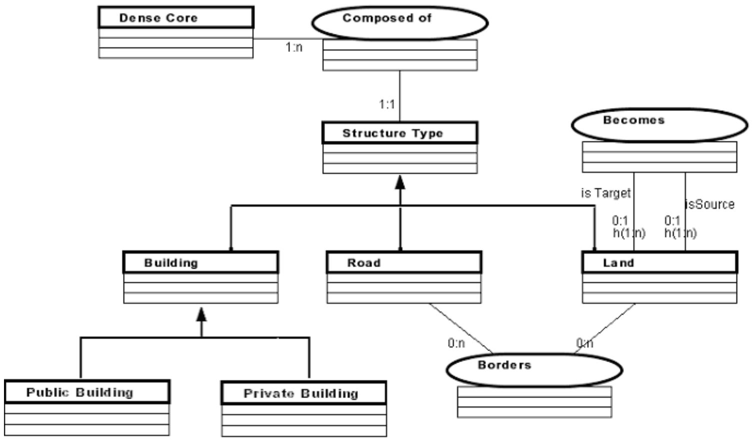


Fig. 10. Dense Core MADS Representation

The MADS relative conceptual model is given figure 1. This MADS model considers seven object types and three relationships (two associations and one multi-association). In order to translate this model into The Ontology Web Language OWL we used using the following transformation rules:

Object Type --> owl:class
Is A Link --> owl:SubClassOf
Relationship Type --> owl:ObjectProperty with range and domain
Simple Attribute --> owl:datatypeProperty
Role Cardinality --> owl:mincardinality, owl:maxcardinality

While doing the transformations we added, to each component, two time attributes designating its activation and deactivation dates using the primitive rdfs:label.

We assume that later on, the designers of the UDP felt the need of defining a new concept "Pedestrian road". This concept is a sub-concept of the Road Class. Its definition is done easily in OWL using the following syntax:

```
<owl:Class rdf:ID="Pedestrian Road">
<rdfs:label xml:activated="06/03/2007" xml:deactivated="">Pedestrian Road</rdfs:label>
        <rdfs:subClassOf>
                <owl:Class rdf:ID="Road"/>
        </rdfs:subClassOf>
</owl:Class>
```

Tracking instances history will be done from within the ontology by implementing a new explicit relationship linking four new classes, one for each domain (temporal, spatial, thematic and conceptual) to their relative object identifier. The Temporal Class has three attributes (TimeID, Date and InstanceID). The Thematic Class has three attributes (ThemeID, InstanceID and TimeID) in addition to all the temporal thematic attributes in the model. The Spatial Class has also three attributes (SpatialID, InstanceID and TimeID) in addition to all the temporal spatial attributes in the model. The Concept class has four attributes (ConceptID, Concept Name, InstanceID and TimeID). These classes hold the attribute values of the instances through the time. An algorithm is developed to construct the explicit temporal relationships considering that each instance has a creation node and a destruction node and can be subject to several changes. The creation/destruction node is a timestamp indicating the time of creation/destruction (TimeID). There are eight possible temporal changes (change in space, in theme, in concept, in space and theme, in space and concept, in concept and theme, in space, theme and concept, and finally no change) [24]. In order to detect these changes, different attributes of each instance are retrieved and compared through time and each instance change implies a temporal relationship creation. An explicit temporal relationship has five attributes:

- Relation Ship ID: the identifier of the Relationship.
- Instance ID: the identifier of the Instance.
- Start Node: Instant where the relationship began.
- Finish Node: Instant where the relationship ended.
- Value: the value of the temporal relationship that will be determined as follows: At object creation, value is set to "Begin". If an object is a new entity derived from a previous existing object the relationship value will be set to "Begin-Replaced" instead of "Begin". Each possible change will have a specific value ("Change in space", "Change in theme", "Change in concept"…). And when an instance is destroyed the value will be set to "End". When no change is detected for the second time consecutively, no new relationship is created; instead the algorithm updates the value of finish node in the previous relationship.

For illustration, consider an instance of land "Land A" that had a rectangular shape[26] in September 2000, and its owner was "Franck". In January 2002, Franck sold a part of this land and the remaining part took the square shape. Then in 2007, Franck sold his land to Alexander. Since then no changes were detected (Figure 2).
The algorithm will create the relationships

Table 2. Representation of the explicit relationships[27]

Relation	Instance ID	Value	Start Node	Finish Node
1	Land A	Begin	T1 [2]	T1
2	Land A	No Change	T1	T3
3	Land A	Change in Space	T3	T4
4	Land B	Begin Replaced	T3	T4
5	Land A	No Change	T4	T5
6	Land B	No Change	T4	T6*
7	Land A	Change in Theme	T5	T6

[26] In this example, the designers give for each land a spatial attribute called "geometrical shape" that can take values like "rectangle", "square" … and a thematic attribute named "owner".

[27] T1 … T7 are instances of the Temporal Class. An example of an instance would be:
TimeID = "T1", Date = "1999.11.05: 05.34.22", Instance ID = "Land A".

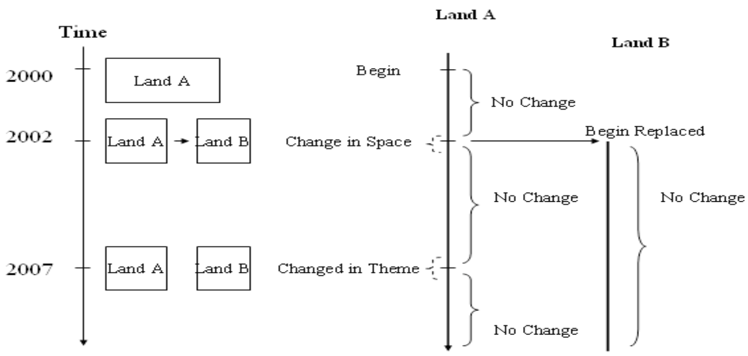


Fig. 2. Temporal Relationship Structure

Instance history tracking provides easy understanding of the transition of a group of instances in a certain domain by answering questions such as: Who was the owner of this land in 1992? What lands changed in theme since 1997? What are the newly defined lands? Etc.

4 Conclusion and Perspectives

In this paper we presented a solution that allows the evolution of spatio-temporal data models in terms of concepts. We discussed multiple models that contributed in this domain and that lead the MADS conceptual model. We studied MADS in detail, and highlighted his shortcomings concerning our needs. Then, we showed how ontologies can overcome MADS' limitations. Based on these studies we proposed combining both models in order to afford the maximum performance. Our approach consists in transforming the MADS model to OWL, then adding temporal attributes to the each component's definition and finally, in order to manage historical changes of attributes in theme, space and concept, we implement an explicit temporal relationship in the ontology model. An example taken from Nantes Urban Development Planning was given as an explanation to our approach. We still have a work to do mainly concerning query redirection and interpretation. And since ontologies are a shared source of knowledge, that can be used by many user communities, we are planning to improve the interoperability and the use of our ontologies by corresponding our ontological concepts to the terms that are already used by other projects like Towntology.

Acknowledgements. We thank Nantes Métropole which has provided the GIS data.

References

1. Pelekis, N., Theodoulidis, B., Kopanakis, I., Theodoridis Y. Literature review of spatio-temporal database models. In: Knowledge Engineering Review 19, 3, 235–274 (2004)

2. Armstrong, M.P.: Temporality in Spatial Databases. In: Proceedings of GIS/LIS '88, American Congress on (1988)
3. Langran, G. E.: Time in Geographic Information Systems. Doctoral Thesis. UMI Order Number: UMI Order No: GAX90-00269., University of Washington (1989)
4. Hunter, G. J., Williamson, I. P.: The Development of a Historical Digital Cadastral Database. Int. Journal of Geographic Information Systems 4, 2, 169 – 179 (1990)
5. Peuquet, D., Duan, N.: An Event-Based Spatiotemporal Data Model (ESTDM) for Temporal Analysis of Geographical Data, Int. Journal of Geographical Information Systems 9, 1, 7-24 (1995)
6. Yuan, M.: Wildfire Conceptual Modeling for Building GIS Space-Time models. In: Proceedings of the GIS/LIS'94, 860-869 (1994)
7. Worboys, M. F., Hearhshow, H. M., Maguire, D. J.: Object-Oriented Modeling for Spatial Databases, Int. Journal of GIS 4, 4, 369–383 (1990)
8. Worboys, M. F.: Event-oriented approaches to geographic phenomena. International Journal of Geographic Information Science, 19(1):1--28 (2005)
9. Parent, C., Spaccapietra, S., Zimányi, E., Donini, P., Plazanet, C., Vangenot, C., Rognon, N., Crausaz, P-A. : MADS, modèle conceptuel spatio-temporel. In : Revue internationale de géomatique 7, 3, 317-352 (1997)
10. Choi, J., Seong, J. C., Kim, B., Usery, E. L.: Innovations in Individual Feature History Management—The Significance of Feature-based Temporal Model. In: Geoinformatica 12, 1 (Mar. 2008), 1-20 (2008)
11. Frank, A. U.: Ontology for spatio-temporal databases. In: Spatiotemporal Databases: The Chorochronos Approach. 9-78 (2003)
12. Gruber, T. R.: A Translation Approach to Portable Ontology Specifications. Knowledge Acquisition. 5, 2 (June 1993), 199–220 (1993)
13. Bateman, J., Farrar, S.: Spatial ontology baseline. SFB/TR8 Internal Report I1-[OntoSpace]:D2, CRCSC, University of Bremen, Germany (2006)
14. Cullot, N., Parent, C., Spaccapietra, S., Vangenot, C.: Des SIG aux ontologies géographiques. In: Revue Internationale de Géomatique 13, 3, 285-306 (2003)
15. Casati, R., Smith, B., Varzi, A. C.: Ontological Tools for Geographic Representation, N.Guarino (ed.), Formal Ontology in Information Systems, IOS Press, 77-85 (1998)
16. Zhou, Q., Fikes, R.: A Reusable Time Ontology. In: Knowledge Systems Laboratory Technical Report KSL-00-01, Stanford University (2000)
17. Khatri, V., Ram, S., Snodgrass, R. T.: ST USM Bridging the Semantic Gap with a Spatio-Temporal Conceptual Model, A TIMECENTER Technical Report 64 (2001)
18. Bittner T., Smith B.: Granular spatio-temporal ontologies. In: AAAI Spring Symposium on Foundation and Applications of Spatio-Temporal Reasoning (FASTR) (2003)
19. Grenon, P., Smith, B.: SNAP and SPAN: Towards Dynamic Spatial Ontology. Spatial Cognition and Computation 4: 69-103 (2004)
20. Hornsby, K. S., Yuan, M.: Understanding Dynamics of Geographic Domains, CRC Press (2008)
21. Cullot, N., Ghawi, R., Yétongnon, K. DB2OWL: A Tool for Automatic Database-to-Ontology Mapping. In: Proceedings of the 15th Italian Symposium on Advanced Database Systems, (Torre Canne di Fasano, Italy, June 2007), 491-494 (2007)
22. Roussey, C., Laurini, R., Beaulieu, C., Tardy, Y., Zimmerman, M.: Le projet Towntology : un retour d'expérience pour la construction d'une ontologie urbaine. Revue Internationale de Géomatique 14(2):217-237 (2004)
23. Parent, C., Spaccapietra, S., Zimányi, E.: Conceptual Modeling for Traditional and Spatio-Temporal Applications: the MADS Approach. Springer-Verlag New York, Inc (2006)
24. Zaki, C., Servières, M., Moreau, G.: An ontologybased model for representing evolution of both data and semantic in GIS. In: GI-Days Proceedings, 157-164 (2008)

Constructors of Geometric Primitives in Domain Ontologies for Urban Environments

J.Finat [1,2], F.J.Delgado [1], R.Martìnez [1], J.J.Fernández [2],
J.I.San José [2] and J.Martìnez [2]

[1] MoBiVA Group, Lab 2.2, Scientific Park, 47011 Valladolid, Spain
[2] Laboratory Architectural Photogrammetry, ETS Arquitectura, 47014 Valladolid
University of Valladolid, Spain

Abstract. A sustainable urban environment requires an integration of documentation and management systems in surveying and planning applications for different kinds of services involving services and processes. Traditional GIS approaches are geo-referenced to 2D information or, more recently, to a virtualized environment arising from the reproj7ection of textured views on 3D models. In this work, a tool for extracting dominant planes from dense 3D range information is applied for the automatic identification of structural elements (façades, ground and roofs) and their automatic labelling in terms of data of dominant planes. This tool is integrated inside the software platform UvaCad.
The constructor of urban geometric primitives is given as a generator of 3D chains corresponding to the grouping of maximal planar quadrilaterals with ”similar properties” for the normal unit vector. As a by-product, one has a semi-automatic volumetric segmentation of small urban environments as support for the development of a Domain Ontology in urban environments. Our methodology is illustrated with some examples corresponding to small villages and typical central squares (Plaza Mayor) of old Spanish cities.

1. Introduction

Urban or architectural surveying, spatial planning and simulation of interactions in complex environments present increasing requirements involving knowledge representation with respect to a well-defined and robust model. Robustness arises from geometric data and it is linked also to the automatic elimination of outliers from the computational viewpoint. An accurate geometry provides a robust support for advanced visualization, including functionalities such as navigation, inspection and interactive simulation of possible interventions. Simultaneously, the information exchange between different experts and reuse of implicit information on augmented 3D models require to solve interoperability issues between different resources with a large heterogeneity for data and metadata of 3D objects. Information reuse in repositories of 3D urban objects must be solved in a semantic framework relative to domain, tasks and/or users to different LoD. This issue poses general problems (see [4], [7] for ”old” references or [14] for more recent results).

It is necessary to design a multilevel and multipurpose software platform able of integrating data, actions and behaviours with their functionalities and constraints (involving goals, roles and agents), and solving the interoperability issues. The integration of modeling and software tools for design, simulation, execution and management of urban resources is an ambitious program which is still far from being achieved. There exists a general agreement about the need of developing a semantic framework for accomplish this program, but there is still the need of developing software tools for processing, analyzing and exchanging information about different kinds of modeling and data treatment. The formalization of urban modeling is a relatively recent area with important contributions in the procedural framework ([8], [5]) or, more recently, in the semantic framework ([12], [13]).

The main challenge is the design and implementation of an integrated 3D information model for urban environments. The pioneering work of T.Kolbe [6], and his collaborators is a reference work in this domain with important contributions of other R& D groups ([1]). The Open Source Software platform CityGML (`www.citygml.org/`) provides a modular approach for integrating surveying, planning and simulating interventions, and with several applications in technological domains. CityGML provides a hierarchical support for integrating every kind of spatial and non-spatial features in a common framework. CityGML is organized following a semantic approach for making easier the interoperability between different sources, tasks and knowledge domains. In this work a special attention is paid to semantic aspects related to Domain Ontologies for Urban Environments.

A *Urban Domain Ontology* concerns to the modeling of urban domains and knowledge about individuals and their relations in their spatial context. The logical support for urban modeling consists of objects and/or concepts, and their properties and relationships which are managed from a logic viewpoint involving attributes in measurable fields (geometric and radiometric properties usually, or in vector and raster data following the usual GIS terminology). Knowledge about relationships and individuals in a spatial context must be translated to interfaces used for generating and communicating knowledge following predicate or descriptive logics involving physical objects. A city is not only a set of buildings, not even can be reduced from the symbolic viewpoint to a network with nodes, paths and zones. The understanding of spatial delimitation is important for functionalities and uses which very often are historically and socially determined, and this involves to relations between different agents which must be translated in terms and logical rules.

A crucial aspect for linking objects and concepts concerns to the design and implementation of software tools for semi-automatic Recognition of georeferenced data. Digital inputs arise from 2D views or 3D representations of volumetric objects. Low- and mid-level Recognition from discrete data follow clustering principles with agglomerative (bottom-up, merging) and divisive (top-down, splitting) strategies. Constructors are nearer to agglomerative strategy from unstructured information, whereas descriptors would be nearer to the identification of attributes (radiometric properties, .e.g) or parameters (numerical data, .e.g) of basic elements which can be found in already available models. Constructors and descriptors are focused towards the construction of a generative grammar for urban environments. Obviously, the solution of this problem depends on the scale. For large scale, a very interested contribution for automating the generation of urban models from aerial skew photography (levels 1 and 2 in the terminology of CityGML) in terms of a constructive grammar following like-fractal models can be seen in [9]. Our work takes a different viewpoint because the inputs arise from range-based information arising from a laser scan device Ilris 3D (Optech). It is mainly focused towards a very limited class of constructors linked to piecewise-linear (PL- in the successive) modeling for boundary surface representations (BSP). Inputs for our modeling arise from merging discrete clouds of points captured with a Laser scan Ilris 3D (Optech).

Some issues regarding to computer implementation of Domain Ontologies including constructors and abstractors remain still open. Both issues are meaningful for Semi-Automatic Recognition and for management of domain ontologies, also. Constructor and abstractor operators are in some sense conjugate between them; they are strongly related to agglomerative and divisive approaches in regard to the physical support linked to a urban GIS. Traditional aspects arising from usual GIS (with their spatial hierarchies organized in different thematic layers) are semantically integrated as a physical support for the development of interactions involving similar functionalities, in despite of their morphological diversity. In this work, we restrict ourselves to specifying relations between vector data of a 3D GIS for buildings and their representation in terms of different geometric representations. Inputs arise from laser scanning and digital cartographic support for urban environments. Some cadastral applications have been developed for conservation interventions of urban neighbourhoods in small villages of Palencia (Spain). A related work can be found in [10]

A large number of old cities in Western Europe are articulated around squares as bounded spaces for organizing activities between individuals in a spatial context. The Plaza Mayor of Valladolid (Spain) is a paradigm because it plays an important role as institutional and political centre from 13th century, and it has been taken as model for designing a large number of squares between 16th and 18th centuries, not only in Spain, but also in Latin America; thus, we have paid some attention to this example (more details in [2] from a GIS viewpoint). Irregularities in old towns are disappearing by modern urban re-design, but they are ubiquitous in smaller villages with vestiges of Renaissance style which are superimposed to an irregular network (sometimes with Arabian influence); to illustrate them, we have included some examples arising from the scanning of historical centre of several small villages of Palencia (Spain), with a semi-automatic detection of dominant planes and boundary surfaces for representing complex objects linked to larger urban spaces (squares, blocks). Our range-based method is explained in [3].

Fig. 1. Pasting dense point clouds.

This paper is organized as follows: Section 2 is focused towards the presentation of some generalities about the application of Ontologies for urban environments. Section 3 deals with the method performed for extracting dominant planes. Section 4 involves to the construction of intermediate geometric entities (connected polygonals of dominant planes) which are meaningful for bounding urban environments (streets, blocks). Some remarks about designing schemata and populating ontologies are developed in section 5. A brief description of an experimental set-up is presented in the section 6. The paper ends with an exposition of work in progress and challenges for the next future.

2. Generalities about Ontological Approach

Morphological urban diversity and the functional heterogeneity involving different agents performing complex tasks require a semantic approach with well specified ontologies for domain, agents and tasks. Following T.R.Gruber, an Ontology is a specification of a conceptualization. This specification can be formulated in different frameworks which involve knowledge fields as diverse as philosophy (a logical language for connecting different services and interaction modalities) or engineering approaches (software design and implementation).

Representations of physical space, relations between different agents and evolving complex systems provide a support for the interplay between GIS, logic and engineering approaches to Urban Ontologies. A first demarcation involves to the specification of lexicon, Thessauri and taxonomies for each Ontology. The specification of appropriate Lexicon and Thessauri for urban environments is being successfully accomplished in the framework of [11]. Thessauri add relations between acknowledged concepts belonging to lexicon. Rules involving such relations are usually verified in an interactive way by validating logical predicates, including uncertainty aspects which are characteristic of human knowledge. Thus, Thessauri and taxonomies include functionalities (relations and operators) of a semantic approach, and must be inserted as an independent module of the database relative to the physical spatial support. The design and implementation of software tools for intelligent identification of such functionalities is a challenge for the next future.

More formally, a *semantic framework for geometric objects* appearing in *urban environments* concerns to specifying the *three levels* of an ontology for urban spaces:

- a *lexicon* given by basic geometric primitives (including dominant planes and simple quadrics, e.g.);
- *thessauri*, i.e. acknowledged vocabularies and relations between terms involving to urban spaces; and
- a *taxonomy*, i.e. a set of rules with the specification of logic types for management of roles played by ”urban spaces” (streets, or squares, e.g.)

The most difficult part concerns to the specification of taxonomies. Troubles arise from two sources: There are two types of logic (propositional and descriptive), and we have not still a well-defined and commonly accepted geometric linguistic for the logic verification of assertions in terms of relators and constructors. Due to this limitation for finding semi-automatic solutions for formalizing our knowledge, most solutions introduce beacons on the scene or tags on data for solving recognition issues (manual labelling, usually).

The articulation between the above three levels must be hold in different types of ontologies involving the domain (urban typologies, mainly), tasks to be performed (surveying, planning, visualizing, modifying, consulting, extracting information) and different kinds of involved users (professionals, civil administration, business or services stakeholders, citizens, etc). The articulation holds at different levels and in terms of different ”events” linked to the interaction.

The automatic generation of Domain Ontologies requires solve Recognition problems from digital support involving 2D views or 3D representations of volumetric objects as independent of user as possible. Recognition problems have a long history in Computer Vision with a feedback between bottom-up and top-down approaches. Query, retrieval, indexing and classification are the main stages for Recognition problems, but there is no still a satisfactory solution for complex objects such that those appearing in urban scenarios. Thus, it is necessary to restrict Recognition to a low number of (geometric or radiometric) primitives which can be grouped in more complex objects to be merged by means of clustering strategies depending on topological (adjacency, proximity, continuity), radiometric (similar grey intensity levels) or geometric (colinearity, coplanarity) constraints.

The application of segmentation and clustering techniques has allowed identify 2D zones, extract and enhance 1D boundaries and close polygonals linked to objects appearing in digital 2D cartography or aerial photography of urban zones. All these data are exported to the 0th level of CityGML and labelled as meaningful data. The same techniques are applied to extract data in orthoviews of façades with a similar strategy for the automatic extraction of simple 2D primitives with their corresponding vector and raster data. These data are projected on 3D models by using the UvaCad software platform for the information insertion on 1st and 2nd levels of CityGML on a toy 3D model. Unfortunately, we have not still enough fine software tools for the automatic identification of simple geometric primitives appearing in buildings (columns, e.g.) in terms of their functionalities or their structural role.

3. Semi-automatic Detection of Dominant Planes. Local Aspects.

The structural role of some components is better understood in terms of simple geometric 3D models, which can support higher resolution models by reprojecting views on 3D models. In this work, accurate discrete 3D models are generated from range information captured with the Ilris 3D laser scan of Optech. The color and additional details arising from high resolution views is reprojected on the resulting dense cloud of points by means of simple geometric transformations between orthoviews and conveniently oriented clouds of points; this solution has been implemented in the framework of UvaCad software platform. A recent extension of this solution allows identify dominant planes in an automatic way. The estimation of dominant planes is performed by constructing a triangulation and by grouping adjacent triangles with a similar unit normal vector, up to a threshold which can be selected by the user.

A dominant plane is the support for a maximal connected planar region (not necessarily simply connected) with a total surface higher than a threshold fixed by the user; most of such planar regions are (unions of coplanar) quadrilaterals with holes corresponding to elements not belonging to the plane or shadows for the scanning (arising from partial occlusions, .e.g). In urban scenes, unions of coplanar quadrilaterals support the visible part of façades, roofs or ground, which are labelled with usual attributes introduced in CityGML. Currently, most of attributes are manually inserted. Only, labels relative to façades, roofs or ground are automatically generated.

Selection of main direction for dominant planes is performed by voting procedures for the list of normal unit vectors linked to a triangulation superimposed to the cloud of points. The distribution of normal unit vectors to the triangles is far from being uniform in the unit sphere. Indeed, there are a finite number of local maxima whose typical (modal) values correspond to dominant planes. So, a) normal unit vectors corresponding to façades are distributed along the equator; b) normal unit vectors corresponding to the ground are concentrated in the North Pole, up to streets with meaningful slopes; c) normal unit vectors corresponding to roofs are distributed along a parallel of the spherical representation whose typical slope depends on climatic or cultural aspects.

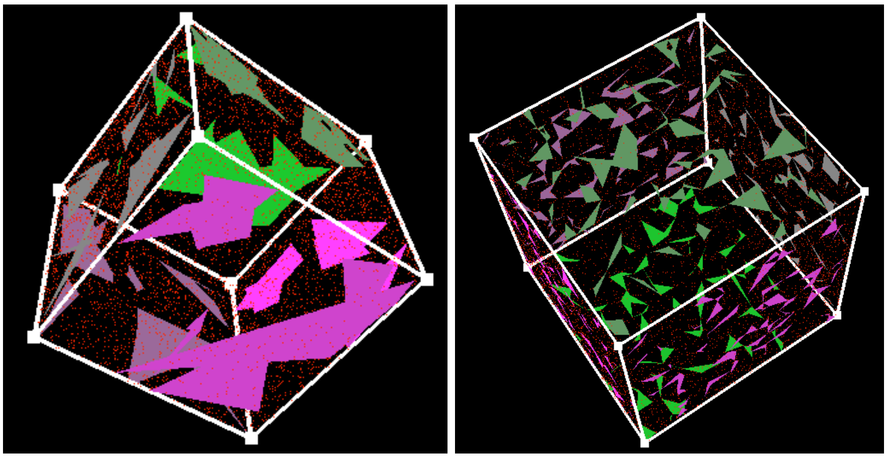


Fig. 2: Different resolution for dominant planes detection in a cube (100 pts. and 600 pts.).

Irregularities (due to the relative orientation or the reflectance variability, e.g.) or even discontinuities (due to partial occlusions, e.g.) in spatial distributions of clouds of points are a source of uncertainty for the automatic generation of 3D models. Thus, we have developed a coarse-to-fine methodology able of identifying coarse models even when one has only a ”soup of triangles” for a well-defined geometric structure. Coarse models correspond to simple geometric primitives (parallelepipeds, spherical or cylindrical components, e.g.), whereas fine models are linked to an estimation of curvatures from discrete clouds of points. For illustrating the developed method, a visualization of dominant planes corresponding to dominant planes of a cube is displayed in the next figure. Our application labels faces and draws theoretical edges corresponding to the intersection of adjacent faces.

Dominant planes in urban environments provide a support for façades, ground and roofs, which are modelled as an union of 3D large quadrilaterals not simply connected (i.e. with holes corresponding to windows, doors, cornices, etc), whose boundaries are given as the intersection of dominant planes with nearest triangles. The visible part of the façade is not necessarily a quadrilateral. However, for automatic identification of structural elements and for simplifying their computer treatment, it is convenient to replace the visible part of the bounding surface by a quadrilateral Q_i. The quadrilateral Q_i can be obtained by a small modification of the algorithm based in the detection of extreme points for convex hulls. Depending on the package corresponding to the normal unit vector $\mathbf{N}_i$, each quadrilateral is automatically labelled with a superindex:

- (f) (façade) if the third component of $\mathbf{N}_i$ is null (up to a threshold)
- (g) (ground) if the first and second components of $\mathbf{N}_i$ are null (up to a threshold)
- (r) (roof), otherwise.

Automatic labelling is a crucial step for generating chains of elements belonging to the same type, and it allows recovering and managing street façades as a whole, e.g.. The selection of transversal sections to a cloud of points allows the automatic management in terms of 2D information. Next example is linked to a very narrow street of a small village (Paredes de Nava, Palencia, Spain), where one can select in an interactive way the allowed maximal number of façades (3 in this case) for obtaining a polygonal where each segment represents the normal section of a façade. Small irregularities w.r.t. dominant planes are detected and displayed in order to correct the initial selection.

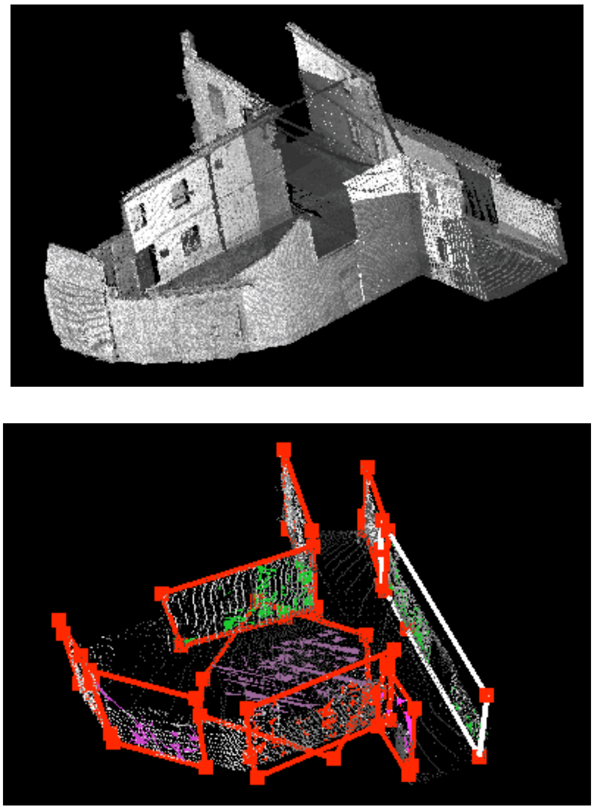


Fig. 3. Polygonal representation of a continuous facades in a Place. Becerril de Campos, Palencia, Spain.

Furthermore, there exists a smart option which avoids the manual selection of the allowed maximal number of façades: it suffices to connect extreme ends of the polygonal by means of a segment $s_{0N} = \overline{\mathbf{P}_0\mathbf{P}_N}$ and evaluate the orthogonal distance of intermediate vertices $\mathbf{P}_i$ w.r.t. the added segment. If the distance

$d(\mathbf{P}_i, s_{0N})$ is higher than a threshold, then replace the segment s_{0N} by the polygonal obtained connecting vertices $\mathbf{P}_0, \mathbf{P}_i, \mathbf{P}_N$. An easy iteration gives an improved polygonal.

Often, it is not possible the scan of roofs, but data arising from aerial photography can be inserted in an interactive way, by selecting an adequate number of control points. By merging maximal parallelograms linked to adjacent façades, we construct a closed 3D polygonal of façades which extends the usual 2D information based in digital cartography with high accuracy (error less than 2 cm for scale 1:100). This approach is compatible with multiresolution approaches. Global management is performed in terms of octrees. This solution has been applied to 4 small villages of Palencia (Spain) with some vestiges of Civil Renaissance buildings.

4. Merging Dominant Planes. Symbolic Representations for Taxonomies

Automatic Detection of Dominant Planes provides a collection $\mathcal{Q}$ of quadrilaterals Q_i with adjacency relations. Hence, an adjacency graph $\mathcal{G}$ is associated to the collection $\mathcal{Q}$ of quadrilaterals. In the adjacency graph $\mathcal{G}$ each quadrilateral Q_i corresponding to a dominant plane DP_i is represented by a node $\mathbf{n}_i$; each segment $Q_i \cap Q_j \subset DP_i \cap DP_j$ of two quadrilaterals corresponds to the edge e_{ij} of the graph $\mathcal{G}$. Each edge is labelled with two superindexes of the collection (*f*), (*g*) or (*r*) depending on the label façade, ground or roof linked to labelling procedures described in the above section. A priori, the graph would contain cycles, but due to incomplete information some edges could be absent. Polygonal of planes with similar characteristics are symbolically represented by maximal connected subgraphs of the adjacency graph $\mathcal{G}$ whose nodes are labelled with the same superindex (*f*), (*g*) or (*r*). The Fig.4 displays a collection of façades belonging to the same tortuous street of Paredes de Nava (Palencia, Spain).

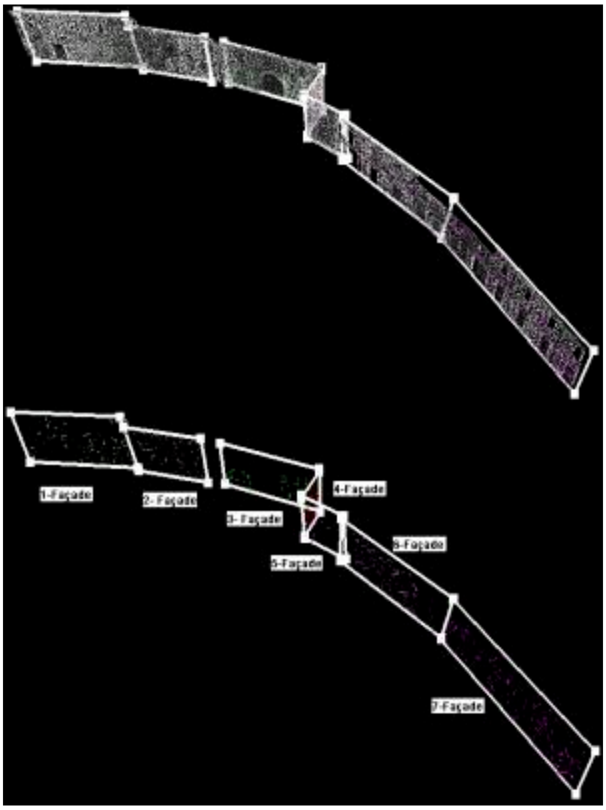
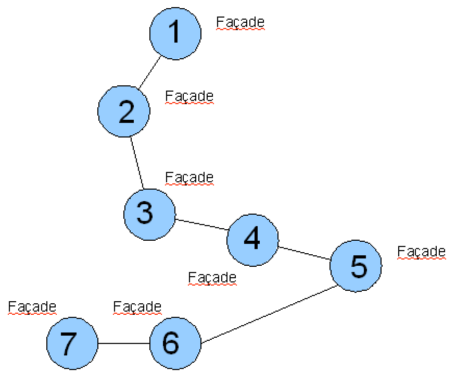


Fig. 4. Recognition and automatic labelling of planes in the streets of Paredes de Nava, Spain.



Often the adjacency graph is not necessarily connected, not even for small urban environments. This can be due to thresholds for radiometric information (non-bounded extremes of streets or doors of country yards are not conveniently detected, e.g.) or to geometric considerations (self-occlusions, incomplete information about roofs or the ground, e.g.). The Fig.3 illustrates both troubles.

To solve these troubles, it is convenient to introduce a spatial or volumetric hierarchy obtained by cutting out the original cloud $\mathcal{C}$ of points. A regular spatial hierarchy is obtained from the superposition of an octree on the cloud $\mathcal{C}$ of points. In view of the large amount of information from original clouds, visualize sampled information of the whole cloud for each cell of the octree, only. However, by selecting the appropriate option in the menu, it is possible to increase the resolution for each octree cell. The recursive information management slows down the execution, but increases the performance of our application. The generation of polygonal of dominant planes corresponding to the street façades or surfaces bounding blocks allows to introduce an adaptive and more "natural" hierarchy attending to structural urban elements, by generating small pieces of a spatial network where urban complex primitives inherits their characteristics from bounding surfaces in an automatic way.

5. Designing Schemata and Populating Ontologies

Specific vocabularies for each knowledge domain provide the lowest level of ontologies corresponding to a typical lexicon; it is commonly managed in terms of relational database involving different kinds of multimedia documents. Labelling of multimedia documents supports the application of different kinds of representations for data semantic. We have used RDF (Resources Description Framework) for information maintenance by means of a relational binary representation. We have initially applied to disperse Cultural Heritage goods in rural environments. In this case, we have populated this ontology by the manual insertion of tags according to Dublin Core Standards. The reuse of already existing DB for achieving a conceptual framework is performed by using local agents. For developing a simple collaborative environment for documentation of strongly damaged vestiges of Cultural Heritage in rural environments. Limitations of inheritance procedures for traditional RDF are solved by the introduction of RDF schemata. However, a simultaneous searching of descriptions and RDF schemata requires a declarative language whose implementation is still in progress for this domain.

Having in mind cadastral applications for local administrations, we have developed an extended 3D Information System in a collaborative framework where several experts can perform different consultations, update information and generate reports according to access protocols. The simplest level for information management is performed in terms of relational database with an extensible collection of terms and relations from users. In this way, we obtain a mid-level structure for the domain ontology. The resulting methodology is being applied to conservation and restoration tasks in isolated buildings (cathedral of Leon), where several instances (General Direction of Cultural Heritage) or individuals (experts in conservation or restoration) insert terminological definitions which are very useful for populating the corresponding domain ontology. However, persists some ambiguity relative to the use of objects as concepts or, alternately, as things. Currently, labelling is performed in a manual way according to an extensible vocabulary on the 3D Information System PINTA (Processing INformation SysTems for Architecture) which has been also developed by the DAVAP cluster.

The current version of the multiplatform software PINTA is focused towards isolated buildings, and it provides a support for collaborative work of experts working on different layers with a common vector reference. Update and revision of daily information can be performed in a remote way, and the organization in different layers allows a remote revision and the insertion of simulated solutions (using VR tools which are being incorporated to UvaCad). To make easier the information management linked to PINTA, a limited lexicon has been used for illustrating the reaching of this 3D Information System. This choice is justified by the need of development of robust solutions for generating the automatic association of labels to objects detected in 3D models arising from discrete clouds of points. The inherent ambiguity of traditional semantic approaches is avoided by taking as reference a hierarchy of 3D Geometric models for each level analysis.

6. Experimental Set-up

Two sensitive parameters in the detection of dominant planes are the chosen resolution for the cloud of points and the tolerance relative to the deviation from a typical value for unit normal vectors. Due to irregularities in the distribution of clouds of points, for detecting dominant planes, it is convenient to work with redundant information; so, for simple geometric primitives (planes, spheres, cylinders, e.g.) it suffices to take some hundreds of points which can be obtained from several sampling procedures based in brute force (decimation) or smart (RANSAC or IMPSAC) algorithms. Tolerance (between 0 and 1) determines the granularity of the estimated plane; a very low tolerance generates too many non-meaningful planes due to noise, whereas a very high tolerance includes many different planes in only one, by losing meaningful information. In our examples, the value 0.04 has been chosen as the most appropriate for the information relative to dominant planes for houses of villages of Palencia. In our application, it is possible also to select the maximal number of triangles, and voting procedures impose strong limitations on the maximal number of dominant planes. This method has been applied to isolated houses, street façades, and blocks of houses.

If the cloud has lesser than 50.000 points the "soap-of-triangles" method combined with a search of nearest neighbours (managed with octrees) provides a collection of dominant planes in less than one second in a Core 2 Duo 2.0 GHz. For larger clouds, the performance is worse; so for instance, for one hundred thousand points with six thousand triangles, our processor spends almost five seconds.

7. Work in Progress and Challenges

Nowadays, we are developing the appropriate interfaces for linking our processing tools with CityGML modules involving the Appearance, Building, and CityObjectGroup as the most meaningful models for levels 1 and 2 where we are working on (see CityGML UML diagrams). In the mid-term, the main trouble is linked to the design and implementation of appropriate taxonomies for populated Ontologies. Nevertheless the persistence of irregularities in façades (ornamental details in towers or cornices) and roofs (dormer windows in attics, e.g.) are not easy to recognise in an automatic way, and requires still a high degree of interaction from user.

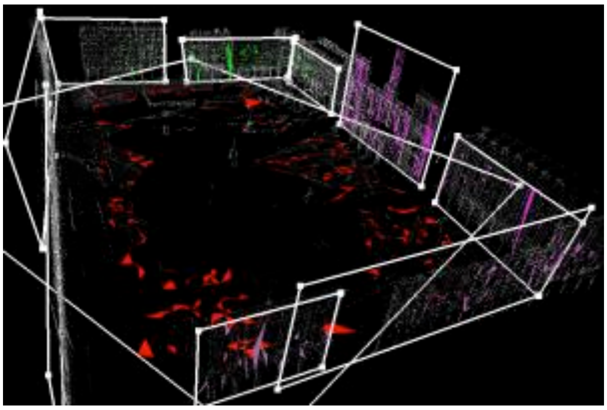


Fig. 5. Plaza Mayor of Valladolid reduced by 1 point for 0.5 m. Automatic recognition is difficult for complex facades.

Design and implementation of taxonomies for urban environments concerns to the automatic extraction of structural elements and their automatic labelling depending on the functional role. An extension of PINTA to urban environments instead of isolated buildings of cultural interest is in development, currently.

8. Conclusions

Semi-automatic Recognition requires the capability of relating structural elements of buildings with their meaning in urban context. Semi-automatic character corresponds to an interactive selection of parameters relative to resolution, tolerance and number of planes, e.g. In this work, we have developed a method for extracting and grouping dominant planes from sparse clouds of points without additional reference to cartographic information, according to levels of proximity and information coherence of small urban environments. Our application allows to obtain information about structural elements of buildings and blocks in regard to their urban environment (streets and squares, mainly) which are represented by means of a 3D graph. The modular design of our application for the detection and grouping of dominant planes allows the adaptation to almost every urban environment, and its integration in any application related to management of architectural information. Furthermore, it is possible to add raster information (textures or roofs from ortho- or skew views), and export the resulting data to geometric formats compatible such as those used in CityGML or X3D.

Acknowledgements. This work is partially supported by the ”Proyecto Singular Estrategico PATRAC (Patrimonio Accesible: I+D+i para una cultura sin barreras)”, PS-380000-2006-2 of the Spanish Ministry of Science and Innovation and by the AIVI (Ambientes Inteligentes para la Vida Independiente) project TSI-020302-2008-89 of the Spanish Ministry of Industry.

References

[1] L.Emgård & S.Zlatanova: ”Design of an integrated 3D information model”, Preprint 2006.

[2] J.Finat, L.M.Fuentes, J.Martinez, J.J.Fernandez, J.I.Sanjose, L.M.Garcia, V.Garcia; ”A MIMO approach for the information management of squares in historical urban centres”, CIPA Symposium, Lisboa 2004.

[3] L.M.Fuentes, J.Finat, J.J.Fernandez: ”Using Laser Scanning fro 3D urban modeling”, UDMS 2006 25th Urban Data Management Symposium, Aalbork, Denmark, 2006.

[4] Guarino N.: ”Formal Ontology in Information Systems”, In Proceedings of FOIS’98, Trento, Italy, 6-8 June 1998. Amsterdam, IOS Press. ,3-15, 1998.

[5] G.Kelly, H.McCabe: ”A Survey of Procedural Techniques for City Generation”, ITB Journal, 2007,

[6] Kolbe, T. & Gröger, G. 2003. Towards unified 3D city models. Proceedings of the ISPRS Comm. IV Joint Workshop on “Challenges in Geospatial Analysis Integration and Visualization II2 September 8-9, 2003 in Stuttgart

[7] Kuhn, W.: ”Ontologies in Support of Activities in Geographic Space”, International Journal of Geographical Information Science, Vol 15, N°7, 613-631, 2001.

[8] Y. I. H. Parish and P. Muller: Procedural modeling of cities. In Proceedings of ACM SIGGRAPH 2001, pages 301–308. ACM, 2001.

[9] R.Saleri Lunazzi: ”Urban and architectural 3D fast processing”, 9th Intl conf. on generative art, Italy (2006)

[10] Stoter, J., and van Oosterom, P.: ”Technological aspects of a full 3D cadastral registration”. International Journal of Geographical Information Science, Volume 19, Number 6, July 2005, pp. 669-696(28), 2005

[11] http://www.towntology.net. 2007-02-13

[12] J.Teller, A.K.Keita, C.Roussey and R.Laurini, "Urban Ontologies for an improved communication in urban civil engineering projects ", Cybergeo, Sélection des meilleurs articles de SAGEO 2005, article 386, mis en ligne le 11 juillet 2007, modifié le 31 août 2007. URL : http://www.cybergeo.eu/index8322.html. Consulté le 26 novembre 2008.

[13] J.Teller, J.R. Lee, C.Roussey (Eds.): ”Ontologies for Urban Development. Studies in Computational Intelligence”, Vol. 61 Springer-Verlag 2007, ISBN 978-3-540-71975-5.

[14] Zlatanova, S. and Prosperi, D. (eds.), Large-scale 3D data integration–Challenges and Opportunities, Taylor and Francis, London, 2006

Ontology-Based Analysis of Chat Conversations An Urban Development Case

Stefan Trausan-Matu[1,2], Traian Rebedea[1]

[1] "Politehnica" University of Bucharest,
Department of Computer Science and Engineering,
Splaiul Independetei nr. 313,
Bucharest, Romania
[2] Research Institute for Artificial Intelligence of the Romanian Academy,
Calea 13 Septembrie nr.13,
Bucharest, Romania
stefan.trausan@cs.pub.ro, traian.rebedea@cs.pub.ro

Abstract. Online collaboration in communities of urbanism experts is enabled by text-based tools, such as instant messaging (chat), discussion forums and web logs (blogs). The paper presents an ontology based system that analyses chat logs. The system integrates knowledge processing with natural language processing and discourse analysis based on Bakhtin's ideas. The system permits to detect the topics of chat, the threads of discussion and the important utterances. It also permits to visualize the graph of the conversation and it allows the inclusion of the detected topics as new concepts in the domain ontology.

Keywords: ontology, semantic closeness, dialogism, chat visualization

1 Introduction

The social interaction tools on the web, like discussion forums, blogs, wikis or instant messaging are basic components of the Web2.0 (the Social Web). Such tools started to be also used by communities of specialists in urbanism[28]. From the above mentioned tools, instant messaging (chat)[29], due to its specific online character, encourages multi-voiced inter-animation for collaboratively building knowledge, in a way similar to classical music polyphony or jazz improvisation [7, 8].

For developing chat analysis programs, human language processing is needed and, therefore, ontologies play a central role, because they constitute a basic knowledge representation framework for semantic analysis. In particular, in the system presented in this paper, ontologies are used for the identification of similar word threads (lexical chains), of the discussion topics and of the important utterances of a chat. The general lexical WordNet ontology (http://wordnet.princeton.edu) is used together with a domain ontology.

---

[28] See, for example http://www.cyburbia.org or http://www.planetizen.com

[29] An example of a chat session in the domain of urbanism may be seen at http://www.planetizen.com/node/30186 or http://www.planetizen.com/node/30813 (last accessed on 28 January 2009).

One important feature of the system presented in this paper is that the domain ontology is extensible. For example, after new concepts are identified as a result of the analysis of the chats, the user may include them in the domain ontology and also introduce relations among them.

In addition to classical ontology-based natural language processing techniques, the polyphonic model of Bakhtin is used in order to identify inter-animation patterns among chats' discourse threads [7]. The same framework may be also used for the analysis of other social interaction tools (forums or blogs).

The paper continues with a section introducing some ideas about ontologies. The third section discusses the socio-cultural and Bakhtin's dialogism paradigms. The next section contains the description of the visualization and ontology expansion tool. The paper ends with conclusions and references.

2 Ontologies and Semantic Closeness

Ontologies are semantic networks modeling human conceptualization, built either manually or automatically, for example, by extracting knowledge from texts (text mining). Ontologies may be also seen as ways of sharing concepts, classifications and inter-relations in communities. Any collaboration using natural language, any discourse needs to start from a common vocabulary or, a more structured alternative, a shared ontology. WordNet or FrameNet (http://framenet.icsi.berkeley.edu) are examples of general ontologies built as extended vocabularies, offering additional linguistic data like related words, or case grammars. In addition to general ontologies, communication in communities of practice needs particular concepts from specific domain ontologies.

The word "ontology" is taken from philosophy, where it denotes the theory about what is considered to exist. Any system in philosophy starts from an ontology, from the identification of the concepts and relations considered as fundamental. Ontologies capture fundamental categories, concepts, their properties and relations. One very important relation among concepts is the taxonomic one, from a more general to a more specific concept. This relation may be used as a way of "inheriting" properties from the more general concepts ("hypernyms"). Other important relations are "part-whole" ("meronym"), "synonym" and "antonym".

Ontologies are very important in knowledge extraction from texts, in general, and from conversations, in particular. For these kind of applications, they offer the substrate for semantic analysis and, very important, the possibility of defining a measure of semantic closeness, based on the graph with concepts from ontologies as nodes and their relations as arcs [2].

The measures of semantic distances allow to identify groups of similar concepts and, therefore, to identify lexical chains of related words. These chains, together with repetitions and anaphors allow to further identify threads of discussions in texts. These threads may interact, according to polyphonic patterns [7].

3 Bakhtin's Polyphonic Theory

In forums and chat conversations, knowledge is socially built through discourse and is preserved in linguistic artifacts whose meaning is co-constructed within group processes [4, 5]. These socio-cultural ideas are based on the work of Lev Vygotsky, who emphasized the role of socially established artifacts in communication and learning [9].

Mikhail Mikhailovici Bakhtin has extended the ideas of Vygotsky, emphasizing the role of speech and dialog in analyzing social life. He remarks that in each dialog and even in written texts there are communities of voices: "The intersection, consonance, or interference of speeches in the overt dialog with the speeches in the heroes' interior dialogs are everywhere present. The specific totality of ideas, thoughts and words is everywhere passed through several unmerged voices, taking on a different sound in each" [1]. This dual nature of community and individuality of voices is expressed by Bakhtin also by the concept of *polyphony*, that he considers the invention and one of the main merits of Dostoevsky novels [1]. The relation of discourse and communities to music was remarked also by Tannen: "Dialogue combine with repetition to create rhythm. Dialogue is liminal between repetitions and images: like repetition is strongly sonorous" [6].

In chat conversations, different voices are obvious recognized. However, starting from Bakhtin's ideas, in our approach the concept of voices is not only limited to the physical vocal characteristics of the participants in the chat. A voice is, from our perspective, something said by a participant in a given moment and it may be reflected in many other subsequent utterances. As a direct consequence, each utterance may contain an unlimited number of voices.

4 Ontology-Based Chat Analysis

The approach presented here integrates Bakhtin's socio-cultural ideas with knowledge-based natural language processing for the identification of the topics discussed in the chat, for the detection of discussion threads and of the most important utterances in a chat. Such a system may be used, for example, for tracking the most important topics discussed by a group of experts in urbanism for solving a given problem. The chat system used in the experiments presented here was ConcertChat [3], which allows the use of explicit references to previous utterances, a facility that enables the existence of multiple discussion threads in parallel and their inter-animation.

Determining the Topics of a Chat

The chat topics are identified as a list of concepts (words) that appeared most frequently in the conversation, by using statistical natural language processing methods. Accordingly, the importance of a subject is considered related to its frequency in the chat.

The first step in finding the chat subjects is to strip the text of irrelevant words (stop-words), text emoticons (e.g. ":)"), special abbreviations used while chatting and other words considered of no use at this stage. Then, the resulted chat is tokenized and each different word is considered a candidate concept in the analysis. For each of these candidates, WordNet and the domain ontology are used for finding synonyms.

The last stage for identifying the chat subjects consists in unifying the candidate concepts discovered in the chat. This is done by using the synonym list for every concept: if a concept in the chat appears in the list of synonyms of another concept, then the two concepts' synonym lists are joined. At this point, the frequency of the resulting concept is the added frequencies of the two unified concepts. This process continues until there are no more concepts to be unified. At this point, we acknowledge the subjects of the chat conversation as the list of resulting concepts, ordered by their frequency.

Figure 1 is a screenshot illustrating some topics identified in an urbanism chat (http://www.planetizen.com/node/30186).

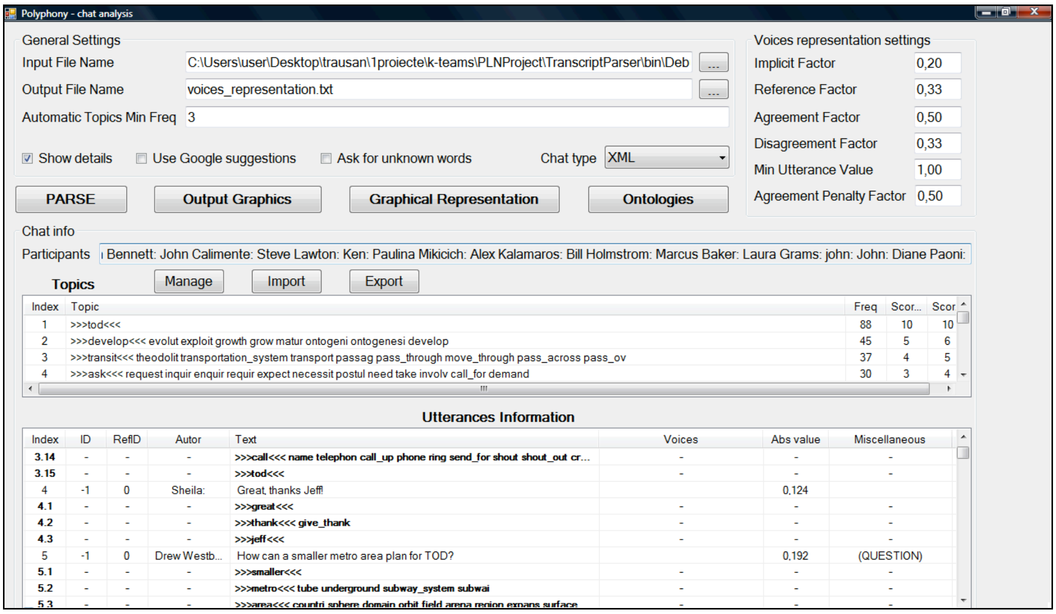


Fig 1. Identification of chat topics in an urbanism conversation

Extending the Domain Ontology with Topics Determined from the Chat

The first three topics identified by the system (see figure 1) are tod ("transit oriented development"), develop and transit. If we don't have these concepts in the domain ontology, the system offers the possibility of adding them. The user may add also relations among new concepts and save the ontology. In figure 2 is an excerpt of the usage of this facility. The concepts are added to the ontology accordingly to their importance in the chat as a topic. The relations that can be added are the ones specific to the WordNet ontology: synonyms, hypernyms and meronyms.

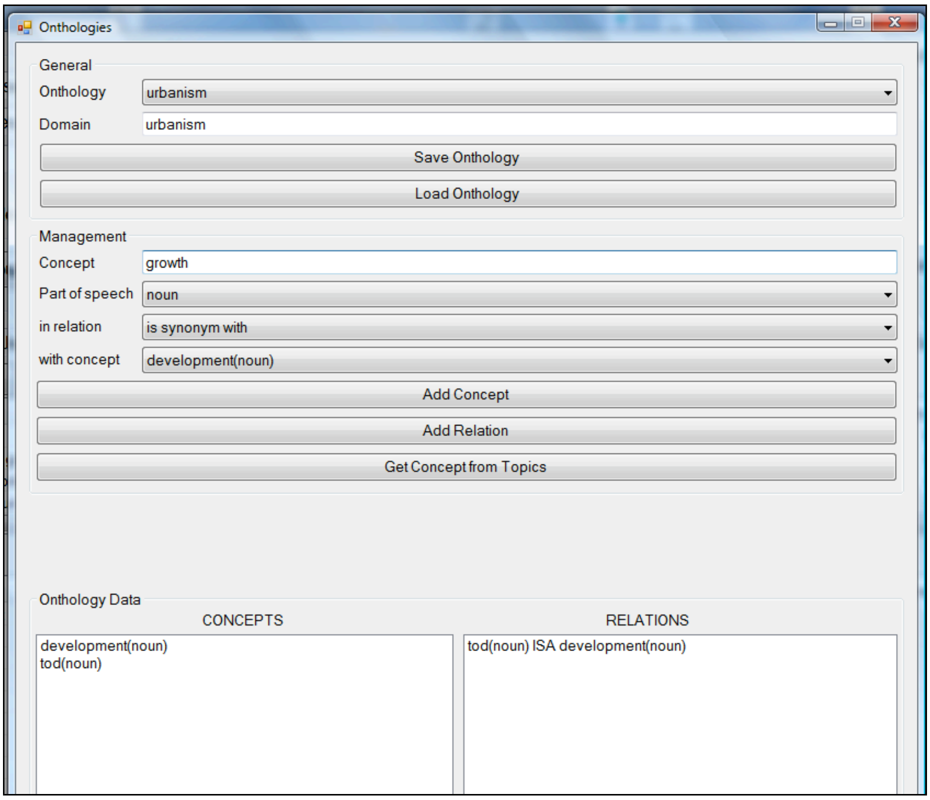


Fig 2. Extending the domain ontology

The Graphical Representation of the Chat

Starting from existing references within the analyzed conversations, both those explicit, allowed by the chat environment (ConcertChat [3]), as well as those implicit, determined by the program, a graph that visualizes the conversation is built. Within this graph, each utterance from the chat is a vertex, while the references between utterances (either explicit or implicit) represent the edges. The output is a directed graph specific to the conversation.

The graphical representation of the chat was designed to permit a better visualization of the conversation, to facilitate an analysis based on the polyphony theory of Bakhtin, and to maximize the straightforwardness of following the chat elements. For each participant in the chat, there is a separate horizontal line in the representation and each utterance is placed in the line corresponding to the issuer of that utterance, taking into account its positioning in the original chat file – using the timeline as an horizontal axis. Each utterance is represented as a rectangle aligned according to the issuer on the vertical axis and having a horizontal axis length that is proportional with the dimension of the utterance. The distance between two different utterances is proportional with the time passed between the utterances. Of course, there is a minimum and a maximum dimension for each measure in order to restrict anomalies that could appear in the graphical representation due to extreme cases or chat logging errors.

The relationships between utterances are represented using colored lines that connect these utterances. The explicit references are depicted using blue connecting lines, while the implicit references that are deduced using the method described in this paper are represented using red lines. The utterances that introduce a new topic in the conversation are represented with a red margin.

Fig. 3. The threads of references in the chat from figure 1

The graphical representation of the chat has a scaling factor that permits an attentive observation of the details in a conversation, as well as an overview of the chat. At the top of the graphical representation of the conversation, there is a special area that represents the importance of each utterance, considered as a chat voice, in the conversation. How this importance is determined is presented further in this paper. Moreover, all the details of an utterance in the chat – the content of the utterance, the implicit and explicit references and other details – can be visualized by clicking the rectangle representing the utterance.

Determining the Importance of an Utterance

An approach for computing the importance of an utterance that is useful in NLP tasks such as chat summarization is to use the length and the correct selection of the words in each utterance – it should contain as many possible key (important) words. Nevertheless, in a social context, another approach is also possible: an utterance is important if it influences the subsequent evolution of the conversation. Using this definition as a starting point, we may infer that an important utterance will be that utterance which is a reference for as many possible subsequent utterances.

Even if this approach could be extended to include the types of subsequent references (implicit or explicit, agreements or disagreements), in the present case we have preferred a more simplistic approach, without making allowances for the types of references to the utterance.

Consequently, the importance of an utterance can be considered as a strength value of an utterance, where an utterance is strong if it influences the future of the conversation (such as breaking news in the field of news). When determining the strength of an utterance, the strength of the utterances which refer to it is used. Thus, if an utterance is referenced by other utterances which are considered important, obviously that utterance also becomes important.

As a result, for the calculation of the importance of every utterance, the graph is traversed in the opposite direction of the edges, as a matter of fact in the reverse order of the moment the utterance was typed. Utterances which do not have references to themselves (the last utterance of the chat will certainly be one of them) receive a default importance – taken as the unit.

Then, running through the graph in the reverse order of references, each utterance receives an importance equal to that of the default plus a quota (subunit) from the sum of the importance of the utterances referring to the current utterance. Another modality to calculate could be 1 plus the number of utterances which refer to the present utterance, but this choice seemed less suitable.

By using this method of calculating the importance of an utterance, the *utterances* which have started an important conversation within the chat, as well as those *utterances* which begin new topics or mark the passage between topics, are more easily emphasized. If the explicit relationships were always used and the implicit ones could be correctly determined in as high a number as possible, then this method of calculating the importance of a voice would be successful.

Identifying Discussion Threads

Using an algorithm for determining the connected components from the conversation graph, we were able to find the utterances connected through at least one relationship. It is normal to assume that all these utterances are part of a single discussion topic.

This method can be used for successfully finding the conversational threads. We have considered that the important topics are those consisting of at least four utterances. For each determined topic, we have highlighted the most frequent concepts (as a synset list) in that topic. This way, each topic is described by the most relevant concepts found in the utterances present in that topic.

An interesting observation to be made is that this method to determine the topics of the conversation produces some remarkable results. Thus, the discussion can have more than one topic at a moment in time – the participants being involved in different topics at the same time. Inter-crossings between different topics can be easily observed on the chat graphics as well as topics started and finished whereas other more important topics are abandoned for a while and then continued.

This method can be improved by considering the similarities between the utterances in closely related topics. We can also combine this solution with an analysis of the time passed between the utterances. Two similar utterances that are separated by a great distance in time can be considered part of different topics.

5 Conclusions

The paper presents an ontology-based application that detects the topics of instant messenger chat conversations, the threads of discussion and the important utterances. The system also visualizes a graph of the conversation. The application may be used for analyzing what was discussed and in what degree participants are implied in the chat conversation.

The paper includes an example of a group of urbanism specialists discussing how to solve a given problem. The automatic analysis emphasized the main topics of discussion. These concepts were introduced in the domain ontology using a specially designed interface of the system.

The application uses the WordNet ontology and a domain ontology. Natural language technology is used for the identification of discussion topics, for the segmentation of the conversation, for identifying implicit references among utterances and for graphical visualization. The generated diagrams allow identifying the participants which had an important contribution in the conversation. The domain ontology may be extended with the new topics identified by the system, as illustrated in the presented example.

Further work will consider more complex semantic distances (than only synonymy) and more elaborated interaction patterns [8]. Machine learning techniques will be used for the identification of discourse patterns. Moreover, a completely automated version for discovering new rules for the implicit relations is in progress.

The system has been used in analyzing chats for Computer-Supported Collaborative Learning in Politehnica University of Bucharest, Romania and Drexel University, Philadelphia, USA. A new version is now under development in the EU FP7 project LTfLL (Learning Technology for Lifelong Learning, see http://www.ltfll-project.org/).

References

1. Bakhtin, M.M., *Problems of Dostoevsky's Poetics*, Ardis (1973)
2. Hirst, G., St-Onge, D., "Lexical chains as representations of context for the detection and correction of malapropisms". in Christiane Fellbaum (ed.), *WordNet: An Electronic Lexical Database*, chapter 13, pages 305 - 332. The MIT Press, Cambridge, MA, (1998).
3. Holmer, T., Kienle, A., Wessner, M. "Explicit Referencing in Learning Chats: Needs and Acceptance," in Nejdl, W., Tochtermann, K., (eds.), *Innovative Approaches for Learning and Knowledge Sharing, First European Conference on Technology Enhanced Learning*, EC-TEL 2006, LNCS, 4227, pp. 170-184, Springer (2006)
4. Koshmann, T., Toward a Dialogic Theory of Learning: Bahtin's Contribution to Understanding Learning in Settings of Collaboration, in C.Hoadley and J. Roschelle (eds.),

Proceedings of the Computer Support for Collaborative Learning 1999 Conference, Stanford, Laurence Erlbaum Associates.
5. Schegloff, E., Discourse As An Interactional Achievement: Some Uses Of 'Uh huh' And Other Things That Come Between Sentences, in Tannen, D. (ed.), Georgetown University *Roundtable on Languages and Linguistics 1981; Analyzing Discourse: Text and Talk,* Georgetown University Press, Washington D.C.
6. Tannen, D., *Talking Voices: Repetition, Dialogue, and Imagery in Conversational Discourse,* Cambridge University Press (1989)
7. Trausan-Matu, S., Stahl, G., Sarmiento, J., Supporting Polyphonic Collaborative Learning, *E-service Journal*, vol. 6, nr. 1, Indiana University Press, 2007, pp. 58-74.
8. Trausan-Matu, S., Rebedea, T., Polyphonic Inter-Animation of Voices in VMT, in Stahl, G. (ed.), *Studying Virtual Math Teams*, Springer, to appear in 2009.
9. Vygotsky, L. (1978). *Mind in Society*, Cambridge, MA: Harvard University Press

Incremental Development of a Shared Urban Ontology: the Urbamet Experience

Jacques Guyot[1], Gilles Falquet[1], Jacques Teller[2]

[1] Université de Genève, Centre universitaire d'informatique,
7 route de Drize, CH-1227 Carouge
jacques.guyot@unige.ch, gilles.falquet@unige.ch

[2] Université de Liège – LEMA, Chemin des Chevreuils, 1 B52/3, B-4000 Liège
jacques.teller@ulg.ac.be

Abstract. Thesauruses are used for document referencing. They define hierarchies of domains. We show how document and domain contents can be used to validate and update a classification based on a thesaurus. We use document indexing and classification techniques to automate these operations. We also draft a methodology to systematically address those issues. Our techniques are applied to Urbamet, a thesaurus in the field of town planning.

Keywords: ontology, thesaurus, text mining, automated classifier, Urbamet,

1. Introduction

We shall briefly remind below the definition of a thesaurus. We shall also describe the concept and history of Urbamet, the collection which was used in our experiment. We shall then explain how we use automated classification with a view to creating or updating thesauruses.

1.1 Thesauruses

For years, thesauruses were the favorite tool used by librarians and documentalists to classify documents. This classification was meant to facilitate document search by users. A thesaurus includes a set of terms used in a given domain, which is distributed among a hierarchy of sub-domains. The set of chosen terms becomes a controlled vocabulary whose meaning is strictly defined by the thesaurus designers. The structure of thesauruses, in particular the types of relationships between terms, has been studied an normalized during the last decades (ANSI/ISO, 2005), (ISO, 1986). A thesaurus is used by the documentalist to assign one or several domain(s) to each document, and to assign keywords chosen from the terms of the assigned domain(s). User requests are then expressed through those keywords and domains.

Indexing documents with the help of thesauruses is a technique that allows building up classifications:

- Without using a computer
- Without knowing the document contents (esp. in the case of multimedia documents).

This is particularly useful when those elements cannot be requested. However, building up a thesaurus requires a major investment in terms of creation, learning and maintenance work. Indexing and search engines that operate directly on document contents have changed the habits of users, who would rather use directly their own vocabulary. Such a method is good enough when the corpus is very large (the complete Web in the case of Google or Yahoo!) but in that case the issue of polysemy (*i.e.* multiple meaning) creates "noise" in the results; this noise would require introducing domains to filter out the results. As for keywords, they are necessary to harmonize the vocabulary in the case of a middle-size corpus (*i.e.* a corpus for which all possible expressions are not available for a given piece of information).

Thesauruses are thus still necessary when it comes to indexing and searching documents on the basis of their content.

1.2 Urbamet

Urbamet (Urbamet, 2009) is a bibliographic database created and maintained by the French Centre for Urban Documentation. The corpus currently includes 280'000 documents and is fed with an additional 8'000 documents each year. Originally designed in 1969 with 2'300 terms, the Urbamet thesaurus currently includes 4'200 terms, which are used to index the document corpus. It is a hierarchy of terms with 24 main themes (top level categories) . Figure 1 shows the main themes and an excerpt of the hierarchy of sub-domains in the field of transportation.

It can be observed on the figure that the terms in Urbamet denote either concepts or sub-domains. For instance, the term “utility vehicle” may denote a concept that has an intension (the properties of a utility vehicle) and an extension (the set of all utility vehicles). Conversely, the term “road and traffic” can hardly denote a concept: it is difficult to figure out what is an instance of “road and traffic”. Moreover “road and traffic” cannot be considered as a specialization of its parent term “land transport”. Hence, the Urbamet thesaurus, at least on the first levels, is mostly a hierarchy of sub-domains. As a consequence, it does not provide a starting point or backbone for the construction of an urban ontology.

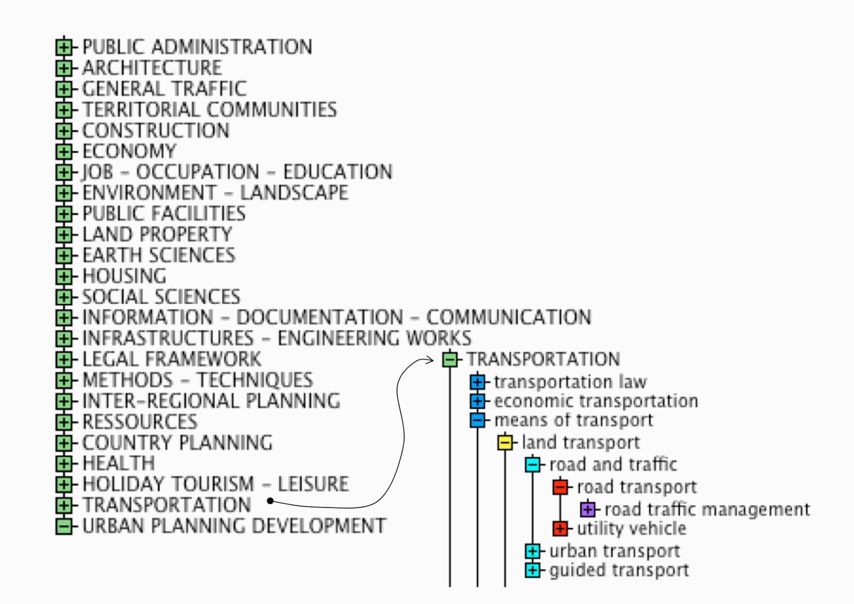


Fig. 1. Urbamet main themes and the *Transportation* sub-hierarchy

1.3 Methodology

Since the thesaurus cannot be directly used to build an ontology, the proposed methodology relies on the existing thesaurus and the indexed document corpus. The document classification induced by the thesaurus is analyzed with an automated document classifier. This tool operates on document contents. Initially a training corpus is used to teach the classifier the class concepts. Then the tool can start classifying other documents. The analysis is performed in the following steps:

1. Extracting the corpus
2. Building up the training catalogue
3. Training and validating the classifier
4. Generating and analyzing the confusion matrix (list of mistakes made by the classifier)
5. Generating the Top-50 terms (list of the most classifying terms)

We shall show how the analysis of the confusion matrix and the Top-50 list helps us understand how the corpus is structured in terms of domains and how the thesaurus may be re-structured on the basis of these indications.

2. Analyzing the Urbamet

We use a classifier based on the neural network technique. The goal of the classifier is, for each document, to predict the class to which it belongs. The input is the list of document terms[30] and the output is a score (between 0 and 1) for each class. In this case the classes are the 24 top-level Urbamet domains.

2.1 Training the Classifier

To build the training and test corpus we extracted from the Urbamet web site about 10'000 information sheets similar to the one shown in 0 (which is mostly in French):


Document < 1 > /9878

Affichage HTML pour impression Affichage PDF pour impression Retour à la liste de résultats

Titre	**Agora 2020 : synthèse à mi-parcours.-**
Auteur(s)	BAIN, Pascal ; CHAPUY, Pierre ; DROUET, Dominique ; FARHI, François ; MAUJEAN, Sébastien ; MIRENOWICZ, Philippe ; THEYS, Jacques
Organisme(s) auteur(s)	CENTRALE MANAGEMENT INTERNATIONAL. PARIS ; FRANCE. MINISTERE DE L'EQUIPEMENT. DIRECTION DE LA RECHERCHE ET DE L'ANIMATION SCIENTIFIQUE ET TECHNIQUE ; FRANCE. MINISTERE DE L'EQUIPEMENT. DIRECTION DE LA RECHERCHE ET DES AFFAIRES SCIENTIFIQUES ET TECHNIQUES. CENTRE DE PROSPECTIVE ET DE VEILLE SCIENTIFIQUES ET TECHNOLOGIQUES ; GROUPE D'ETUDES RESSOURCES PLANIFICATION AMENAGEMENT. PARIS ; RECHERCHE DEVELOPPEMENT INTERNATIONAL. PARIS
Source bibliographique	*Paris : DRAST, CPVS, fév. 2005.- 150 p., ann., tabl., graph.*
Collection	Dossiers du CPVS, n° 8 - février 2005
Notes	Coll. Les Dossiers du CPVS n°8. Coll. Dossiers du CPVS
Cote	**CDU 58938 ; RST RCPVS05-001**
Résumé	A l'été 2003, la DRAST a engagé une vaste consultation prospective sur les attentes en matière de recherche dans ses différents domaines d'intervention - en y incluant le logement, la ville, la gestion des risques et l'observation des milieux. Cette consultation, intitulée Agora 2020, entre dans sa phase finale. Le dossier constitue une première synthèse des travaux réalisés en 2003 et 2004. Il est divisé en deux grandes parties : une synthèse à mi-parcours qui reprend sous forme de " messages clefs " les deux premières phases d'Agora 2020 ; un dossier complémentaire rassemblant quelques comptes-rendus représentatifs de l'ensemble des travaux déjà réalisés (présentation d'Agora 2020, les attentes des acteurs, observation de la terre des milieux et gestion des risques, des visions du monde qui ne communiquent pas : une illustration à travers le thème de la ville).
Lieu(x)	France
Thème(s)	Aménagement urbain ; Habitat - Logement ; Infrastructures - Ouvrages d'art ; Sciences de la terre ; Transports
Mot(s)-clé(s) français	aménagement du territoire ; besoin ; habitat - logement ; infrastructures - ouvrages d'art ; ministère équipement logement ; prospective ; recherche ; transports ; ville
Mot(s)-clés anglais	forecast ; housing ; infrastructures - engineering works ; inter-regional planning ; need ; public works housing ministry ; research ; town ; transportation

Fig. 2. Sample Urbamet information sheet

[30] Terms are the text words and co-occurrences of two words appearing with a given frequency. For instance, "engineering works" is recognized as a single term. We can also use term stemming techniques by grouping under a single tag all words from the same family. For example, "manage, managed, management, managing" can be considered as a single term. A list of so-called "stop-words" is used to get rid of noisy words such as "the, it, and, …"

We used the Winnow algorithm to set the neuron weights that link the terms to the classes. Initially all the weights have the same value. For each document in the training corpus, and according to the terms used in the text, we sum up the values in each class. Then we sort out the results according to the sums. For the classes which are above a given threshold and which are not correctly predicted, we lower the weight of the document terms. Conversely, for the classes which are below the threshold and which are correctly predicted, we raise the term weights. Thus by simulating a "punishment/reward" heuristic, we produce a neural network that has learned the underlying classification of the training corpus.

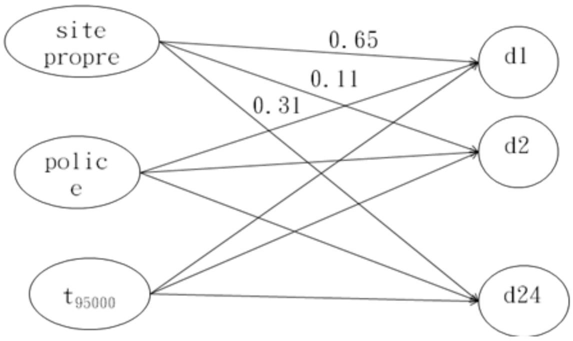


Fig. 3. Neural network with weights linking terms to classes

2.2 Testing the Classifier

The quality of the classification must be tested. It depends essentially upon whether the condition of class inclusion can be deducted from the document content, and thus from the terms (for instance, dividing a corpus in two parts on the basis of whether the ideas in a document are ethical or not is an example where the terms are of little use). To perform such a test, we separate a part (20%) of the training corpus which is not used in the training phase; it is used later on to evaluate the classifier performance. In the table below, we see that the first class predicted by the classifier is correct in 59% of cases, while it would have been predicted in only 4% of cases had the prediction been made randomly. The following lines show results where the second and third choices were added up to the first one.

Table 1. Performance of the classifier

Number of Predicted Classes	Classifier Performance	Random Choice Performance
1	59 %	4 %
2	75 %	12 %
3	81 %	23 %

From the results above, we can see that the Urbamet classifier is effective and that the Urbamet classification can be deduced from the document contents. Therefore there is a relationship between the document terms and the document classes.

2.3 The Confusion Matrix

It is interesting to examine the classifier's mistakes. These mistakes are due to the fact that it is difficult to distinguish the classes on the basis of their respective vocabulary. A confusion matrix may be built up on these mistakes: Each row of the matrix represents the class which should be found, while each column represents the predicted class. Ideally, only the diagonal should be filled up to 100%. The complete confusion matrix can be found in appendix. Let us take a look at its exceptions.

2.4 Vocabulary Shared between Two or Several Domains

Although the *Transportation* and *Traffic* domains are relatively well separated from the other domains, 24% of the documents which should have been classified in *Transportation* were actually classified in *Traffic* and 10% of documents which should have been in *Traffic* were in *Transportation*.

This is due to the fact that the vocabulary is common to both the *Transportation* and *Traffic* domains and thus makes the separation difficult.

Table 2. Confusion matrix for *Transportation*, *Traffic* and *Tourism*

In \ out	Transportation	Traffic	Tourism	…
Transportation	45%	24%	3%	
Traffic	10%	40%	1%	
Tourism	1%	1%	49%	

2.5 Orthogonality of Domains

The *Legal* and *Methods* domains are not well distinguished from the other domains. The documents that should have been classified in those two classes were in fact scattered across all domains.

Table 3. Confusion matrix for *Legal*, *Methods*, *Urbanism*, and *Infrastructure*.

In \ out	Legal	Methods	Urbanism	Infra…
Legal	8%	3%	5%	3%
Methods	2%	4%	4%	13%
Urbanism	17%	14%	24%	4%
Infrastructure	2%	11%	1%	22%

This is due to the fact that *Legal* and *Methods* are not domains of urbanism, but rather aspects of it. One could equally speak of the legal aspects of transportation, of real estate, or of environment. Such domains are said to be "orthogonal" to the other ones.

2.6 Top-50 Term List of a Domain

Neural networks can be criticized because of the lack of explanation on the fact that the classifier chose a particular class (as compared to rule-based engines which can explain their reasoning). However, it is always interesting to analyze, for each class, the list of the most heavily weighted terms. This list is a selection of the "champion terms" of the domain.

We shall only analyze here one domain, namely *Environment* (which includes 326 documents). The list of the most heavily weighted terms (in French) is the following:

paysagiste, écologique, paysagères, écologiques, biodiversité, jardins, paysagistes, marais, parcs-naturels, jardin, directive, environnementales, naturel, paysages, pnr, protection, espèces, berges, paysagère, naturels-régionaux, paysage, arbres, précaution, faune, éco, forestier, protection-nature, environnemental, environnementale, green, pédagogiques, charte, écologie, patrimoine-naturel, vertes, ceinture, naturelles, verts, landscape, utilisé, principe-précaution, ceinture-verte, empreinte-écologique, durables, littoral, parcs, baie, conservation, participer, plans-programmes [31]

These terms are the "champions" of this specific class for this specific corpus.

[31] In this experience we did not use any stemming technique, which explains why some terms are found both in singular and plural forms.

2.7 Top-50 Terms versus Thesaurus Terms

We then compared the top-50 terms with the terms of the Urbamet thesaurus related to the *Environment* domain:

paysagiste, ***écologique***, paysagères, ***écologiques***, ***biodiversité***, ***jardins***, paysagistes, ***marais***, parcs-naturels***, jardin, directive, environnementales***, naturel, paysages, ***pnr***, protection***, espèces, berges***, paysagère, naturels-régionaux, paysage, ***arbres, précaution, faune, éco, forestier***, protection-nature, ***environnemental, environnementale***, ***green, pédagogiques, charte, écologie***, patrimoine-naturel, ***vertes, ceinture***, naturelles, ***verts, landscape***, utilisé, ***principe-précaution, ceinture-verte, empreinte-écologique***, durables, ***littoral***, parcs, ***baie, conservation, participer, plans-programmes***

The terms that were not included in the Urbamet thesaurus are displayed in bold and underlined. It appears that 34 terms out of the 50 were not in the thesaurus. The hypothesis we make to explain this fact is the following:

- The documents which are classified in the *Environment* domain are correctly classified
- The *Environment* domain has changed since 1969
- The thesaurus updates do not reflect those changes
- The *Environment* domain includes in fact two domains: one is *Urban Environment* and the other one is *Ecology*.

3. Towards a Methodology to Update Thesauruses

The examples provided in the previous sections show that automated analysis tools are relevant. Yet a detailed analysis of the confusion matrix and the Top-50 list definitely requires a corpus expert (a documentalist who knows well his/her corpus and thesaurus).

We suggest some elements of methodology when using an automated classifier to validate the domains. It should be assessed initially that the classifier is globally able to reach a given level of efficiency. Then the confusion matrix allows to:

1. Analyze the domains that are not clearly separated (such as *Traffic* and *Transportation*). In such a situation the following steps should be applied:
   - Check out the quality of the classification for both domains
   - Possibly merge both domains into a single one and then separate them into two sub-domains.
2. Look for orthogonal domains that would be distributed across all domains (such as *Legal* and *Method*). In that case it could be necessary to:
   - Build up a hierarchy of domains. For example *Legal* and *Method* are sub-domains of all the other domains.

A hierarchy of domains may be used to train a classifier by building up a neural network (*i.e*. a classifier) for each node in the hierarchy. In our example, we can see that confusion may be avoided by removing the *Legal* and *Method* domains from the first level of the hierarchy. Indeed, the terms related to legal and methodological issues will be scattered across the various other domains and as such will be lightly weighted. Conversely, at the second level of the classification, the domain-related terms will be lightly weighted while the legal and methodological terms will be heavier.

With the Top-50 list of terms we can:

1. Analyze the highly classifying terms:
   - To discover an emerging new domain which was covered by another one (such as *Urban Environment* and *Ecology*)
   - To discover an emerging new domain which was distributed among several other domains (a typical example is computer science, which before the 1970s did not exist as an independent domain but was considered to be either mathematics, automatics or electronics)
2. Turn the classifying terms into concepts of an ontology:
   - These concepts are the seeds on which an ontology may be grown up.
3. Repeat the previous steps on a regular basis (every *x* years):
   - Repeating these steps allows monitoring how the confusion matrix and the Top-50 list evolve, which is an indication of how the domains themselves evolve.

4. Conclusion

Thesauruses such as the Urbamet thesaurus are not ontologies. In particular, their hierarchical structure is not an "*is a*" relationship: therefore they should be considered as a hierarchy of domains which is connected with a corpus of documents.

We have shown that in such situations, text mining techniques, and more particularly automated document classification techniques may be used to support the following actions:

- Analyzing the thesaurus
- Maintaining the thesaurus and restructuring domains
- Finding new domain terms to build up ontologies.

Typical text mining based ontology extraction tools, such as the OntoLearn system (Velardi et al., 2001), rely on statistical analysis of the corpus terms, together with syntactic analysis. The methodology we propose takes advantage of an existing classification scheme and, to a certain extent, discovers how and why it works (e.g. it finds the most classifying terms). This discovery process yields insights into the structure of the domain and thus provides a basis for building an ontology.

The methodology we propose must still be evaluated on other test cases. However, as far as we know, the evolution of knowledge resources has not been documented in the urban field. Thus we intended to test our approach on a physics corpus for which there exists a classification (the Annual Classification of the Physikalische Berichte) that has already been studied (Hurni, 2009).

References

1. ANSI/ISO (2005). Guidelines for the Construction, Format, and Management of Monolingual Thesauri. ANSI/NISO Z39.19-2005, American National Standards Institute (ANSI). Revision of Z39.19-1983.
2. Gómez-Pérez A, Fernández-López M, Corcho O (2003). Ontological Engineering, chapter Methodologies and Methods for Building Ontologies. Springer-Verlag, London (United Kingdom).
3. Gómez-Pérez A, Manzano-Macho D (2003). A survey of ontology learning methods and techniques. OntoWeb Consortium, Deliverable 1.5.
4. Hurni, J.-P (2009). Cornelius Lanczos et les Physikalische Berichte (1923–1933), ISRI Technical Report ISRI-Ge-09-01, Genève. (in preparation).
5. ISO (1986). Guidelines for the establishment and development of monolingual thesauri. ISO 2788, International Organization for Standardization (ISO).
6. Kietz JU, Maedche A, , Volz. R (2000). A method for semi-automatic ontology acquisition from a corporate intranet. In Proceedings of Workshop Ontologies and Text, EKAW'2000.
7. Littlestone, N. (1988) "Learning Quickly When Irrelevant Attributes Abound: A New Linear-threshold Algorithm" Machine Learning 285-318(2)
8. Urbamet (2009). Urbamet home page. http://www.urbamet.com/ Retrieved on 10 Feb. 2009
9. Velardi P, Fabriani P, Missikoff M (2001). Using text processing techniques to automatically enrich a domain ontology. In Proceedings of the international conference on Formal Ontology in Information Systems, FOIS 2001, pp 270–284.

Appendix

Confusion Matrix for the 24 Domains

	TRA0	JUR0	METC	CIR0(	RESC	COLC	ADM(	HABC	FONC	PLA0	HUM(	URBC	TOUC	IDC0(	GEO(	INF0C	ECO(	CTR0	EQU(	EMP(	SAN(	ARC(	RUR(	ENV0
TRA0(	45%	3%	5%	24%	6%	2%	2%	1%	3%	6%	1%	4%	3%	7%	4%	26%	4%	0%	7%	3%	1%	0%	1%	2%
JUR0(	1%	8%	3%	1%	5%	6%	5%	5%	9%	5%	1%	5%	2%	3%	1%	3%	2%	4%	3%	3%	6%	1%	6%	6%
MET0	2%	2%	4%	5%	5%	2%	6%	1%	1%	2%	2%	4%	3%	10%	15%	13%	2%	12%	2%	1%	1%	3%	1%	3%
CIR00	10%	1%	5%	40%	1%	1%	1%	0%	0%	1%	1%	2%	1%	1%	0%	5%	0%	0%	3%	0%	1%	0%	0%	0%
RES0	5%	7%	8%	2%	39%	4%	3%	2%	1%	2%	2%	3%	1%	2%	19%	4%	2%	14%	3%	0%	9%	2%	4%	10%
COL0	1%	7%	2%	1%	3%	25%	16%	2%	3%	6%	4%	4%	4%	7%	0%	2%	3%	0%	4%	3%	3%	1%	3%	2%
ADMC	0%	3%	4%	1%	1%	8%	27%	0%	0%	3%	2%	1%	1%	2%	1%	1%	1%	0%	1%	1%	1%	0%	0%	1%
HAB0	1%	12%	4%	1%	3%	4%	2%	45%	20%	1%	7%	8%	0%	3%	1%	1%	3%	16%	4%	3%	4%	10%	1%	3%
FON0	1%	3%	0%	0%	0%	1%	0%	4%	20%	1%	0%	2%	1%	1%	3%	0%	1%	0%	0%	0%	0%	0%	4%	1%
PLA0(	3%	5%	3%	1%	2%	6%	5%	1%	4%	26%	1%	6%	2%	6%	1%	3%	4%	1%	0%	0%	1%	0%	4%	4%
HUMC	2%	5%	7%	5%	4%	11%	14%	11%	2%	3%	52%	8%	6%	7%	12%	2%	7%	1%	3%	11%	8%	12%	10%	4%
URB0	6%	17%	14%	10%	6%	10%	4%	11%	19%	19%	7%	24%	3%	10%	3%	4%	10%	7%	9%	5%	7%	23%	23%	15%
TOU0	1%	1%	2%	1%	0%	2%	1%	0%	1%	1%	1%	1%	49%	5%	0%	1%	1%	0%	6%	1%	0%	0%	4%	2%
IDC0C	1%	1%	4%	0%	0%	2%	1%	0%	0%	2%	1%	1%	3%	20%	1%	1%	1%	0%	0%	1%	2%	1%	0%	0%
GEOC	0%	0%	2%	0%	2%	0%	0%	0%	1%	0%	0%	0%	0%	1%	8%	1%	0%	0%	0%	0%	0%	0%	0%	2%
INF00	9%	2%	11%	4%	2%	1%	1%	0%	0%	2%	0%	1%	1%	1%	5%	22%	1%	2%	1%	0%	0%	1%	0%	2%
ECOC	5%	5%	6%	2%	3%	8%	7%	4%	7%	11%	6%	9%	6%	8%	3%	5%	43%	9%	4%	10%	2%	2%	6%	8%
CTR0	0%	2%	6%	0%	4%	0%	0%	3%	0%	0%	0%	1%	0%	1%	0%	1%	1%	15%	2%	0%	2%	6%	1%	1%
EQU0	1%	1%	1%	1%	1%	1%	0%	0%	0%	0%	0%	1%	3%	0%	0%	1%	0%	1%	33%	0%	1%	1%	1%	1%
EMPC	2%	3%	1%	0%	0%	3%	2%	2%	0%	0%	4%	2%	2%	2%	1%	1%	5%	1%	1%	51%	13%	3%	0%	1%
SAN0	0%	2%	0%	0%	2%	1%	0%	1%	0%	0%	1%	1%	0%	2%	0%	0%	0%	1%	1%	3%	36%	0%	1%	0%
ARC0	0%	1%	2%	0%	1%	1%	0%	3%	0%	0%	3%	5%	0%	2%	0%	1%	1%	11%	4%	2%	1%	21%	2%	4%
RUR0	0%	2%	0%	0%	1%	1%	0%	0%	3%	1%	1%	2%	2%	0%	0%	0%	1%	0%	1%	0%	1%	1%	23%	2%
ENV0	2%	9%	6%	1%	10%	3%	2%	2%	3%	6%	2%	7%	7%	2%	20%	4%	5%	3%	6%	1%	2%	9%	9%	26%

ONTONEOLOGY: New Words for the City

Marc Ellenberg[1], Claire Jafflin[2]

[1] CERTU, 9 Rue Juliette Récamier, 69006 Lyon, France
marc.ellenberg@equipement.gouv.fr
[2] LET-ISH, Université Lumière Lyon 2
14 avenue Berthelot, 69363 Lyon Cedex 07, France
claire.jafflin@let.ish-lyon.cnrs.fr
Members of the Specialised Commission for Terminology and Neology
Equipment, Transport and Housing - France

Abstract. The Cost action C21 Towntology, aims at deploying ontologies in the urban domain, including together transports, urban planning, activity schemes, facilities, and so on, in order to enhance communication among actors of civil engineering, and all private and public partners of the urban development. This paper presents the objectives and methods of the French Specialised Commission for Terminology and Neology (CSTN) Equipment, Transport and Housing, in its work for enriching the French language. The collaborative process for the identification of the terms to be examined, in a terminological perspective, the process of naming and defining, frequently in relation with foreign languages, seems to offer possibilities of contribution to the establishment of these urban ontologies, thanks to a practice of operational ontoterminology (neologism installing ontology in the center of terminology, Roche, 2007). A tenth of practical cases will illustrate this.

Keywords: vocabulary, terminology, ontology, urbanism, transport

1 Introduction

Since languages aren't frozen things, they permanently evolve, and thus modify their words and meanings. While neologisms spontaneously created from French or foreign languages allow to add new words in the daily speaking, no matter whether new notions are associated with these words, the technical domains raise linguistic and specific problems. In fact, with the growing of knowledge and practices, number or terms and expressions appear, representing new notions and realities which must be designated in order to be understood and shared in a context of growing diversity and complexity.

The present paper is aimed at describing the experience of linguistic work of the Specialised Commission of Terminology and Neology for the French language, and to show how this work can be inserted into an ontologic process. After a presentation of the history and the objectives of the Commission, the process adopted to reach a consensus on terminology will be explained through the following steps : - detection of the terms to be studied, - selection criteria of the words to be worked out, - decision on the terms, - definition, and at last additions to ontology.

2 The State Commitment to Enriching Scientific and Technical Language

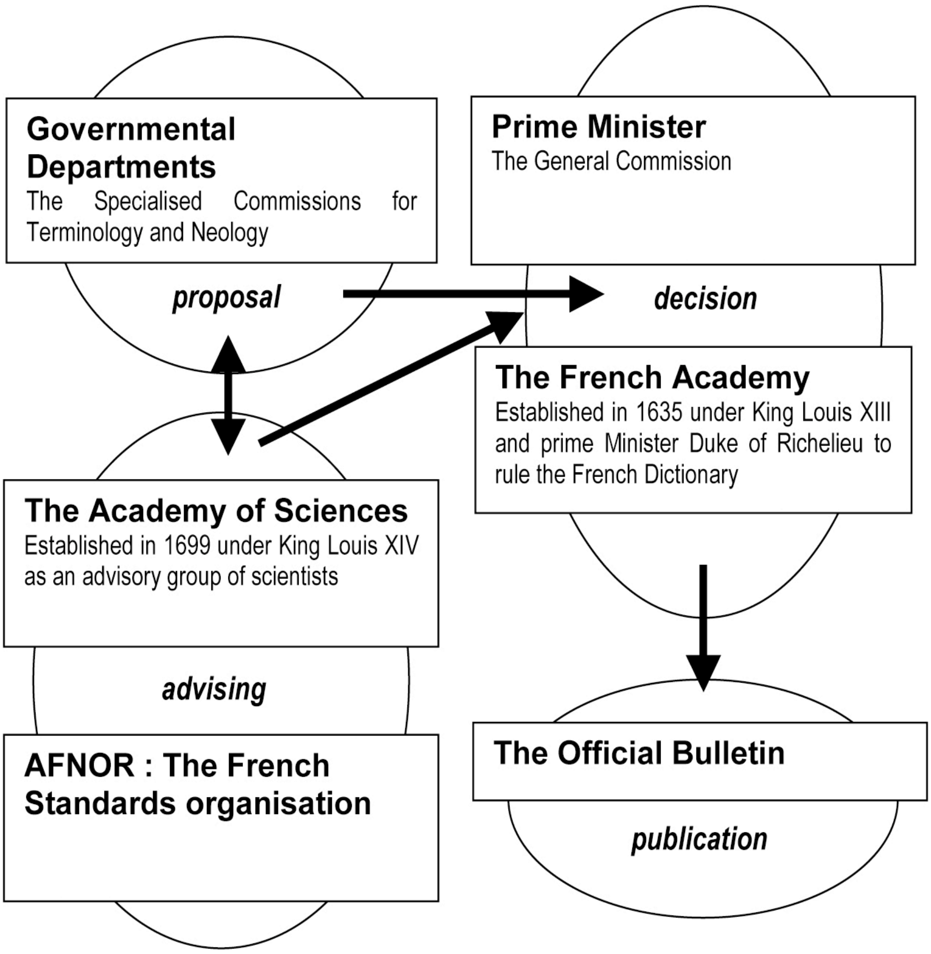
Facing the need and willingness to clarify and standardize the words and their meaning, since the seventies, the French legislation has decided to give to governmental services an essential role for the production of new words and for the dissemination of the terminology. Governmental Commissions were created, but not much coordinated at these times. That is why the idea of a general commission for filtration and validation of the terms emerged, proposed by the various ministerial Commissions and then, for publication of information and recommendation.

Therefore, the decree of 3rd of July 1996 concerning the enrichment of the French language defines precisely the ways and means for the production and dissemination of the official terminology. It requires the General Commission to “favour the enrichment of the French language, to expand its use, mainly in the economic life, scientific works and technical and juridical activities, to facilitate its dissemination by proposing new terms and expressions able to stay as references, to contribute to the shining of the Francophony, and promote the multilingualism”. This General Commission for terminology and neology, appointed by the Prime Minister, is composed of 19 members, and revised every 4 years. Its objective implies developing translation activities to and from French, notably by encouraging the terminological data collection and their structuration into quick and accessible databases, which are able to respond to the international cooperation needs, for professionals and for general public.

When published in the Official Bulletin of the French Republic, the adopted words and definitions are of mandatory use for the state public bodies and services. The issue is to take advantage of the normative role ruling the large written production of the French administration, on both national and regional levels..

Meanwhile, the General Delegation for French Language and Languages of France (DGLFLF) contributes to the thinking on the computerised use of the language and encourages actions aiming at developing and strengthening the position of the French language in the networks

In order to fulfill the needs for technical and specialised terms, the General Commission relies on specialised domains commissions, working in full assembly or by sub-groups. Different scientific and technical domains and sub-domains have been determined. Each ministry indeed must establish one or several specialised commissions for terminology and neology to deal with some of the defined scientific and technical domains and sub-domains. In this perspective, aeronautics, architecture, motor cars, housing, civil engineering, navy, fuel engines, petroleum come under the specialised commission for Equipment, Transport and Housing together with other commissions. On the other hand, tourism, transport and urbanism come only under the commission for Equipment, Transport and Housing.



Each specialised commission is composed of 20 to 30 members who are representatives of the different services of the ministry together with external prominent experts of the domain. These experts are chosen amongst professionals of the sector (representatives of universities, companies, associations, specialised press…) and amongst language specialists (interpreters, terminologists, linguists, writers, and so on) The commissions must pay attention and watch to the current situation, and establish the list of cases where it is opportune to extend the French vocabulary in consideration of the expressed needs. Then they propose to the General Commission of Terminology and Neology the necessary and suitable terms and expressions, particularly the equivalents of foreign terms and expressions, together with their definition.

The General Commission examines the terms, expressions and definitions proposed by the specialised commissions, taking care of their harmonisation and adequacy, and asks for the judgement of the French Academy. The terms and definitions which gained agreement from the Academy are then transmitted to the Official Bulletin to be published, provided that they didn't encounter any objection from the minister, within one month.

The Academy of Sciences and the French Standards Organisation (AFNOR) representatives are entitled to be members of the General Commission of terminology and neology, and also of all the specialised commissions which deal with scientific type vocabulary. Moreover it must be stressed that some of the chairmen or members of the specialised commissions, mainly in the technical domains, are also experts for the standardisation committees established by AFNOR.

Besides that, a laboratory of the National Center for Scientific Research encompasses a department which assists the specialised commissions and the general commission for the review of the terminology documentation.

And last, all the terminology organisations of the foreign French speaking countries cooperate as needed with the general and specialised commissions.

3 How to Reach the Terminological Agreement

3-1 Detection of the Notions to be Examined

Detection of new notions is not a formal process. The chart of the notions to be studied is set up according to the observed and expressed needs together with reviewing texts from all origins (newspapers, parliamentary reports, companies advertising,...) Every member of the commission is entitled to propose the study of such or such word encountered in his professional or personnel life and appearing to justify a terminological work. Several situations can exist :

It can be a new word, or rather recently used, for which the meaning stays inaccurate.

It can also be an English word or an Anglicism. This is not meant to be a fight against the English language, but on one hand to favour multilingualism by the development and quality of the French language, and on the other hand to prevent misleading of the general public and associated exclusions. A kind of Anglo-saxon snobbery sometimes reflects the economic and cultural domination of the North-American continent. An English taste is sometimes given to words which are not English but derived from English. This is the case of the French word "*parking*" : a "*parking*" is not an English word, representing a "*car park*" or a "*parking lot*", it is a French creation based on an English word. In numerous scientific domains, English is dominant, and within the detection process of the terms to be studied, the specialised commissions are aware of the difficulty to produce a new term able to be substituted to the English one, and a Sisyph work to impose it. Numerous proposals of new words will stay unused... but the detection of the notions for which a terminological work would be worthwhile is not to be rejected in order to clarify and define them in an understandable way for everybody and publish them in the official bulletin.

Another case for which a terminological work is needed is related to a term well installed in the language, but the meaning of which requires a clarification. For example a maritime "*rail*", or "*dispositif de separation du trafic*" (in English "*traffic separation system*"), or again "*autoroute de la mer*" (in English "*sea highway*"), or *merroutage"* (in English "*sea/road transport*")

Eventually it can be a request coming from a Ministry who detects a non stabilised notion, or from professionals unable to agree on a common concept. For example specialists of road accidents frequently use "*accidentology*" either to name the statistical methods (which seems generally accepted) or to name the set of statistical results (for which other people use "*accidentalness"*")

3-2 Selection Criteria of Words To Be Studied for Naming and Defining a Notion

The various proposals for the study of detected terms don't yet form a classification. The first step is then to select the notions to be studied.

Some criteria are driving the selection :

1- Urgency concerning professional needs, or risk to anchor in the common language a commercial brand (such as "*pedibus*", or "*busway*")
2- Need of understanding a foreign acronym
3- Need of a conceptual standard allowing interoperability among current research and studies, and more generally more efficiency in communication
4- Need to react against a wrong use ("*degazage*" wrongly used for "*rejet en mer*" in English "*degassing*" against *"discharge"*)
5- Existing term in the database "France-Terme", needing a revision due to the evolution of its meaning, or to a partial or even wrong definition
6- Verification that the term belongs really and only to the domain covered by the specialised commission, or needs a common work together with another commission

When the need of definition of a term comes from a social explicit demand, generally administrative, or when an English word is making roots from the absence of a spontaneous French feeling for an equivalent word, the specialised commission is then expected to react. In 2006 an emergency procedure has been created which allows to examine a term, disrupting with the established agenda, as a matter of priority. In this case, the choice of the word is made according to a pragmatic way and refers to a non hierarchical approach of the system of notions.

In general, the work of the commission relies on a system of non-homogeneous notions of which the type is a mix of hierarchic (generic or split) or non-hierarchic (sequential or pragmatic) structures, taking into account the appearance of new words.

The decision to examine such notion is then made sometimes voluntarily in a hierarchical system, deriving from subordination or superordination relations; this is the case of the multimodal transport which leads to define also intermodal transport, combined transport, rail-road transport, in a type-kind relation, but also by developing a partitive classification of the types of wagons, such as wagons for containers, wagons for trucks, and so on, accompanied or not by the drivers, in a type-product relation. In this perspective, the commission decided to examine the term "optimodality", introduced by the public authority. The work is in progress.

This hierarchical approach implies also the systemic approach, which leads, about a word examined in the scope of a specific transportation mode, to look in any case for the existence of similar or equivalent terms among the different other transportation modes. This has been the case for the word "delestage", published in 2000 for road and air transport, "delestage" meaning for road transport "rerouting a traffic flow in order to reduce congestion" while for air transport it means "cancelling for a period one or several scheduled flights". These both definitions didn't raise any problem. But the English translation mentioned in 2000 was wrong and was related to another meaning of "delestage" named in English "jettisoning" which means "alighting of fuel" which is an operation sometimes needed before an emergency landing to prevent fire. Thus this term has been reviewed in 2007.

This review also leads to look after the meaning of the word in the maritime domain where it is also used. The first meaning being also used in the rail domain, the commission decided to take it for a definition. It must be reminded here that the transport domain is split into sub-domains (air, car, rail, inland waterways, maritime, road, surface) for which the ontological links are obvious. For each word it is frequent that the notions are the same in different sub-domains, but the attributes are often different.

When the definition is a question of opportunity, as it happened for the decision to examine "car pooling", the approach is frequently realised by a system of non hierarchical notions. The willingness to address a classification and a hierarchical process will be raised when the decision to examine also "public cars" or "shared taxis" will be taken, because "car sharing" and "share of cars" have been already published, and "car pooling" is on the way to be published, and is a subdivision of the term "*share of vehicles*" (as "*carsharing*" and "*bikesharing*")

In fact, this combination of hierarchical and non-hierarchical approaches cannot be avoided, due to the requested objectives, to the users needs and to the large quantity of relevant notions. In the long run of the work of the commission, the vertical and horizontal links can be explored in an ontological perspective, but it is not the case for the short term. This explains that the commission's work is more terminological than ontological.

During the study of a notion, it may happen to be necessary to examine different categories of terms representing elements belonging to the notion. Let us take the example of terms representing material and immaterial elements, namely objects and uses. The hierarchical and non-hierarchical links between the objects on the one hand, and between their uses on the other hand can be analysed in a common process or in a separate process. When the resulting classifications are examined, and when the relations objects-uses are analysed, it may happen that the trees obtained are asymetrical and mixed, one object being used for several actions and one action being possibly accomplished by several objects.

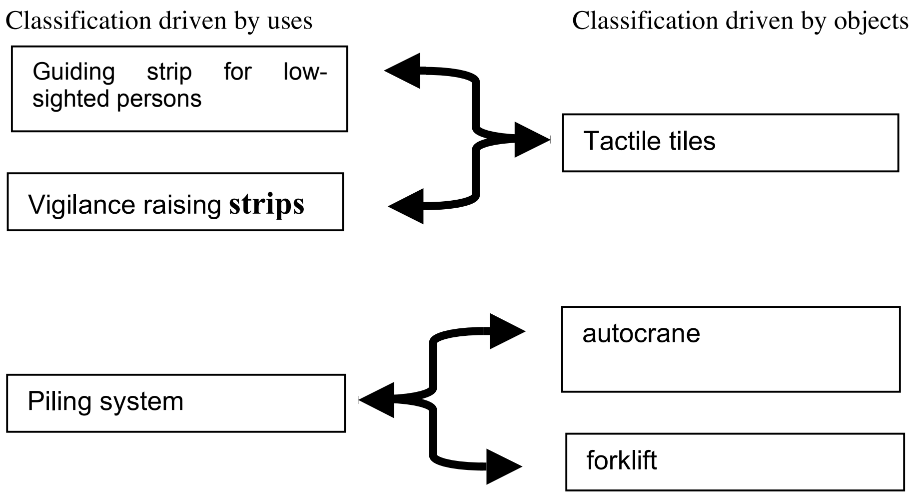
Let us compare two methods of classification, one based on objects and the other based on the functions. First let us take the example of the "tactile tiles". It is a thin and streched set of any material which can be sensible by the feet through the sole of the shoes, due to ts geometry. This doesn't explain what is its use, but it is enough for manufacturers to realise an object of this nature after having discussed with the cutomer its detailed characteristics (dimensions, type of material, …). But on the users side we can notice a diversity of situations for its use that lead to a more sophisticated vocabulary :

- Guiding strip for low-sighted persons (installed along an itinerary)
- Vigilance raising strip (installed to signal pedestrian crossing, stairs,…)

At he opposite a "piling system" in a harbour can be realised by different objects : autocrane, forklift, …

Generaly speaking, there is no systematic similarity between a classification made according to functions and a classification made according to the objects. One object can be used in order to fulfill several fuctions, according to the way it is used, and a function can be fulfilled by several objects. As the classification criteria for functions (what is the purpose? who is acting? how is it done ?…) are different from the classification criteria for the objects (what dimensions? what material? how is it handled? …) there is no reason why the resulting hierarchies should be similar.

In an ontological perspective, the work of the commission, on the long run, should be organised in order to cope with different types of classifications, for instance in this example according to the uses and according to the objects. But this has not yet been put into application.



3-3 Naming and Selection of Terms

To respond to the objectives assigned to the CSTN, it is necessary to allow that the dynamism of the language, following the socio-technical evolutions, be understandable by anybody, if possible in an intuitive way with easy pronunciation of the word (shortest words being the best). Simplicity in the wording together with simplicity in the definition are considered as necessary to keep French language living and France economy competitive on the technical and technological standpoint, with young or elder professionals mastering the usage terms, and facilitating the communication between the general public and the professionals.

Before giving some examples on the urban domain, we will detail the methodology of the CSTN for this stage of the terminological work.

First of all, the Commission check the France-Terme database to ensure that the word hasn't yet been studied. France Term is a computerised database accessible on Internet giving the scientific and technical terms recommended in the official bulletin by the General Commission of terminology and Neology (30.000 terms)

After this checking**,** two types of cases can be raised : one where the problem is to define a word already in use but the signification of which is to be agreed, and another one where the question is to create a new word in order to bring in a term corresponding to a notion that is up to now represented by a circumlocution, or a foreign word or an acronym.

The first case doesn't need a specific work for the naming stage, except in one specific case: the case where the word commonly used constitutes against the notion a misuse or an anachronism. The commission must then decide whether it has rather to let it go, or it must react. Take for example the word “accidentology”. Normally, all the French words ending by “-logy” represent a science or a method or a process. But the commission noticed that this word is frequently used to name the results of the study of the accidents, not only the studying methods. Consequently it has been discussed and decided to create the word “accidentality” to represent the results of the accidents analysis, altogether with keeping the word “accidentology” only for the activity of searching for the results, the corresponding sciences and methods.

It can be also mentioned here that euphemisms are nowadays expanding, since they correspond to a need of “political correctness”, and prevent to offend general public or categories of public sensitivity. This is the case of “lowsighted” which includes “blind” or “reduced mobility persons” which includes “handicaped”. The question is here the hierarchical inclusion with which the commission decided not to interfere.

On the other hand, the second case leads to a creative activity, driven by a certain number of considerations : easy and intuitive understanding, easy pronunciation and hearing, grammatical and etymologic validity,…

The intuitive understanding is facilitated by the existence of terms belonging to the same family of terms, and by the creation of words composed of a set of frequently used etymological roots. The French language does not allow easily, as it is possible in German, to associate a group of substantives to create a new word. The creation of a unique word to represent a “fare supplement ticket for high speed train” is hopefully not on the table, while it constitutes one on the longest words of the German language : Eisenbahnschnellzugzuschlagshein. At the opposite, the commission validated the word “autocariste” that stays in English for a “motorcoach operator. The commission dares (and it is a frequent situation) to validate terms composed of two or more words : “entrée de ville” (“city entrance”), “friche industrielle” (“industrial derelict land”), …

In order to facilitate reminding, pronunciation and hearing, the words have to be composed of a maximum of 3 syllables, soft sound for concepts, hard one for technical things, alliteration rather than combination of consonants in reverse order (try to pronunce fluently “un gradé dragon dégrade un dragon gradé” !)

Difficulties raise for some acronyms or substantivised adjectives : “GPS” and “major” for example. The term “GPS” is of current use to represent in-vehicle systems that give on the panel a map of the zone where is the vehicle, the exact location of the vehicle, and some advice concerning the way to destination. But technically, “GPS” only represent the satellite detection and transmission system that localize the vehicle. The commission thought it relevant to create the word “geonavigator”, terms that encompass all the functions and rather easy to memorize. The question is now to obtain that the providers use this term. Concerning “major”, a word representing large and well known companies, the difficulty is that in French the acronym PME (i.d. SME small and medium enterprises) exist, but nothing exists for the big ones. The commission resigned itself to create the term “compagnie majeure” that is in two words “major company”.

Each specialised commission selects the words to be examined with the ambition and belief to be a servant of the language.

As we mentioned before, the commission for terminology and neology has always the right to create a new word. It is even a very general tendency is a case of translation of a foreign expression. Each member checks then the new word among his personal and professional circle and can propose variations. The debate is initiated about the new term, and if a consensus cannot be reached then we proceed to vote. Considering that the creation of a neologism is always a perilous exercise, the commission always examines the possible use of word in use in other countries of the francophony, mainly Belgium and Quebec.

3-4 Definition of a Selected Term

After having chosen the word, the work will be to look for a suitable definition, taking into account at least with the same strength the use of the word in the speech than the notion which is meant. The practical way of the commission is a sound work about the significance, as expressed in the terminology of Ferdinand de Saussure (Dubois *et alii*, 2001). It does not neglect at the same time the signifying, the acoustic image of the word, essential for its adoption by largest public.

The work of the commission for equipment, transport and tourism, as for the other commissions, is much collaborative and must find consensus in most cases. This work needs then a rather long time. Knowing that every member is a specialist of his domain, and more aware than the average of other members about the terms coming on the table, the other members demand to understand in detail the reasons for the proposal for all examined terms. The discussion leads then to build a very rich set of ideas which encompass much more than the initial proposal, and the result by itself is richer again than the discussed ideas. The discussions exploit all the possibilities of the French language, its faculty to take into account all the prominent technical innovations, and the linguistic evolution. In fact thoughts begin by looking to concepts and notions. These debates, using as much etymology as epistemological process, are productive in their own course. In this way, the work of the commission takes care of the classical vision of the terminology as the vehicle of understanding. At the same time, the commission keeps an eye opened as much as possible on the faculty for the definition to evolve together with the expected or foerseen evolution of the notion.

Nevertheless, since the objective of the commission is in fact the understanding of the term by the largest possible public, in order not to obstruct the interaction, the selection of the word, and following its definition, are clearly meant to participate to the usage language, and to the representative language as well.

The definition must be given in one unique sentence, and must deliver the mention of the relevant descriptive characteristics, and also the formulation of the relations existing among these characteristics. The technical level and the wording of the definition must be adapted to the needs of users, specifically when the commission detected terminological misuses in the large public.

The research of equivalents in foreign languages, mainly English and German, when the origin of the notion makes it useful, is based on the parallelism between the relations term(s)-notions existing in each of the studied languages.

The General Commission for Terminology and Neology, is requested by the '96 decree to « promote plurilinguism ». This objective implies favouring the translation activities from and towards the French language, mainly by encouraging terminology data collection and structuration towards databases easy and quick to access and able to cope with the needs of international cooperation for professionals and the general public.

One of the objectives of the commission is thus to specify the « foreign equivalents » by determining the translation of the studied notion or word. The main question is about the way in which the culture associated to the language is taken into account in this exercise. Two options are possible :

- either a difference of interpretation is accepted for the notion, depending on the foreign language and the associated culture
- or a note can explain in details the cultural differences related to the studied word

In a first part of its history, the commission proposed only English terms under the section "foreign languages equivalent". This is in contradiction with non obsolete administrative principles which request that, when a translation of the French official documents is needed, it must be done at least in two foreign languages. In general the choosen languages are English and Spanish, or English and German. We will not go through the reasons of these choice. But we will try to show how the cultural heritage can influence the understanding of these translation.

In the times being there is a concern about the setting of the schedules of the railway passenger services. In general, these schedules were set by operators that introduced progressively new services according to the growing of the transportation demand. The efficiency of the network exploitation was the main element to determine the time of departure and the time of servicing the different stations. It appeared recently that customers (or users ?)[32] put the stress on the easyness of memorising the times, whatever the period of the day when they have the need to travel. This happens when the frequency is less than 4 trains per hour. A complex study is aimed at reorganizing the schedules of the trains. This study (as well as its result !) is named "cadencement". This word is derived from the word "cadence" which origin is a musical term for the time period characteristic of the rhythm. This is perfectly convenient in German where the word "takt" gives on one hand in music the term "Taktmesser" (metronome), and in the other hand in transport "Taktfahrplan" (rythmed schedule). But when you look to the studies of the German operators, you see that they normally use the locution "integrierte Taktfahrplan" or "integraler

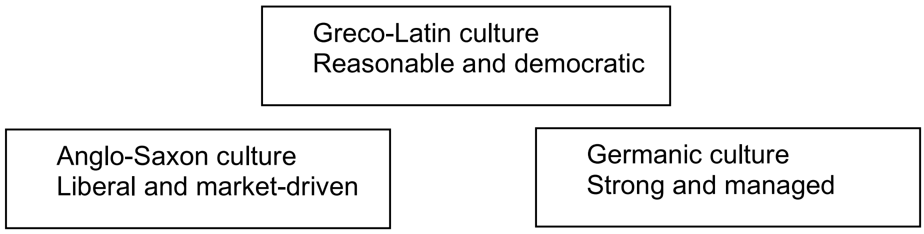
[32] Here is an example of different cultural concepts for public facilities : are travellers in a train customers (economic and market culture) or users (social culture) ?

Taktfahrplan”, in which the added adjective means that they don’t only study the schedules of one line, but also the network schedules including connections, that are part of the process. At the opposite, in English, it seems that translations are less structures or imaged, or more diverse, as « regular interval schedules » or « clockface timetables ».

For this example of « cadencement », we took the case of railway system, in which the schedules concern services with frequency expressed by hourly elements. In parallel in maritime transport the harbour servicing can be scheduled “frequently and regularly” on the weekly basis. The notion is the same, but the attributes can be different.

A second characteristic example is that of « gated communities » which are appearing already for a long time in the US, and which are being copied in the old continent, and in France, where a specific word lasted to appear. A large number of research studies were launched about this phenomenon, that seemed so specific that the researchers didn’t find it useful to propose a translation. It can be even considered that the translation could have been contradictory with the objective of the research. The researcher must describe in details all the aspects of the notion and not create a global “symbol” which could let the reader free to create himself a set of impressions. It was then the role of the commission to create a French word for that. It stayed at the term “protected residence”, with a definition as “closed residential zone responding to certain requests of safety and facilities”. You can feel immediately the frustration of the researchers who can consider this as a betrayal of their work. In fact the matter is not to translate, but to transpose. The spirit in which the inhabitants live in a “gated community” in the US is certainly different of the “protected residence” inhabitants in France. Relations between people, families, and to services, social organisations and priorities are different on both side of Atlantic ocean. In consequence, the literal translation in “community with gate” would not have been satisfactory. The “communiarism” in France is something dubious and disturbing, it represent groups of persons willing to be apart from normal living habits. In another way the term “gate” is only a part of the system which organises the protection of the zone, and is used to represent all the elements of fences and walls, together with the gate itself. This way of using “a part for the whole thing” is more in use in the anglo-saxon languages than in latin ones, where it sounds as a sophisticated or poetical way of expression. Eventually there stays a difference between the notions studied among the word and its translation, which does not raise real problems for the commission. Let us mention that in French “protected residence” is fully different of “observed residence” which is the legal notion for a person assigned to stay in a certain town during the preparation of after a trial to be controlled by the police.

Let us mention also that a large public in France uses the term “*ghetto de riches*” for the “*gated communities*”. This term has been rejected by the Commission, but shows that different cultural paradigms make difficult the translation process. It shows also that the terminological and ontological classifications can be different in various languages.



4 Discussion : from Terminology to Ontology, or Ontoterminology

The question is to determine in which measure the process adopted by the Commission of Terminology and Neology is to prepare an ontology and what is the threshold between these two domains.

The scientific process leading to create ontologies is rather recent. Its roots can be found in the researches about artificial intelligence, twenty years ago. According to Thomas R. Gruber (1993), an ontology is the specification of a conceptualisation of a knowledge domain, independantly from any specific human language : "*« An ontology is a specification of a conceptualization. […] A conceptualization is an abstract, simplified view of the world that we wish to represent for some purpose. Every knowledge base, knowledge-based system, or knowledge-level agent is committed to some conceptualization, explicitly or implicitly. An ontology is an explicit specification of a conceptualization. The term is borrowed from philosophy, where an Ontology is a systematic account of Existence. For AI systems, what "exists" is that which can be represented. […] Pragmatically, a common ontology defines the vocabulary with which queries and assertions are exchanged among agents. Ontological commitments are agreements to use the shared vocabulary in a coherent and consistent manner. The agents sharing a vocabulary need not share a knowledge base; each knows things the other does not, and an agent that commits to an ontology is not required to answer all queries that can be formulated in the shared vocabulary. In short, a commitment to a common ontology is a guarantee of consistency, but not completeness, with respect to queries and assertions using the vocabulary defined in the ontology. »*

Ontology thus presents analogies with standardisation of the traditional terminology. Nevertheless, as it is shown in the papers of Chantal Berdier and Catherine Roussey (2007) and in the works of LIRIS[33], ontology deals mainly with the conceptual representation of objects in order to allow their computerised treatment and is connected to the knowledge management. Ontology is not really aimed at "telling"; its heart and objective is in the research of notional links between terms. It

[33] Laboratoire d'InfoRmatique en Image et Systèmes d'information (UMR 5205 CNRS / INSA de Lyon / Université Claude Bernard Lyon 1 / Université Lumière Lyon 2 / Ecole Centrale de Lyon)

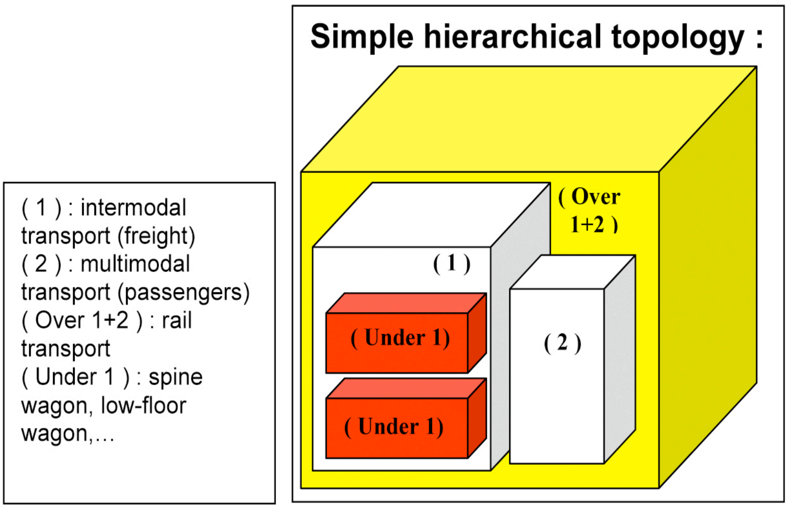
can effectively create only if there is agreement between users about a terminology and thus about notions described by the terms.

If the research on ontology expands, it is related to the heavy and strong development of information systems and computerised tools allowing a broader and quicker sharing of the knowledge. Over the abilities of the computerised tools, the very strong development of ontological researches seems to be linked to the need of transversal and collaborative approaches in numerous domains. In this sense, terminology and ontology are pillars of the knowledge management.

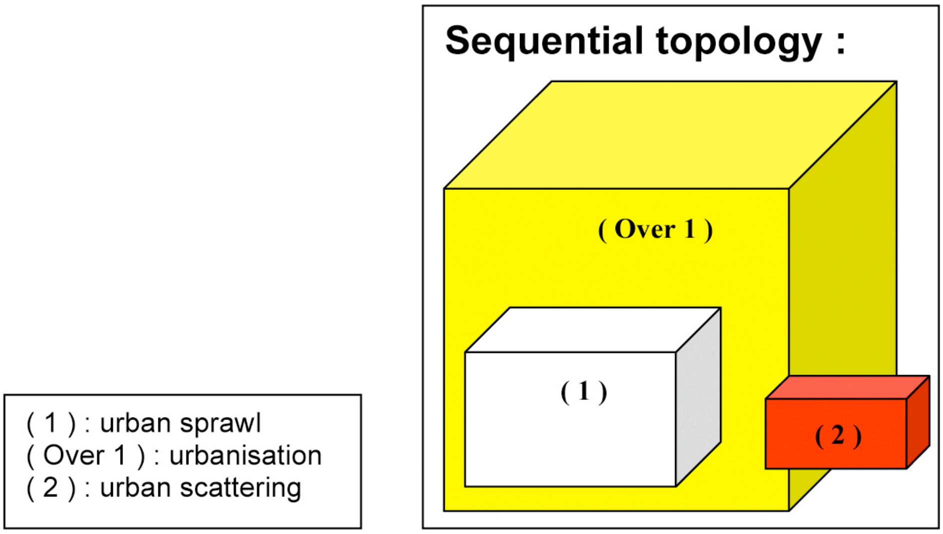
The above presentation of the activity process of the Commission for terminology and neology and its specialised commissions shows that its work cannot be considered as establishing a set of standardizing principles, or as a simple elaboration of a technico-scientific dictionary. As we have tried to show it, the terminological work of the Commission deals with specialists' words according to a notional approach of the language objects. This terminological approach is double sided, according to the distinctions made by Loïc Depecker and Christophe Roche, specifically during the colloquium *"terminologie et ontologie : description du réel" (2006)*. On one hand it is semasiological (starting from the word to study the meaning) and socio-linguistic or socio-terminologic (as we have shown hereabove for "*gated communities*" and "*tactile tiling*"). On the other hand it is onomasiologic (starting from the idea, the notion, to go towards its expression into the language, and thus the word), making standards and stabilizing concept (as we have shown for "*accidentality vs. accidentology*" and "*multimodal transport*"). In this way, the work of the Commission seems to be in the beginnings of an ontological process.

The following diagrams show an attempt of design for different topologies concerning the notions and the words studied by the commission.

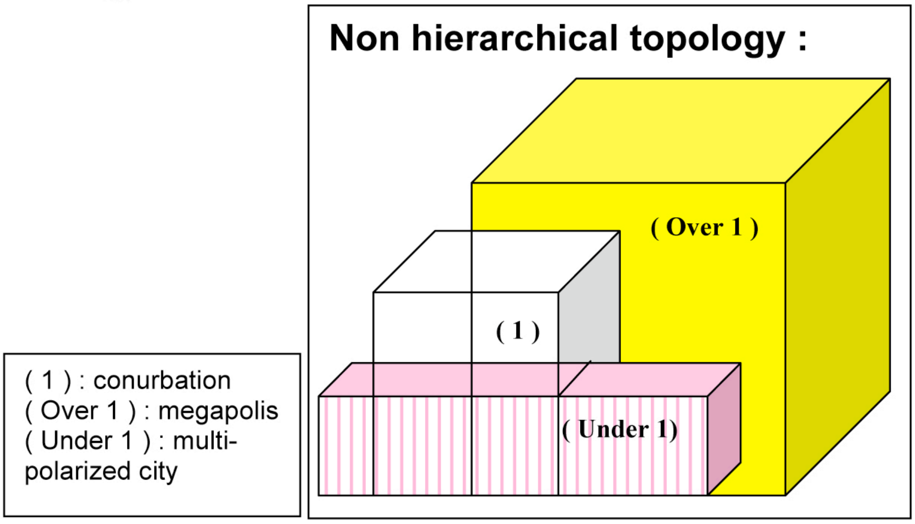
The first one is a simple hierarchy : the intermodal transport (concerning freight) and the multimodal transport (concerning passengers) are fully included in the rail transport system (although they can need also a reference to other type of transport, road transport for instance). Inside intermodal transport different types of wagons exist : those which can take on board the trucks with their loads, those which can only take the containers, and so on.



The second one is of sequential type. Urban sprawl and urban scattering are both linked to urban extension, but the first one is considered to represent new elements of the urbanized area, while the second is not (yet) within the urbanized area. The link between urbanisation and urban scattering can be defined as a diagonal ontological tree as presented by Wüster (1974) and Van Campendoudt (1997)



The third one shows a non hierarchical topology. A conurbation (meaning the merging of neighbouring cities due to their extension) can be a type of megapolis, but the phenomenon of conurbation can happend in average-sized cities, not only megapolis. On the other hand, a conurbation can contain several dense zones which make it a multi-polarized city, but a multipolarized city can exist without conurbation or inside a megapolis.

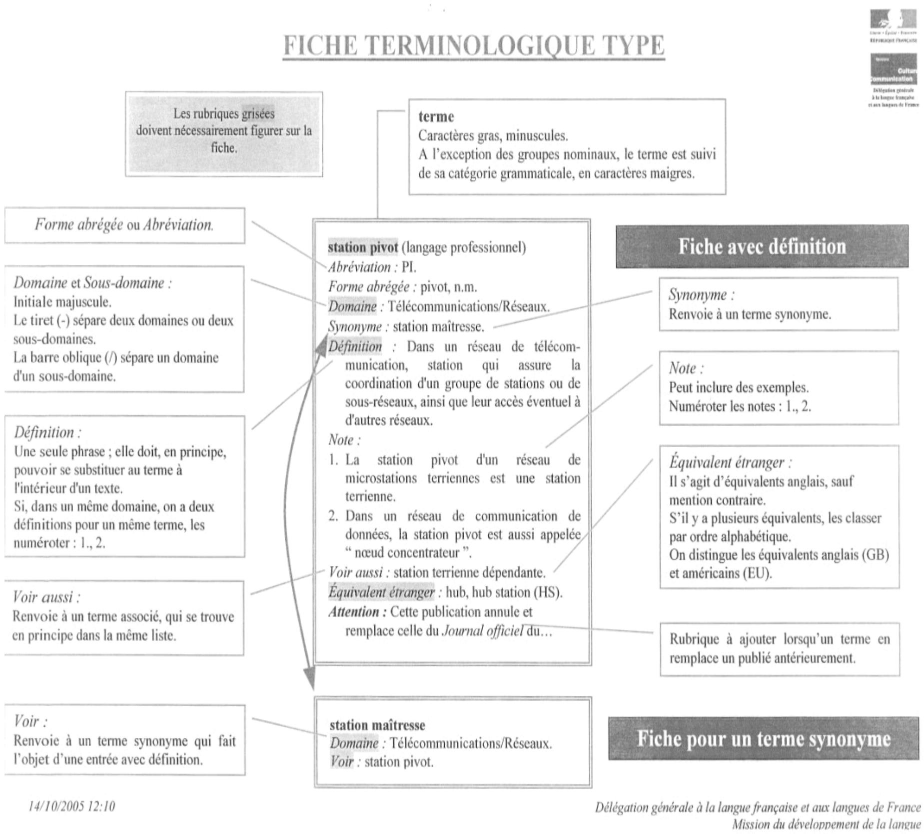


We have shown here above that the approach with a non hierarchical set of notions is frequent. It was the case for the choice of the definition of "*car sharing*". But this choice has also been made because the evolution of the notion according to the times, frequent in the scientific and technical vocabularies, is pertaining to the mission of the commission. The research, even implicit, of a classification which addresses different dimensions of a concept and of its dynamics, refers to an ontoterminological consensus, as said by Christian Roche (2007, 7). It is a situation of cross-creation from the terminological and ontological processes.

Within the terminological practices of the commission, this plurality of approaches both linguistic and logic, but also hierarchical and non-hierarchical, is imposed by the assigned objectives, by the needs of users, and by the vast quantity and diversity of relevant notions as well.

The terminographic practice leads the commission to adopt a system of non-homogeneous notions, of which the type can be together hierarchical (generic and partitive) and non-hierarchical (sequential and pragmatic) taking into account the spontaneous apparition of new terms. This mixity is imposed by the needs of the users and by the fixed objectives as well. According to the terms, and to their belonging rather to a hierarchical structure or rather to a non-hierarchical structure, the definition will include the mention of the super-ordinated notion, together with the specific characteristics of the object, or the mention related to the function of the object. The supplementary notional information can be mentioned in technical notes.

The typical terminological form of the commission is shown hereafter :



English translation of the typical terminological form on the following page :

Grey items are mandatory

Term
Bold, low case
Except for nominal groups, the term is followed by its grammatical category, in small prints

Abbreviated form or abbreviation

Domain and sub-domain :
Initial high case
Dash (-) separates 2 domains or 2 sub-domains
Slash (/) separates domain and sub-domain

Definition : one single sentence. It can normally be substituted to the word inside a text
If there is 2 definitions for one term in one domain, they must be numbered : 1, 2,

See also : link to an associated term, normally in the same list

See : link to a synonym term which has an entry and a definition

Form with definition

Hub station (professional language)
Abbreviation : HS
Abbreviated form : hub, m.n.
Domain : telecommunicatications/networks
Synonym : master station
Definition : in a telecommunication network, station which coordinates a group of stations or sub-networks, and their potential access to other networks as well.
Note :
1- The hub station of a land micro-stations network is a land station
2- In a data communication network, the hub station is also named “concentrating node”
See also : land sub-station
Foreign equivalent : !!!
Attention : This form deletes and replaces that one published in the Official Bulletin

Synonym : link to a synonym term

Note : can include examples. Number the notes : 1,2,

Foreign equivalent : normally English equivalents, mention if other. If several equivalents, class acc to alphabetical order. Make distinction between Brittany (GB) and States (EU)

To be added if the term replaces a formerly published one

Form for synonym

Master station
Domain : telecommunicatications/networks
See : hub station

When it appears that the number of terms concerning a domain constitutes a sufficient and homogeneous set, these forms are published as small handy booklets, which gather the relevant production of the specialised commissions. These booklets are distributed to the concerned administrations and medias. Another way of dissemination has been determined with France-Terme after an initiative of the Ministry of Culture as an interactive WEB site. This database of scientific and technical terms recommended by the general commission for terminology and neology and published in the Official Bulletin cannot be considered as a general linguistic dictionary. Nevertheless some of the terms are of current use. The use of the recommended terms is mandatory for public administrations, but anybody can adopt them. It is possible to search the database by domains and to download the retrieved words into a .rtf file. The WEB site presents also a news section, a suggestion mailbox, a library with downloadable documents, and among them a guide : "Write… Simply : Principles and recommendations for a quality administrative writing". This document notably has been realised in the frame of the work of a sub-committee concerning the quality of the French language in public administrations which belongs to the Franco-Quebec working group on the government modernisation. France-Terme also integrates a terminology survey about the additions made to the database. RSS feeds also exist, and subscription to the news bulletin is open for free.

5 Conclusion

Even if a word can be created by a small group of individuals, which illustrates the role of the nomothet in Platon's Cratyle, even if the political powership deals with the language and its influence, the vocabulary development is collective, subconscious, versatile. The role of the specialised commission for neology and terminology can be on one hand to be a barrier against terms coming from other linguistic contexts, and on the other hand to drive and control among existing proposals to give them a social acceptance level which these terms didn't reach and make them visible for a larger audience.

The practice of the commission looks after the word together with the meaning, even if the current work of the specialised commissions is sometimes softened by the general commission, which is concerned also by the sentential use of the words, in link with the objectives of the French Academy. This global effort in meaning, naming and defining new concepts, together with the concern of making the language understandable by the largest audience, place our work in the frame named by Christian Roche "ontoterminology" (2007).

But the work of the commission doesn't attempt to create ontologies because the approach isn't systemic, over the discussions about a single notion inside the considered domain and sub-domains. The commissions don't look at first to create classifications of notions. In fact, the ontological activity could only belong to the long times. This is not in contradiction with a strict ontological process, but this cannot at short or mid-term create a systemic model.

References

1. BERDIER, C., ROUSSEY, C., 2007. Urban Ontologies : the Towntology Prototype towards Case Studies, *in* Ontologies for Urban Development: Interfacing Urban Information Systems, Teller J., Lee J., Roussey C. ed. University of Geneva 6,7 November 2006 ; *Studies in Computational Intelligence 61*. Springer Verlag, pp. 143-156.
2. DEPECKER, L., 2007. Linguistique, terminologie, et ontologie, Actes du Colloque Terminologie et ontologie : descriptions du réel, 1er décembre 2006, Société française de terminologie, *Le savoir des mots n°4*, pp. 13-20.
3. DUBOIS, J., MEVEL, J.-P., GIACOMO, M., 2007. Dictionnaire de linguistique et des sciences du langage, Larousse, 576 p.
4. FELBER, H., 1994. Terminology Research : Its Relation to the Theory of Science, in *ALFA*, vol. 7/8, pp. 163-172.
5. GRUBER, T. R, 1993. A Translation Approach to Portable Ontology Specifications. *Knowledge Acquisition*, 5(2), pp. 199-220.
6. GRUBER, T. R, 1995. Toward principles for the design of ontologies used for knowledge sharing. Presented at the Padua workshop on Formal Ontology, March 1993, later published in *International Journal of Human-Computer Studies*, Vol. 43, Issues 4-5, November 1995, pp. 907-928.
7. ROCHE, C., 2005. Terminologie et ontologie, Larousse, *Revue Langage, n°157*, mars 2005, 11 p.
8. ROCHE, C., 2007. L'ontologie comme principe terminologique, Actes du Colloque Terminologie et ontologie : descriptions du réel, 1er décembre 2006, Société française de terminologie, *Le savoir des mots n°4*, pp. 21-42.
9. ROCHE, C., 2007. *Le terme et le concept : fondements d'une ontoterminologie*, in Actes de la Conférence TOTh 2007 : Terminologie et Ontologie : Théories et Applications, Annecy 1er juin 2007, pp. 1-22.
10. TELLER, J., LEE, J. R, ROUSSEY, C. (Eds.), 2007. Ontologies for Urban Development. *Studies in Computational Intelligence 61*. 210 p.

References - Webography

1. VAN CAMPENDOUDT, M. 1997. Abrégé de terminologie multilingue, Centre de recherche TERMISTI, Institut supérieur de traducteurs et interprètes, Bruxelles, Belgique.
   http://www.termisti.refer.org/theoweb1.htm#table
2. GRUBER, T., 2008. Centre de recherche Knowledge systems, AI Laboratory, Standford university, US. Sur le site http://www-ksl.stanford.edu/kst/ontology-sources.html, http://www-ksl.stanford.edu/kst/what-is-an-ontology.html