

Semantic Representation of Geospatial Objects Using Multiples Knowledge Domains

Rainer Larin Fonseca and Eduardo Garea Llano

Advanced Technologies Application Centre, 7^{ma} A # 21406, Playa, Havana - 12200, Cuba
{rlarin, egarea}@cenatav.co.cu

Abstract. Geographical data is obtained through abstractions made from objects in the real world. Generally, each of these abstractions is obtained by taking into account only one point of view about the object being analyzed. When different abstractions are made on the same object different data sources regarding to it are produced. These data sources are generally heterogeneous. Thus the semantic processing of these objects become challenge since different data sources must be combined to obtain good results in tasks such as information retrieval and analysis for decision-making. This paper presents an approach based on ontologies to enrich the semantic representation of geospatial objects taking into account different abstractions made on them. The experimental results show the usefulness of this approach and how it is possible to make a multidimensional semantic representation automatically using classification algorithms and search techniques on trees.

Keywords: Ontology, Classification, Semantic Representation, Geospatial Data.

1 Introduction

Usually the objects in the geospatial field can be analyzed from different semantic points of view. Thus the semantic processing of these objects becomes a challenging task, since heterogeneous data sources regarding to these objects must be combined to obtain results in tasks such as data retrieval and analysis for decision-making.

There are several approaches regarding to the representation of different semantic point of view about a same object [1-7]. A conceptual spaces theory is presented in [1] by Peter Gärdenfors. The Gärdenfors' idea proposes that the different domains can be defined as a set of quality dimensions with a geometrical or topological structure. Then, an object in the conceptual space can be seen as a vector $v = (d_1, d_2, \dots, d_n)$ where each $d_i \forall i = 1 \dots n$ is a quality of this object referred to the related domain. In this approach the object is formed with qualities from different domains.

This paper presents a new ontology-based approach for making a semantic representation of geospatial objects taking into account several knowledge domains. Thus, is enabled the representation of different semantic of point of views about geospatial objects. It also enables the semantic representation of objects stored in heterogeneous data sources. This paper continues with Section 2 which deals about main elements of the proposed approach. The Section 3 presents a method for reducing the search space in the classification process of these objects. Section 4 presents experimental results that validate the proposed approach and finally Section 5 presents the main conclusions.

2 The Fifth Dimension of Geographical Objects

Geospatial objects are usually described based on four conventional dimensions. These dimensions are referred to the location and behavior over time of these objects. In the last years there has been a new trend based on describe semantically geospatial objects. This new type of description can be named as the fifth dimension of these objects which is referred to the semantic nature of geospatial objects and the way which they should be understood by both humans and machines.

Geospatial objects have different meanings for different specialist. This fact implies that they assign different terms to the same object; e.g. “*The Zapata Swamp*” can be defined as “*Nature Reserve*”. But this designation does not take into account all the semantic nature of this object, since it also can be defined as “*Historic Place*”. When we are assigning a meaning to an object we are making a semantic representation of this object. This representation takes place in the “*Semantic Space*”. The fact that the semantics of a particular object covers several disjoint domains can be defined as “*Multidimensional Semantics*” (MS). With the word “*Semantics*” we refer to the meaning of the object taking into account all the knowledge domains that define it. Then, we can say that the semantics is the set of all possible definitions related to the object nature. Taking into account whole the semantics of an object is hard enough; it usually takes into account only a subset of these definitions. In this way a “*Semantic Abstraction*” (SA) of the object is made. Formally we can define the Multidimensional Semantics (MS) of an object as follows:

Let $D_o = (d_1, d_2, \dots, d_n)$ be the set of all domains of knowledge that define the semantic nature of an object “ o ”, which satisfies that $d_i \cap d_j = \emptyset$ with $i \neq j$ and $n \geq 1$. Thus we can define the Multidimensional Semantic Abstraction (MSA) of an object as a subset of D_o . The representation of semantic abstraction of geographical objects brings great benefits since it enables different systems to process these objects more efficiently avoiding problems such as incompatibilities between heterogeneous formats.

Geospatial objects can be represented in the “*Semantic Space*” (SS) based on their semantic abstraction which is defined as “*Context*”. The Context defines the knowledge area over which these objects are processed from semantic point of view.

Each object represented in the Semantic Space has only one context and a context can contain more than one object. Geospatial objects belonging to different contexts can be linked together through relationships between them, see Fig. 1.

Explicitly, an object in the Semantic Space can be seen as a vector $e = (s_1, s_2, \dots, s_n)$ where each $s_i \forall i = 1..n$ is the similarity value regarding to the domain d_i .

To formally define the Semantic Space, it is necessary to state some definitions:

- Let $A = (a_1, a_2, \dots, a_k)$ be the finite set of attributes.
- Let $B = (b_1, b_2, \dots, b_h)$ be the finite set of axioms.
- Let \hat{a} be an attribute that contains the geo-reference value of the geographical objects.
- Let $C = (c_1, c_2, \dots, c_p)$ be the finite set of concepts where for each $c_i \subseteq A \forall c_i \in C$.

- Let $X = (x_1, x_2, \dots, x_n)$ be the finite set of concepts referring to geographical objects where $x_j \subseteq A \cup \{\hat{a}\} \forall x_j \in X$.
- Let $U = X \cup C$ be the finite set of all possible concepts.
- Let $R = (r_1, r_2, \dots, r_t)$ be the finite set of Relationships where $r_s \subseteq U \times U \forall r_s \in R$.
- Let $r' \subseteq X \times C$ be relationships of instantiation.
- Let $\mathbb{O} = (O_1, O_2, \dots, O_q)$ be the finite set of Ontologies where $O_1 = (\mathcal{U}_1, \mathcal{R}_1, \mathcal{B}_1)$ such that $\mathcal{U}_1 \subseteq U, \mathcal{R}_1 \subseteq R, \mathcal{B}_1 \subseteq B$.

For each $x_i \in X$ the context t_j is defined as follows:

- Let $t_j = (N, E)$ be the j^{th} context, where N is the finite set of concepts that have relationships of instantiation with the geographical objects x_i and E is the finite set of Ontologies linked with these objects, defined as below:
 - $N = (c \in C : x_i r' c)$. (1)
 - $E = (O_t = (\mathcal{U}_t, \mathcal{R}_t, \mathcal{B}_t) \in \mathbb{O} : \exists c \in N \wedge c \in \mathcal{U}_t)$. (2)

Then we can define the Semantic Space as follows:

- Let $T = (t_1, t_2, \dots, t_n)$ be the Semantic Space defined by a finite set of contexts where t_j is the context referring to $x_j \in X$. Note that x_i, x_j with $i \neq j$ can have the same context.

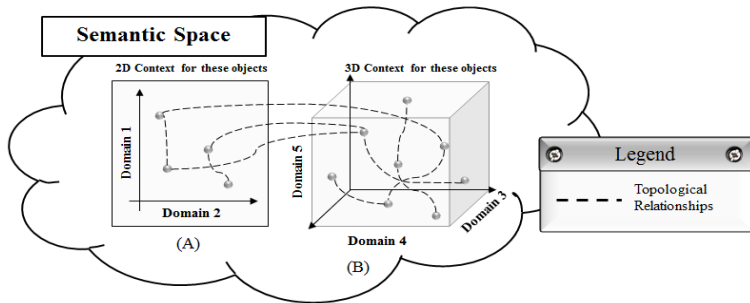


Fig. 1. Two different contexts represented in the Semantic Space. In (A) a 2D context defined by two domains is shown, in (B) a 3D context defined by three domains is shown.

In this approach, the semantic abstraction of geospatial objects is made by Data-Representation Nodes and together form the Data-Representation Ontology (DRO)[8]. Based on the Semantic Space definition we can define the DRO as a graph $G = (V, Z)$ where one hand we have that $V = (v_1, v_2, \dots, v_n)$ is the finite set of DRN such as for each geographical object x_i the corresponding DRN v_i exists. On the other hand we have that Z is the finite set of edges where $z_{ij} \in Z$ exists among the DRN v_i and v_j if and only if the geographical objects x_i and x_j are linked by a relationship $r \in R$. Similarly, the Higher-Level Ontologies (HLO) [8] in this space are defined by the expression 2 (E). The main advantage of this approach is the representation of more information about the geographical objects, which implies obtaining better results in tasks such as information retrieval and analysis for decision making.

3 Search Space Reduction in DRN Classification

As is described in [8] each geospatial object in the DRO is represented by a DRN. To set an interrelationship among the DRO and the domain ontologies (HLO) is necessary to classify the DRN regarding to existing concepts in HLO based on semantic similarity criteria. To automatically achieve this, a method based on the paradigm of the tree search with heuristic using nearest neighbor rule with rejection (1-NN_R) is presented. This new method is a variant of the method proposed in [8] which take advantage of the underlying taxonomy in the HLO to reduce the search space in the classification process. Thus, the HLO taxonomy can be defined as a finite set of concepts such that there is at least one (c^*), which can be named as “*Root*” because it is the most abstract concept from which other concepts are defined. The remaining concepts in the taxonomy are descendants of the Root; they can be named as “*child-concepts*” for those who have a relationship of type “*subclass_of*” with others including the Root. These concepts can be divided into p disjoint sets named as “*sub-taxonomies*”. Those concepts that do not have descendants can be named as “*leaf concepts*”.

From this abstraction, the DRN classification ($x_i \leftrightarrow \text{drn}_i$) can be faced using search trees methods. Thus a concept $c_i \in C$ defined by drn_i can be obtained without having to process whole concepts in C , see pseudo code (A) in Fig. 2.

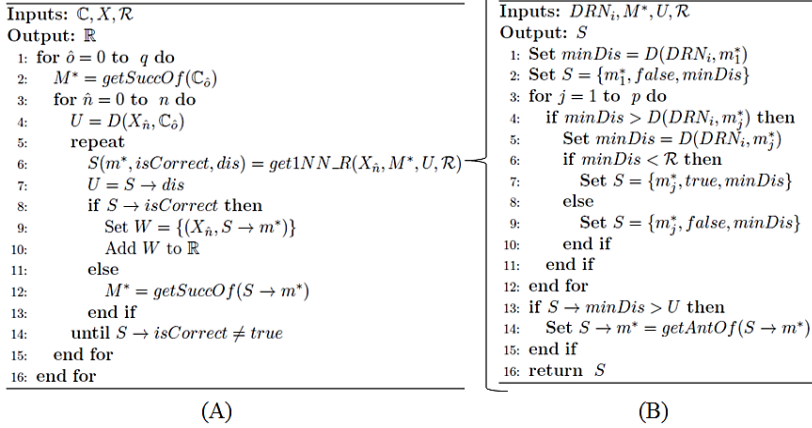


Fig. 2. (A) Pseudo code of the algorithm 1-NN_R-BFS. (B) Pseudo code of the algorithm 1-NN_R. Algorithms used for classification of NRD.

Where:

Let M^* be the finite set of concepts $m_j^* \in C$ that are input parameters in the proposed method. M^* is defined as below:

$$M^* = \{m_1^*, m_2^*, \dots, m_p^*\} / M^* \subseteq C \quad (3)$$

Likewise S is the result of the 1NN_R algorithm described in the pseudo code (B) of Fig. 2. Is defined as below:

$$S = (m^*, \text{isCorrect}, \text{dis}) \quad (4)$$

Where:

- m^* : It is the concept that better defines to drn_i .
- isCorrect : Variable that denote if the nr_i was classified or not.
- dis : It is the computed dissimilarity between nr_i and m_j^* .

In the pseudo code (B) the 1NN_R modified algorithm is shown. This algorithm is used by the proposed algorithm (1NN_R-BFS), which is presented in pseudo code (A), see Fig. 2.

For the pseudo code B we have that:

- $D(\text{drn}_i, m_j^*)$: Function that computes the dissimilarity among drn_i and m_j^* .
- \mathcal{R} : Denotes the rejected threshold for the DRN classification.
- $\text{getAntOf}(m^*)$: Function that returns the concept m^{**} which is an antecessor of m^* in the taxonomy.
- U : Denotes the computed dissimilarity to the concept m^{**} whose descendant are contained in M^* . U is used as stopping criterion of the algorithm.

The 1-NN_R algorithm let to know if the drn_i was classified or not at each iteration. S contains the concept m_j^* which better defines to drn_i , the value of if the classification was correct and the computed similarity value. This paper proposes the Jaccard distance [9] to determine the similarity between both DRN (x_i) and Concepts (c_i) based on their features. For a better understanding of this pseudo code (A) of Fig. 2 it is necessary to define that:

- m^* : Denotes the Root /Sub-root concept in the HLO taxonomy for all HLO processed.
- \mathbb{C} : Denotes the finite set of Root /Sub-root taken into account by the $1 - \text{NN}_R - \text{BFS}$ algorithm, see the expression (5):

$$\mathbb{C} = \{c_1^*, c_2^*, \dots, c_q^*\} \quad (5)$$

- X : Denotes the finite set of DRN to be classified, see the expression (6):

$$X = \{\text{nr}_1, \text{nr}_2, \dots, \text{nr}_n\} \quad (6)$$

- $\text{getSuccOf}(c^*)$: Is the function that returns in M^* those concepts $m_j^* \forall j = 1 \dots p$ which are successors of m^* in the HLO taxonomy.
- $\text{get1NN}_R(\text{nr}_i, M^*, U, \mathcal{R})$: Is the function that executes the $1 - \text{NN}_R$ algorithm.
- \mathbb{R} : Denotes the finite set W_i of pairs of m_j^* and nr_i that have a taxonomic relationship between them, see the expressions (7) and (8):

$$\mathbb{R} = \{W_1, W_2, \dots, W_k\} \quad (7) \quad W = \{(\text{nr}_i, m_j^*)\} \quad (8)$$

The underlying idea behind this method is to take the shortest path to the concept m^* that better defines to the nr_i that is being processed. This method follows the heuristic that the similarity increases when the drn_i being close to the concept m_j^* with which it can be classified and decreases when the drn_i being away from it.

4 Experimental Results

4.1 Automatic Representation of Geographical Data in the Multidimensional Semantic Space

To validate the proposed method, the use of geographical objects that can be seen from different semantic domains was taken. Each of these domains is represented by a different ontology. Three domains for representing the set of objects have been selected. These domains are: *Nature Reserve (NR)*, *Tourist Attraction (TA)* and *Historic Place (HP)*. The selected objects are: *Playa Girón (PG)*, *Castillo de los Tres Reyes del Morro (CM)*, *Parque Baconao (PB)*, *Sierra del Rosario (SR)*, *Ciénaga de Zapata (CZ)*, *Sierra Maestra (SM)*, *Cayo Coco (CC)*, *Catedral de La Habana (CLH)*, *Sierra del Escambray (SE)*, *Centro Histórico de Camagüey (CHC)*. Each of these objects contains features that are taken for the classification process. The Table 1 shows the binary vectors that represent the existence or absence of these features for both objects and domains.

Table 1. The obtained vectors of common features for both concepts in HLO and DRN and the ground true of the objects classification taking into account expert criteria is presented

Domains	Binary Vectors for the Domains														
NR	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0
TA	1	1	0	0	0	0	0	0	1	1	1	1	1	1	0
HP	1	1	0	0	0	0	0	0	0	0	0	0	0	1	1

Objects	Binary Vectors for the Objects															Ground True			
PG	1	1	1	0	1	1	1	0	1	0	1	1	0	1	1	1	1	0	NR, TA, HP
CM	1	1	0	0	0	0	0	0	1	0	1	1	1	1	1	1	1	1	-, TA, HP
PB	1	1	1	1	0	1	1	1	1	1	0	1	1	1	1	1	1	0	NR, TA, HP
SR	1	1	1	1	0	1	0	1	0	0	0	0	0	0	0	0	0	0	NR, -, -
CZ	1	1	1	1	0	1	0	1	0	0	1	0	1	1	1	0	0	0	NR, TA, -
SM	1	1	1	1	0	1	1	1	0	0	0	0	0	0	1	1	0	0	NR, -, HP
CC	1	0	0	0	0	0	0	0	1	1	0	1	1	0	0	0	0	0	-, TA, -
CLH	1	1	0	0	0	0	0	0	1	1	0	1	1	0	1	1	1	1	-, TA, HP
SE	1	1	1	1	1	1	1	0	0	0	0	0	0	0	1	1	1	1	NR, -, HP
CHC	1	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	1	-, -, HP

The classification was made in MATLAB[10] using the Jaccard Distance implementation of PRTools toolbox [11, 12]. The Fig. 3-A shows the belonging values obtained for each geospatial object regarding to the domains. Axes are indicating the semantic domains. The objects are represented based on the value of belonging to each of these domains. In this figure can be shown as geospatial objects semantically different are distant between them while similar objects are close. The Fig. 3-B shows these geospatial objects represented in the multidimensional Semantic Space using the viewer Protégé 4.0.

The main contribution of the proposed approach is that it provides more information about the objects represented in the Multidimensional Semantic Space. This fact enables the data-analysis from different points of view improving conventional tasks such as information retrieval and analysis for decision-making. The experiment also shows the feasibility of carrying out this type of representation automatically taking into account different domains of knowledge from which geospatial objects can be defined.

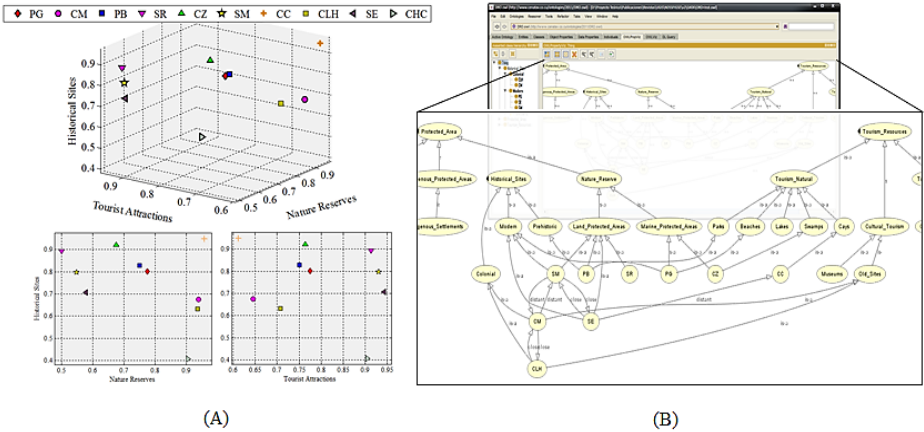


Fig. 3. (A) Graphic Representation of the belonging values of the geographical objects regarding to the domains. (B) Graphical Representation of the Generated Ontology (ORD + ONS) using the Protégé viewer.

4.2 Comparative Results between 1-NN_R and 1-NN_R-BFS Algorithms

The main objective of this experiment is compare the performance of the 1-NN_R-BFS algorithm regarding to 1-NN_R algorithm based on the number of transactions in the classification process. Thus, the search space is reduced taking advantage of the taxonomic structure of the ONS. The used datasets have the structure shown in expression 11 and the results are shown in Table 3.

$$DS = (G_{ds}, O_{ds}) \tag{11}$$

Where:

- G_{ds} : It is the geospatial dataset that contains geospatial objects (X) to be used.
- O_{ds} : It is the ontological dataset that contains the HLO (E) to be used.

For this experiment has been used three datasets (DS) described to below:

- DS_1 : For this dataset we have that:
 - G_{ds_1} : It is the set of geospatial data layers that contains information about two types of objects: soil and geology.
 - O_{ds_1} : For this dataset we have used a *Land Cover Ontology*. This HLO is a domain ontology which contains the semantic abstraction of the geospatial dataset explicitly defined by the Geology and Soil concepts.

The goal of this dataset is to show the performance of both algorithms taking into account a high amount of geospatial objects. The geospatial objects are semantically defined by one domain ontology (*Land Cover Ontology*).
- DS_2 : For this dataset we have that:
 - G_{ds_2} : This geographical dataset is the same of the dataset DS_1 , but taking one data layer for each type of objects (geology and soil).

- O_{ds_2} : For this dataset we have used three HLO, they are: The Land Cover Ontology (previously used in the dataset DS_1), the Rocks Classification Ontology and the Soil Classification Ontology.

The goal of this dataset is to show the performance of both algorithms taking into account a multidimensional context into the Semantic Space. The context is defined by three different Ontologies. Some of these objects belong to one or several domain.

- DS_3 : For this dataset we have that:
 - G_{ds_3} : It is the same geographical dataset used in the dataset DS_1 .
 - O_{ds_3} : It is the same ontological dataset used in the dataset DS_2 .

The goal of this dataset is to show the performance of both algorithms taking into account many geospatial objects over a multidimensional context.

Table 2. Transactions effected by each algorithm on each dataset

Datasets	Algorithms		
	\hat{A}	\check{A}	Saved Transaction
$DS_{1(D^+,O^-)}$	194940	58482	70 %
$DS_{2(D^-,O^+)}$	114855	30870	73,11 %
$DS_{3(D^+,O^+)}$	604314	337896	44,08 %

\hat{A} : 1-NN_R Algorithms.

\check{A} : 1-NN_R-BFS Algorithms.

D^+ : A lot of data in the geographical dataset.

D^- : Few data in the geographical dataset.

O^+ : Multidimensional ontological dataset.

O^- : One-dimensional ontological dataset.

The performance results of each algorithm for each of the datasets are shown in Table 2. Each of these datasets is focused on different variants of processing that might be encountered in real applications, these are: A large amount of objects in a one-dimensional context (D^+, O^-), a few objects in a multidimensional context (D^+, O^-) and a large amount of objects in a multidimensional context (D^+, O^-). As result we can see that in all cases the number of transactions performed by the 1-NN_R algorithm is bigger than those made by the 1-NN_R-BFS algorithm. This is because on one hand the 1-NN_R algorithm processes all the objects in each data set, making unnecessary transactions. On the other hand we have that the 1-NN_R-BFS algorithm optimizes the processing through reducing the search space taking advantage of the taxonomic structure of the ONS.

5 Conclusions

The proposed approach seeks to lay the groundwork for representing geographical objects from semantic point of view automatically. The main advantage of this approach lies in the proposal for a mechanism that can contain different semantic abstractions made from geographical objects. This makes it possible to represent these objects closer to their semantic nature. This results in the representation of more information about these objects. Thus, tasks like information retrieval and analysis for

decision making can be improved. In addition the use of the algorithm 1-NN_R-BFS decreases the number of transactions made over the sets of data. This fact brings the additional advantage that these methods can be used in real applications.

References

1. Gärdenfors, P.: Conceptual spaces as a framework for knowledge representation. *Mind and Matter* 2, 9–27 (2004)
2. Raubal, M.: Formalizing Conceptual Spaces. *Formal Ontology in Information Systems*. In: *Proceedings of the Third International Conference (FOIS 2004)*, p. 114 (2004)
3. Carmagnola, F., et al.: A Semantic Framework for Adaptive web-based Systems. In: *Proceedings of the 9th Conference on Advances in Artificial Intelligence*, pp. 370–380 (2005)
4. Balley, S., Parent, C., Spaccapietra, S.: Modelling geographic data with multiple representations. *International Journal of Geographical Information Science* 18, 329–354 (2004)
5. Adnani, M.E., Yétongnon, K., Benslimane, D.: A multiple layered functional data model to support multiple representations and interoperability of GIS: application to urban management systems. In: *Proceedings of the 9th ACM International Symposium on Advances in Geographic Information Systems*, pp. 70–75 (2001)
6. Strang, T., Linnhoff-Popien, C., Frank, K.: CoOL: A Context Ontology Language to Enable Contextual Interoperability. In: Stefani, J.-B., Demeure, I., Zhang, J. (eds.) *DAIS 2003*. LNCS, vol. 2893, pp. 236–247. Springer, Heidelberg (2003)
7. Ahlqvist, O.: A Parameterized Representation of Uncertain Conceptual Spaces. *Transactions in GIS* 8(4), 493–514 (2004)
8. Larin-Fonseca, R., Garea-Llano, E.: Automatic Representation of Geographical Data from Semantic Point of View throughout a New Ontology and Classification Techniques. *Transaction in GIS* 15(1) (2011)
9. Jaccard, P.: Étude comparative de la distribution florale dans une portion des Alpes et des Jura. *Bulletin de la Société Vaudoise des Sciences Naturelles* 37, 547–579 (1901)
10. MATLAB, version 7.10.0 (R2010a). The MathWorks Inc., Natick (2010)
11. Duin, R.P.W., et al.: *PRTools4 A Matlab Toolbox for Pattern Recognition Version 4.1.5*. Delft Pattern Recognition Research, Faculty EWI - ICT, Delft University of Technology, The Netherlands (2009), <http://prtools.org>
12. Duin, R.P.W., et al.: *DisTools A Matlab Toolbox for Pattern Recognition* Delft Pattern Recognition Research, Faculty EWI - ICT, Delft University of Technology, The Netherlands (2009), <http://prtools.org>