GEOINFO 2010
XI Simpósio Brasileiro de Geoinformática
XI Brazilian Symposium on GeoInformatics

Proceedings of the XI Brazilian Symposium on Geoinformatics

Campos do Jordão, SP, Brazil

November 28 to December 01, 2010

Vania Bogorny

Lúbia Vinhas

Eds.

Dados Internacionais de Catalogação na Publicação

SI57a Simpósio Brasileiro de Geoinformática (11. : 2010: Campos do Jordão,SP)

Anais do 11º Simpósio Brasileiro de Geoinformática, Campos do Jordão, SP, 28 de novembro a 01 de dezembro, 2010. / editado por Vania Bogorny (UFSC), Lúbia Vinhas (INPE). – São José dos Campos, SP: MCT/INPE, 2010.
CD + On-line
ISSN 2179-4820

1. Geoinformação. 2.Bancos de dados espaciais. 3.Análise Espacial. 4. Sistemas de Informação Geográfica (SIG). 5.Dados espaço-temporais. I. Borgorny, V. II. Vinhas, L III. Título.

CDU: 681.3.06

Organized by

INPE

ine
CTC • UFSC
Informática e Estatística

Sponsored by

INPE

Supported by

SELPER
SOCIEDADE DE ESPECIALISTAS
LATINOAMERICANOS EM
SENSORIAMENTO REMOTO

Conference Organization

General Chair

Lúbia Vinhas
Instituto Nacional de Pesquisas Espaciais, INPE

Program Chair

Vania Bogorny
Universidade Federal de Santa Catarina, UFSC

Local Organization

Daniela Seki
INPE

Janete da Cunha
INPE

Luciana Moreira
INPE

Terezinha Gomes dos Santos
INPE

Program Committee

Vania Bogorny, UFSC, Program Chair

Alejandro C. Frery, UFAL
Ana Paula Afonso, Universidade de Lisboa, Portugal
Andrea Rodriguez, Universidad de Concepción, Chile
Antonio Carlos Rocha Costa, FURG
Antonio Miguel Vieira Monteiro, INPE
Armanda Rodrigues, Universidade Nova de Lisboa
Camilo Daleles Rennó, INPE
Claudio Almeida, INPE
Claudio Baptista, UFCG
Claudio Esperança, UFRJ
Clodoveu A. Davis-Jr, UFMG
Christopher B. Jones, Cardiff University, UK
Eliana M. Kalil Mello, INPE
Flavia de Fonseca Feitosa, INPE
Frederico Fonseca, The Pennsylvania State University
Gilberto Câmara, INPE
Gilberto Ribeiro de Queiroz, INPE
Gennady Adrienko, Fraunhofer Institute IAIS, Germany
Jorge Campos, UNIFACS
Jugurta Lisboa Filho, UFV
Julio C. L. D'Alge, INPE
Jussara de Oliveira Ortiz, INPE
Holger Schwarz, University of Stuttgart, Germany
Karla A. de V. Borges, PRODABEL
Karine Reis Ferreira, INPE
Laercio M. Namikawa, INPE
Leila M. G. Fonseca, INPE
Lúbia Vinhas, INPE
Luis Otavio Alvares, UFRGS
Marcelino P. dos Santos Silva, UERN
Marcelo Tilio M. de Carvalho, Tecgraf, PUC-Rio
Marco Antonio Casanova, PUC-Rio
Marcus Vinicius A. Andrade, UFV
Maria Isabel S. Escada, INPE
Raul Queiroz Feitosa, PUC-Rio
Renato Fileto, UFSC
Ricardo Rodrigues Ciferri, UFSCar
Ricardo da Silva Torres, UNICAMP
Sergio Donizete Faria, UFMG
Sergio Rosim, INPE
Silvana Amaral, INPE
Stephan Winter, The University of Melbourne, Australia
Tiago G. de Senna Carneiro, UFOP
Valeria Cesário Times, UFPE
Valeria Gonçalves Soares, UFPB
Yannis Theodoridis, University of Piraeus, Greece

Additional Reviewers

Lin Tzy Li, UNICAMP
Leone Masiero, Tecgraf, PUC-Rio
Marcelo G. Metello, Tecgraf, PUC-Rio

Preface

The GEOINFO series of scientific symposia is an annual forum for exploring ongoing research, development of innovative applications on geographic information science and related areas. After a sabbatical break in 2009, GEOINFO returns to the friendly atmosphere of Campos do Jordão, SP, Brazil, site of many memorable editions.

A total of 59 articles were submitted for the 2010 edition, and a total of 21 were accepted, thus defining an acceptance rate of 35.59%. Maintaining its tradition, GEOINFO 2010 presents two highly regarded keynote speakers, Fosca Giannotti and Edzer Pebesma, to present to the local community a perspective of the state-of-the-art in the area and motivate very productive exchanges and discussions.

This year the main discussion topics at GEOINFO are Mobility Data Analysis and Semantic Issues in GIS.

The General Chair and the Program Chair wish to thank specially Terezinha, Daniela, Janete and Luciana from INPE, for their help with the symposium organization.

Vania Bogorny
Program Chair

Lúbia Vinhas
General Chair

Contents

Looking inside the Stops of Trajectories of Moving Objects

Bruno N. Moreno[1], Valéria C. Times[1], Chiara Renso[2], Vania Bogorny[3]

[1] Centro de Informática, UFPE, Brazil
[2] KDDLAB, ISTI-CNR, Pisa, Italy
[3] Departamento de Informática e Estatística, UFSC, Brazil

{bnm,vct}@cin.ufpe.br, chiara.renso@isti.cnr.it, vania@inf.ufsc.br

Abstract. *Trajectory data are normally generated as sample points, which are very difficult to understand and to analyze because they are often collected with no semantic information. Several studies have been developed for trajectory data analysis. Recently, a new model was designed to reason over trajectories as stops and moves, where stops are the important parts of trajectories. Based on this work, different methods have been developed to instantiate this model, based on different characteristics like speed and direction, aiming to give more semantics to trajectories. In this work we go one step forward to existing works that compute stops of trajectories. We evaluate the behavior of a trajectory considering first its geometric properties like velocity and direction change, and then, based on this analysis we propose to use domain knowledge that describes some characteristics of the application domain to infer the goal of the stops. To validate the proposed method we present some experiments over real trajectory data.*

Keywords
Moving objects, automatic semantic trajectory annotation, movement behavior, stops and moves, trajectory patterns

1. INTRODUCTION

The price reduction of mobile devices like mobile phones, GPS, and RFID has significantly increased their use for several objectives. This has generated large amounts of data that can be explored for several application domains, like traffic management, animal migration, human behavior in a shopping mall, etc.

The mobile devices leave behind spatio-temporal traces that characterize the trajectory of a moving object. Trajectory data are normally generated as sample points, which are very difficult to understand and to analyze because they are often collected with no semantic information. It is even more difficult to extract implicit and previously unknown patterns from this data.

Several works have been developed for trajectory data analysis. One group of works has developed methods to generate patterns focusing on the geometrical properties of trajectories and defining types of trajectory patterns like convergence, encounter, flock, leadership, etc [12]. Another group of works has focused on the analysis and mining of trajectory sample points, basically considering time and space. Some examples include the extraction of clusters of trajectories located in dense regions

[15], groups of trajectories that move between regions in the same time interval [8], patterns of trajectories with similar shapes [11], or with similar distances [18].

More recently some works started focusing on the analysis of trajectories from a semantic point of view, trying to add context information. Guc *et. al.* [9], for instance, developed a method where the user manually annotates the trajectories with the interesting points.

In 2008, Spaccapietra proposed the first data model looking at trajectories from the conceptual point of view [20]. In this approach, a trajectory is a set of important places called *stops*. From this starting point, different works have been proposed to instantiate the model of stops, like [1], [12], and [17]. Alvares [1] proposed an approach that identifies the important parts of trajectories considering as context information a set of geographic information available for the region where the trajectories were collected. In [12] an algorithm is proposed to instantiate stops based on the variation of the direction of the trajectory. The work of Palma [17] computes the important places (stops) of trajectories by finding the regions where the velocity is lower than the average speed of the trajectory.

Figure 1 shows three examples of trajectories. In Figure 1.1, a trajectory sample point is presented. In Figure 1.2 a semantic trajectory is presented for a tourism application, and in Figure 1.3 a semantic trajectory is presented for a traffic management application. In Figure 1.2 stops were computed by the method IB-SMoT, proposed by [1], where the only semantic information considered is the geographic location. In this work, a trajectory must intersect a previously defined geographic location for a minimal amount of time. In Figure 1.3, the method CB-SMoT was used to find low speed stops, among which, some are *known geographic places* (e.g. airport), but two are clusters that do not intersect any geographic location given a priori.

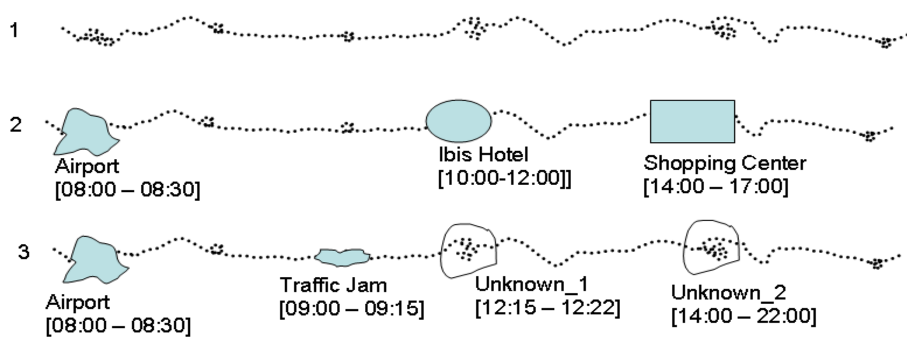


Figure 1 - Examples of trajectories

One problem of the method CB-SMOT is that it finds the regions with low velocity, but if these regions do not intersect a known geographic location, these places are labeled as *unknown stops,* as shown in Figure 1.3. Besides this problem, how could we distinguish for instance, in two animal stops if they are either feeding or resting? How can we distinguish stops at an open shopping in downtown or in a big shopping mall if people are shopping, eating at a restaurant, watching a movie, or working?

In this work we propose to go one step forward to existing works that have focused on the generation of stops as the identification of important parts of trajectories. We propose a novel method to look *inside the stop* analyzing the behavior of the moving object to infer the goal of the stop. This method performs basically two main steps: first, it evaluates the behavior of the trajectory considering the geometric properties like *velocity* and *direction* change, and second, it makes use of domain knowledge to infer the goal of the stop based on the pattern of speed and direction change.

The remainder of this paper is organized as follows. Section 2 presents some related works and the main contribution. Section 3 presents the basic definitions of the proposed method and an algorithm to infer the goal of the stop. Section 4 gives experimental results and Section 5 concludes the paper and suggests future work.

2. Related Works and Contribution

Existing works for adding more semantic information to trajectories can be classified in two groups: the works like [3] that annotate, normally manually, information to trajectories, and the works that follow the model of stops and moves [19]. We follow the second approach, which allows us to automatically discover context information to enrich trajectories, independently of application domain, since this approach is more generic.

The method presented in this article gives meaning to certain points of a trajectory, which correspond to the important places and are called stops. Stops are application dependent, and are automatically generated by a method that is the most appropriate for the domain.

In this section, we summarize three methods to find stops in trajectories that are closely related to our work and that deal with single trajectories: IB-SMoT, CB-SMoT and DB-SMoT. IB-SMoT [1] generates groups of trajectory points based on the intersection of these points with geographic objects defined as relevant to the application domain. This intersection must meet a minimum time threshold, such that the subtrajectory should continuously intersect the geographic object for the minimum time. These intersected places can be hotels, schools, etc. The main problem of this method is that for several applications there might be no geographic information.

CB-SMOT [15] is a clustering method based on the variation of trajectory speed. This method has basically two steps: first it evaluates each trajectory and generates clusters formed by subtrajectories in which the speed is lower than a given threshold, called *avgSpeed,* for a minimal amount of time (*minTime*). For example, if the average speed of the trajectory is 100 km/h and the *avgSpeed* is specified as 50 km/h, all subtrajectories in which the speed is lower than 50 km/h for at least *minTime*, will be labeled as *unknown* stop. In a second step, the method tests the intersection of the clusters with a set of user defined candidate stops, which are geographic objects relevant to the application. All clusters (unknown stops) that intersect the geographic objects for a minimal amount of time will be labeled with the name of the geographic object, otherwise they remain as unknown stops.

The algorithm DB-SMoT [13] is also a clustering method, but clusters on single trajectories are generated based on the variation of the trajectory direction. This method is interesting in specific domains where the direction variation has a greater impact then speed. Clusters are generated for subtrajectories where the direction variation is lower than a given threshold *minDir* and for a minimal amount of time *minTime*.

Existing methods to find stops are unable to either discover the behavior of the moving object or the goal of the trajectory or a stop. One main reason is because only objective measures are used, like speed, direction and time. The main contribution of our proposal is to go one step forward by looking inside the stop, considering not only objective measures, but semantic information stored into a knowledge base about the domain to infer the behavior.

In summary, the main contributions of the paper are (I) an algorithm that is generic enough to enrich trajectory important places with semantic information based on the behavior of the moving object, in different application domains; (II) the use of domain knowledge to interpret and understand traces of moving objects in order to use this information for decision making processes in applications like urban planning, animal migration, marketing, etc.

3. The Proposed Method

In this section we first present some new concepts and definitions that may be useful to understand the proposed approach and then we present the algorithm that makes use of these concepts.

3.1 Basic Definitions

According to [20], a trajectory is the user defined record of the evolution of the position (perceived as a point) of an object that is moving in space during a give time interval in order to achieve a given goal.

Definition 3.1 (Trajectory Context): Trajectory Context is a set of conditions and influences used to identify *why* a mobile object has stopped while it is moving in space during a given time interval.

Context information provides the ability to discriminate what is important or not at any given time [20]. Context information can be geographic (where and when the object has gone) or behavioral (how and why the object executed the movement) or about recognition (who is the mobile object or what has moved) [5] [16] [22]. The context information about trajectories allows the movement of the mobile object to be tracked and understood in order to enable the planning of future actions for certain types of events or situations, or even find groups of objects with similar behavior.

In this paper, context information is seen as the knowledge that indicates the reason or the purpose of the movement of the mobile object because in this work, as indicated in Definition 3.1, we are interested in defining *why* a mobile object has stopped. The proposed method investigates the goal of the trajectory by analyzing its stops. The definition of stop given in [17] is therefore extended to include the goal of a stop at a given time and location.

Definition 3.2 (Contextualized Stop): A contextualized stop represents an important place of a trajectory in which the mobile object has been for a minimal amount of time and for a given reason.

The proposed method makes use of contextual information about the stops and spatio-temporal data on the movement of mobile objects to infer the reason *why* the mobile object has performed a given stop. To perform this type of inference, we consider that each stop has a sub-trajectory. In order to identify different subtrajectories inside stops we apply again the clustering algorithm used to compute stops (based on direction or speed variation), in order to lead to new stops. These new stops are defined as *contextualized substops*.

Definition 3.3 (Contextualized Substop): A contextualized substop is a stop of a subtrajectory such that:

(i) Its goal is derived from a set of rules; and

(ii)It is part of the goal of the contextualized stop that represents the subtrajectory.

The proposed method accesses a knowledge base represented by a set of rules and checks if the subtrajectory of each substop satisfies one or more rules. Each rule of the knowledge base represents a possible goal. The set of goals inferred by the method for all substops of the subtrajectory summarize the purpose of the contextualized stop. Figure 2 shows an example of a knowledge base describing the behavior of pedestrians inside a shopping center. The first rule expresses that when a moving object stays for at least 2 hours without moving (speed zero and direction variation zero), the object is at a cinema. The third rule expresses that if the moving object stays for at least 8 hours, with speed lower than 0,5km/h and little direction variation no more than 10 degrees, the goal of the stop is *working*.

minTime	maxSpeed	maxDirection	goal
2 hours	0	0	cinema
1 hour	1,5 km/h	20 degrees	shopping
8 hours	0,5 km/h	10 degrees	working

Figure 2. Example of a knowledge base

The objective of this work is not limited to detect the interesting places visited by a moving object, but what he/she was doing at these places. Figure 3 gives a more clear idea of this work. Each stop has a purpose, a goal, for a given moving object (person, animal, etc.), and this stop occurs at some place. This means that the goal of a stop, is one more aspect of the stop concept. The goal of the stop can be an activity that is assumed to be done by the moving object at a specific place. The specific activity depends on the application and can be found in a knowledge base, such as an ontology. The same occurs for a substop, it has a goal and occurs at some place. For instance, the most generic goal of a stop at a shopping center can be for entertainment, and what we intent to discover is what the purpose/goal of each substop is (e.g. watching a movies, eating at a restaurant, shopping).

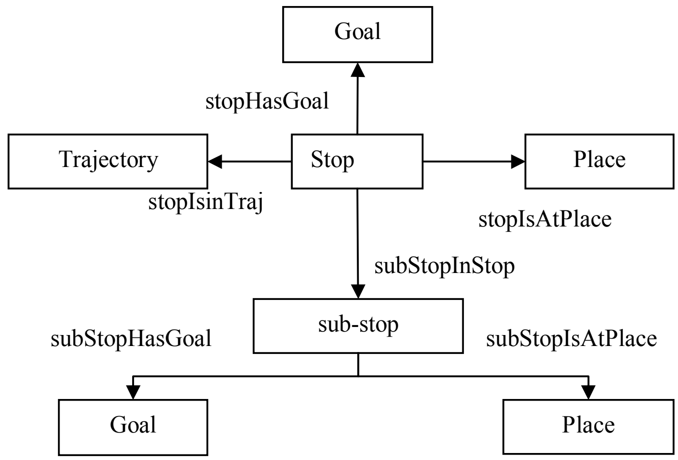


Figure 3 – Semantic Trajectory Conceptual Representation

3.2 The Proposed algorithm

Listing 1 presents the algorithm that illustrates the proposed approach. The algorithm receives as input a set of stops *S* computed by any of the existing methods in the literature, that the user chooses as the most interesting for the application domain, and a knowledge base about the domain, *kbase*.

```
1 INPUT:
2    S //set of stops
3    kBase //Knowledge Base
4
5 OUTPUT: B // a set of contextualized substops
6
7 METHOD:
8    subStops -> CB-SMoT(S);
9     subStops -> subStops  + DB-SMoT(S);
10    IF (subStops!= {})
11       B -> executeInference(subStops, kBase);
12   END
```

Listing 1 – Computing Substops psudo-code

The output of the method is a set of contextualized substops. In a first step the algorithm computes substops based on both speed variation and direction variation. For this, it calls the algorithm CB-SMoT (line 8) to compute low speed clusters/substops for all stops *S,* generating substops. In a second step, substops are computed over the same set of stops *S* with the method DB-SMoT (line 9), which finds clusters/substops where the direction of the moving object has changed. Of course both methods require input parameters as *minDir*, *avg,* and *minTime*, but we consider this as part of the methods CB-SMOT and DB-SMoT that are well known in the literature. These parameters can also be obtained by the methods from the knowledge base.

If a list on non-empty set of *substops* is generated, then these substops will be analysed using a knowledge base, through the method named *executeInference* (line 11), that is explained in Listing 2. If a stop has no substops, the stop itself is inserted in the set of substops.

The inference method presented in Listing 2 receives as an input a set of substops and the knowledge base. The output is a set of contextualized stops/substops *C.* For each stop *s* in *substops* (line 9) the method recovers the duration of the stop (timeStop) and the speed and direction variation of the substop (lines 11 and 12), previously computed by the methods CB-SMOT and DB-SMOT. Then for each rule/row in the knowledge base (line 13) the speed and direction variation of the stop are then compared with the maximum speed and maximum direction variation stored in the rule (line 17). If the speed variation of the substop is lower or equal to the speed variation of the rule and the direction variation is lower or equal to the direction variation of the rule, than the time is tested.

If the stop duration is equal or greater to the minimal time defined in the rule (line 18), than the goal of this substop is found (line 29) in the knowledge base and a contextualized substop is inserted in the set of contextualized substops (line 20).

```
1 INPUT:
2  substops //set of substops
3  kBase //Knowledge Base
4
5 OUTPUT: C //Set of contextualized substops
6
7 METHOD:
8
9  FOR each stop s in substops DO
10   timeStop = endTime(s) - startTime(s); //stop duration
11   directionStop = getDirectionVariation(s);//average dir. of the stop
12   speedStop = getSpeedVariation(s); //average speed of the stop
13   FOR each rule r in kBase DO
14      maxDirectionOfRule = getMaxDirection(r);//min direction of this rule
15      maxSpeedOfRule = getMaxSpeed(r); //min speed of this rule
16      minTimeRule = getMinTime(r); //min time of this rule
17      IF (speedStop<=maxSpeedOfRule AND directionStop<=maxDirectionRule)
18           IF (timeStop >= minTimeRule)
19               s.addGoal(r.getGoal);//add the goal of rule r as goal of s
20               C -> C + s; //adds s to list of contextualized stops
21           ENDIF
22      ENDIF
23   END FOR
24  END FOR
25  END METHOD
```

Listing 2 – executeInference pseudo-code

In the following section we present some initial results obtained with the proposed method.

4. EXPERIMENTS AND EVALUATION

We performed experiments with two different datasets: a bird dataset and a dataset of pedestrians in a park, as explained in the following sections.

4.1 Bird dataset

A first experiment was performed over the ceconia bird's trajectory dataset. These data were collected during the migration of *Ciconias,* being most birds fitted with geographical positioning devices. The acquired data were transmitted to a group of researchers[1] who gave a name to each bird. The whole dataset has only 1886 records. As a first experiment, we chose the trajectory of the bird *Prinzesschen,* which has more points. From the total of 1886 records, this bird has 528 points.

Four stops were generated for this trajectory, and considered as input for the proposed method, as shown in Figure 4(1). For two stops (stop 2 and 3), one substop was generated, as shown in Figure 4(2). After the substops generation, the next step is to infer the objective of each substop. Using a fictitious knowledge base, shown in Figure 5, we discovered the goals of two substops, as shown in Figure 4(2). It is important to note that the information contained on this base was prepared without the aid of an expert in the area of birds, such as an ornithologist. However, assuming that such information will be defined by a domain expert, this does not affect the experimental results.

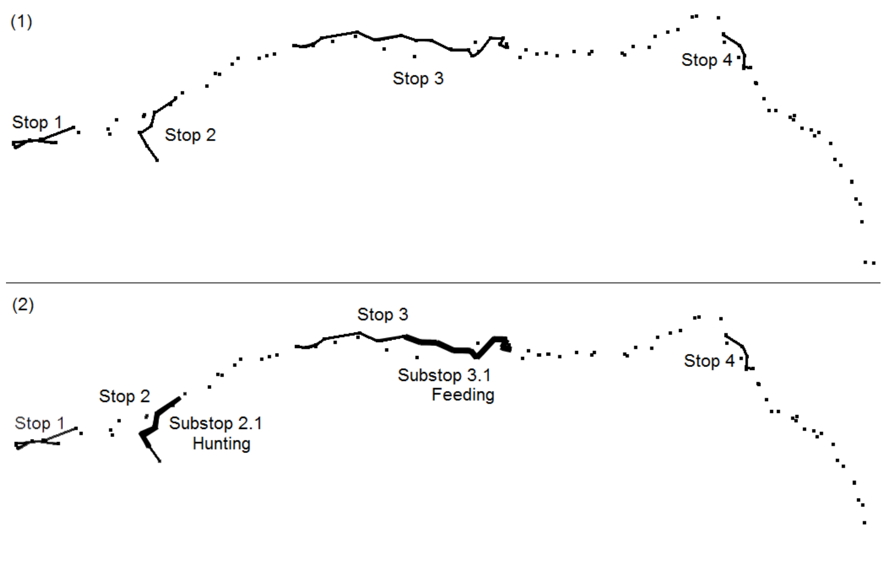


Figure 4 - Trajectory stops and their respective sub-stops

minTime	maxSpeed	maxDirection	goal
5 hours	0 km/h	0 degrees	Resting
2 hours	6,5 km/h	30 degrees	Feeding
3 hours	10 km/h	60 degrees	Hunting

Figure 5. Fictitious knowledge base

The bird trajectory dataset is not very interesting for evaluating the proposed method because there are only a few points, with long gaps in time. Therefore, we evaluated the method with another real dataset as shown in the following section.

[1] http://www.storchenhof-loburg.info, http://www.fr.ch/mhn/.

4.2 Pedestrian dataset

A second experiment was performed over a pedestrians dataset generated at a park in The Netherlands [21] . A set of people were equipped with a GPS device and each person was asked with the activity that he/she would do at the park. Among these activities, some were walking, running, walk the dog, picnic, etc. These data, differently from the birds dataset has the points very close in time, in an average of about every 10 seconds. Based on the metadata send with the data, we created a knowledge base with some activities that the pedestrians would perform in the park, as shown in Figure 6. Basically, what characterizes the behavior of the moving object is the speed variation and direction change. For instance, two activities have as minimal time 15 minutes (walking and cycling), but we suppose that if a person is walking the maximal speed is 7km/h, while cycling would be at 36 km/h. The knowledge base is fictitious, but the objective is to show that the method is able to give more meaning to trajectories considering prior knowledge.

minTime (min)	maxSpeed (km/h)	maxDirection	Goal
15	7	50	walking
15	36	80	cycling
5	15	90	dog letting
5	4	45	photo
60	2	20	picnic
30	2	20	relaxing
30	20	50	runnig

Figure 6. Knowledge base of possible activities in a park

This experiment was performed over 246 trajectories. We generated *stops* with the method CB-SMoT, in order to give the input of our method. A total of 148 stops were generated, considering 30 minutes as the minimal time and the speed should be half of the average speed of the trajectory. So the input of our method were 148 stops. Among these stops, a total of 494 substops were generated using the knowledge base parameters to generate them. Among the 494 substops, 160 had their objective inferred.

Because of space limitations, we show the result of only two contextualized trajectories, which are shown in Figures 7 and 8. Figure 7 left shows an example of a trajectory that had one big stop as input and 5 substops were generated (Figure 7 right). From the 5 substops, 4 have been contextualized, in *taking pictures* (*photo*) *and walking*.

Figure 8 left shows an experiment on a single trajectory that also had one big stop as input and 3 small substops were generated with our method, among which one was contextualized (Figure 8 right). Additional experimental results are available at http://cin.ufpe.br/~bnm/lookingInsideStops.

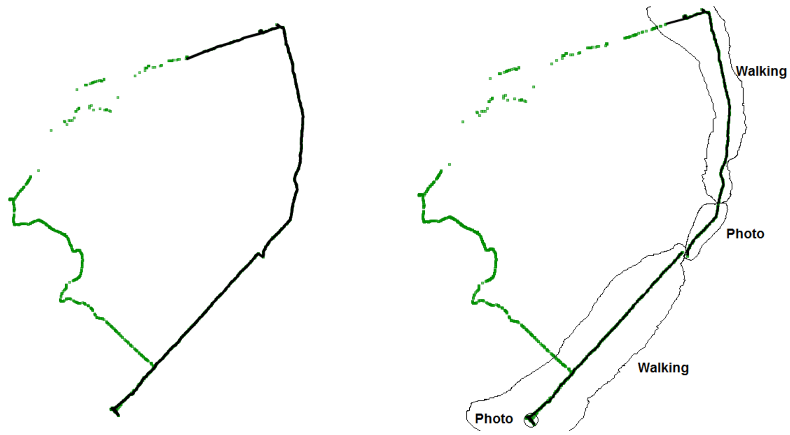


Figure 7. (left) Stops of one trajectory and (right) Contextualized substops of the same trajectory

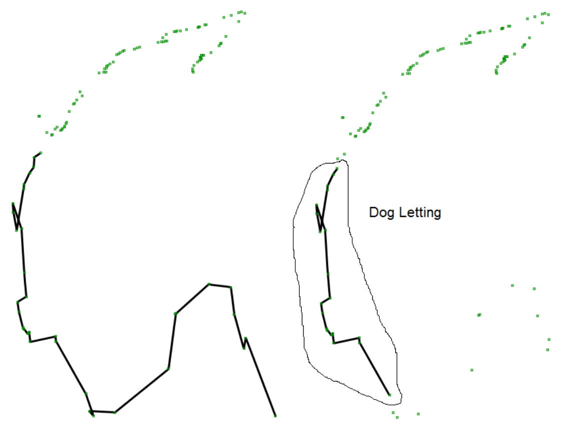


Figure 8. (left) Stops of one trajectory and (right) Contextualized substops of the same trajectory

4.3 Discussion

It is obvious that from the semantic point of view the results of the experiments would have to be evaluated by a specialist in the application domain in order to semantically validate them. The knowledge bases would also have to be build by a domain expert. However, the experiments presented in this paper have the objective to show the effectiveness of the method, that the proposed algorithm is able to infer the activities/goals of individual trajectories through the analysis of the behavior of the moving object and the use of domain information, what is novel in this research field.

So far there are no similar methods that infer the goal of a trajectory through the analysis of its direction and speed variation, and its interpretation using domain knowledge. Existing works use only the spatial intersection of stops with geographic information that for several applications is not available or may not help to infer the goal of trajectories for decision making processes.

5. CONCLUSIONS AND FUTURE WORKS

Several studies have focused on the semantic properties of trajectories. Most of them propose different objective measures to instantiate the well known model of stops and moves. The novelty of this paper is the analysis of speed and direction variation, for a certain time, and the use of context information stored in a knowledge base to infer new and more knowledge about important places of trajectories (stops). With this discovery we are able to infer the behavior of the moving object and understand the goals of his/her trajectory. Among other objectives, this method can be used for two main reasons: to discover the meaning of an *unknown stop* or to discover the activity of the moving object inside a known stop. The possibility of using a knowledge base on the analysis of stopping points of a moving object brings great benefits to trajectory data mining. Semantic trajectories can be used to discover common group behavior patterns.

The main drawback of this work is that for one stop we may infer more than one goal. For instance, if a moving object is at a restaurant in shopping mall or at a cinema, both stops have speed zero and direction variation zero. Future ongoing works include the discovery, based on a set of trajectories, the probability of stop goals that satisfy more than on rule.

ACKNOWLEDGMENTS

The authors thank to MODAP project, CNR (Italy) CNPq, FAPESC (CP 005/2009) and UFSC that partially provided support for this research. Thanks also to Wageninghen for the trajectory data

6. REFERENCES

[1] Alvares, L. O., Bogorny V., Kuijpers B., Macedo, J. A. F, Moelans B., Vaisman A. 2007. A Model for Enriching Trajectories with Semantic Geographical Information. In: ACM-IS, Seattle, USA, New York, NY, USA: ACM Press, 162-169.

[2] Backus, J. 1978. Can Programming be liberated from the von Neumann style: a functional style and its algebra of programs. Communications of the ACM, Vol. 21, No. 8.

[3] Baglioni M., Macêdo J. A. F., Renso C., Trasarti R. e Wachowicz M. 2009. Towards Semantic Interpretation of Movement Behavior, AGILE, pp. 271-288, 2009.

[4] Brézillon, P. 1999. Context in Artificial Intelligence: IA Survey of the Literature. Computer&Artificial Intelligence, v. 18, pp. 321-340.

[5] Bulcão Neto, R. F. Um processo de software e um modelo ontológico para apoio ao desenvolvimento de aplicações sensíveis a contexto. Tese de Doutorado, Instituto de Ciências Matemáticas e de Computação – ICMC-USP, 2006.

[6] Dey, A. K., Abowd, G. D. 2001. A Conceptual Framework and a Toolkit for Supporting the Rapid Prototyping of Context-Aware Applications. Human-Computer Interaction, v. 16, n. 2-4, pp. 97-166

[7] Ester, M., *et al*., Kriegel H. P., Sander J., Xu X. 1996. A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise. *In:* E. Simoudis, J. Han and U. M. Fayyad, eds. *Second International Conference on Knowledge Discovery and Data Mining* AAAI Press, 226-231.

[8] Giannotti, F., Nanni, M., Pinelli, F., Pedreschi, D. 2007. Trajectory pattern mining. In Berkhin, P., Caruana, R., Wu, X., eds.: KDD, ACM (2007) 330–339

[9] Guc B., May M., Sygin Y., Körner C. 2008. Semantic Annotation of GPS Trajectories. 11th AGILE International Conference on Geographic Information Science (AGILE), 2008.

[10] Hazin, F. H. V. A pesca na Zona Econmica Exclusiva -ZEE: sua importância para o Brasil. Revista Brasileira de Engenharia de Pesca, 1, 10-18. 2006

[11] Kuijpers, B., Moelans, B., de Weghe, N.V. 2006. Qualitative polyline similarity testing with applications to query-by-sketch, indexing and classification. (2006) 11–18

[12] Laube P., Imfeld S. and Weibel R. 2005. Discovering relative motion patterns in groups of moving point objects. International Journal of Geographical Information Science, vol. 19, no. 6, pp. 639–668, 2005.

[13] Manso, J. A. ; Times, V. C. ; Oliveira, G. ; Alvares, Luis Otavio ; Bogorny, V. DB-SMoT: A Direction-Based Spatio-Temporal Clustering Method. In: IEEE International Conference on Intelligent Systems (IS), 2010, Londres. Proceedings of the IEEE International Conference on Intelligent Systems.

[14] Moreira A., Santos M. Y. From GPS tracks to context: Inference of high-level context information through spatial clustering. In: International Conference and Exhibition on Geographic Information, 2, Estoril - "Proceedings". ISBN 972-97367-5-8. 2005;

[15] Nanni, M., Pedreschi, D. 2006. Time-focused clustering of trajectories of moving objects. Journal of Intelligent Information Systems 27(3) (2006) 267–289

[16] Nunes, V. T., Santoro, F. M., Borges, M. R. S. Capturing Context about Group Design Processes. In: Proc. of the 11th International Conference on Computer Supported Cooperative Work in Design (CSCWD'07), pp. 18-23, Melbourne, Australia, 2007.

[17] Palma, A. T., Bogorny V., Kuijpers B., Alvares L. O. 2008. A Clustering-based approach for discovering interesting places intrajectories. In: ACMSAC, Fortaleza, Brazil NewYork, NY, USA: ACM Press, 863-868;

[18] Pelekis, N., Kopanakis, I., Ntoutsi, I., Marketos, G., Theodoridis, Y. 2007. Mining trajectory databases via a suite of distance operators. In: ICDE Workshops, IEEE Computer Society(2007) 575–584

[19] PostgreSQL. 2010. Open-source Object-Relational Database Management System – web site: http://www.postgresql.org/ . Visited at 14/03/2010.

[20] Spaccapietra, S., Parent C., Damiani M. L., Macedo J. A., Porto F., Vangenot C. 2008. A Conceptual View on Trajectories. 2008. Data and Knowledge Engineering (DKE)

[21] van Marwijk, R., Pitt, D. G. Where Dutch Recreationists Walk: Path Design, Physical Features and Walker Usage. In: Fourth International Conference on Monitoring and Management of Visitor Flows in Recreational and Protected Areas Management for Protection and Sustainable Development Montecatini Terme, Tuscany/Italy , 14/19 October 2008

[22] Vieira, V., Mangan, M. A. S., Werner, C. M. L., Mattoso, M. L. Q. Ariane: An Awareness Mechanism for Shared Databases. In: Proc of the X International Workshop on Groupware (CRIWG'04), pp. 92-104, San Carlos, Costa Rica, 2004.

[23] Vieira. V., Tedesco P., Salgado A. C. 2009. Modelos de Processos para o Desenvolvimento de Sistemas Sensíveis ao Contexto. Texto JAI, 2009

[24] Wolf J., Guensler R. e Bach-man W. 2001. Elimination of the travel diary: An experiment to derive trip purpose from GPS travel data. Notes from Transportation Research Board, 80th annual meeting, Washington,D.C., 2001

Proceedings XI GEOINFO, Campos do Jordão, Brazil, Nov. 28 to Dec. 1, 2010, INPE, p. 21–32.

An Ontological Gazetteer for Geographic Information Retrieval

Ivre Marjorie R. Machado[1], Rafael Odon de Alencar[1], Roberto de Oliveira Campos Junior[2,3], Clodoveu A. Davis Junior[1]

[1]Dep. de Ciência da Computação – Universidade Federal de Minas Gerais (UFMG)
Av. Antônio Carlos 6627 – ICEx – 31270-010 – Belo Horizonte – MG – Brasil

[2]Prog. de Pós-graduação em Engenharia Elétrica – Pontifícia Universidade Católica de Minas Gerais (Puc Minas) – Belo Horizonte – MG – Brasil

[3]Serviço Federal de Processamento de Dados (SERPRO)

{ivre,odon,clodoveu}@dcc.ufmg.br,roberto.oliveira@ieee.org

Abstract. *The volume of spatial information on the Web grows daily, added to that, the problems to recognize references to spatial relationships and to deal with ambiguous names. This article presents a gazetteer, which has a structure different from conventional gazetteers. The ontological gazetteer will not only identify the names of places, but also record concepts and terms related to a place, as in an ontology in which concepts are the main places and features. A case study showed good results for detection of place names and inference implied by news Web sites based on content of ontological gazetteer.*

1. Introduction

The volume of information currently available on the Web is very large, and grows daily. Retrieving such information requires systems that are capable of understanding the needs of the users, locating relevant documents, and present such documents under a relevance ranking. This is the task associated to information retrieval systems, which also deal with issues regarding indexing and storage of documents.

Users manifest their retrieval needs in many ways, but mostly in the form of sets of keywords submitted to a search engine. Previous work (Sanderson and Kohler 2004; Wang, Wang et al. 2005; Delboni, Borges et al. 2007; Backstrom, Kleinberg et al. 2008) has shown that a significant portion of the queries involve terms or expressions with spatial meaning, including place names and natural language expressions that denote positioning. However, getting significant results out of such queries is often difficult, because geographically relevant keywords sometimes are not understood as such by information retrieval systems. Geographic information retrieval techniques have important limitations in the recognition of spatial references and in dealing with ambiguous names (e.g., "São Paulo" can be a Brazilian state, a city, or a soccer team). There are also difficulties in the retrieval of information constrained to a geographic context. For instance, if a set of places associated to a document can be determined, it would be possible to modify the document's position in a ranking, or to filter out documents that refer to undesired locations. Recognizing a term as a possible reference to a place is usually done with the help of a gazetteer, dictionary of place names (Hill 2000).

Current gazetteers are available on the Web, and are based on very simple data structures, with just three components: the name of the place, its type (as defined in a feature type hierarchy), and its footprint (a simple pair of coordinates indicating its location). With this kind of structure, gazetteers present several limitations, which make them harder to employ in information retrieval problems. Furthermore, gazetteer contents usually do not include intra-urban place names, such as street names, neighborhoods, landmarks or tourist attractions, and there are no resources with which to record and use the spatial relationship among its elements, other than estimating their proximity based on footprints. Gazetteers are notoriously hard to maintain and to expand, and as a result their coverage is usually irregular: although some include urban details on U.S. or European cities, Brazilian places are not as well covered. Some of these difficulties can be overcome by using geocoding services such as the ones available in the Google Maps API, which do not make the gazetteer entries explicit, but are able to supply a pair of coordinates that corresponds to a textual description.

Regardless of the limited structure, several Web-based geographic applications use information from gazetteers, as demonstrated by Goodchild and Hill (2008). We believe that gazetteers, as sources of organized information on places, can decisively contribute with the solution of geographic information retrieval problems. Therefore, this paper presents a novel conceptual schema for an enhanced gazetteer, in which the semantic connections among places can be recorded along with the usual topological connections, in order to support geographic information retrieval tasks. Such an *ontological gazetteer*, as proposed here, can go beyond the recognition of geographic names, allowing a more complete view of each place's semantic significance, expressed using its connections to other places and to terms and expressions that characterize it. Using this enhanced structure, we expect to support research initiatives towards solving difficult problems such as place name disambiguation, geographic text classification, and geographic context recognition (Silva, Martins et al. 2004; Wang, Xie et al. 2005; Adriani and Paramita 2007; Overell and Ruger 2007).

The remainder of this paper is organized as follows. Section 2 presents related work. Section 3 describes the conceptual database schema used to create the ontological gazetteer. Section 4 presents strategies for the application of the proposed gazetteer in the most important geographic information retrieval problems. Section 5 presents a case study in one of such problems. Finally, Section 6 presents our conclusions and a rather extensive list of future work.

2. Related work

Hill (2000) presents the basic elements of digital gazetteers: the place name (toponym), its type, and a footprint, which indicates its location. Such components are typical of conventional (i.e., the toponymical dictionaries usually found in atlases) gazetteers, and have been used as the basis for the development of the Alexandria Digital Library (ADL) Gazetteer. Since this pioneering initiative such basic structure has also been used in other Web-based gazetteer projects.

Uryupina (2003), Toral and Munoz (2006), and Popescu, Grefenstette et al. (2008) present proposals for populating and automatically maintaining gazetteers. These works extract data from the Wikipedia, which is a large knowledge base in different languages. Gouvêa, Loh et al (2008) propose a strategy for the identification of entities found in news texts, to be used in the development and updating of gazetteers. Alencar, Davis-Jr. et al. (2010) describe a strategy for classifying text into geographic categories through data extraction from Wikipedia to find evidence of place names in texts.

Lopez-Pellicer, Silva et al. (2010) present Geo-Net-PT 02, a geographic ontology of Portugal, an evolution of Geo-Net-PT 01 (Rodrigues, Chaves et al. 2006). This ontology has been developed using a vocabulary, called Geo-Net, proposed by the same research group. Geo-Net uses a conceptual schema to describe places, using their name, type, relationships and footprint. It uses URIs, RDF and OWL to describe, share and codify the ontology. The initial application of the ontology is the discovery of geographic characteristics based on an attribute of a place.

Several information retrieval tasks can be performed with the aid of gazetteers, such as named entity recognition, place name disambiguation, geotagging, document classification, and others. Amitay, Har'El et al. (2004) present Web-a-Where, a system that identifies geotags for Web pages with the support of a gazetteer. Souza, Davis-Jr et (2005) and Souza, Delboni et al. (2004) developed Locus, a geographic locator built around a gazetteer and based on a previously created ontology, OnLocus (Borges 2006; Borges, Laender et al. 2007). Overell and Ruger (2007) describe a model based on co-occurrence to solve the place name ambiguity problem, which uses a combination of heuristics and gazetteers.

Our work proposes changes in the structure of the gazetteer and demonstrate that it can be used in problems of Geographic Information Retrieval. We have used several sources to populate the gazetteer, including data from Wikipedia. This structure is an ontological construct, which enables the understanding and expansion semantics between entities. The *ontological gazetteer* is presented in the next section.

3. Ontological Gazetteer

A gazetteer is a geospatial dictionary of place names, also known as a toponymical dictionary. Current digital versions are analogous to the toponymical indices usually found in printed atlases. While in the atlas each place name is associated to a generic type, a map number and a grid coordinate, in digital gazetteers a pair of geographic coordinates (lat-long) is used as a footprint. There are also known variants of each place name, such as abbreviations and popular names, as well as language-specific versions.

The place type comes from a previously determined hierarchy, which varies among gazetteers. For instance, the Alexandria Digital Library Gazetteer (ADL)[1] (Hill 2000) has a top-level definition of feature types that includes *administrative areas*, *hydrographic features*, *land parcels*, *manmade features*, *physiographic features*, and *regions*. These in turn get more specialized, up to three more levels. On the other hand, the GeoNames[2] gazetteer defines feature codes, with the first level consisting of nine classes, with a single level of further specialization. Other digital gazetteers include TGN[3] (Getty Thesaurus of Geographic Names) and GKB[4] (Global Knowledge Base).

[1] http://www.alexandria.ucsb.edu/
[2] http://www.geonames.org
[3] http://www.getty.edu/research/conducting_research/vocabularies/tgn/
[4] http://xldb.fc.ul.pt/wiki/Grease

Previous works (Fu, Jones et al. 2005; Souza, Davis-Jr et al. 2005; Borges 2006; Borges, Laender et al. 2007) point out some of the limitations of current online gazetteers, seen here as possible support tools for geographic information retrieval. The main limitations are (1) the limited spatial representation (a point or a rectangle) and absence of support for spatial relationships, (2) the absence of support for semantically complete, but geographically imprecise locations, such as "south of France" or "upstate New York", (3) the lack of intra-urban detail, including places often mentioned in natural language text and possibly know by non-residents, such as monuments or tourist attractions. Furthermore, the level of detail available in Web-based gazetteers seems to be lower in developing countries, such as Brazil (Gouvêa, Loh et al. 2008).

Souza, Davis-Jr et al. (2005) developed Locus, a geographic locator that uses a gazetteer as its main component. Results obtained from designing Locus suggested the creation of an ontology of places, named *OnLocus* (Borges 2006; Borges, Laender et al. 2007). OnLocus describes spatial and semantic relationships between locations, distinguishing between the actual place and its name, a *place descriptor*. However, OnLocus was designed as part of an effort to extract geographic knowledge from Web pages, so it focused on indirect references to places, such as postal codes and telephone area codes (Borges, Laender et al. 2007). In turn, the good performance of such indirect references in geographic information retrieval tasks suggested that a gazetteer might be much more helpful if it could record the various types of relationships that exist between places, going beyond the topology of geographic objects and allowing the inclusion of other types of semantic relationships. In order to implement this kind of semantically richer relationships, the gazetteer's design needs to include the flexible structure often found in ontology creation tools, such as Protégé[5], thus becoming what we call an *ontological gazetteer*, or *ontogazetteer*.

In order to accomplish this enlarged role, the ontogazetteer must be able to record various types of relationships between places, including spatial (proximity), topological (adjacency, containment), hierarchical (territorial subdivisions) and semantical, also recording the motivation behind each relationship. It should be possible to infer relationships between places, using the semantic properties of existing relationships. We also propose to expand the spatial representation of each place to a complete geometry, so that spatial and topological relationships can be established as needed, or recalculated as a result of data maintenance. The ontogazetteer also must be able to record alternative names to a place as synonyms, also adapting the notions of hyponymy and hyperonymy for territorial subdivisions.

Another proposed enhancement for the ontogazetteer is the association of natural-language terms and expressions to each place. The idea is to improve the available information resources for performing typical geographic information retrieval tasks, such as disambiguation and geographic context recognition. An experimental procedure for obtaining these terms has been presented in (Alencar, Davis-Jr. et al. 2010), along with a classification procedure.

[5] http://protege.stanford.edu/

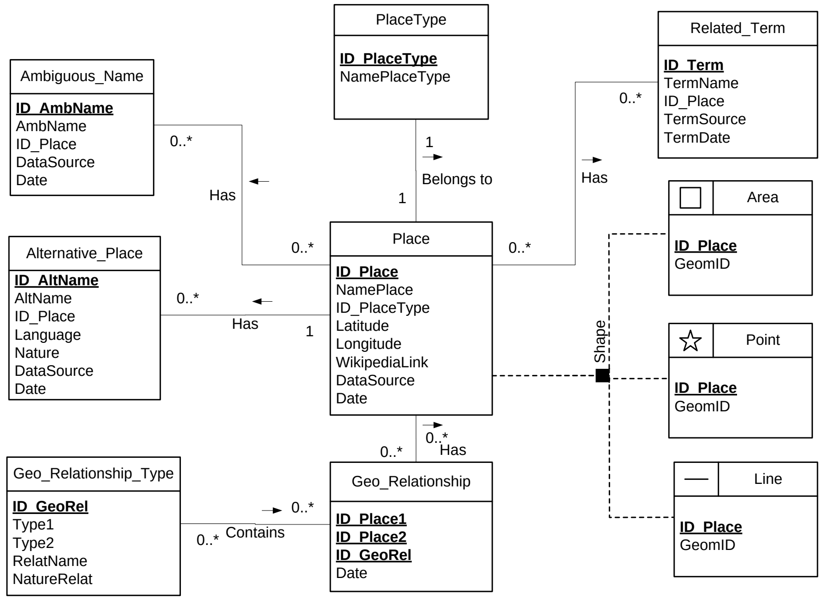


Figure 1. Gazetteer conceptual schema

Figure 1 presents the OMT-G (Borges, Davis-Jr et al. 2001) schema proposed for the ontological gazetteer. The schema represents all place names using the `Place` class. The `Alternative_Place` class maintains alternative names, abbreviations, acronyms, popular names and other variations. Each place belongs to a `PlaceType`; we initially based our place type definitions on the feature code thesaurus from ADL. Alternative names, abbreviations, acronyms and others are kept in the `AlternativePlace` class. Relationships between places are maintained by the `GeoRelationship` class; notice that two places can have various relationships between them, each one of a different `Geo_Relationship_Type`. This feature allows the gazetteer to record and use a number of different geographic and semantic connections between places. In order to support disambiguation, there is a class that keeps known ambiguous place names (`Ambiguous_Name`). Also for disambiguation and to support other geographic information retrieval applications, there is a class that stores lists of terms related to the place (`Related_Term`). One possible source for such names is the Wikipedia (Alencar, Davis-Jr. et al. 2010) (that is also the reason for keeping an attribute in the `Place` class to store the URL of its Wikipedia entry). For instance, the `Related_Term` class can contain the term "acarajé" associated to the place "Bahia". Finally, each place can have one or more than one geographic representation, as a point, a line or a polygon (Davis-Jr and Laender 1999). Places represented by more complex geometries will also have a point representation, as in current gazetteers.

The class diagram in Figure 1 was detailed and mapped to a geographic database. It is currently being populated, using existing geographic data and other gazetteers as primary sources. From the OnLocus ontology, we derived several types of geographic relationships between places. Special procedures and triggers have been created for each these relationship types, so that the relationships could be materialized in the `Geo_Relationship` table and kept up-to-date whenever new places are added. Next section describes how the features of the ontological gazetteer can be used to fulfill geographic information retrieval tasks.

4. Applications for the Ontological Gazetteer

Information Retrieval (IR) has been the focus of much recent research, due to the explosive growth of the Web. Geographic Information Retrieval (GIR) expands and focuses IR techniques on problems such as the detection of references to places, or to the association of locations to Web documents. Some of these problems have been highlighted by Jones and Purves (Jones and Purves 2008) in a research agenda for GIR:

1. Detection of geographic references in the form of place names;
2. Disambiguation of place names;
3. Geographic interpretation of vague place names, such as “South of France”;
4. Document indexing according to the geographic context and non-spatial content;
5. Geographic relevance ranking of documents;
6. Search interface improvement;
7. Evaluation of methods for comparing GIR systems and techniques.

Gazetteers can be used as components of the solution for most of these problems. We argue that our proposed ontogazetteer can provide a better support for solving these and other GIR problems, since it goes beyond a simple georeferenced list of place names and introduces richer geographic and semantic relationships, related terms, and a record of ambiguous place names. In the following subsections, we will describe more specifically how the ontogazetteer can contribute in many different GIR problems.

4.1. Detection of geographic references

Geoparsing is the process of analyzing a text in order to identify references to places, in the form of place names and other space-related terms (Jones and Purves 2008). *Geotagging*, on the other hand is the process of identifying geographic entities mentioned directly or indirectly in the text and creating tags that allow the document to be linked to a location or set of locations (Amitay, Har'El et al. 2004; Teitler, Lieberman et al. 2008). Both geoparsing and geotagging require the recognition of geographic references found in text, and if this task is fulfilled adequately, the geographic context of the document can be established.

The ontogazetteer maintains lists of official and alternative place names that facilitate the identification of candidate names contained in the text. Distinguishing between actual references to places and other uses of the same words can be done by determining spatial relationships among candidate names. Since the ontogazetteer also maintains information on spatial hierarchies and adjacent places, it is possible to infer, from the co-occurrence of related places, which candidate names should be disconsidered. Furthermore, the actual context of the document can reside in some higher level of the spatial hierarchy; e.g., a text that mentions several cities in a state actually refers to the state itself.

Notice that the proposed structure of the ontogazetteer, in which relationships are materialized beforehand, was conceived as such in an effort to expedite relationship queries, by avoiding spatial queries during a GIR-related process. Therefore, applications can decide on the types of relationships that are to be considered, and which entities are to be taken into consideration. Since the full geometric shape is available, more complete and refined analyses can be performed, either in specific cases or as an additional filter.

4.2. Place name disambiguation

Place names are frequently ambiguous. For instance, the name "São Paulo" exists in 6,522 different GeoNames records. According to Smith and Crane (2001), 92% of TGN's toponyms are ambiguous. Several different types of ambiguities have been described in previous research (Amitay, Har'El et al. 2004; Volz, Kleb et al. 2007).

When humans read a text, ambiguities are resolved using their previous knowledge and subtle hints found in the text itself, or in elements that surround it, such as the section of a newspaper in which the text appears. Place name disambiguation, also known as toponym resolution, tries to imitate these methods (Jones and Purves 2008). The ontogazetteer can help in this task by offering lists of ambiguously named places, alternative names and related places. These additional pieces of information can be used in heuristics designed to establish which one of the ambiguously named places is the most likely to be the one the text refers to. The list of related terms included in the ontogazetteer can contribute as well. If one or more of the candidate places has a weak relationship to other elements found in text (other place names, natural language terms), it can probably de disregarded.

4.3. Interpretation of vague place names

People often use vague or approximate references to places in natural language, as in "downtown" or "Northern Italy". In spite of the likely mention to a definite place, the geographic scope of such a reference is rough and imprecise (Jones and Purves 2008). Gazetteers usually do not include references to vague places, and the limited spatial representation keeps them from being located adequately. Using the complete geographic representation available in the ontogazetteer, it becomes possible to infer a subdivision of the place mentioned using clues provided by the associated natural language expressions. The usefulness and interpretation of space-related expressions for GIR has been demonstrated in previous work (Delboni, Borges et al. 2007).

4.4. Spatial and textual indexing

One of the techniques for indexing the contents of a text document is the creation of an inverted index file for the words contained in the document. This index provides, therefore, an association of each word to the list of documents that contain it. In the case of geographic references, this idea can be expanded using a list of places in addition to the list of words. The source for the list of places can naturally be a gazetteer (Jones and Purves 2008). After the identification of places related to each document, a spatial index can be generated, using positions (footprints) or minimum bounding boxes of the full geographic representation, so that documents can be retrieved using spatial relationships, such as proximity and containment.

4.5. Geographic relevance ranking

Ranking according to geographic relevance requires a measurement of the relative importance of a document for a given query. Usually, documents are selected according to the occurrence of the query terms, and ranked according to a measurement that takes into consideration the existing links to candidate documents. In the case of a geographic ranking, there must be an association of the query terms (or query region) to the places referred to by the document, and ranking needs to combine both geographic and keyword-based criteria (Jones and Purves 2008). Since the ontogazeteer keeps lists of relevant terms, a ranking strategy can determine how specific certain query terms are in relation to places, helping to narrow down the results and assigning more importance to documents in which both terms and places are related to the query. From footprints and geographic representations, proximity relationships can be determined, so that aspect can influence the ranking as well.

5. Case study

We exemplify the use of the proposed ontogazetteer with a case study. Consider Web news sources. Usually, news texts contain one or more locations related to the facts, as part of the news reporting technique. Therefore, in this case study we put together a collection of news texts, detect the occurrence of place names, and infer the geographic context of each of them. We are able to recognize both explicitly and implicitly mentioned places. The latter are those whose relationship to the facts in the text is implied by the contents, but which are not directly mentioned.


Jovem é encontrado morto em lote vago no Bairro Ouro Preto

Estado de Minas -
Publicação: 03/08/2010 08:45 Atualização: 03/08/2010 08:58

A polícia investiga um homicídio ocorrido no fim da noite de segunda-feira no Bairro Ouro Preto, Região da Pampulha. Jonny Moreira da Silva, de 23 anos, foi encontrado morto com dois tiros na cabeça em um lote vago na Rua Luiz Lopes. Segundo militares do 34º Batalhão, o rapaz estava com cinco pedras de crack.

A autoria e a motivação do crime são desconhecidas, mas a polícia suspeita de envolvimento com tráfico de drogas. O corpo do jovem foi levado para o Instituto Médico Legal (IML). O registro foi feito na Seccional Noroeste.

Figure 3. Example (source: Uai – August 3, 2010)

Figure 3 presents a sample news text (in Portuguese) from the *Uai Minas*[6] news source, published August 3, 2010. Two place names have been identified, "Ouro Preto" and "Pampulha". In this case, "Ouro Preto" is ambiguous: the text refers to a neighborhood in Belo Horizonte, not to the famous historical city; nevertheless, both are obtained as candidate places. Consulting gazetteer data, other places related to the ones explicitly mentioned can be obtained, including "Belo Horizonte" (city), "Belo Horizonte Metropolitan" (micro-region), "Ouro Preto" (micro-region), "Metalurgical and Campos das Vertentes" (macro-region), and "Minas Gerais" (state). The latter places are implicit in the text. The text also includes a reference to a street ("Rua Luiz Lopes"), but the gazetteer currently does not include street data.

[6] http://www.uai.com.br/htmls/app/noticia173/2010/08/03/noticia_minas,i=172053/index.shtml

5.1. Creation of the news texts collection

The collection of news texts was performed from June to August 2010 (Table 1). For each of the news sources, a collector was developed in order to extract and store its title and body text, using XPath. Only news about the state of Minas Gerais were collected, because most of the gazetteer data put together so far refer to this state. In order to ensure that, news were obtained from local- or state-related sections of the news sites.

Table 1. Web news sources

News Web	News Web Site	Local News Section Name	# Docs
Globominas	http://globominas.globo.com/	General News	139
O Tempo	http://www.otempo.com.br/	Latest News (cities)	75
Uai	http://www.uai.com.br/	Minas	71
Terra	http://www.terra.com.br	Latest News (Brasil)	11
		Total	**296**

5.2. Detection and inference of place names

After the news documents were collected, a pre-processing step removed stopwords (except for "de", "da(s)", "do(s)", which are quite common in Brazilian place names). Next, candidate names were extracted, using regular expressions that were designed to identify single or composite proper nouns.

The recognition of place names from the news documents was supported by the ontological gazetteer. A simple string matching was performed between candidate names and place names from the gazetteer, including alternative names. Instances from the `Geo_Relationship` class were used to infer implicit references to other places. This inference procedure identified places whose names did not appear in the text, but were related most of the explicitly mentioned names. Typically, names of places that are higher in a territorial hierarchy were found.

5.3. Experimental evaluation

In order to verify whether the place names recognized from the news documents (both explicitly and implicitly) were valid, a manual verification of each document was performed by a group of volunteers, composed by people with various backgrounds. A Web interface was developed[7], showing the text's title and body, along with a list of the implicit and explicit places, generated as described in the previous section. For each of these places, we asked the volunteers to determine the degree of its relationship to the news, using a scale with three levels: (0) unrelated, (1) slightly related, and (2) strongly related. Furthermore, the volunteer had the possibility of skipping the evaluation, if she did not know the place or if the determination could not be made from the existing information. We also included a text field so that volunteers could record observations or indicate difficulties and special cases. From the 296 news documents originally collected, we were able to find place names in 267, which were then used for this experiment. An average of 8,46 places were found per document.

[7] http://94.229.77.252/AvaliacaoCGN/

5.4. Experiment results

From the 267 news documents containing place names, 100% were verified by volunteers. Overall, 2,244 relationships between places and documents were evaluated, of which 72% were considered to be valid, and most of those were considered strong relationships (Table 2).

Table 2. Evaluation according to degree of relationship

Degree of relationship	Count	%
0- Unrelated	609	27,14
1- Slightly related	106	4,72
2- Strongly related	1.455	64,84
3- Indifferent	74	3,30
Total	**2.244**	**100,00**

Table 3. Evaluation according to type of place

Type	Total	Implicit refs	Avg.	Std.dev.
Macrorregion	184	171	0,36	0,75
Microrregion	298	129	0,98	0,94
River	2	0	1,00	1,41
Mesorregion	246	0	1,17	0,95
Municipality	446	35	1,50	0,85
State	341	233	1,61	0,75
Neighborhood	546	0	1,73	0,67
Highway	107	0	1,91	0,42

Table 3 shows the evaluation according to the type of place. The average and standard deviation columns refer to the degree of relationship, i.e., averages approaching 2 indicate consensus that there is a strong relationship between places of a type and the documents. We observed that places whose names are more widely known ranked better in the evaluation. Place types that belong to an administrative territorial hierarchy, such as macrorregion, mesorregion and microrregion were included mostly as implicit context, but people had difficulties recognizing their names. As an example, few volunteers apparently knew that the macrorregion in Minas Gerais to which Belo Horizonte belongs is called "Metalúrgica e Campos das Vertentes". Naturally, such names are hardly ever mentioned in journalistic texts.

We also noticed a high number of explicit mentions to neighborhoods, reinforcing the idea that gazetteers should cover intra-urban detail. Highways were also frequently cited, and from this fact we surmise that the relationships between places along a highway can be semantically very important. Notice also the high incidence of implicit references to the state, something that was expected, due to the fact that the sources of news were state-related sections of news sites. On the down side, the very small number of references to rivers indicates that the gazetteer's coverage of hydrography features is currently deficient.

6. Conclusions and Future Work

This paper proposed a new structure for gazetteers that seeks to diminish their limitations as components of geographic information retrieval systems. Our proposal uses ontology concepts to define a flexible way to establish and maintain semantically richer relationships between places, and adds resources for keeping alternative names and lists of place-related terms. Relationships go beyond the geographic or topologic ones, and can be used to create semantic connections between geographically unrelated places. The paper also described ways in which the semantically enhanced gazetteer can be used in typical geographic information retrieval tasks.

Naturally, the usefulness of the ontogazetteer is a direct function of the quality and comprehensiveness of its contents. Therefore, our first task the near future is to expand the gazetteer's contents as much as possible, using information already available in geographic databases. From geographic features found in databases, we can easily derive geographic and topologic relationships. Semantic relationships are being expanded initially considering indirect geographic relationships; e.g., two municipalities through which the same river runs are considered to be related, even though they are not adjacent to each other. Place-related terms are the focus of some parallel work in our group, using the Wikipedia as a knowledge base with promising results (Alencar, Davis-Jr. et al. 2010).

A case study implemented a GIR task, namely the identification of geographic context in news documents, and showed that the ontogazetteer can be a valuable resource for solving that problem. Furthermore, using relationships recorded in the ontogazetteer, we were able to infer the connection between many documents and places that are not explicitly mentioned in their text. Placename recognition achieved good results, mainly for more the types of places that are more usually found in news, such as municipality, neighborhood, highway and state.

Future work includes developing an extension of the case study with a broader base of documents. We also intend to develop a service-based interface to the gazetteer, so that remote applications can retrieve data and execute queries, without direct access to the gazetteer's database. Finally, the expansion of the gazetteer's contents, including related term lists, is our hardest but more important goal.

7. Acknowledgements

This work was partially supported by the Brazilian research funding agencies CNPq and FAPEMIG.

References

Adriani, M. and M. L. Paramita (2007). Identifying location in indonesian documents for geographic information retrieval. In Proc. of the 4th GIR'07. Lisbon, Portugal, ACM.

Alencar, R. O., C. A. Davis-Jr., et al. (2010). Geographical classification of documents using evidence from Wikipedia. In Proc. of the 6th Workshop on GIR. Zurich, Switzerland, ACM.

Amitay, E., N. Har'El, et al. (2004). Web-a-where: geotagging web content. In Proc. of the 27th SIGIR Conf. on Research and Development in Information Retrieval. Sheffield, UK, ACM.

Backstrom, L., J. Kleinberg, et al. (2008). Spatial variation in search engine queries. Proceeding of the 17th International Conference on WWW. Beijing, China, ACM.

Borges, K. A. V. (2006). Uso de uma Ontologia de Lugar Urbano para Reconhecimento e Extração de Evidências Geo-espaciais na Belo Horizonte, UFMG. Doutorado: 195.

Borges, K. A. V., C. A. Davis-Jr, et al. (2001). "OMT-G: An object-oriented data model for geographic applications." GeoInformatica 5(3): 221-260.

Borges, K. A. V., A. H. F. Laender, et al. (2007). Discovering geographic locations in web pages using urban addresses. In Proc. of the 4th Workshop on GIR. Lisbon, Portugal, ACM.

Davis-Jr, C. A. and A. H. F. Laender (1999). Multiple representations in GIS: materialization through map generalization, geometric, and spatial analysis operations. In Proc. of the 7th ACM GIS. Kansas City, Missouri, USA.

Delboni, T. M., K. A. Borges, et al. (2007). "Semantic expansion of geographic web queries based on natural language positioning expressions." Transactions in GIS 3: 377-397.

Fu, G., C. B. Jones, et al. (2005). Ontology-based Spatial Query Expansion in Information Retrieval Lecture Notes in Computer Science, On the Move to Meaningful Internet Systems: ODBASE.

Goodchild, M. F. and L. L. Hill (2008). "Introduction to digital gazetteer research." Int. J. Geogr. Inf. Sci. 22(10): 1039-1044.

Gouvêa, C., S. Loh, et al. (2008). Discovering Location Indicators of Toponyms from News to Improve Gazetteer-Based Geo-Referencing. GEOINFO Rio de Janeiro, RJ.

Hill, L. L. (2000). Core Elements of Digital Gazetteers: Placenames, Categories, and Footprints. Proceedings of the 4th ECDL, Springer-Verlag.

Jones, C. B. and R. S. Purves (2008). "Geographical information retrieval." Int. J. Geogr. Inf. Sci. 22(3): 219-228.

Lopez-Pellicer, F. J., M. J. Silva, et al. (2010). Linkable Geographic Ontologies. GIR'10, Zurich, Switzerland.

Overell, S. E. and S. Ruger (2007). Geographic co-occurrence as a tool for gir. Proceedings of the 4th ACM Workshop on GIR. Lisbon, Portugal, ACM.

Popescu, A., G. Grefenstette, et al. (2008). Gazetiki: automatic creation of a geographical gazetteer. Proc. of the 8th ACM/IEEE-CS JCDL. Pittsburgh, PA, USA.

Rodrigues, C., M. Chaves, et al. (2006). Uma Representação Ontológica da Geografia Física de Portugal. IX Encontro de Utilizadores de Informação Geográfica, ESIG - Oeiras, Portugal.

Sanderson, M. and J. Kohler (2004). Analyzing geographic queries. GIR'04. Sheffield, UK.

Silva, M. J., B. Martins, et al. (2004). Adding Geographic Scopes to Web Resources. SIGIR'04 - Workshop on GIR, Sheffield, UK, ACM.

Smith, D. A. and G. Crane (2001). Disambiguating Geographic Names in a Historical Digital Library. Proceedings of the 5th ECDL, Springer-Verlag.

Souza, L. A., C. A. Davis-Jr, et al. (2005). The Role of Gazetteers in Geographic Knowledge Discovery on the Web. Proceedings of the 3th LA-Web Congress, IEEE Computer Society.

Souza, L. A., T. M. Delboni, et al. (2004). Locus: Um Localizador Espacial Urbano. VI GEOINFO, 22-24, Campos do Jordão, SP, Brazil.

Teitler, B. E., M. D. Lieberman, et al. (2008). NewsStand: a new view on news. Proceedings of the 16th ACM SIGSPATIAL international conference on Advances in geographic information systems. Irvine, California, ACM.

Toral, A. and R. Munoz (2006). A proposal to automatically build and maintain gazetteers for Named Entity Recognition by using Wikipedia. EACL 2006.

Uryupina, O. (2003). Semi-supervised learning of geographical gazetteers from the internet. In Proc. of the HLT-NAACL 2003 Workshop on Analysis of geographic references (1) ACL.

Volz, R., J. Kleb, et al. (2007). Towards ontology-based disambiguation of geographical identifiers. WWW2007. Banff, Canada.

Wang, C., X. Xie, et al. (2005). Web resource geographic location classification and detection. Special interest tracks and posters of the 14th in WWW. Chiba, Japan.

Wang, L., C. Wang, et al. (2005). Detecting dominant locations from search queries. Proceedings of the 28th annual international SIGIR conference on Research and Development in Information Retrieval. Salvador, Brazil, ACM.

Proceedings XI GEOINFO, Campos do Jordão, Brazil, Nov. 28 to Dec. 1, 2010, INPE, p. 33–44.

A Geographic Annotation Service for Biodiversity Systems

Fabiana B. Gil[1,2], Nádia P. Kozievitch[2], Ricardo da S. Torres[2]

[1] CPqD Foundation
Rod. Campinas - Mogi-Mirim km 118.5 Campinas, SP, Brazil

[2]Institute of Computing – University of Campinas
Av. Albert Einstein, 1251 Campinas, SP, Brazil

fbellette@gmail.com, nadiapk@ic.unicamp.br, rtorres@ic.unicamp.br

Abstract. *Biodiversity studies are often based on the use of data associated with field observations. These data are usually associated with a geographic location. Most of existing biodiversity information systems provides support for storing and querying geographic data. Annotation services, in general, are not supported. This paper presents an annotation Web service to correlate biodiversity data and geographic information. We use superimposed information concepts for constructing a Web service for annotating vector geographic data. The Web service specification includes the definition of a generic API for handling annotations and the definition of a data model for storing them. The solution was validated through the implementation of a prototype for the biodiversity area considering a potential usage scenario.*

1. Introduction

The term biodiversity – or biological diversity – describes the richness and variety of biological organisms in a given habitat. Some known issues in biodiversity are irreversible loss of species, loss of environmental services (production of oxygen by plants, the hydrological balance, soil fertility, and climate balance), and biopiracy.

Computer Science can be a great allied of Biologists, providing them with tools to analyze and report findings on species and their behaviors. Some challenges for the Biodiversity Information Systems (BISs) are [Torres et al. 2006]: (i) handle large volumes of information, (ii) integrate information from different sources and formats (heterogeneity), (iii) manipulate data and images, (iv) manipulate geospatial information reference.

In this paper we address the forth challenge. Biodiversity studies are often based on the use of data associated with field observations, later matched with geographic locations. Most of existing biodiversity information systems provides support for storing and querying geographic data. Annotation services, in general, are not supported.

Annotation has been recognized as one of the most important services in digital library systems to foster the cooperation among users and the integration of heterogeneous information resources [Agosti and Ferro 2008]. In this paper, we describe a new geographic data annotation Web service that can be easily integrated with other applications. The specification and implementation of the proposed Web service relies on two main contributions: (a) the proposal of a data model based on superimposed information [Maier and Delcambre 1999a] to manage geographic annotations; and (b) the definition of a generic API to manipulate annotations.

The rest of this paper is organized as follows: section 2 briefly presents the related work; section 3 describes the proposed geographic data annotation service; section 4 describes the use of the proposed annotation service in a Biodiversity Information System; and finally section 5 presents conclusions and summarizes future work.

2. Related Work

2.1. SuperImposed Information

Superimposed information (SI) [Maier and Delcambre 1999a] is the new information (or new interpretation) associated with the existing information or base (as for example, an annotation, term or keyword, a comment about a base document). In general, SI is not structured, does not have a pre-defined schema, and allows different types of SI under the same base.

Superimposed information has been used in several applications, named *Superimposed applications (SAs)* [Maier and Delcambre 1999b]. They allow the manipulation and creation of superimposed information, typically to highlight, annotate, draw, select, organize, connect, or reuse the information. Examples include SLIMPad [Delcambre et al. 2001], and RIDPad, Schematics Browser, SIMPEL, and Mash-o-matic from [SI]. None of these SAs, however, support interactions with geographic data.

The research described in this paper differs from related research in the sense that it takes advantage of superimposed information concepts to provide mechanisms for geographic data annotation in biodiversity applications. Furthermore, the use of a Web service implementation ensures appropriate encapsulation of data and procedures, which allows reuse of the developed component in other Digital Library (DL) initiatives involving geographic data annotation.

There are some other DL initiatives for the biodiversity domain. One example concerns floristic digital libraries (FDLs) [Sanchez et al. 1999]. FDLs use an agent-based infrastructure to manage botanical information about taxonomic keys, distribution maps, illustrations, and treatments (morphological descriptions). Another example is the Taiwanese digital museum of butterflies [Zhu et al. 2000]. This digital library offers a set of XML-based modules to manage data on butterflies. Geographic data annotation services, however, are not supported in these systems.

More recent DL-oriented initiatives include SuperIDR [SI], Ecopod [Yu et al. 2006] and BDiG-PELD [Barros et al. 2008]. The *Superimposed Image Description and Retrieval Tool* (SuperIDR) [SI] is designed to work with Tablet PC. With SuperIDR, users can navigate through the collection by species or taxonomic organization. Users can perform queries using part of an image or a textual description and recover the original content. Both systems work with annotation, but they still do not support geographic annotation.

Ecopod [Yu et al. 2006] aimed to replace the paper field guides with a mobile computing platform for identifying plants and animals in the field. In this tool, geographic information is used as an evidence in the species identification process. For example, some species are observed more frequently in specific regions. Annotation facilities, however, are not supported. Barros et al. [Barros et al. 2008] have proposed a georeferenced digital library, named BDiG-PELD that integrates network of ecological sites using OAI-PMH protocol and the ODL (Open Digital Library) framework. Even though geo-referencing facilities are used to determine species location, annotation services are not provided.

There are also examples of applications which work with geographic data annotation[Macário et al. 2009]. Besides having several standards for services and implementations, these applications still have lack of support for many features, like annotation of objects that are result of a query.

2.2. Geographic Data

The term spatial data denotes any type of data that describe phenomena associated with some spatial dimension. Geo-referenced data belong to a particular class of spatial data, describing events, objects and phenomena on the globe associated with their location on the surface, at a certain moment or period of time [Câmara et al. 1996].

2.2.1. Data Models

Within the context of applications which deals with geographic information, the real world is often modeled under two complementary views: the field-based model and object-based model. The field-based model (usually implemented according to the structure called *raster*) deals with a continuous surface on which the geographic phenomena varies according to different distributions (like atmospheric pressure). Each layer corresponds to a different theme (like vegetation, soil) [Gomes Jr 2007].

The object-based model (usually implemented by using vector representation) represents the world as an area occupied by identifiable objects that have their own geometry and characteristics. In the object-based model, a geographic object is typically represented in the vector format, that is its geometry is described using points, lines, and polygons. Lines are formed by sequences of points and polygons (open or closed), by a sequence of lines [Gomes Jr 2007].

The current version of the proposed annotation service is focused on vector data.

2.3. Spatial Operators

Spatial operators are used in our Web service to define which geographic objects (or parts of them) are being annotated. Several spatial operators have been proposed to process geographic objects based on their relationships. These relationships can be divided into three categories: metric, topological, and directional [Güting 1994]. Examples of topological relationships include *equals*, *disjoint*, *intersects*, *touches*, *crosses*, *overlaps*, and *contains*. Directional relationships include, for example, *above* and *north_of*. Metric relationships include *distance* and *area*.

Spatial operators are used in our annotation service to define regions which will be later associated with annotations.

3. Web Service for Annotating Geographic Data

This section describes the proposed Web service. First the data model is described. It includes annotation and geographic concepts, followed by the operations of the proposed

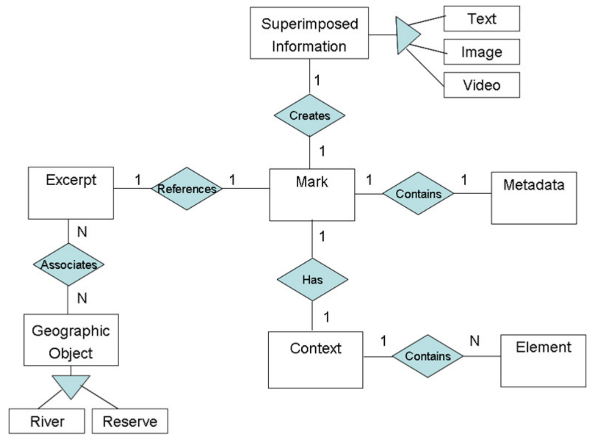


Figure 1. Proposed entity-relationship diagram.

annotation Web service.

3.1. Data Model

The data model is based on SI concepts, and the entity-relationship diagram diagram is presented in Figure 1. The entities are represented by rectangles; relationships are represented by diamonds; lines represent the connections between the entities and relationships; and the cardinality of the relationship is showed at the side of the diamonds.

The entities *Excerpt*, *Mark*, *SuperimposedInformation*, *Context*, and *Element* are inspired by the SPARCE model [SI].

When an annotation is inserted, a mark is created. A mark is used to link the superimposed information with the base information [Archer et al. 2008]. Different types of marks can be created when dealing with maps. Figure 2 illustrates four examples of marks associated with geographic data. The mark showed in Figure 2(a) refers to a rectangular region that includes several geographic objects represented as lines and points. In Figure 2(b), the mark includes parts of Brazilian states (i.e., parts of polygons). Figure 2(c) illustrates the definition of a mark associated with the result of applying a spatial operator (*intersection*) on two regions defined by the coverage of two antennas (i.e., intersection of two polygons). Finally, Figure 2(d) illustrates how to define a new mark associated with the intersection of the regions defined by two other marks.

An *Excerpt* represents the content of a marked region. In the case of vector data, the content may be a point, a line (or part of a line), a polygon (or part of a polygon), or sets of objects with different geometries. Figures 3 and 4 illustrate how excerpts are created, given defined marks.

Context entity defines a set of *elements* which provides contextual information associated with a mark. In the case of textual base document, context data may include, for example, information about font type, font size, or page number. For geographic data, context may involve information related to geographic coordinate system or visual properties (e.g., color, size, length) used for visualization purposes.

Metadata attributes relies on *Dublin Core Metadata Initiative* elements (at

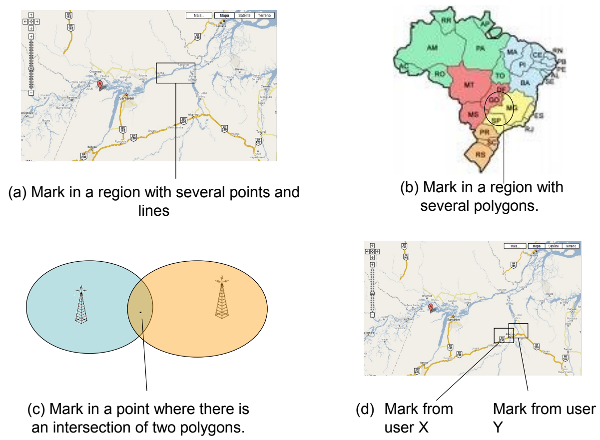


Figure 2. Examples of marks.

http://dublincore.org). The metadata includes information about the annotation creator, language, annotation date, and rights held in and over the geographic data being annotated.

Superimposed Information refers to a new information associated with the information base (vector geographic data). In the current version of our annotation service, SI refers to textual annotations.

The spatial components of the data model (not shown in Figure 1 for clarity reasons) follow the Open Geospatial Consortium (OGC) model (available at http://www.opengeospatial.org/standards.). OGC defines the *Geometry* as the most generic entity. The geometry is associated with a Spatial Reference System, which specifies the coordinate system.

The basic representations of geometry are *point*, *curve*, *surface*, and *collection*. A *GeometryCollection* can be used to represent a collection of distinct geometries. *MultiPoint*, *MultiLineString*, *MultiPolygon*, *MultiCurve*, and *MultiSurface* are homogeneous geometry collections of *Point*, *LineString*, *Polygon*, *Curve*, and *Surface*, respectively. A *LineString* represents a sequence of connected points. A *Point* represents a simple location in the coordinate space. A *Curve* is a generalization of *LineString*. A *Line* is a specialization of *LineString* which has only two points. A *LinearRing* represents a *LineString* which is closed and complete. A *Polygon* is a close plane figure, defined by one exterior boundary and zero or more interior boundaries. A *Surface* is a generalization of a *Polygon*.

3.2. Web Service API

With the data model and annotation and geographic concepts presented in the previous section, we now present the interfaces for the insert, update, delete, and query operations.

A generic interface was first defined to manage annotations:

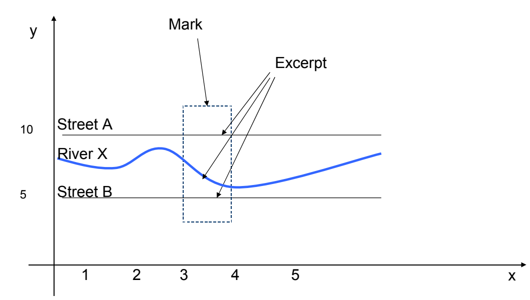
$< return >$ *operation* (**Text** Annotation, **Query** Query, **Metadata** Metadata, **Context** Context) where *operation* can assume the insert (*InsertAnnotation*), update (*UpdateAnnotation*), delete (*DeleteAnnotation*), or query (*QueryAnnotation*) operations. The *Annotation* parameter refers to the annotation content. The *Query* parameter defines a query whose results include the set of geographic objects to which the target operation will be related. The *Metadata* parameter refers to attributes associated with the mark metadata. The *Context* parameter includes elements of the mark context.

Note that since several parameters are managed, the interface was later specialized for geo-queries (regions, excerpts, etc.) and non-spatial queries (metadata, annotation, etc.) and their combination. One remarkable aspect of the proposed API relies on the support of operations which are defined over geographic objects that are result of queries.

3.2.1. Annotation Insertion

The interfaces for annotation insertion are specializations of the generic interface, considering regions, list of objects and results of geographic queries. Figure 3 illustrates the insertion of an annotation associated with the region defined by the rectangular region (mark). As it can be observed, this annotation is linked to part of Street A, part of River X, and part of Street B (*Excerpt*). Table *IdGeographicObject* contains records related to geographic objects (e.g., Street A, Street B, and River X). Table *IdGeographicObject_has_Excerpt* contains records related to the relationship of geographic objects and excerpts, after inserting a mark.

- **boolean** *InsertAnnotation* (**Text** Annotation, **Geometry** Region, **Metadata** Metadata, **Context** Context)
  This operation inserts an annotation associated with a set of geographic objects within a given region.

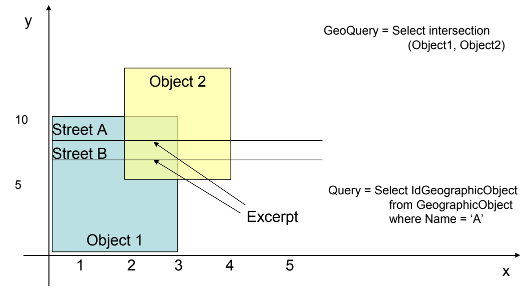


IdGeographic Object	Name	Type	The_Geometry
1	A	Street	0,10; 10,10
2	B	Street	0,5; 10,5
3	X	River	0,7; 10,7

IdGeographic Object_has_ Excerpt	IdExcerpt	Id Geographic Object	The_ Geometry
101	100	1	3,10; 4,10
102	100	2	3,5; 4,5
103	100	3	3,7; 4,6

Figure 3. Example of the execution of the operation *boolean InsertAnnotation (Text Annotation, Geometry Region, Metadata Metadata, Context Context).*

- **boolean** *InsertAnnotation* (**Text** Annotation, **Geometry** Region, **List** {ObjectType Type}, **Metadata** Metadata, **Context** Context)



IdGeographic Object	Name	Type	The_Geometry
1	A	Street	0,9; 10,9
2	B	Street	0,8; 10,8

IdGeographic Object_has_ Excerpt	IdExcerpt	Id Geographic Object	The_ Geometry
101	100	1	2,9; 3,9
102	100	2	2,8; 3,8

Figure 4. Example of the execution of the operation *boolean InsertAnnotation (Text Annotation, GeoQuery GeoQuery, List {ObjectType Type}, Metadata Metadata, Context Context)*.

This operation inserts an annotation associated with a set of geographic objects that can be found in a given region and that are of a given type.

- **boolean** *InsertAnnotation* (**Text** Annotation, **Geometry** Region, **Query** Query, **Metadata** Metadata, **Context** Context)
  This operation inserts an annotation associated with a set of geographic objects defined as the result of the query specified in the parameter *Query*.
- **boolean** *InsertAnnotation* (**Text** Annotation, **GeoQuery** GeoQuery, **Metadata** Metadata, **Context** Context)
  This operation inserts an annotation associated with geographic objects defined as the result of the geographic query specified in parameter *GeoQuery*.
- **boolean** *InsertAnnotation* (**Text** Annotation, **GeoQuery** GeoQuery, **List** {ObjectType Type}, **Metadata** Metadata, **Context** Context)
  This operation inserts an annotation associated with a list of geographic objects found in a given region and that satisfy the query defined in parameter *Query*.
  Figure 4 illustrates the insertion of an annotation associated with the region defined by the rectangular region (mark). The annotation is linked to parts of Streets A and B (*Excerpt*), which are the result of the query specified by parameter *GeoQuery*.

3.2.2. Query

This section describes the interfaces for querying annotations, considering a list of keywords, an object, a region, or results from a query.

- **List** {**Text, Metadata** } *QueryAnnotation* (**List** {**MetadataField** Field, **Operator** Operator, **MetadataValue** Value, **Conector** Conector})
  This operation returns the annotations (and associated metadata), which satisfy the given parameters.

- **List** {**Text, Metadata** } *QueryAnnotation* (**List** {**String** Keywords})
  This operation returns the annotations which contain a given list of keywords.
- **List** {**Text, Metadata** } *QueryAnnotation* (**String** ObjectName)
  This operation returns the annotations which are associated with a given geographic object.
- **List** {**Text, Metadata** } *QueryAnnotation* (**Geometry** Region)
  This operation returns the annotations found in a given region.
- **List** {**Text, Metadata** } *QueryAnnotation* (**Query** Query)
  This operation returns the annotations associated with objects defined by the results of the query defined in *Query*.
- **List** {**Text, Metadata** } *QueryAnnotation* (**GeoQuery** GeoQuery)
  This operation returns the annotations associated with geographic objects defined by the results of the geographic query *GeoQuery*.
- **MarkGeometry** *QueryMark* (**Text** Annotation)
  This operation returns the geometry of a mark whose annotation includes terms found in parameter *Annotation*. This operation can be used, for example, to obtain the geometries of the marks defined by users X and Y, as illustrated in Figure 2(d).

4. Case Study: Biodiversity Information System

This section presents a case study concerning the use of our annotation Web service to handle biological data. In our case study, species field observations are associated with geographic information.

4.1. Data Source

The data source contains observations of butterflies collected during field trips to the Municipal Reserve of Santa Genebra, a remnant forest fragment in Campinas, Brazil.

The biodiversity database model proposed in [Malaverri et al. 2009] was used for storing field observations. This data model considers the following elements: species taxonomic information, description of collection methodology and data on how, where, when, and by whom butterflies were collected.

In our prototype, we consider the data provided by the study described in [Filho 2003]. The main objective of that study is to describe quantitatively the community of frugivorous butterflies with an emphasis on populations of *Anaea ryphea (Cramer) (Nymphalidae: Charanidae)*. Samples were held for 26 months. The data acquisition procedure considered the distribution of 12 traps in three areas with different levels of disturbance: 4 at the reserve boundaries, 4 in the forest interior and 4 near a central path (abandoned road), which crosses the whole reserve. The objective of this study is to evaluate butterfly features, given different luminosity, temperature, and wind conditions.

In this prototype, only text annotations were considered. Initially, the database was loaded with data obtained from digital maps of Brazil in Shapefile format (available at http://www.gismaps.com.br/). The following natural and environment resources were chosen for evaluation: Geography, Preservation Areas, and National and State Parks.

4.2. Architecture of the Biodiversity Information System

The architecture of the implemented Biodiversity Information System is composed of three layers: Application, Services, and Persistence (Figure 5).

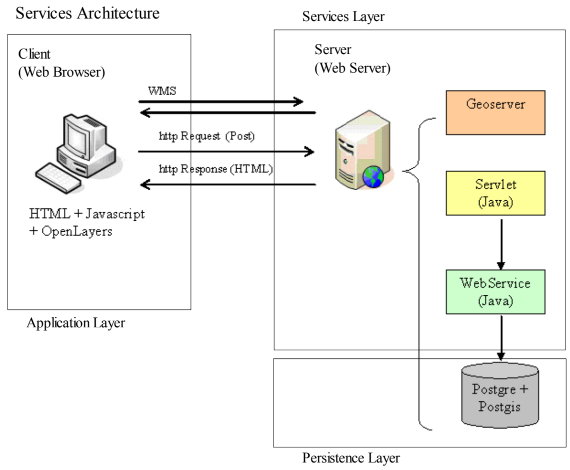


Figure 5. Architecture of the implemented Biodiversity Information System.

The Application layer is responsible for the interaction with the user and the service layer. Usually, it can be a Web interface or a component of another system. In the case of our prototype, the client application layer was composed by HTML pages and Javascript running in a Web Browser. The rendering of the maps was performed on the server. The OpenLayers framework was used for map presentation. OpenLayers is an Open Source JavaScript library used to display spatial data in web pages. The OpenLayers API allows several tools to interact with the visual maps or performing actions on certain events. The HTML page invokes a servlet passing the desired operation and the required parameters.

The Service layer provides the APIs, which define how features can be invoked by other applications. This layer is composed by two components: a map server that is used to compose and render a map to be published on the web; and the proposed geographic annotation service that is used to insert annotations associated with geographic data. Geoserver (available at http://geoserver.org/display/geos/welcome) was used to implement the map server. GeoServer is compliant with the OGC open standards such as Web Map Service (WMS), Web Coverage Service (WCS) and Web Feature Service (WFS-T).

The Persistence layer includes database facilities to manage data about geographic objects, annotations, and marks. This layer uses PostgreSQL database system (available at http://www.postgresql.org) and PostGIS (available at http://postgis.refractions.net), whose provides mechanisms for processing spatial operators.

4.3. Usage Scenario

Figure 6 illustrates the creation of an annotation. Initially, a polygon is drawn within an ecological reserve (orange polygon). Later an annotation is inserted, indicating that the defined polygon contains special traps for capturing butterflies. In this case, the excerpt is a subregion of the ecological reserve.

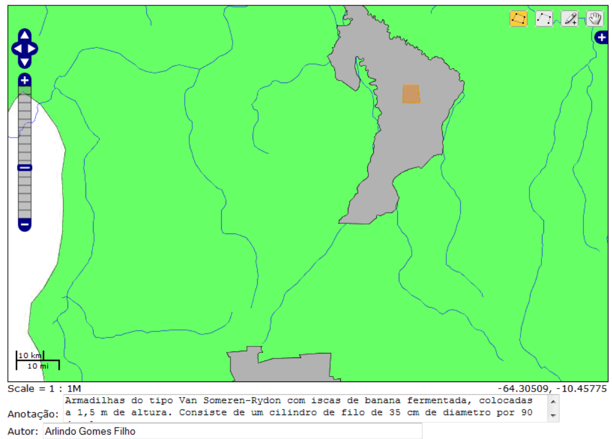


Figure 6. Example of annotation of a region within an ecological reserve.

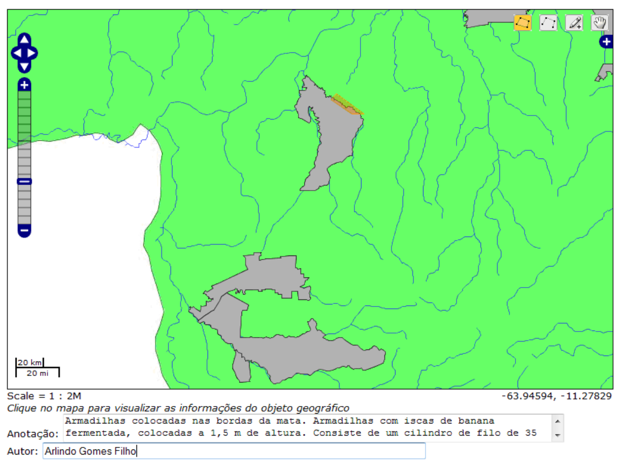


Figure 7. Example of annotation around the border of an ecological reserve.

A Biologist can insert annotation on traps located at the boundaries of the reserve (illustrated in Figure 7). In this example, the polygon was drawn around the border and is partly contained in the area of the ecological reserve. The excerpt is a subregion of the ecological reserve which *intersects* the defined mark region. The intersection operation takes advantage of the *InsertAnnotation* operation that uses a geographic query. A Biologist can also query annotations whose marks are contained within an ecological reserve (illustrated in Figure 8). In the example, the object was the geographical reserve REEXTE near the JACI-PARANA river.

A simple scenario is: suppose that a department has several biologists which worked in different Brazilian areas along the years, annotating reserves, rivers and regions. Using a simple map, it is easier to visualize which area was more explored, or visualize the intersection of regions which were explored by different researchers, for example.

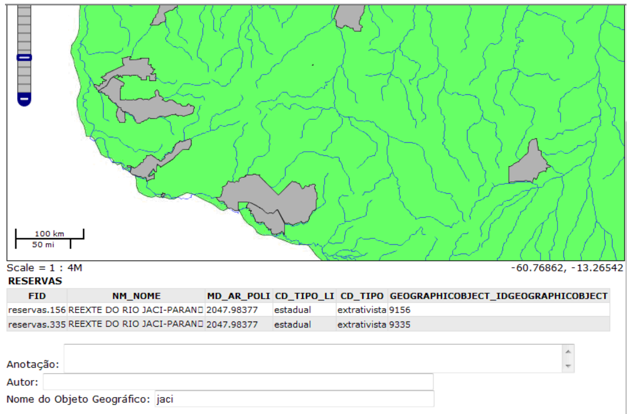


Figure 8. Example of an annotation query by textual term.

5. Conclusions

We have presented the specification and implementation of a Web service for managing geographic data in Biodiversity Information Systems. The service handles information in text format, referred as annotations.

The specification and implementation of the proposed Web service rely on two main contributions: the definition of a data model based on Superimposed Information and geographic concepts and the specification and implementation of a set of operations to manage them. One remarkable aspect of the proposed API relies on the support of operations (e.g., insertion, deletion, query, update, and deletion) which are defined over geographic objects that are result of queries. To validate our ideas, we developed a Biodiversity Information System that takes advantage of the proposed annotation service to correlate ecological data with geographic information.

Future work includes the extension of the proposed API to handle raster data, the extension for the SuperIDR tool [SI] and BioCore (available at http://www.lis.ic.unicamp.br/projects/biocore/). Another future work consists of applying the proposed service in other domains (e.g., Agriculture).

6. Acknowledgments

We would like to thank CAPES(4479-09-2), FAPESP and CNPq (in special, the BioCORE Project).

References

Agosti, M. and Ferro, N. (2008). Annotations: A Way to Interoperability in DL. In *ECDL '08*, pages 291–295.

Archer, D. W., Delcambre, L., Corubolo, F., Cassel, L., Price, S., Murthy, U., Maier, D., Fox, E. A., Murthy, S., Mccall, J., Kuchibhotla, K., and Suryavanshi, R. (2008). Superimposed Information Architecture for Digital Libraries. In *ECDL '08*, pages 88–99.

Barros, E. G., Laender, A. H. F., Gonçalves, M. A., and Barbosa, F. A. R. (2008). A digital library environment for integrating, disseminating and exploring ecological data. *Ecological Informatics*, 3(4-5):295–308.

Câmara, G., Casanova, M. A., Magalhães, G. C., and Medeiros, C. M. B. (1996). *Anatomia de Sistemas de Informação Geográfica*. 10a Escola de Computação, Instituto de Computação, Unicamp, Campinas.

Delcambre, L., Maier, D., Bowers, S., Weaver, M., Deng, L., Gorman, P., Ash, J., Lavelle, M., and Lyman, J. (2001). Bundles in Captivity: An Application of Superimposed Information. *17th ICDE*, pages 111–120.

Filho, A. G. (2003). A Comunidade de Borboletas Frugívoras da Reserva de Santa Genebra, Campinas, SP, Com Ênfase na Flutuação Populacional de Anaea Ryphea (Cramer) (Nymphalidae: Charanidae) e Sua Relação com as Plantas Hospedeiras. Master's thesis, Unicamp.

Gomes Jr, L. C. (2007). Uma arquitetura para consultas a repositórios de biodiversidade na web. Master's thesis, Unicamp.

Güting, R. H. (1994). An introduction to spatial database systems. *The VLDB Journal*, 3(4):357–399.

Macário, C. G. N., de Sousa, S. R., and Medeiros, C. B. (2009). Annotating geospatial data based on its semantics. In *GIS '09: Proceedings of the 17th ACM SIGSPATIAL*, pages 81–90.

Maier, D. and Delcambre, L. (1999a). Superimposed Information for the Internet. In *Proceedings of the ACM SIGMOD Workshop on the Web and Databases*, pages 1–9.

Maier, D. and Delcambre, L. M. L. (1999b). Superimposed information for the internet. In *WebDB (Informal Proceedings)*, pages 1–9.

Malaverri, J. G., Vilar, B. S. C. M., and Medeiros, C. B. (2009). A tool based on web services to query biodiversity information. In Filipe, J. and Cordeiro, J., editors, *WEBIST*, pages 305–310.

Sanchez, J. A., Flores, C. A., and Schnase, J. L. (1999). Mutant: Agents as Guides for Multiple Taxonomies in the Floristic Digital Library. In *Proceedings of the Fourth ACM JCDL*, pages 244–245, Berkeley, California, USA.

SI. Superimposed Information Projects at Virginia Tech, available at http://si.dlib.vt.edu/, last accessed on 01/10/2010.

Torres, R. S., Medeiros, C. M. B., calves, M. A. G., and Fox, E. A. (2006). A Digital Library Framework for Biodiversity Information System. *International Journal on Digital Libraries*, 6(1):3–17.

Yu, Y., Stamberger, J. A., Manoharan, A., and Paepcke, A. (2006). Ecopod: a mobile tool for community based biodiversity collection building. In *JCDL '06*, pages 244–253.

Zhu, B., Ramsey, M., and Chen, H. (2000). Creating a Large-Scale Content-Based Airphoto Image Digital Library. *IEEE Transactions on Image Processing*, 9(1):163–167.

Using Semantic Similarity to Improve Information Discovery in Spatial Data Infrastructures

Fabio Gomes de Andrade[1], Cláudio de Souza Baptista[2]

[1]Departamento de Informática – Instituto Federal de Educação Ciência e Tecnologia da Paraíba (IFPB)
58900-000 – Cajazeiras – PB – Brazil

[2]Departamento de Sistemas e Computação
Universidade Federal de Campina Grande (UFCG) – Campina Grande, PB – Brazil

fabio@ifpb.edu.br, baptista@dsc.ufcg.edu.br

Abstract. *In the recent years, several works have been proposed with an approach to the use of semantics to improve the process of discovering geographic resources offered by spatial data infrastructures. However, semantic queries may return a large number of results, what causes the necessity for efficient ways to evaluate the relevance of each result retrieved. This paper proposes a framework that uses ontologies and thematic relevance to suggest a measurement that allows evaluating how relevant is each resource offered by the infrastructure to the user's query. This feature allows the results retrieved in a query to be organized through a ranking, in such a way that the most relevant resources are presented to the user first.*

1. Introduction

In recent years, spatial data infrastructures (SDIs) [Williamson et. al 2003] have been developed in order to ease the discovery of spatial data and improve the interoperability of spatial data supplied by different information sources. The development of open standards to the geospatial domain has reduced the problems concerning data interoperability. However, discovering information which is already available at spatial data providers is still a hard task.

A limitation of current SDIs is that their catalog services perform their queries based uniquely on keywords. This characteristic leads to the execution of queries with low recall, as the resources described with terms related to the keywords used to generate the query are not retrieved. Also, low precision is obtained, since many irrelevant resources which have the terms of the query in their description end up being retrieved. The difficulty in locating existing information makes many companies to spend much time and money with the production of data already made available by other providers and which could be made at lower costs or, in some cases, with no cost.

As a way to overcome the limitations of the present catalog services, it is increasingly common the application of semantic web concepts to improve the discovery of spatial data. The objective of the semantic web [Berners-Lee et. al 2001] is to use formal means to describe the semantics of resources published in the web, improving the data sharing among applications. The semantic web principles have been implemented through ontologies [Guarino 1995]. Ontologies are formal conceptualizations of an application domain, which makes its semantic understandable to both human and machines.

Usually, applications that use ontologies to discover information use an approach based on a semantic relationship known as *subsumption*. This kind of solution consists in locating all resources whose description is *subsumed* by a certain search concept. The great advantage of this kind of solution is that it improves the recall of queries, since inference rules may be used for runtime inference of new knowledge. Nevertheless, this kind of solution considers that all of the retrieved results have the same relevance to the user. As a semantic query may return a large number of results, information which is more relevant to the user may be shown at the end of the result and, eventually, may not even be analyzed. For example, if a user requests feature types about a concept *WaterCourse*, feature types associated to subclasses of this concept (which offer only a part of the information requested), can be presented earlier than types linked exactly to the concept search, which are probably more relevant to user. This feature produces the necessity for the development of mechanisms that permit the evaluation of the relevance that each retrieved result has to the user.

To tackle this limitation, this paper proposes an approach that uses notions of similarity as a way to improve the discovery of information in spatial data infrastructures. The main contribution consists of the development of a similarity measurement that enables to evaluate the relevance of each resource featured by the SDI to the user's query. Still, the paper develops a new measurement to evaluate similarity among concepts defined in ontologies, and shows how some ideas of classic information retrieval can be reused and adapted to the spatial domain.

The remaining of the paper is organized as follows. Section 2 discusses related works. Section 3 addresses the process of semantic annotation used to describe semantics of the feature types. Section 4 describes the approach used to evaluate the relevance of each resource offered to user query. Section 5 shows the implementation and the results obtained. Finally, section 6 concludes the paper and highlights further work to be undertaken.

2. Related work

Over the years, several works have been proposed to solve the problem of the discovery of information in SDIs. Though these works are related to the same research area, they differ from each other with respect to the type of resource discovered and the approach used to discover these resources.

[Smits and Friis-Christensen 2007] developed a work for discovery of data in SDIs. In that work, the semantic annotation of the resources is done by associating them to concepts defined in a thesaurus. After that, the user can browse the terms of the thesaurus to discover resources associated to it. In another work [Stock et. al 2010], the feature types of the SDI, as well as the operations that can be performed with them are described through a feature type catalog (FTC), which maintains links to the services that implement each operation defined in the FTC. The retrieval of information is done by browsing this FTC. [Lutz and Kolas 2007] developed a work that uses rules to perform the semantic annotation and the retrieval of spatial data spread on different data sources. In another work, [Lutz et al. 2008], WFS data types are semantically annotated through mapping records. After that, a reasoner using Description Logic is used to retrieve feature types which are subsumed by the user's search concept. In another work [Klien et al 2006], the authors use Comprehensive Source Descriptions to describe the semantic features of geoservices used in the discovery of data to solve the management of disasters. All of the works cited above improve the discovery of resources in SDIs by using semantics to describe their resources. However, they no not offer the means to evaluate the relevance of each retrieved resource. Other important works present the same limitation, such as [Athanasis et. al 2009], [Lutz and Klien 2006] and [Wiegand and Garcia 2007].

Janowicz et al [Janowicz et al, 2008] developed a similarity-based solution to discover spatial data supplied by SDIs. The proposed work uses a framework to evaluate the similarity between a concept defined in the user's query and the concepts associated to spatial data types offered by the infrastructure. However, this work is directed towards a very specific ontology language. Besides, it does not take into account the relevance of the theme during the retrieval process. This way, resources annotated with the same concept are judged with the same relevance. This is a drawback, especially in queries that return a large number of results.

The analysis of related work shows that the discovery of information in SDIs is still an open problem. The use of ontologies enables to explore semantics to enhance the quality of queries. However, queries can produce a large number of results which need to be evaluated by the user before being completely retrieved. When this happens, data that are potentially more relevant to the user can be shown at the end of the result, and cannot be judged by the user. This problem leads to the necessity to develop efficient mechanisms to evaluate the relevance of each result retrieved. Such solution is based on a ranking approach, in which more relevant resources are presented first. This ranking reduces the time spent during the result evaluation process and facilitates both data discovery and reuse.

3. The semantic annotation proccess

Before describing the proposed approach for evaluating the relevance of spatial resources supplied by SDIs, it is necessary to understand the kind of information that can be discovered. Aiming to standardize the access to geographic data offered by several spatial data sources, SDIs offer a set of web services that enables the access to spatial data in several different formats. Examples of such services are: Web Map Service (WMS), for the access to vector maps layers; Web Feature Service (WFS), for the access to spatial data in GML format; and Web Coverage Service (WCS), to provide access to raster data. The approach proposed by this paper focuses on discovery of feature types offered by geospatial services, in which each feature type can be a vector map layer, a GML feature type or a raster image, depending of the kind of the offered service.

The framework has a relational database which contains information about all spatial data services offered by information sources registered in the infrastructure. This information is registered at the time a data source registers its resources in the infrastructure. When the service is registered, the framework stores on its database the information about each data type it offers. For each data type offered, the framework stores information such as name, title, textual description, the type (vector map layer, feature type or coverage) and the bounding-box of the geographic region it covers. All this information is retrieved automatically at the time the service is registered, through the execution of its *getCapabilities* operation.

After this information is retrieved, the framework shows to the user, who is registering the service, a page containing the information about all data types available. Then, the framework asks the user to perform the semantic annotation of all types offered. For that, the user must choose, among the existing concepts in the domain ontologies used by the infrastructure, the one that better represents the information supplied by that data type. The URI of the concept chosen for annotation of each data type is stored together with its information and used during the process of information discovery. For example, a feature type that offers information about water reservoirs may be annotated with the concept *WaterCourse*, while one that describes only rivers may be annotated with the concept *River*. After all data types are annotated the registering is finished, and the information concerning the service and the data types it offers becomes available to the discovery process.

4. An approach to evaluate relevance

The main objective of the work described in this paper is the description of a measurement that permits to evaluate how relevant each data type offered by the infrastructure is for the user's query. The verification is done in three steps:

(i). to verify how much the concept used to annotate the data type which is under evaluation is similar to the search concept defined by the user;

(ii). to verify how relevant the theme requested by the query is to the spatial service that offers the data type under evaluation; and

(iii). to combine the values of both measurements to evaluate the relevance of the data type under evaluation for the query.

4.1 The information discovery process

The information discovery process occurs in the following manner as depicted in Figure 1. In a graphic interface, the user selects a theme and a geographic region of interest. This theme represents one concept defined in one of the domain ontologies used by the infrastructure. During the information discovery process, this concept is called search concept (SC). After, a spatial query is executed to filter, among the available data types, those whose bounding-box intersects the geographic region defined for the query. After that, the relevance of each filtered data type for the user's query is evaluated. Once a query can produce too many results, a threshold value describing the minimum desired relevance is defined, and the next stage in the discovery process is to filter the data types that have this degree of relevance greater than the threshold. Finally, the final results are organized in descending order of relevance and presented to the user.

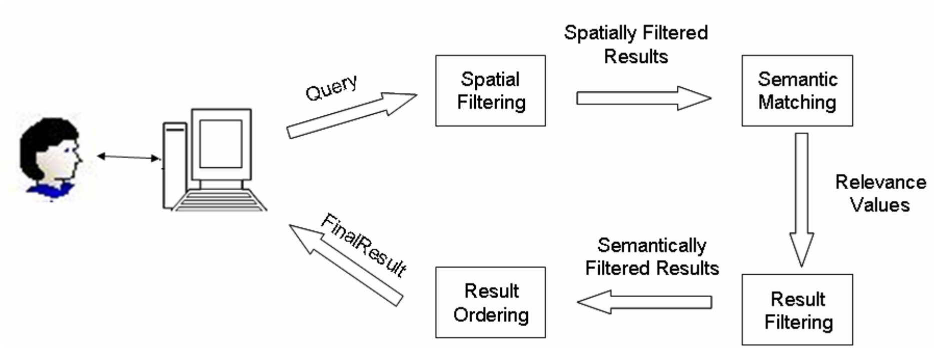


Figure 1. The information discovery process

4.2 Evaluating the similarity between the concepts

The first stage of the process used to measure the relevance consists in evaluating the similarity between the search concept of the user's query and the concept chosen for the semantic annotation of the data type under analysis. The approach used to evaluate this similarity has the following characteristics:

- **support to other types of relationships:** many works that propose to evaluate the similarity between concepts consider just the inheritance relationship. However, other relationship types, such as composition, cannot be neglected, since they denote an association of the concepts involved. As the composition is not such a strong relationship as the inheritance, weights are necessary to distinguish the relationship types. For example, two concepts associated by an inheritance relationship must have a degree of similarity greater than that existing between two concepts associated by other relationship type.
- **asymmetry:** symmetrical similarity measurements consider that the similarity between two pairs of concepts is the same, independently of the comparison order. However, for the problem studied in this paper, it was considered that the symmetry is not an interesting feature. For example let us suppose that a concept B is a sub-concept of a concept A. We may state that all data associated to B are relevant to the user, since every instance of concept B is also an instance of concept A. Nevertheless, if the user is looking for data associated to concept B, not all data associated to concept A are relevant to the user, since not all instances of concept A are instances of concept B. This characteristic requires, in the second case, that the symmetry must be smaller than in the first case, due to the existence of information of no interest to the user. The same idea is applied to the composition relationship;
- **degree of generalization:** ontologies are described through concepts that are organized in a hierarchical form, through inheritance relationships. Let us suppose that there is a hierarchy from a search concept SC in the ontology under analysis. As we go through this hierarchy, we find concepts that are more and more specialized with respect to SC and, consequently, have more difference with respect to it. Thus, the similarity of concepts must decrease gradually as the deepness of the concept increases.

The evaluation of the similarity between concepts is performed through a semantic network, generated from the parsing of the ontology at the time it is added to the SDI. The construction of this network takes into consideration two types of semantic relationship existing between the ontology concepts: inheritance and composition. The following algorithms in Table 1 present how a semantic network may be generated. The first algorithm is used to start the process of generation of the network and the second one to expand the production of the network to new concepts obtained from new concepts which are processed by the algorithm. In the first algorithm, the network is generated from each root concept (RC) in the ontology. A root concept is a concept that has no superclass in the ontology.

Table 1: Semantic Network Generation Algorithm

```
generateSemanticNetwork(O: Ontology): SemanticNetwork;
begin
      SN = new SemanticNetwork();
      rootNode = createNode(“Thing”);
      SN.addNode(rootNode);
      for each RCi in O do
      begin
           newNode = createNode(RCi)
           SN.addNode(newNode);
           SN.addSubclassEdge(rootNode, newNode);
           expandSemanticNetwork(SN, newNode, O);
      end;
      return SN;
end;
expandSemanticNetwork(sn:SemanticNetwork,currentNode:Node,
O:ontology): void;
begin
      SC = O.getSubClasses (currentNode.getConcept());
      for each SCi in SC do
      begin
          newNode = createNode(SCi);
          sn.addNode (newNode);
          sn.addSubclassEdge(currentNode, newNode);
      end;
      OP = O.getObjectProperties(currentNode.getConcept());
      for each OPi in OP do
      begin
          newNode = createNode(OPi.getRange());
          sn.addNode (newNode);
          sn.addAssociationEdge(currentNode, newNode);
      end;
end;
```

The result of the execution of the algorithms above is a semantic network which contains all of the concepts defined in the ontology and the semantic relationships existing between these concepts. In such network, nodes represent concepts, and arrows represent semantic relationships. The network produced has two kinds of arrow: one to define inheritance relationships and other to define composition relationships. Figure 2 shows a semantic network produced for the hydrographic domain, extracted from the GEMET ontology.

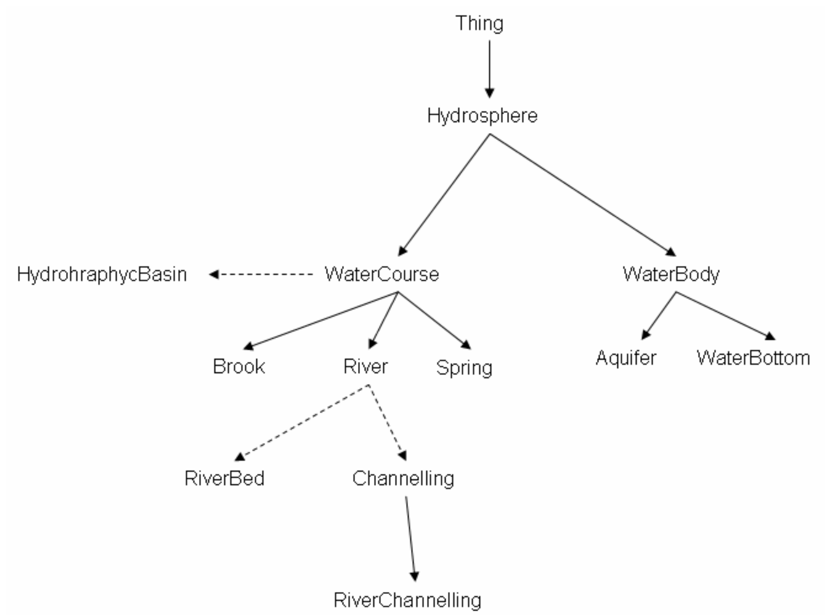


Figure 2. Semantic network produced for a hydrographic ontology

After the semantic network is produced, the framework evaluates the degree of similarity for all combinations of pairs of concepts defined in the ontology. This similarity is calculated through the analysis of the path that connects the two concepts under evaluation in the semantic network. The calculation of this similarity is performed taking into consideration two kinds of variables: the semantic relationship between the concepts and the distance between them in the network.

The objective of the semantic relationship between the concepts is to assign a greater degree of similarity to pairs of concepts that have stronger semantic relationships. As inheritance is a semantic relationship stronger than composition, concepts associated by an inheritance relationship must have a degree of similarity greater than that concepts associated by a composition relationship. To implement this constraint, a weight is assigned to each arrow of the network. For each node, two weights are possible: a normal weight and an inverse weight. The weight used to evaluate the similarity depends on the order of the concepts involved. This constraint is used to keep the asymmetry requirement. The use of two kinds of weight enables to ensure asymmetry, keeping the simplicity to discovery paths in graphs. Currently, the weight used for normal and inverse weights are, respectively, 0.8 and 0.6, for inheritance relationship, and 0.6 and 0.4, for composition relationship.

To perform the comparison between the two concepts, the first step consists in locating, in the network, the path that connects the two concepts. To allow the comparison between concepts Q and D, there must be at least one path from node Q that leads to D, or vice-versa. When none of these paths exist, the concepts are considered disjoint and the degree of similarity between them is assumed to be zero. When any of these paths can be found, the framework uses weights of the nodes in this path to evaluate the semantic relationship between them. Let W={w_1, w_2, ,,,, w_n} be the set of the weights of each arrow in the shortest path that connects the concepts Q and D in the semantic network that represents the ontology in which these concepts were defined. The value of the semantic relationship can be formally defined in Equation I. In order to ensure asymmetry property, the weights values depends on the order of the concepts in the path. If the path starts with the search concept Q and ends with the concept D, the normal weights are considered. However, if the path starts with the concept D and ends with the concept Q, the inverse weight are considered.

$$sem\mathrm{Re}l = \min\{w_1, w_2, \ldots, w_n\} \quad \text{(I)}$$

The second variable used to compare the similarity between the concepts is the distance between them. Rada et al [Rada et al 1989] introduce the semantic distance as a metric to evaluate similarity among concepts in semantic networks. The goal of this variable is to guarantee that pairs of concepts which are closer in the network have a greater degree of similarity compared to more distant pairs. We use this metric to implement the constraint to the degree of generalization between the concepts (Section 4.2). This measurement is inversely proportional to the degree of similarity, that is, as increases the distance between the concepts, the similarity between them diminishes.

After evaluating the semantic relationship and the distance between the two compared concepts, the values of these variables are used to measure the similarity between the concepts. To calculate these values, a weight is assigned to each of these variables. W_1 represents the weight assigned to the semantic relationship, while w_2 represents the weight of distance. The use of these weights makes the similarity between the concepts to be evaluated through the Equation II:

$$sim(Q,D) = w1 * sem\mathrm{Re}lationship(Q,D) + w2 * \left(\frac{1}{dist(Q,D)}\right) \quad \text{(II)}$$

The solution evaluates the degree of similarity between all pairs of concepts defined in the ontology (in both directions), generating a similarity matrix. The values of these similarities are stored in a relational database. Table 2 shows the similarity matrix for an excerpt of the concepts of the semantic network depicted in Figure 2. Concepts marked with S represent the concept defined in the user query, whereas concepts marked with D represent the ones used to annotate the feature type that is being evaluated.

Table 2. Similarity matrix

	Hydrosphere(D)	**WaterCourse (D)**	**River(D)**	**RiverBed (D)**
Hydrosphere (S)	1	0.84	0.74	0.54
WaterCourse (S)	0.68	1	0.84	0.58
River (S)	0.58	0.68	1	0.68
RiverBed (S)	0.38	0.42	0.52	1

4.3 Evaluating the degree of thematic relevance

Besides the degree of semantic similarity among the concepts involved in the query, we consider the degree of thematic relevance to improve the discovery process. The objective of this measurement is to evaluate how relevant is a theme requested in a query to the service which offers the data type under analysis. Through this measurement, data types offered by services in which the theme has more relevance are shown first to the user during the presentation of results. The value of this measurement is very important, since many data types are offered by several services, especially if the user's query requests a very common theme.

The degree of relevance that a certain theme has to a service is calculated through the normalized frequency, which is a measurement used in the classical information retrieval [Baeza-Yates and Ribeiro-Neto 1999] to evaluate the relevance of a certain term in a document. To evaluate this degree, the framework registers, at the time the service is registered, the normalized frequency of each theme offered by it. This way, the degree of relevance (*relDegree*) of a theme C to a service is calculated through the proportion of the number of times the theme occurs in the service (ni) and the number of data types offered by the service (N). The value of ni for a certain concept C is calculated based on the semantic relationships defined in the ontology. Such calculation comes from Equation III. In this equation, fi(C) is the number of occurrences of the concept under evaluation, and fi(S) and fi(SC) represent, respectively, the number of occurrences of a synonym concept of C and the number of occurrences of a concept that is a sub-class of C.

$$relDegree(C,S) = \frac{fi(C) + \sum fi(S) + \sum fi(SC)}{N} \quad \text{(III)}$$

The information of relevance of each concept is stored in the database of the infrastructure. The major advantage of keeping this information stored is that this allows us to accelerate the response time of the query. Such feature also helps us to keep the scalability of the solution for large amount of data.

4.4 Calculating the relevance

After defining the metrics used to calculate the degree of relevance of a data type to a user's query, the next step consists in defining how the values of these metrics will be used for that purpose. One possibility to solve this problem would be the representation of the user's query and the data type under evaluation as vectors in a bi-dimensional space and use the Euclidian distance to evaluate the similarity. However, this kind of metric represents similarity through a real number, corresponding to the distance between the points. Thus, in order to solve the problem, we adopted the sum of the values of the metrics, where a weight is assigned to each of the metrics. This technique, besides offering flexibility, since all weights may be altered to perform new queries, also offers similarity values between 0 and 1, which makes the evaluation of similarity more intuitive for the human being.

Thus, given a theme Q defined in the user's query and the theme D associated to the feature type under evaluation, the degree of relevance of this type for the query is calculated through Equation IV. In such equation, *sg* represents the degree of similarity of concepts Q and D, whereas *relDeg* represents the degree of relevance that the theme D has to the service by which the feature type under evaluation is offered. Finally w_1 e w_2 represent the weights that each type of measurement has to the calculation of spatial similarity. Each weight must have a value between 0 and 1, and their sum must always be equal to 1:

$$semanticSim(Q,D) = w_1 \times sg(Q,D) + w_2 \times relDeg(D,S) \quad \text{(IV)}$$

5. Implementation and results

To evaluate the proposed approach, a prototype was developed. The first step in this implementation was to define the domain ontologies that would be used for semantic annotation and discovery. In our experiments we have used ten domain ontologies, which were created from data models according to the Brazilian Spatial Data National Infrastructure. These ontologies are represented in OWL and the Jena framework is used to parse them. After that, we gathered several spatial services (WMS and WFS) offered by several providers throughout Brazil. Each service was processed and their information was stored in a database. Besides, we registered information concerning each feature type they offer. Each type was semantically annotated through the domain ontologies defined in the infrastructure. Currently, this database stores about 457 feature types, distributed among 21 geospatial web services, from 16 service providers. All this information is stored in a PostgreSQL/PostGIS database server.

To illustrate the results obtained during the evaluation process, let us suppose a simple query in which the user wants to find feature types regarding *WaterCourses* in Brazil. In the database used for evaluation, there are 70 feature types directly related either to this concept or to one of its subclasses. These types are distributed among the services offered by 8 different Brazilian sources: the National Water Agency (ANA), the Executive Agency of Water Management of the State of Paraíba (AESA), the National Agency for Electrical Energy (ANEEL), the Brazilian Institute for Environment (IBAMA), the Ministry of Fisheries and Aquaculture (MPA), the Protection System for Amazon (SIPAM), the Department of Water Resources of the State of Santa Catarina (SIRHESC) and the Federal University of Minas Gerais (UFMG), according to the Table 3. For each entry in the table, we have the provider name, the thematic relevance of the concept search (*WaterCourse*) to the service, the number of feature types annotated with the search concept, the number of feature types annotated with *WaterCourse* subclasses and the number of feature types annotated with concepts which are related to the search concept through a composition relationship.

Table 3. Data providers example concerning the hydrography concept

Provider	Thematic Relevance	Concept Search	Subclasses	Composition
ANA	1	2	4	0
AESA	0.5625	3	6	1
ANEEL	0.1621	5	5	0
IBAMA	0.1250	3	0	0
MPA	1	4	0	0
SIPAM	0.0517	2	2	0
SIRHESC	0.7045	29	2	0
UFMG	0.3076	0	3	1

After executing the query, 36 features types obtained a relevance degree greater than or equal to 90%. In this category were all feature types exactly annotated with the search concept. In this category, the types are listed in descendant order of thematic relevance. The types offered by the ANA and the MPA are listed first, with relevance of 100 %. After, the result shows the 29 types of data provided by the SIRHESC with relevance around 94% and the types offered by the AESA, with relevance around 91 % .

The second category contains feature types that have a relevance value between 80% and 90%. In this case there were two types of results. The first one contains data types that are associated with exactly the search concept, but are offered by services with low thematic relevance. Hence, there were the other two feature types offered by ANA, which had relevance of 81%. The second one is composed of data whose services have high thematic relevance, but they are annotated with concepts that represent subclasses of the search concept. In this case, we have data types offered by ANEEL, IBAMA and SIPAM. These types have gained importance around 83%, 82% and 81%, respectively.

The third category of results includes 18 feature types that have a relevance value between 60% and 80%. This category contains data types offered by services in which the *WaterCourse* theme is highly relevant, but have been annotated with subclasses of the search concept. The remaining data types offered by SIRHESC, AESA, UFMG, ANEEL and SIPAM are listed, in this order. Finally the last category contains the data types of services offered and AESA and UFMG that were annotated with concepts that have a composition relationship with the search concept. The relevance values were to 58% and 53% respectively.

6. Conclusion and further work

SDIs play an increasingly important role in the dissemination of geographic information offered by several organizations. However, locating geographic data offered by these infrastructures in an efficient and precise manner is still a hard task. Though ontology-based solutions have improved the discovering process, there is still the need to evaluate the relevance of the retrieved results to the user, such that the more relevant results can be exhibited first.

This paper described a framework that combines the notion of semantic similarity between concepts defined in ontologies and ideas applied to the classical information retrieval to evaluate how relevant are the spatial data offered by the infrastructure to an end-user's query. Though the present results have shown that the approach is interesting, some future works are still necessary.

On of the works necessary in the future consists in extending the notion of semantic similarity to treat more complex concepts and relationships, as, for example, concepts defined through conjunction, disjunction and negation of other concepts. Another important future work is to evaluate the user preferences once the result is presented. Besides to validate our approach, this work will enable to improve the weights used to calculate semantic similarity. Still, there is the need to evaluate the similarity between concepts defined in different ontologies, what can give an even better recall for the user's queries.

References

Athanasis, N., Kalabokidis, K., Vaitis, M. and Soulakellis, N. (2009) "Towards Semantics-based Approach in the Development of Geographic Portals". In: Computers & Geosciences, Volume 35(2), pages 301-308.

Baeza-Yates, R. and Ribeiro-Neto, B. (1999) "Modern Information Retrieval". Addison Wesley

Berners-Lee, T, Hendler, J. and Lassila, O. (2001) "The Semantic Web". http://www.scientificamerican.com/article.cfm?id=the-semantic-web, May

Guarino, N. (1995) "Formal Ontology, Conceptual Analysis and Knowledge Representation". In: International Journal of Human Computer Studies, Volume 43(5-6), pages 625 – 640.

Janowicz K., Wilkes M. and Lutz, M. (2008) "Similarity-based Information Retrieval and its Role within Spatial Data Infrastructures", International Conference on Geographic Information Science, Cova, T. V., Miller, H. J., Beard, K., Frank, A. U. and Goodchild, M. F., Germany, Springer, p. 151-167.

Klien, E., Lutz, M., and Kuhn, W. (2006). "Ontology-Based Discovery of Geographic Information Services - An Application in Disaster Management". In: Computers, Environment and Urban Systems, Volume 30(1), pages 102-123.

Lutz, M. and Klien, E. (2006) "Ontology-based Retrieval of Geographic Information". In: International Journal of Geographical Information Science, Volume 20(3), pages 233-260.

Lutz, M. and Kolas, D. (2007) "Rule-Based Discovery in Spatial Data Infrastructure". In: Transactions in GIS, Volume 11 (3), pages 317-336.

Lutz, M., Sprado, J., Klien, E., Schubert, C. and Christ, I. (2008). "Overcoming Semantic Heterogeneity in Spatial Data Infrastructures". In: Computers & Geosciences, Volume 35(4), pages 739-752.

Rada, R., Mili, H., Bicknell, E. and Blettner, M. (1989) "Development and Application of a Metric on Semantic Nets". In: IEEE Transactions on Systems, Man and Cybernetics, Volume 19 (1), pages 17-30.

Smits, P. and Friis-Christensen, A. (2007). "Resource Discovery in a European Spatial Data Infrastructure". In: IEEE Transactions on Knowledge and Data Engineering, Volume 19 (1), pages 85-95.

Stock, K., Atkinson, R., Higgins, C., Small, M., Woolf, A., Millard, K. and Arctur, D. (2010) "A Semantic Registry Using a Feature Type Catalogue instead Ontologies to Support Spatial Data Infrastructures". In: International Journal of Geographic Information Science, Volume 24 (2), pages 231-252.

Wiegand, N. and Garcia, C. (2007) "A Task-Based Ontology Approach to Automate Geospatial Data Retrieval". In: Transactions in GIS, Volume 11(3), pages 355-376.

Williamson, I., Rajabifard, A. and Feeney, M. F. (2003), Developing Spatial Data Infrastructures: From Concept to Reality, Taylor & Francis.

Proceedings XI GEOINFO, Campos do Jordão, Brazil, Nov. 28 to Dec. 1, 2010, INPE, p. 57–66.

Low-cost Satellite-based Products for the Web - the Example of Fire Web Service

Theodor Foerster[1], Thore Fechner[1], Holger Fritze[1], Frederik Loock[1] & Albert Remke[2]

[1]Institute for Geoinformatics (ifgi) - University of Muenster
Muenster, Germany

[2]52° North - Initiative for Geospatial Open Source software GmbH
Muenster, Germany

{theodor.foerster,t.fechner,h.fritze,f.loock}@uni-muenster.de

a.remke@52north.org

Abstract. *GEONETCast is a satellite-based dissemination system for a range of in-situ and remote sensing products. Enabling such products for the Web is interesting to ease a broad data access for industry and academia. In this context, standardized Web Service interfaces and data formats play an important role. In this article, we present an architecture for accessing satellite-based MODIS fire data products through the so-called Fire Web Service. A browser-based application allows users to query up-to-date but also historic data about fire events. The presented architecture has been implemented based on Open Source software.*

1. Introduction

Broadcasting in-situ and remote sensing data through a satellite network for low-cost is important to support developing countries in planning and monitoring their environment with up-to-date data. One example of such low-cost satellite-based broadcasting system is GEONETCast [Wolf and Williams 2008]. It provides space-based, air-borne and in-situ data, metadata and products usable by diverse communities through a global network of satellite-based data dissemination systems. Consequently, GEONETCast products can be received through standard satellite dishes and are directly accessible through the connected computer. However, these products are only available locally (i.e. on the PC, to which the dish is connected) through mostly proprietary data formats and specific software. With the increasing availability of the Web an interoperable, easy and worldwide access to this data is promising but has not been realized yet.

This paper describes an architecture to serve GEONETCast products on the Web using standardized data formats and Web Service interfaces. The architecture is demonstrated by the use case of a Fire Web Service. Detecting fires through satellite data such as Moderate Resolution Imaging Spectroradiometer (MODIS) is one of the important tasks when monitoring large-scale rural areas. The architecture allows users to query up-to-date and historic fire events through a browser-based application. The presented architecture is based on Open Source software, which also motivates the use of open standards.

Section 2 gives an overview of the products available in GEONETCast and how those products can be received. The architecture and its implementation to serve these

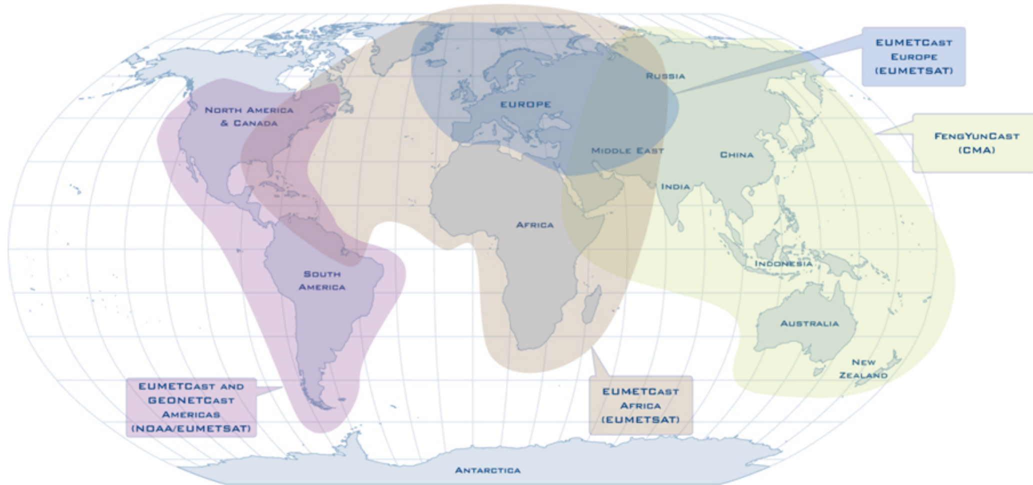


Figure 1. Overview of GEONETCast satellite systems.

products on the Web for the example of the Fire Web Service is presented in Section 3. The use case of the Fire Web Service is then presented in 4. The paper ends with an outlook and conclusion of the presented findings.

2. GEONETCast - an Overview

GEONETCast is a satellite-based dissemination system for environmental data created by remote and in-situ sensors. GEONETCast is thereby a part of the Global Earth Observation System of Systems (GEOSS). In particular, GEONETCast is a task in the GEO Work Plan and is led by EUMETSAT, the United States, China, and the World Meteorological Organization (WMO). Many GEO members and participating Organizations contribute to this task. The dissemination of GEONETCast products is based on the following satellites:

- FENYUNCast(Asia)
- EUMETCast (Africa & Europe)
- GEONETCast Americas (North and South America)

An overview of the different satellite systems and their coverage is also depicted in Figure 1.

Section 2.1 will describe the different thematic types of products and their application. The technical setup of a system for receiving such GEONETCast products is presented in Section 2.2.

2.1. GEONETCast Products

With over 180 products GEONETCast offers a broad thematic range from spectral transmission and climate measures (i.e. surface temperature, precipitation) to disaster management (e.g. fire monitoring). The data is free for use in research and education.

One of the available products is MODIS data, which has been selected as an example for the presented architecture. MODIS is part of Nasa's Earth Observation System. MODIS provides a set of land surface products, which are described in

[Justice et al. 1998]. In particular, MODIS fire data (MOD14) has been selected for this study and also serves as a basis for other derived products of MODIS.

2.2. Technical Setup

This section describes the common technical setup to access GEONETCast products. The setup is depicted in Figure 2 and consists of three steps: receive, store and serve.

The GEONETCast groundstation receives the data from the GEONETCast satellites through a standard TV dish (connected through TV-card) and decodes the received data stream. The GEONETCast Toolbox [Maathuis et al. 2008] acts as a data manager to filter the desired products, which are then stored in a structured way on the data server. The data server is able to manage user access and serves the data through the file system.

It is important to note, that in some setups the data server and the ground receiving station are hosted on the same computer. For the technical setup as used for this implementation these two components are separated on different machines for scalability and maintenance reasons (e.g. the data server has a periodic backup).

Based on this technical setup the architecture is designed and implemented (Section 3). In particular, the data server is the entry point to access the MODIS data for the Fire Web Service.

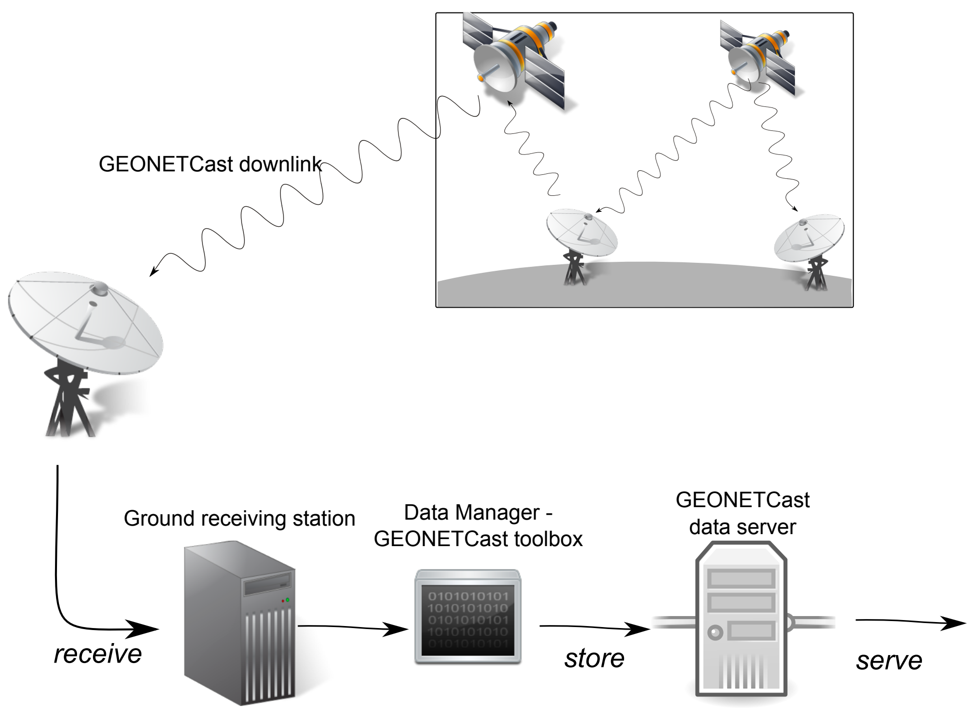


Figure 2. Overview of the technical setup for receiving GEONETCast products.

3. Fire Web Service Architecture

This section describes the architecture of the Fire Web Service and its implementation. Related work has already been reported by [Davies et al. 2009]. They also used MODIS data for monitoring fires and served this data on the Web. However, their setup involved a lot of manual steps and the access to the data was limited to portrayal (no querying possible). Additionally, the Center for Weather Forecast and Climatic Studies (CPTEC) of the National Institute for Space Research in Brazil (INPE) provides a web portal to access data about fire events. This portal only provides limited querying capabilities and does not allow users to integrate the data available in the portal into other applications. The presented architecture in this article is more comprehensive as it is based on live streaming of MODIS data through GEONETCast using Web Service interfaces. These Web Service interfaces allow users to query the data and also to integrate it into other applications such as Google Earth. Finally, the described implementation is based on common Open Source software and can be re-built with low cost, if required.

3.1. Design of the Architecture

For data access on the Web, interoperability enabled by standards plays an important role. Common standards such as established by the Open Geospatial Consortium (OGC) allow users to share data and services out-of-the-box without knowing the implementation details. Therefore, the presented architecture makes excessive use of standards to support the integration of the offered data and services in other applications.

The fire events served by the Fire Web Service are modeled as point-based features. This allows users to integrate low volume data and to query it easily for specific attributes. Both are advantages over the bulky and less structured raster-based data, which requires high network bandwidth for data transmission and advanced tools to extract the required attributes.

For the Fire Web Service and considering the feature-based data for the fire events, the OGC Web Feature Service (WFS) [OGC 2005] has been identified as the appropriate interface. WFS interface allows users to query and access feature data consisting of (multiple) points, lines or polygons. The communication with WFS is based on the Internet Protocol HTTP using an XML-based encoding the so-called WFS Filter encoding. To retrieve specific features from a WFS, the *getFeature* operation is used, which receives messages as WFS Filters. WFS returns feature data (as result of the WFS Filter query) in the Geography Markup Language (GML) or also for instance in KML. KML is the data encoding established and used by Google-based applications. To also create and store new features on a WFS over the Web, a transactional interface has been developed. The additional operations of the so-called WFS-T (T stands for transactional) are insert, update, delete.

The architecture of the Fire Web Service is depicted in Figure 3. The Fire Web Service is accessible via a WFS-T interface as well as through a browser-based application to portray and query the data. Internally, the Fire Web Service uses a spatial database to store the fire data.

The Fire Web Service polls periodically the MODIS data from the GEONETCast data server by searching through the available products and copying the required files locally for further processing. Additionally, the business logic transforms the MODIS data from a raster-based format to the vector-based format and inserts it through the WFS-T interface into the database. The inserted data is then available to the browser-based application of the Fire Web Service but also to external users who can query the fire data

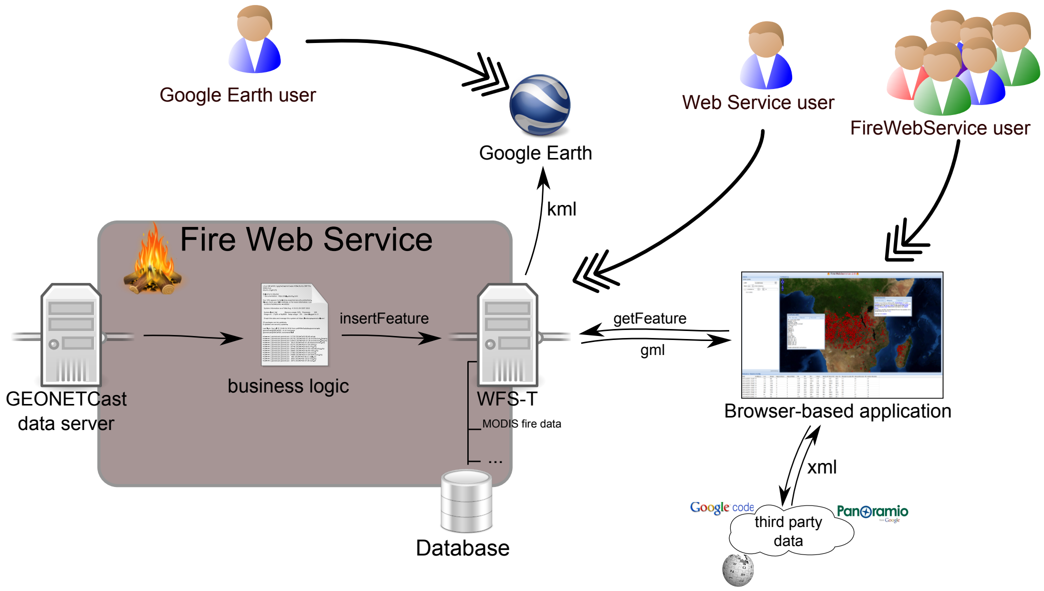


Figure 3. Architecture of the Fire Web Service.

directly and integrate it in their application using WFS interface.

The browser-based application links besides the Fire Web Service other additional data sources. In particular, the browser-based application links data from Wikipedia, images from Panoramio and third-party geocoding functionality. This allows users to get comprehensive information for a particular area of interest. The functionality of the browser-based application is described in Section 4.

3.2. Implementation

Based on the design of the architecture (Section 3.1), this section describes the implementation. The implementation is based on open source tools to meet the requirements of building a low-cost infrastructure. As the applied open source tools come from different sources and initiatives, standards and interoperability are essential, to make these tools work together.

The Fire Web Service polls the MODIS data from the GEONETCast data server every five minutes by a Cronjob. Every new dataset is transformed from raster data into vector data using the fire detection algorithm for MODIS data [Giglio et al. 2003]. This algorithm transforms the raster data into an ASCII or a shape file. The transformed data is then sent in GML format via a WFS Filter Request to the Fire Web Service using the transactional interface of the WFS. An example of such a request is presented in Listing 1. For this implementation, Geoserver has been chosen as the approriate implementation due to its WFS-T interface and its support for various formats such as KML [Deoliveira 2008]. Geoserver stores the data internally in for instance a PostGIS database[1]. The PostGIS database serves as a backend for the Geoserver to compensate the frequent updates of approximately 15000 fire events per day worldwide.

[1]POSTGIS website: postgis.refractions.net/

Listing 1. Example of WFS-T insert request to store fire data.

```
<wfs:Transaction service="WFS" ...">
    <wfs:Insert>
        <de.fws:modisfiredata>
            <de.fws:the_geom>
                <gml:Point>
                    <gml:coordinates decimal="." cs="," ts"">
                    168.143,-16.252</gml:coordinates>
                </gml:Point>
            </de.fws:the_geom>
            <de.fws:appearancedate>2010-06-21T15:13:21Z</de.fws:appearancedate>
            <de.fws:line>1177</de.fws:line>
            <de.fws:sample>506</de.fws:sample>
            <de.fws:confidence>100.0</de.fws:confidence>
            <de.fws:adjcloud>0</de.fws:adjcloud>
            <de.fws:adjwater>0</de.fws:adjwater>
            <de.fws:r2>-1.0</de.fws:r2>
            <de.fws:t21>334.9</de.fws:t21>
            <de.fws:t31>288.4</de.fws:t31>
            <de.fws:power>51.9</de.fws:power>
            <de.fws:mediandt>0.0</de.fws:mediandt>
            <de.fws:meant21>290.3</de.fws:meant21>
            <de.fws:meant31>288.4</de.fws:meant31>
            <de.fws:stddevt21>1.3</de.fws:stddevt21>
            <de.fws:stddevt31>0.5</de.fws:stddevt31>
            <de.fws:stddevdt>1.2</de.fws:stddevdt>
        </de.fws:modisfiredata>
    </wfs:Insert>
</wfs:Transaction>
```

Since the goal was not only to provide access to the data through Web Service interface, but also to visualize it and to combine it with public available data, a browser-based application was created based on OpenLayers[2] and GeoExt[3]. OpenLayers builds a map of the fire data coming from WFS and links it with other sources. The GeoExt library allows us to build a rich web application around those maps including functions such as the built-in support of tables, drop down menus, action bars and pop ups.

The Wiki API[4] was used to provide additional information from Wikipedia about locations with fire events. Every coordinate of a fire is translated on demand into an address via reverse geocoding by the Google Geocoder [5]. The derived address is then used to query Wikipedia for a related article if one is available. If an articles for this specific region is present it is displayed as additional information for the specific fire. The same approach is used for obtaining photos from Panoramio and displaying them in relation to the specific fire.

The described implementation is hosted on a Linux operating system for easy maintenance and high availability. In particular, the Fire Web Service is running on a virtual machine with access to a 2 GHz single core processor with 2 Gigabyte RAM and 7 Gigabyte of storage. Based on our experience, this light-weight configuration provides suf-

[2]OpenLayers website: www.openlayers.org

[3]GeoExt website: www.geoext.org/

[4]Wiki API website: jwikiapi.sourceforge.net/

[5]Google Geocoder website: code.google.com/intl/en-US/apis/maps/documentation/geocoding/

ficient computational performance since the transformation process from MODIS raster data (0.5 MB) to vector data (max. 200 points) takes less than 10 seconds as well as the insert operations into the WFS-T. The transformation and the insert operation are performed every 5 minutes. The system is scalable since every component (e.g. database, WFS-T, business logic) can be hosted on a separate computational node. Some operations such as rendering the map, the requesting geocoding functionality and additional information from wikipedia are performed by the client (without interaction with the Fire Web Service). Consequently, the only task of the Fire Web Service is to transform the data and to provide this data through WFS-T interface. In the future, if the number of users increases, the number of requests to the WFS-T would increase. This can be compensated by a more powerful server configuration hosting the WFS-T. The architecture with its components does not need to be changed.

4. Fire Monitoring Use Case

Fire is a natural phenomenon and poses a threat especially when reaching built-up areas. Although it is a natural hazard a rigorous supression leads to even more severe fires. Using fire for cultivating agricultural land has become an established element. However, an excessive application also leads to severe problems. Finally, monitoring such fires is required and different data is already available such as the MODIS fire products [Justice et al. 2002].

For easy and customized access of this data, Web Service technology can be used. To demonstrate the use case, the described architecture of Section 3 is applied to monitor fire events and to integrate the fire data with other third party sources such as Wikipedia or images from for instance Panoramio. Integrating the data with other third party sources is necessary to provide comprehensive information to the user and is possible due to established standards for data and Web Services. In particular, users can access and query the data through a browser-based client and can use additionally Google Earth to combine the data with other sources.

To provide an easy and user-friendly access, a browser-based application has been designed. An example of the browser-based application is depicted in Figure 4. The different parts of the interface are attached to the several tasks, which are described in the following:

a) *Query data* - The user is able to query the fire data available on the Fire Web Service based on its attributes (including time). By querying the temporal dimension, it is possible to not only receive up-to-date data about fire events, but also historic data. Another important attribute is the confidence value of a fire event, which indicates the possibility of a fire. The confidence value is a result of the applied fire detection algorithm performed on the MODIS data.
b) *Map data* - The result of the query can be inspected in the map view, in which the fire events are located on freely-available imagery such as from Blue Marble [Stockli et al. 2005]. Each fire event on the map can be clicked for further details.
c) *Inspect attribute table* - Based on the query, the data can also be inspected using a tabular view.

The fire data can also be integrated into Google Earth by KML format as supported by the WFS interface. An example of such integration is depicted in Figure 5. The

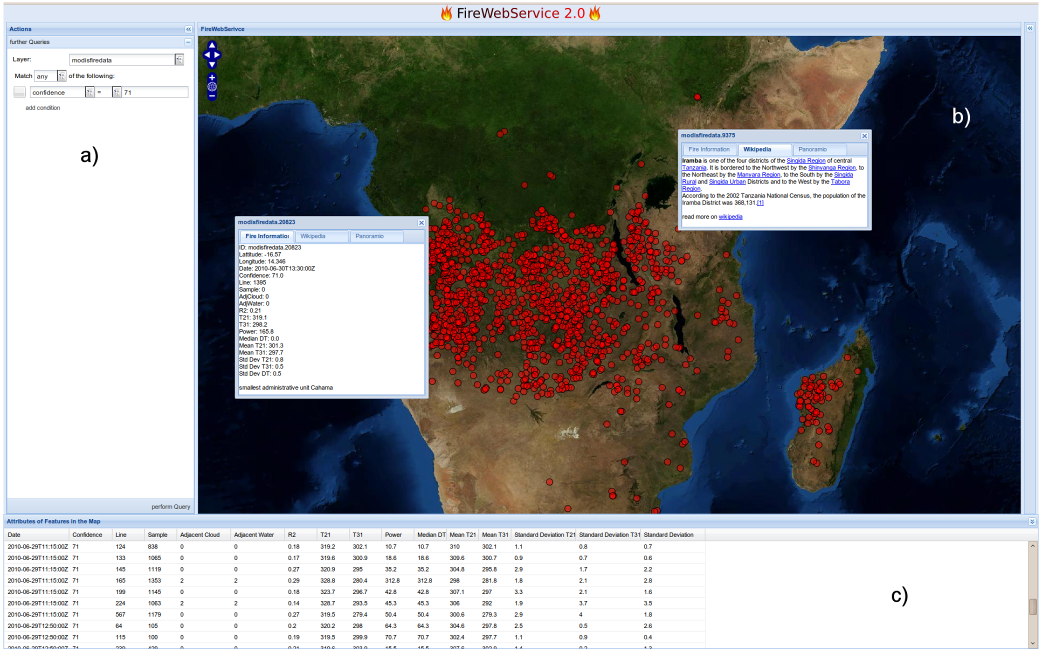


Figure 4. Screenshot of the browser-based application for the Fire Web Service with map of sub-Sahara region - different views: a) query view, b) map view, c) tabular view.

attributes of the fire can be inspected directly by the user. However, a query mechanism for the data such as provided by the browser-based application is not available.

5. Discussion & Conclusion

This article describes a web-based architecture to access satellite-based products for environmental monitoring such as wild fires. In particular, it describes how data products available from the GEONETCast data stream can be integrated into web-based applications. The architecture is demonstrated by the application of a Fire Web Service and the use case of fire monitoring. As the architecture uses low-cost satellite-based products such as GEONETCast and is implemented using Open Source software, it can be re-built with low cost. Additionally, as this architecture is based on open standards for data and Web Services such as GML, WFS and KML, it can be applied easily in Spatial Data Infrastructures, if required. The presented architecture is not exclusively suitable for infrastructure with web access, but also suitable for the intranet. Therefore, this architecture supports also large organizations in development countries with only limited web access.

The GEONETCast data stream and the presented architecture offer various applications from meteorology and hydrology, which can be realized in the future. Especially, serving the GEONETCast products directly on the Web seems to be promising in the future. *In-situ sensor data* from GEONETCast might become available through the Sensor Web [Botts et al. 2008] and thereby be easily integrated with other in-situ data stemming for instance from local organizations. One example, in which MODIS fire products (not accessed through GEONETCast) are accessed through the Sensor Web, is the Advanced Fire Information System (AFIS) of South Africa. The AFIS use case might be enhanced

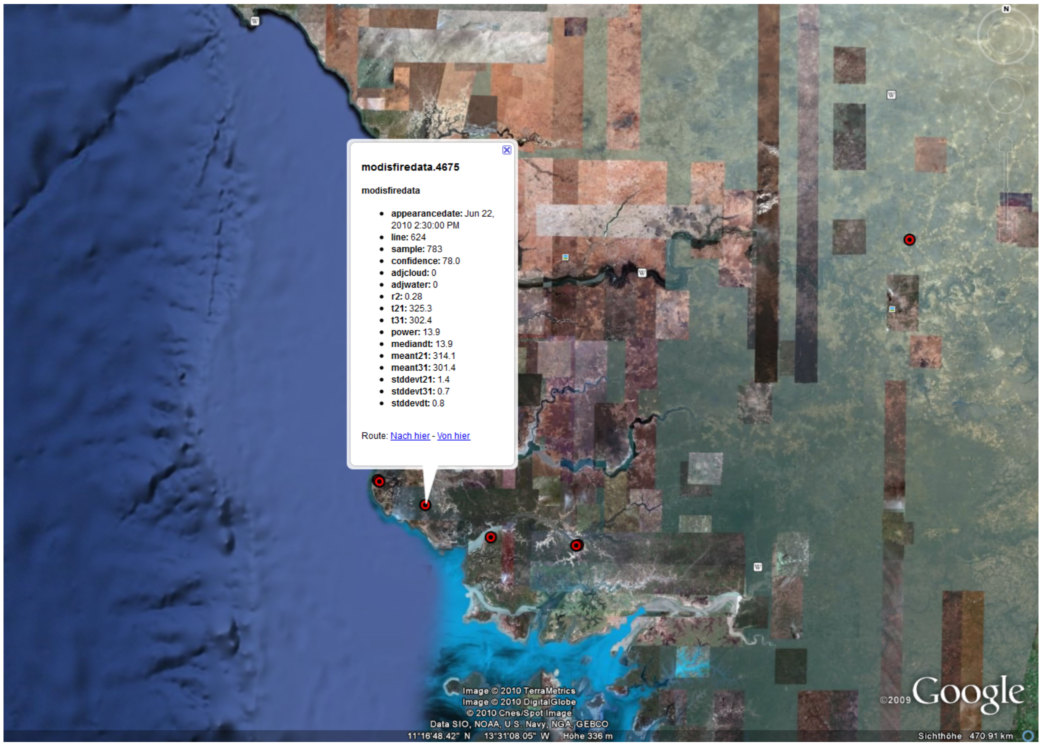


Figure 5. Screenshot of Google Earth with data from the Fire Web Service.

with GEONETCast capability to provide real-time fire data [McFerren et al. 2007]. *Remote sensing data* of GEONETCast should be available through OGC Web Coverage Service, which provides customized access to coverage data such as Eumetsat data. Users are then able to receive specific bands of the coverage for a specific area of interest. These customized coverages are then ready for instance to be processed using a web-based geoprocess model, typically accessible as OGC Web Processing Service.

The next step of enabling web-based access is to provide a notification mechanism, whenever the GEONETCast data stream contains a significant event such as a large fire. Therefore an event-based architecture is required to register for specific events (described as patterns) and to notify accordingly.

Overall, data and processing functionality will become available on the Web as Web Services through architectures as described in this article to enhance the sharing of data and functionality. Thereby the web will become a platform to support environmental monitoring and to improve the awareness of global change. In this context, low-cost satellite-based products will play an important role for development countries in the future.

References

Botts, M., Percivall, G., Reed, C., and Davidson, J. (2008). OGC sensor web enablement: Overview and high level architecture. In *GeoSensor Networks*, number 4540 in Lecture Notes in Computer Science, pages 175–190. Springer, Heidelberg.

Davies, D. K., Ilavajhala, S., Wong, M. M., and Justice, C. O. (2009). Fire Information for Resource Management System: Archiving and Distributing MODIS Active Fire Data. *IEEE Transactions on Geoscience and Remote Sensing*, 47:72–79.

Deoliveira, J. (2008). GeoServer: uniting the GeoWeb and spatial data infrastrucutures. In *Proceedings of the 10th International Conference for Spatial Data Infrastructure*, St. Augustine, Trinidad.

Giglio, L., Descloitres, J., Justice, C. O., and Kaufman, Y. J. (2003). An enhanced contextual fire detection algorithm for modis. *Remote Sensing of Environment*, 87(2-3):273 – 282.

Justice, C. O., Giglio, L., Korontzi, S., Owens, J., Morisette, J. T., Roy, D., Descloitres, J., Alleaume, S., Petitcolin, F., and Kaufman, Y. (2002). The modis fire products. *Remote Sensing of Environment*, 83(1-2):244 – 262.

Justice, C. O., Vermote, E., Townshend, J. R. G., Defries, R., Roy, D. P., Hall, D. K., Salomonson, V. V., Privette, J. L., Riggs, G., Strahler, A., Lucht, W., Myneni, R. B., Knyazikhin, Y., Running, S. W., Nemani, R. R., Wan, Z., Huete, A. R., van Leeuwen, W., Wolfe, R. E., Giglio, L., Muller, J., Lewis, P., and Barnsley, M. J. (1998). The Moderate Resolution Imaging Spectroradiometer (MODIS): land remote sensing for global change research. *IEEE Transactions on Geoscience and Remote Sensing*, 36:1228–1249.

Maathuis, B., Mannaerts, C., and Retsios, B. (2008). The ITC geonetcast-toolbox approach for less developed countries. In *Proceedings of Commission VII*, volume XXXVII of *The international archives of the photogrammetry, remote sensing and spatial information sciences*, pages 1301–1306, Beijing, China. International Society for Photogrammetry and Remote Sensing.

McFerren, G., Roos, S., and Terhorst, A. (2007). Fire alerts for the geospatial web. In Scharl, A. and Tochtermann, K., editors, *The Geospatial Web*, Advanced Information and Knowledge Processing Series. Springer, London, UK.

OGC (2005). Web feature service implementation specification. Implementation specification OGC 04-094.

Stockli, R., Vermote, E., Saleous, N., Simmon, R., and Herring, D. (2005). The blue marble next generation - a true color earth dataset including seasonal dynamics from MODIS. Technical report.

Wolf, L. and Williams, M. (2008). GEONETCast - delivering environmental data to users worldwide. *IEEE Systems Journal*, 2(3):401–405.

Proceedings XI GEOINFO, Campos do Jordão, Brazil, Nov. 28 to Dec. 1, 2010, INPE, p. 67–75.

The role of digital generalization in image segmentation

Mario R. Pettinati[1], Lúbia Vinhas[1], Gilberto Câmara[1], Julio C. L. dAlge[1]

[1]National Institute For Space Research INPE
Av. Dos Astronautas, 1758 12227-010 São José dos Campos SP - Brazil

mario@funcate.org.br, {lubia,gilberto,julio}@dpi.inpe.br

Abstract. *The use of remote sensing imagery to obtain land cover and land use maps is a common task in GIS applications. Segmentation techniques identify closed regions in images, producing vector datasets composed by polygons. Since segmentation is a bottom-up technique, the resulting vector datasets are often too detailed. Thus, we need to use generalization techniques to reduce data storage and generate maps with different degrees of detail at different scales. This paper proposes new method for polygon generalization, useful for vector data sets drawn from remote sensing data.*

1. Introduction

This paper discusses the problem of generalizing land use and land cover maps obtained by the segmentation of remote sensing images. Image segmentation methods are important for remote sensing image analysis. Segmentation divides an image into continuous, disjoint and homogeneous regions. Segmentation algorithms have many advantages over pixel-based image classifiers. The resulting maps are usually much more visually consistent and more easily converted into a geographical information system. However, segmentation techniques tend to produce regions with excessive detail. These polygons (vectors) need to be simplified and generalized to help their use.

Generalization is a practice that originates in Cartography. It calls for selecting which objects will be present in a map, and simplifying shapes and structures, based on criteria of relative importance (Robinson et al. 1995). Automated map generalization is an active research area, focusing mainly on topographic map data (e.g. buildings, road networks etc.) produced by national mapping agencies Mackaness et al. (2007); Stoter et al. (2009); Oosterom (2009). In this paper, we consider generalization techniques that are suited for polygons obtained when a remote sensing image is segmented and then classified. Figures 1 shows an example of a remote sensing image.

Due the raster structure of remote sensing images, segmentation algorithms produce a set of jagged border polygons that are derived from the pixel structure. These polygons have high geometric complexity. The algorithms also make up small regions which are not compatible with the scale suitable to derive maps from a particular image resolution. Figures 2(a) and 2(b) respectively, illustrate these problems.

When we segment a set of images to cover a large geographic region, the result is a large vector mosaic with unnecessary detail. To solve this problem we need an automated digital generalization method. This work aims to review and suggest improvements on methods for polygon generalization, based on line simplification, when applied to remote sensing derived vector data sets.

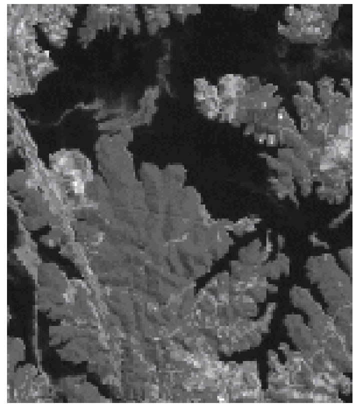
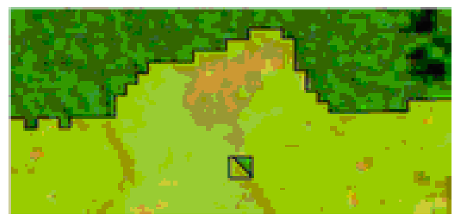


Figure 1. Example of a remote sensing image

(a)



(b)

Figure 2. Examples of problems in image segmentation: (a) irrelevant features and (b) complex vectors

This works is organized as follows. In the section 2 we review some related work. In the section 3 we propose the enhancements. Section 4 shows an example of generalization of a remote sensing derived data and the result of some experiments. Finally, section 5 presents the conclusions of this work.

2. Related work

2.1. Generalization

In the context of digital cartography and GIS, map generalization involves two broadly distinct types of tasks: cartographic generalization and database generalization (Weibel and Jones 1998). Cartographic generalization aims to derive graphic products from a source database. Database generalization deals with the production of multi-level databases that contain diverse data sets at different scales. Database generalization methods do not consider artistic or intuitive components, neither deal with symbolization problems, but prioritize spatial accuracy and completeness. Jones and Ware (2005), considering the prevalence of geographical information access on the internet, recognizes two types of tasks associated to map generalization: semantic generalization and geometric generalization. Semantic generalization deals with the choice of information categories that should be represented, while geometric generalization is concerned with the simplification of shapes and structures that represent individual features.

Geometric generalization can be considered as a sub-process of the overall map generalization process and is done mainly through the application of generalization operators that represent single actions or atomic generalization functionalities. Typical generalization operators include simplification, exaggeration, aggregation, elimination, and displacement. They can be realized with different algorithms, and a reasonable set of tools have been implemented as part of GIS functionality (Foerster et al. 2008). A review about operators and their application to different types of features can be found in Choe and Kim (2007).

Cartographic constraints have been established as a concept to select the appropriate generalization operator or, to control the sequencing of them, in the automation of the generalization process. Constraints are factors, such as topology, proximity, size or shape, that are used to describe object characteristics and relationships required to produce the best result for a specific map scale and type (Neun et al. 2009). A constraint can be described by an appropriate measure that captures the property it expresses, for example, the area of a parcel is a measure for the size constraint (Steiniger and Weibel 2007).

Categorical data is a common data set used in GIS applications. Usually it refers to some spatially continuous phenomenon discretized into a manageable representation. For example, thematic maps of geological units or land use. One possible digital representation for categorical coverages is to use vector structures, more specifically a polygonal subdivision, where each polygon has an attribute value to represent a category or theme associated to that region. There is a lack of methods designed specifically for the generalization of categorical coverages in commercial GIS and cartographic systems, and usually line simplification is the method used. In order to obtain better results the special topological structure of categorical data should be considered (Galanda 2003).

Simplification is an operator used for linear or areal features to simplify unnecessarily detailed geometric data without fundamentally altering the basic shapes. It does not affect the non-spatial component of the data, and should preserve topological and spatial relationships between features. This operator can be implemented by different algorithms and there is no general theory that explains which algorithm is more convenient for the overall map, as well as for individual features. D'alge (2007) also addresses the process of generalization in the digital domain, especially considering categorical data. He performed a serie of generalization experiments for a dataset consisting of vegetation maps for the Brazilian Amazon using an adaptation of the model proposed by Mc Master and Shea (1992) and concludes that operators such as line simplification could be used to generate vegetation maps at different scales, although some further improvement should be done in the algorithms in order to solve potential topological problems.

2.2. Line simplification algorithms

One of the most cited algorithms for line simplification is the Douglas & Peucker algorithm (Douglas and Peucker 1973). The purpose of the algorithm is, given a curve composed of line segments, to find a similar curve with fewer points. It is a recursive algorithm. Initially it selects the first and last points of the curve and considers the line segment between these points and mark them as to be kept. It than finds the furthest point from the line segment. If the point is closer than a given tolerance (an input parameter of the algorithm) to the line segment then any points not currently marked to keep can be discarded. If the point is not closer to the line segment then that point must be kept. The algorithm recursively calls itself with the first point and the furthest point and then with the furthest point and the last point.

Another simplification method is the one based on the concept of effective area (Visvalingam and Whyatt 1993). Their method builds triangles from each three consecutive vertices of the curve and calculates its area. The central vertex of the triangle with the smallest area is eliminated and the algorithm recursively calls itself considering the remaining points. The recursion stops when a given condition is reached, for example, a given number of points are removed.

The methods cited above do not guarantee the maintenance of the original topology of a polygonal subdivision. This is due to the fact that they process each polygon independently, not considering its topological relationships with other geometries of the dataset. This might generate inconsistencies such as polygon self intersection, polygons overlapping or generation of areas not covered by any polygon. These inconsistencies can be fixed by a post processing step (Falls et al. 2005) (Muller 1990).

In this work, we agree that line simplification should be followed by a post-processing step to remove the inconsistencies in the generalized polygonal division. But we also propose enhancements that can be applied to different line simplification algorithms in order to reduce the number of further inconsistencies that might be generated.

3. Enhancements to line simplification algorithms

The focus of this work is on categorical data derived by automated segmentation and classification algorithms applied to remote sensing images, resulting in a contiguous collection of polygons that are topologically consistent, that is, there are no polygons in with self-intersections neither overlapping neighboring polygons in the collection. Each polygon of the collection follows a series of conditions that define their validity according to the geometry model proposed by the Open Geoespatial Consortium (Ryden 2005). However, the polygons usually have an unnecessarily high geometrical complexity and possibly very small artifacts, inconsistent with recommended cartographic scales for the vector products.

Considering the characteristics of the vector categorical coverages previously described, we propose two enhancements that can be applied to different line simplification algorithms in order to achieve better results. The first enhancement refers to the concept of anchor vertices. Anchor vertices are defined as the vertices that are part of three or more distinct segments. As an example, consider the Figure 3, which represents a small portion of a categorical coverage containing polygons *P1*, *P2* and *P3*. The vertices *v1*, *v2*, *v5* and *v6* are considered anchors, meaning that they should not be deleted or removed during the line simplification phase. In this example, vertices *v3* and *v4* can be removed during the simplification phase.

The second enhancement represents a way of propagating the simplification of a polygon to its neighbors: every time a vertex is removed from a polygon it is also removed from any other polygon that includes the same vertex. Figure 4 illustrates the propagation of the simplification. Figure 4(a) shows that vertex *V* is present in both Polygons *A* and *B*. Figure 4(b) shows a simplification step of polygon *A*, which removed the Vertex *V* from it. Figure 4(c) shows the propagation of the simplification, which propagated the remove of Vertex *V* to polygon *B*.

The generalization process can be summarized as follows:

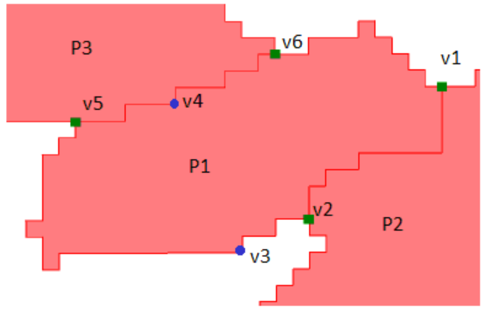


Figure 3. Anchor Vertices (*v1*, *v2*, *v5*, *v6*)

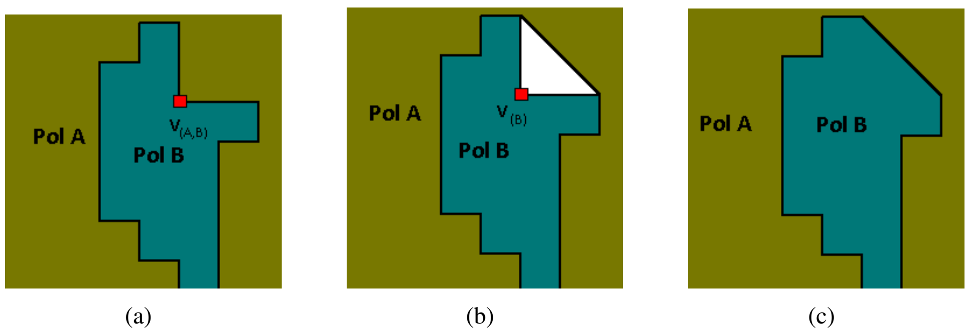


Figure 4. Propagation of the simplification: (a) Vertex *V* existing in polygons *A* and *B*; (b) *V* removed from polygon *A* and (c) *V* should be removed from polygon *B*

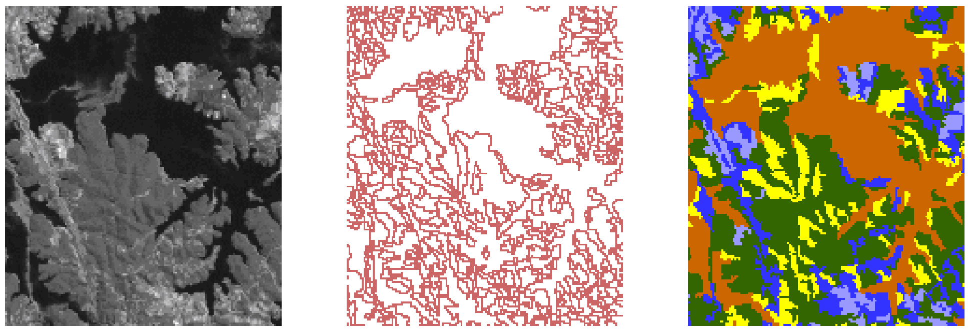
- Step 1: detect and label the anchor vertices;
- Step 2: apply the modified version of a line simplification algorithm on each polygon of the categorical coverage;
- Step 3: remove the inconsistencies that might have been generated during Step 2 (two types of consistencies are considered: polygons with self intersection lines and polygons overlapping).

4. Example of generalization of remote sensing derived data

This section illustrates the geometric simplification that can be performed using the proposed enhancements. The algorithms were implemented in C++ language and tested on real data. We used a Landsat-TM image of a region in São Paulo state, in Brazil as shown in Figure 5(a). Figures 5(b) and 5(c) show, respectively, the result of a segmentation and subsequent classification in 5 land cover classes. The segmentation and classification were performed using SPRING GIS (Camara et al. 1996). It should be noticed that it is not in the scope of this work to evaluate the quality of the segmentation and classification performed, since it depends on various factors, which do not affect the main motivation of this work.

Users of remote sensing imagery usually have to deal with the issue of choosing the most appropriate cartographic scale to generate products derived from a given image, or an image with a given spatial resolution, and there it not a definitive way to relate image spatial resolution to a maximum cartographic scale. Possible approaches are discussed in

(Boggione et al. 2009). During thematic mapping, object location accuracy requirements are usually milder than those for topographic maps. For example, LANDSAT-class images (30 m resolution) can be used for thematic mapping up to 1:60,000 scales, depending on the map theme. In this experiment the data original scale is set at 1:60,000, thus guiding the selection of the parameters used in the line simplification phase.



(a) (b) (c)

Figure 5. Example of a categorical map derived from a remote sensing image: (a) original image (b) segmented image and (c) classified image

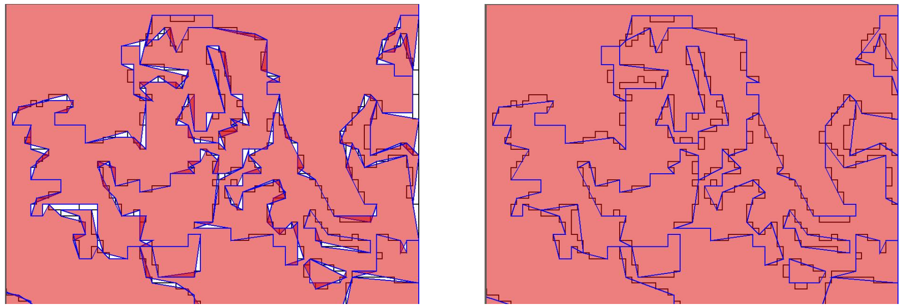
As described in section 3, two line simplification algorithms were modified in order to consider the list of anchor vertices: the Douglas & Peucker (DP) and the Effective Area (EA). The enhanced versions of the algorithms were applied to the categorical coverage shown in Figure 5(c). The dataset consisted of 1,464 polygons with 113,627 vertices.

Figure 6(a) shows the result of simplifying the categorical coverage using the DP algorithm without the anchor vertices and in Figure 6(b) with the anchor vertices (for legibility only a small part of the coverage is shown). The input polygons are the ones with a black border and the simplified ones are in red with blue border. The original DP algorithm removed 76,978 vertices (67.7% of the total), while DP plus anchor vertices removed 79,072 vertices (69.6% of the total). A possible explanation for the increasing number of removed vertices by the enhanced version is that the DP algorithm is dependent on the initial vertex of the segment that it uses to make the simplification. So, as the enhanced version also simplifies the adjacent polygons, the simplification power of the algorith might have been increased.

Figure 7(a) shows the use of the EA without the anchor vertices and Figure 7(b) with the anchor vertices. The input polygons are the ones with a black border and the simplified ones are in red with blue border. In this case, the original EA algorithm removed 63,286 vertices (55.7% of the total), while EA plus anchor vertices removed 62,008 vertices (54.6% of the total).

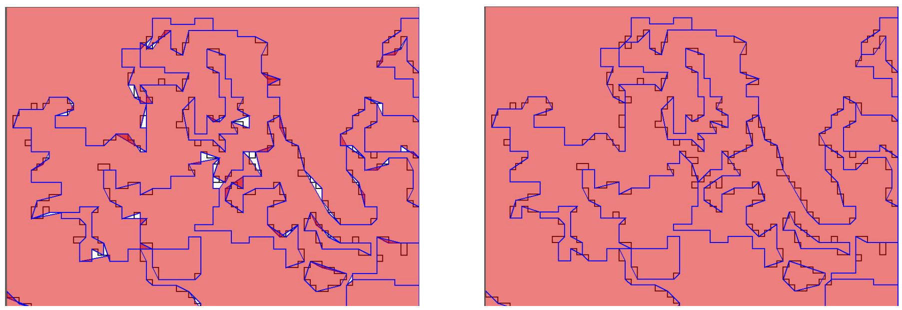
Although the reduction of vertices was similar using DP and EA algorithms, their original versions introduced topological inconsistencies such polygon overlapping and areas that are not covered by any polygon. The introduction of anchor vertices solved such inconsistencies.

This works deals only with pure geometrical generalization, based on line simplification algorithms applied to a polygonal division. A complete generalization process



(a) (b)

Figure 6. Simplification using the DP algorithm: (a) simplification without anchor vertices and (b) simplification with anchor vertices



(a) (b)

Figure 7. Simplification using the EA algorithm: (a) simplification without anchor vertices and (b) simplification with anchor vertices

would have also to consider the semantic structure of the categorical data. An indicative of how the simplification affected the semantic of the data can be the area of each class before and after the simplification. For the five classes in this categorical coverage, simplification using the enhanced versions of the algorithms produced a variation of less than 1%, comparing to the area calculated in the original raster categorical map.

5. Conclusions

Due the raster nature of remote sensing images, segmentation algorithms produce regions that are closely linked to the pixel structure, that is, a set of jagged border polygons. That implies in polygons with a unnecessarily high geometric complexity. Segmentation can also generate very small regions, associated with a single pixel for example, which are not necessarily compatible with appropriate scales of maps derived from a particular image resolution. Geometric generalization has to be considered as part of this process in order to overcome these problems.

In this type of data, an important constraint to be considered is the maintenance of topological consistency. In order to do that, we have introduced some important enhancements to line simplification algorithms. The first enhancement is the labeling of some vertices as anchors which should not be removed. The second enhancement is that every time a vertex if removed from a polygon it is also removed from any other polygon that include the same vertex. This last enhancement represents a way of propagating the simplification of a polygon to its neighbors. These two enhancements are the key factor to maintain the topological consistency after the simplification process.

We presented some experimental results showing the application of our method. Initially we intended to simplify the raw result from segmentation/classification to generate a vector categorical coverage consistent with the scale associated to the spatial resolution of the images, without unnecessary geometric complexity. A further development of this work is to apply the enhanced versions of the algorithms to generalize categorical coverages at smaller resolutions for different purposes, for example to facilitate its access on the internet.

References

Boggione, G. A., Silva, M. V. A., Carvalho Junior, N. R., Teles, T. L., and Nazareno, N. R. X. (2009). Definição da escala em imagens de sensoriamento remoto: uma abordagem alternativa. In *Anais XIV Simpósio Brasileiro de Sensoriamento Remoto, Natal, Brasil, 25-30 abril 2009, INPE*, pages 1739–1746.

Camara, G., Souza, R. C. M., Freitas, U. M., and Garrido, J. (1996). Spring: Integrating remote sensingand gis by object-oriented data modelling. In *Computers & Graphics, 20(3)*, pages 395–403.

Choe, B. N. and Kim, Y. G. (2007). Framework and workflows for spatial database generalization. In *Transactions in GIS, 11(1)*, pages 101–104.

D'Alge, J. C. L. (2007). Generalização cartográfica em sistemas de informação geográfica: Aplicação aos mapas da vegetação da amazônia brasiliera.

Douglas, D. and Peucker, T. K. (1973). Algorithms for the reduction of the number of points required to represent a digitized line or its caricature. In *The Canadian Cartographer*, volume 10, pages 112–122.

Falls, C., Liu, Y., Snoeyink, J., and Souvaine, D. (2005). Testing shortcuts to maintain simplicity in subdivision simplification. In *Canadian Conference on Computational Geometry (CCCG'05)*, pages 35–38.

Foerster, T., Morales, J., and Stoter, J. E. (2008). A classification of generalization operators formalised in ocl. In *Proceedings of GI-days 2008*, pages 141–156. Munster, Germany: Ifgi Prints.

Galanda, M. (2003). Agent based generalization of polygonal data.

Jones, B. C. and Ware, M. J. (2005). Map generalization in the web age. In *International Journal of Geographical Information Science , 19 (8-9)*, pages 859–870.

Mackaness, A. W., Ruas, A., and Sarjakoski, L. (2007). Generalisation of geographic information: cartographic modelling and applications. *Elzevier*.

Mc Master, R. B. and Shea, S. K. (1992). Generalization in digital cartography. *Washington: Assoc. of American Geographers*.

Muller, J. C. (1990). The removal of spatial conflicts in line generalization. In *Cartography and Geographic Information Science*, volume 17, pages 141–149(9).

Neun, M., Burghardt, D., and Weibel, R. (2009). Automated processing for map generalization using web services. In *Geoinformatica, 13*, pages 425–452.

Oosterom, P. V. (2009). Research and development in geo-information generalisation and multiple representation. In *Computers, Environment and Urban Systems , 33*, pages 303–310.

Robinson, H. A., Morrison, J. L., Muehrcke, P. C., Kimerling, A. J., and Guptill, S. C. (1995). In *Elements of Cartography.* New York: John Wiley & Sons Inc.

Ryden, K. (2005). Opengis® implementation specification for geographic information - simple feature access - part 1:common architecture. *Version 1.1.0, OGC 05-126.*

Steiniger, S. and Weibel, R. (2007). Relations among map objects in cartographic generalization. In *Cartography and Geographic Information Science, 34 (3)*, pages 175–197.

Stoter, J., Burghardt, D., Duchêne, C., Baella, B., Bakker, N., Blok, C., and et al. (2009). Methodology for evaluating automated map generalization in commercial software. In *Computers, Environment and Urban Systems , 33*, pages 311–324.

Visvalingam, M. and Whyatt, D. J. (1993). Line generalisation by repeated elimination of points. In *Cartographic J.*, volume 30, pages 46–51.

Weibel, R. and Jones, C. B. (1998). Computational perspectives on map generalization. In *Geoinformatica , 2 (4)*, pages 307–314.

Proceedings XI GEOINFO, Campos do Jordão, Brazil, Nov. 28 to Dec. 1, 2010, INPE, p. 76–86.

Change Allocation in Spatially-Explicit Models for *Aedes aegypti* Population Dynamics

Raquel M. Lana[1], Tiago G. S. C. Carneiro[1], Nildimar A. Honório[2], Cláudia T. Codeço[3]

[1]Computer Science Department, ICEB, UFOP, 35400-000, Ouro Preto- MG

[2]Laboratory of Hematozoa, Oswaldo Cruz Institute, Fiocruz, [3] Program for Scientific Computing, Fiocruz

Manguinhos, CEP, Rio de Janeiro- RJ

raquelmlana@gmail.com, tiago@iceb.ufop.br, nahonorio@gmail.com, codeco@fiocruz.br

Abstract. *This work proposes a new approach to couple Aedes aegypti population dynamic models with local scale spatially-explicit computational models for the Geographical Space. A geographic database was developed for a neighborhood in Rio de Janeiro, RJ, and used to estimate the spatial pattern of mosquito infestation to estimate areas of epidemic risk.*

1. Introduction

Models describing the population dynamics of *Aedes aegypti*, classified as deterministic [Ferreira and Yang 2003] or stochastic [Otero et al. 2006; Focks et al. 1993A], share a common structure based on System Theory [Bertalanffy 1975]. A typical example [Ferreira e Yang 2003] is shown in Figure 1. In this model, the dynamics of a mosquito population is modelled as the flow of individuals between stocks, denoted in the diagram by four rectangles: E(t) - eggs, L(t) – larvae, P(t) – pupae, and W(t) – adults. Stocks are connected by arrows, *f1*, *f2* and *f3*, representing the flow of individuals between life stages: $\sigma 1$ - egg to larva, $\sigma 2$ – larva to pupa, and $\sigma 3$ – pupa to adult. Flow occurs at rates that are temperature dependent. New individuals enter in the population by birth (at a rate *ovip*) and mortality rates are stage-specific (*m1*, *m2*, *m3* and *m4*). This model structure represents the main demographic events in the life cycle of an *Aedes aegypti* population and its formulation in the form of dynamic equations allows the simulation of its temporal dynamics. Since space is not explicited, however, the model does not describe the distribution of *Aedes aegypti* through space.

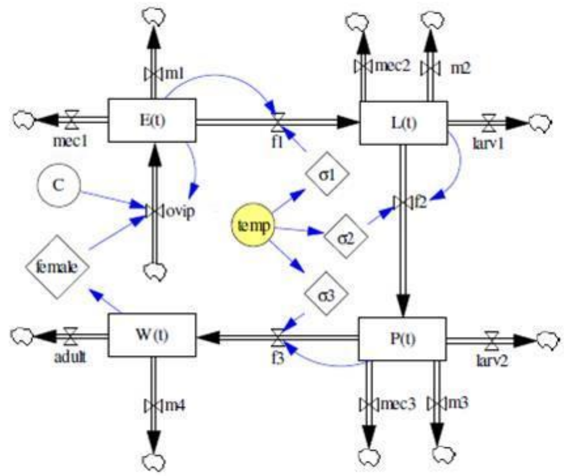


Figure 1. Flow diagram describing *Aedes aegypti* life cycle (adapted from Ferreira e Yang, 2003). Temperature (*temp* in yellow circle) controls three development rates: *σ1* - egg to larva, *σ2* – larva to pupa, and *σ3* – pupa to adult; *ovip* is the oviposition rate; *m1*, *m2*, *m3* and *m4* are natural stage-specific death rates, *mec1*, *mec2*, *mec3* is a death by breeding site removal (mechanical control); *Larv1* and *larv2* are death rates induced by larvicidal control. The arrow "Adult" indicates the death rate by adulticide. *C* is the carrying capacity of the area.

To understand the spatial-temporal dynamics of these populations, this work proposes a new approach to couple *Aedes aegypti* population dynamic models with local scale spatially-explicit models, which are integrated with geographical databases. The goal is to calculate, at each simulation time step, the variation in population size given by the dynamic models and allocate it in a grid of regular cells that represents the Geographical Space.

2. Theoretical Foundations

Few computational models are capable of simulating *Aedes aegypti* population spatial-temporal patterns. Otero et al. (2008) proposed a stochastic spatially-explicit model in which changes are modelled considering cells as occupied by autonomous mosquito populations interconnected by flows of flying individuals. A similar approach was used by Magori et al. (2009). However, more realistic simulations of *Aedes aegypti* life cycle can be achieved when population dynamic models [Focks et al. 1993a] considers the spatial distribution of breeding sites in their formulation as well as the dynamics of the aquatic stage of the mosquitoes (larvae and pupae). The breeding site density per house and the house density per area are model parameters. However, in both studies the simulation experiments were conducted in artificial spaces where the breeding site density and local temperature were also synthetic. In other words, the models were not integrated with geographical databases.

Remote sensor images and digital maps were used by Tran and Raffy (2005) to develop a model to assess Dengue transmission processes in the municipality of Iracabouro, French Guiana. Chang et al. (2009) also used geographical data to help Dengue control specialists to prioritize specific neighborhoods for targeted control interventions.

3. Methodology

A deterministic *Aedes aegypti* population dynamic model, modified from Ferreira and Yang (2003) [Lana 2009] was implemented in the TerraME modeling environment [Carneiro 2006]. The implemented model was calibrated using data from a real urban area, whose socioeconomic and biophysical properties were organized into a geographical database implemented in TerraLib [Camara et al. 2000]. The allocation procedure for spatialization was also implemented in TerraME. The kernel estimator provided by TerraLib was used to parameterize the allocation procedure.

3.1. Data

The data used in this work was collected by Honório et al., (2009), who weekly monitored the *Aedes aegypti* population in Higienópolis district (Figure 2), Rio de Janeiro, RJ, Brazil, during 1.5 years, using ovitraps. Ovitraps are traps that attract mosquito females looking for places to lay eggs. [Fay and Eliason 1966; Reiter et al., 1991]. Forty ovitraps were randomly placed in a 0,25 km^2 area. Each week, the ovitraps' contents were taken to the laboratory and number of *Aedes aegypti* eggs was counted. After that, traps were cleaned and returned to the houses. Week mean air temperature during the period was obtained from the nearest meteorological station, located at the Rio de Janeiro's international airport.

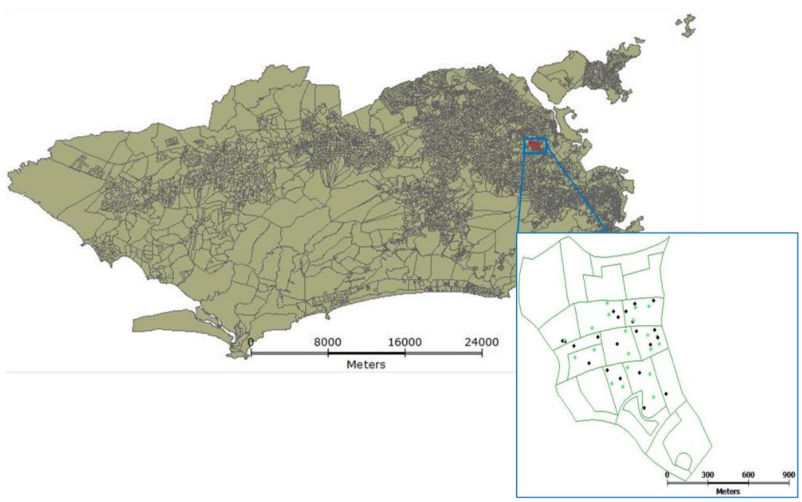


Figure 2. Study area and ovitrap locations – Higienopólis, Rio de Janeiro, RJ.

3.2. Population Dynamic Model

Four differential equations describe the rate of change of mosquito abundance, per life stage: eggs, larvae, pupae and adult (Figure 1).

$$\frac{dE}{dt} = ovip(t)W(t)\left[1 - \frac{L(t)}{C}\right] - \left[\sigma_1(t) + m_1(t) + mec_1(t)\right]E(t),$$ **Equation 1**

$$\frac{dL}{dt} = \sigma_1(t)E(t) - \left[\sigma_2(t) + m_2(t) + larv_1(t) + mec_2(t)\right]L(t),$$ **Equation 2**

$$\frac{dP}{dt} = \sigma_2(t)L(t) - \left[\sigma_3(t) + m_3(t) + larv_2(t) + mec_3(t)\right]P(t),$$ **Equation 3**

$$\frac{dW}{dt} = \sigma_3(t)P(t) - \left[m_4(t) + adult(t)\right]W(t).$$ **Equation 4**

Equation 1 describes the dynamic of the egg stock. Eggs are layed at a temperature and density-dependent rate. *ovip(t)* is a quadratic function describing the effect of temperature on oviposition rate. Individuals leave the egg stage by either natural death, induced death (by mechanical control) or by ecloding into larvae. Equation 2 describes the dynamic of the larva stock. Larvae eclode at a temperature-dependent rate. Individuals leave the larva stage by either natural death, induced death (by mechanical control and larvicide) or by evolving into pupae. Equation 3 describes the dynamic of the pupa stock. Pupa emerges at a temperature-dependent rate as well. Individuals leave the pupal stage by either natural death, induced death (by mechanical control and larvicide) or by emerging into adult. Equation 4 describes the dynamic of the adult female stock that lay eggs. Female adults also emerge at a temperature dependent rate and die by either natural death or induced death (by adulticide).

In comparison to Ferreira and Yang (2003), this model has the following modifications:

1- It uses the equation proposed by Sharpe and DeMichelle (1977) to describe the temperature-dependent developmental rates. This equation describes the temperature dependent rate of development of a poikilothermic organism as the temperature dependent rate of activation and deactivation of an enzyme.

2- Eggs are layed at a temperature and density-dependent rate. A quadratic relation between oviposition and temperature sampled was found to Higienópolis district (Figure 3).

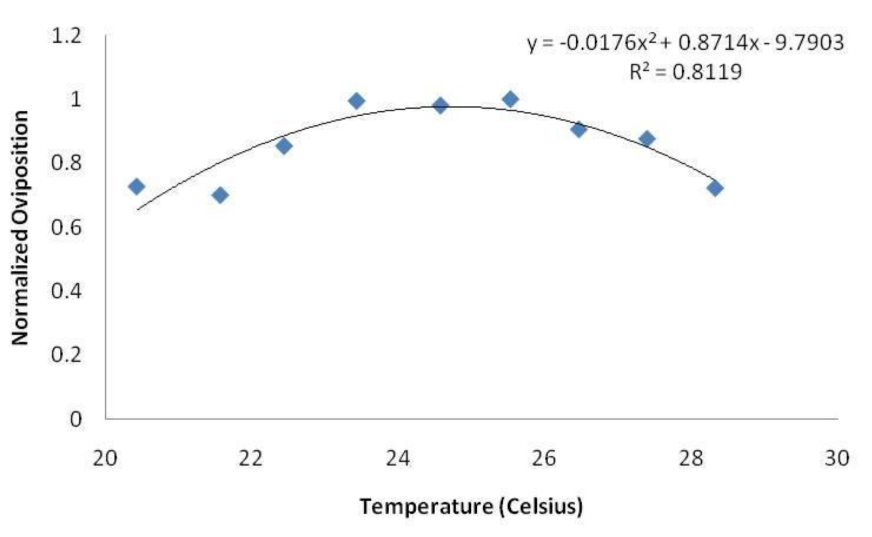


Figure 3. Quadratic function describing the relationship between oviposition rate and air temperature. The source of data is of Honório et al. (2009).

3.3. Calibration and Validation

The model presents only one free parameter, the carrying capacity *C*. The other parameters are maintained fixed (Table 1).

Table 1: Parameters used in the dynamic model

Parameter	Value
ovip (t)	(Quadratic function in Figure 3)
σ1(t), σ2(t), σ3(t)	Fixed (equation proposed by Sharpe e DeMichelle, 1977)
m1(t), m2(t), m3(t)	Fixed (1/100, 1/3, 1/70 respectively)
mec1(t), mec2(t), mec3(t)	Fixed (0)
larv1(t), larv2(t)	Fixed (0)
adult(t)	Fixed (0)
C	Fitted

To calibrate the carrying capacity *C* to the Higienópolis area, the ovitrap data was divided into two subsets. The first subset (green points in Figure 2) was used to calibrate the single free parameter using a Monte Carlo method (Rubinstein and Kroese 2007) to minimize the mean quadratic error. 2000 iterations were performed in 10000 Monte Carlo experiments.

Once calibration was achieved, the second subset was used to simulate the model again and another error value was obtained and compared to the error obtained by the calibration process. Since this error was lower than the calibration error, we considered that the model calibration was adequate. The division into subgroups was done to avoid clusters and guarantee a similar temporal distribution of the two sets of sample points.

3.4. Geographical Database

Several layers of information for the study area were integrated into a geographical database developed in the TerraView geographic information system (GIS), version 3.2.0. Informations included point maps with ovitrap locations, number of eggs collected per ovitrap per week, census tracts in the area, census data, and spatial location of schools, houses, water reservoirs (Pereira Passos Institute - Rio de Janeiro Prefecture, 2000).

For simulation purpose, a grid of regular cells (10 x 10 meters) was created and its cells were used to store model's inputs and outputs.

3.5. Scale Issues and Estimation of the Infestation Spatial Pattern

To generate maps of the spatial distribution of the *Aedes aegypti* population in Higienópolis, we used two approaches:

3.5.1. Census tract scale

In the first approach, the ovitrap data was aggregated by census tract. Higienópolis is divided in 22 census tracts and the study area has 10 census tracts (Figure 4a). A census tract with only one trap was excluded from the analysis (marked with a white X in Figure 4a). The carrying capacity of the dynamic population model was calibrated separately for each census tract. Figure 4b shows that estimated carrying capacity and the mean number of eggs collected per census tract. We observed that, discounting a scale factor (5.387), the carrying capacity captures the spatial variation in egg density.

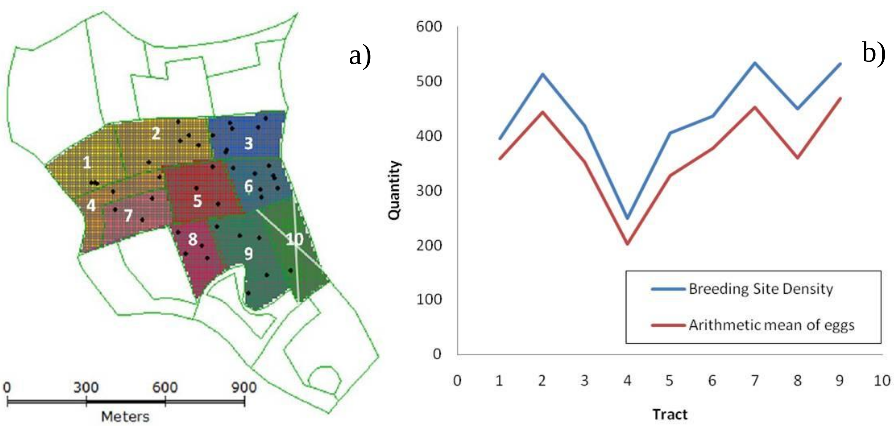


Figure 4. (a) The Higienópolis district divided in census tracts (a). The census tracts within the study area are colored and enumerated. The 10th census tract was excluded from the analysis. (b) Comparison between the estimated carrying capacity per census tract and the mean number of eggs. The red line was increased five times to facilitate comparison with the blue line. Applying a linear regression we obtain the equation *C = 37.48 + 5.387*mean (Eggs),* with r2 = 96.5%.

3.5.2. Kernel Estimator

Our second approach aimed at producing a spatially continuous estimate of mosquito abundance (defined by our grid). To achieve this goal, we used a variation of the Kernel estimator for point events with an associated real value (Bailey and Gatrell, 1995). The method smooths the surface interpolating the density of eggs in each location, without modifying the data statistical characteristics and variability. In this work, the Kernel estimator with adaptive radius provided by TerraView software was used to generate 78 weekly maps of egg density from the observed data. These maps could be used to parameterize the allocation model proposed below. Since models with many parameters are difficult to calibrate, as a first approximation, the 78 maps were summarized into an unique average map of egg density. Therefore, the final kernel map is an aggregation of all 78 weeks into a single map (Figure 5). This map was used as input to the spatially-explicit allocation model described below.

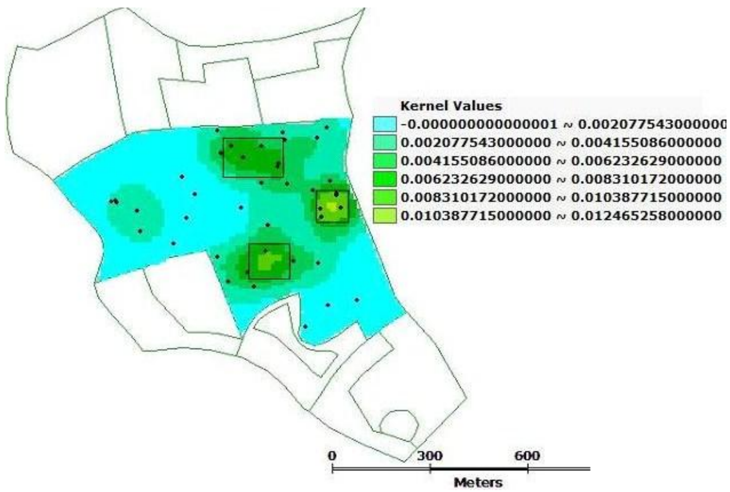


Figure 5: Average kernel map of egg density.

3.6. Allocation model for spatialization of the *Aedes aegypti* population

Some assumptions were considered in order to develop an *Aedes aegypti* population allocation procedure.

a) Cells of 10 by 10 meters were generated and adopted as the spatial scale for this approach.

b) The estimated egg population is distributed through space according to the kernel map of egg density (Figure 5). It is important to note that we found the carrying capacity to be proportional to the mean egg density, so the underlying assumption is that eggs are distributed according to the carrying capacity. For example, if the calculated average egg population is 400 eggs, some cells will have a null quantity of eggs, others can have 100, 200, 400 eggs, or even a higher concentration of eggs.

The resulting algorithm used to allocate egg populations is shown in Figure 6. It traverses the cellular space allocating the *Aedes aegypti* population. The cells are visited in a decreasing order of egg density estimated by the average kernel map of egg density. At each cell, the algorithm deposits a quantity of eggs that is proportional to the average capacity of an adult female to lay eggs [Otero et al. 2006] and proportional to the local egg density estimated in the average kernel map.

```
for each time step t do
    estimatedPop = DynamicModel (t)
    allocatedPop = 0
    while (allocatedPop < popEstimated) do
        for each cell in decresingOrder ( averageKernelMap )
                quantity = 63 * cell.KernelIntensity
                cell.eggPop = cell.eggPop + quantity
                allocatedPop = allocatedPop + quantity
        end for each cell
    end while
    t = t + 1
end for each time step
```

Figure 6. *Aedes aegypti* population dynamic allocation algorithm.

4. Results and Future Works

This work presents an approach to allocate the *Aedes aegypti* population on the real space. The allocation algorithm uses a Kernel estimator based map and a ranking mechanism to traverse the space allocating the mosquito population in a 10x10m cell grid (change). In this study, the model was parameterized and integrated to a geographical database for the Higienópolis district from Rio de Janeiro city, RJ, Brazil.

The population dynamic model, parameterized for the study site in Rio de Janeiro, presented some problems to fit to the data (Figure 7). Contrary to our expectations, the observed time series appears to be less responsive to temperature than expected by the model. This result suggests that other variables may have a bigger effect on the control of the week oviposition rates, for example, the rain regime or air relative humidity. In the winter, we observed the largest discrepancy between simulated and observed oviposition. Most of the time, the model underestimates the quantity of weekly deposited eggs. Several factors can be contributed for this imperfection. The oviposition statistic is just based on 1.5 years of sampling. Besides, the Higienópolis district is not an isolated place as the model assumes, thus the mosquito population can receive and lose individual for neighborhoods.

Figure 7. Graph of comparing between Observed oviposition (OO) and Simulated oviposition (SO). The blue line, *temp*, is the temperature time series.

Despite the simplifications introduced in the spatialization of the model, the model was capable of capturing the spatial pattern of mosquito abundance, with four hotspots that vary in intensity through time. Despite this spatial similarity, though, simulated and observed maps differ in the intensity of the mosquito abundance (Figure 8). During the warm seasons, mosquito abundance is less intense in simulated maps (black background) than in the observed maps (white background). The opposite occurs during the cold seasons. These discrepancies occur due to the errors in the estimation of population size by dynamic model discussed above.

The allocation procedure introduced here is a simplification. The method neglects the interactions between spatial heterogeneity and the growth of the mosquito population. It considers the whole district as a homogeneous area during computation of the population size and, then, it distributes the individuals over the space. It does not consider the spread of mosquito by flight. Other simplification is the use of an egg density average map to base the allocation. The average map fixes the spatial structure while the intensity of eggs changes during the time. Hence, we consider that the average map is only an indicator of average risk.

Future works will investigate integrated methods to develop spatial dynamic models for the *Aedes aegypti* life cycle. In this new approach, the spatial structure will be dynamic and population dynamics will be governed by autonomous populations located in each cell, as in Figure 9. Dispersion of mosquitoes by flight will be also considered. These improvements will allow for the simulation of control strategies to evaluate their efficiency. For instance, strategies as the use insecticides in risk areas or the elimination of breeding sites of certain regions can be evaluated through simulated scenarios.

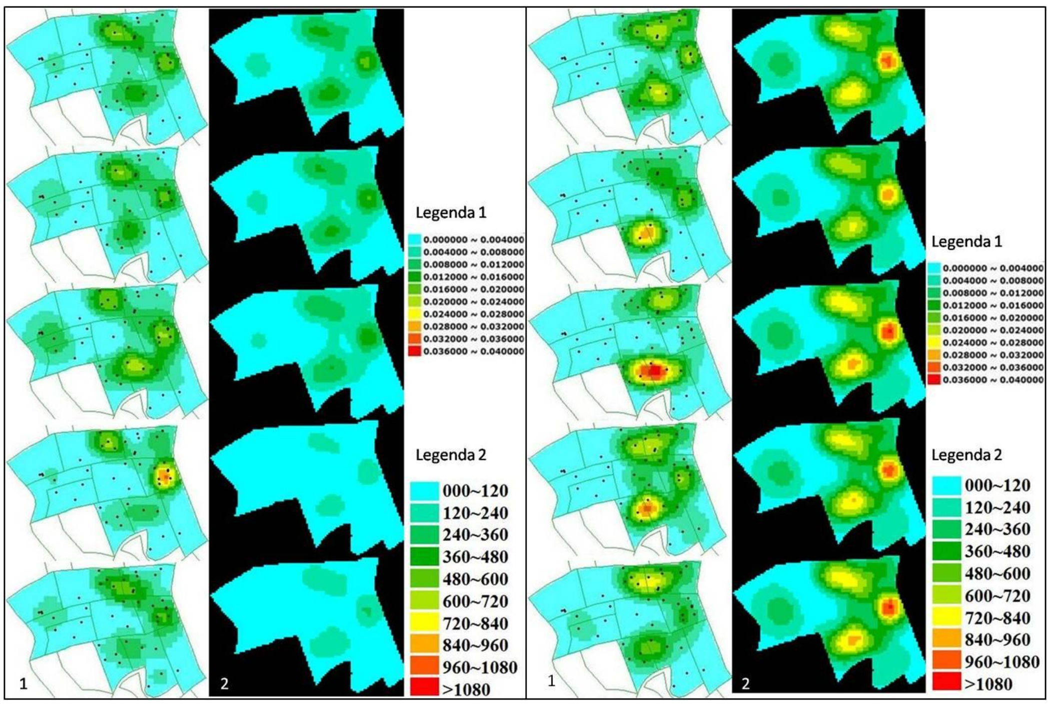


Figure 8. Comparing observed (white background) and simulated (black background) infestation maps – left maps show results for the winter season, right maps show results for the summer season.

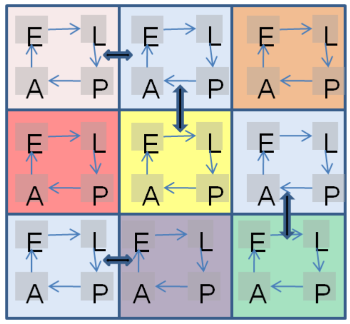


Figure 9. Autonomous *Aedes aegypti* populations occupy each space cell. Mosquitoes may fly to the neighbor cells indicated by blue arrows. *E*: egg, *L*: larva, *P*: pupa and *A*: adult.

Acknowledgments

I thank the Pos-Graduate Program in Ecology of Tropical Biomes and Laboratory of Modelling and Simulation of Terrestrial Systems in Federal University of Ouro Preto (UFOP), Minas Gerais, Brazil. Reviewers for contributions. This work was supported by CNPq and UFOP.

5. References

Bailey, T.C. and Gatrell, A.C., (1995) "Interactive Spatial Data Analysis", Essex: Longman.

Bertalanffy, L. V. (1975) "Teoria Geral dos Sistemas", Ed. Vozes.

Câmara, G., R. Souza, B. Pedrosa, L. Vinhas, A. Monteiro, J. Paiva, M. Carvalho and M. Gattass (2000), TerraLib: Technology in Support of GIS Innovation. II Workshop Brasileiro de Geoinformática, GeoInfo2000, São Paulo.

Carneiro, T.G.S., (2006) "Nested-CA: a foundation for multiscale modeling of land use and land change", São José dos Campos, INPE.

Chang, A. Y., Parrales, M. E., Jimenez, J. Sobieszczyk, M. E., Hammer, S. M., Copenhaver, D. J., Kulkarni, R. P., (2009) "Combining Google Earth and GIS mapping technologies in a dengue surveillance system for developing countries", International Journal of Health Geographics, p.1-11, http://www.ij-healthgeographics.com/content/8/1/49.

Fay, R.W., Eliason, D.A., (1966) "A preferred oviposition site as surveillance method for Aedes aegypti", Mosq. News 26, p.531-535.

Ferreira, C.P., Yang, H.M., (2003) "Estudo Dinâmico da População de Mosquito Aedes aegypti", TEMA Tend. Mat. Apl. Comput. 4(2), p.187-196.

Focks, D.A., Haile, D.C., Daniels, E., Moun, G.A., (1993a) "Dynamics life table model for Aedes aegypti: Analysis of the literature and model development", J. Med. Entomol, 30, p.1003-1018.

Honório, N.A., Codeço, C.T., Alves, F.C., Magalhães, M.A.F.M., Lourenço-de-Oliveira, R., (2009) “Temporal Distribution of Aedes aegypti in Different Districts of Rio de Janeiro, Brazil, Measured by Two Types of Traps”, J. Med. Entomol. 46(4), p.000-000.

Lana, R. M. (2009). Modelos dinâmicos acoplados para simulação da ecologia do vetor Aedes aegypti. Master’s thesis, Universidade Federal de Ouro Preto, Programa de Pós-Graduação em Ecologia de Biomas Tropicais, Ouro Preto, Brasil.

Magori K, Legros M, Puente ME, Focks DA, Scott TW, et al., (2009) “Skeeter Buster: A Stochastic, Spatially Explicit Modeling Tool for Studying Aedes aegypti Population Replacement and Population Suppression Strategies”, PLoS Negl Trop Dis 3(9): e508. doi:10.1371/journal.pntd.0000508

Otero, M., Schweigmann, N., Solari, H.G., (2008) “A Stochastic Spatial Dynamical Model for Aedes aegypti”, Bulletin of Mathematical Biology.

Otero, M., Solari, H. G., Schweigmann, N., (2006) “A Stochastic Population Dynamics Model for Aedes aegypti: Formulation and Application to a City with Temperate Climate”, Bulletin of Mathematical Biology, 68, p.1945-1974.

Reiter, P., Amador, M.A., Colon, N., (1991), “Enhancement of the CDC ovitrap with hay infusions for daily monitoring of Aedes aegypti populations”, J. Am. Mosq. Control Assoc. 7, 52-55.

Rubinstein, R. Y. and Kroese, D. P. (2007). Simulation and the Monte Carlo Method. John Wiley & Sons, New York, 2nd edition.

Sharpe, P.J.H., DeMichele, D.W., 1977. Reaction kinetics of poikilotherm development. J. Theor. Biol. 64, 649-670.

Tran, A., Raffy, M., (2005) “On the dynamics of dengue epidemics from large-scale information”, Theoretical Population Biology, 69, p.3-12.

Proceedings XI GEOINFO, Campos do Jordão, Brazil, Nov. 28 to Dec. 1, 2010, INPE, p. 87–97.

Integrating Business Processes into GIS-based Simulations

Marcelo G. Metello, Eduardo David, Marcelo T. M. de Carvalho, Marco A. Casanova

Grupo de Tecnologia em Computação Gráfica – Tecgraf
Departamento de Informática – PUC-Rio
Rio de Janeiro, RJ – Brazil

{metello,david,tilio}@tecgraf.puc-rio.br, casanova@inf.puc-rio.br

Abstract. *This work proposes a mechanism to integrate different types of processes into a common gis-based simulation engine and, in particular, how to handle business processes described as workflows. In this sense, the paper presents the simulation process approach adopted, describes a method for mapping a workflow into the simulation process and it identifies which information needs to be added to the workflow description to make it possible to simulate it. To illustrate the ideas proposed, the aspects regarding management of an emergency situation are discussed. The discussion considered the different processes involved with it and the use of computational simulation to help manage the emergency situation.*

1. Introduction

In the Geographical Information Systems (GIS) field, a considerable amount of research has been devoted to developing dynamic models of man-driven and earth phenomena, such as physical processes, land use cover change, traffic control and socio-economic dynamics, among others. Many formalisms were developed to express these dynamic models. Among the most popular, we may quote cellular automata [von Neumann 1966], system dynamics [Forrester 1961] and multi-agent systems [Michel et al. 2009]. These models are used to predict the outcome of given scenarios and can be used to help planning, for example, urban activity, logistics, emergency response or marketing strategies.

Organizations have also worked out ways to formally express their activities. The adoption of the so-called Business Process Management (BPM) systems has grown significantly over the last few years [Weske 2007]. The most common practice of BPM tools is to express business processes as workflows. However, most BPM solutions focus on analyzing, publishing and monitoring business processes. There has been little research on simulating these processes, which help verifying and improving them.

Sometimes it is relevant to consider interaction between external processes, such as physical simulations, with organizations business processes, such as technical procedures. One example occurs during an emergency situation, such as an oil leak accident on the sea, where the behavior of the oil spill is influenced by the response team actions and the way the command of the emergency handle the situation depends on how big the oil spill is. In order to simulate such situations and obtain more realistic results, this type of interference must be considered.

Based on this idea, a simulation engine has been developed to provide such mechanism to deal with different processes that can interact with each other. In a previous work [Metello et al. 2008], it was presented an earlier version of this engine, which could simulate some kinds of processes in a GIS environment, such as physical dispersion of leaked products and transportation of equipment. Even though this engine could simulate a large variety of processes commonly modeled in GIS, it could not simulate workflows based processes. This paper addresses a newer version of the simulation engine, which can simulate business processes described as workflows. The paper has two main objectives. First, it describes a method for mapping a workflow into a simulation process to be executed in the engine. Second, it identifies the information required to be added to the workflow description that enable it to be simulated.

This paper is organized as follows. Section 2 illustrates in more detail the motivating example of handling emergency situations. Section 3 describes a formal base for integrating processes of different nature together in a simulation environment. Finally, section 4 describes how a process described by a workflow can be simulated in that environment.

2. A Motivating Example: Emergency Planning

An emergency occurs when a sudden and unexpected situation may cause damage to human health and/or to the environment and properties and which, therefore, requires immediate action and specific resources. In this case, it is not only important to respond quickly, but the response must be conducted in a well organized manner, in order to minimize the damages. The complexity of emergency management, coupled with the growing need for multi-agency and multi-functional involvement on these situations, has increased the need for methodologies that follow standards and that can be used by all emergency response disciplines. In this sense, the use of the *Incident Command System* (ICS) [Bigley and Roberts 2001] standard by public safety and private sector organizations around the world has been increasing.

When the incident begins, the first steps are notifications, initial assessment, command meeting, initial response and incident briefing using specific ICS forms. After this initial response period, the emergency handling process becomes cyclic. Each cycle, called *Planning "P"*, consists of a planning and an operational phases. The planning phase include meetings for assessing the current situation, update objectives, tactics definition, preparing and approving the incident action plan (IAP). The operational phase is where the response plan is executed and its progress is evaluated, after which starts a new cycle.

Use of information systems, such as InfoPAE [Carvalho et al. 2001] help to improve the management of this process. InfoPAE is an automated system designed to improve the response to emergency situations. It also proved to be a valuable training tool. The system offers sophisticated action plans and easy access to vital information and resources allocated for different types of scenarios. The system is developed by Petrobras (the Brazilian oil company) and the Computer Graphics Technology Group (Tecgraf) at PUC-Rio.

One of the difficulties of such systems is that, even though it is possible to describe an emergency action plan at a reasonably detailed level, this is somewhat limited with respect to the representation of dynamic aspects. The action plan is usually described using workflows and most workflow representations do not specify the time at which an action should be executed, how long it will take to finish and how the environment is affected by its execution. For instance, if a plan has instructions to install barriers to contain an oil spill at some point, it should take into account where and how the spill moves and how long it will take to have the barriers installed at the correct point. Otherwise, the plan is useless.

Testing the quality of an action response plan and the performance of the emergency response team is mandatory to ensure minimum impact of the incident. Beside others initiatives such as field exercises, use of computational simulation can be a cost-effective mechanism to improve it. Simulation can be used for different purposes as to estimate, in advance, if there will be enough resources and time for executing a specific action or support complex decisions during the planning phase of emergency response. One of the common used for it is to simulate the behavior of physical phenomena such as those involving dispersion of chemical products in the environment. However, the pure simulation of physical dispersion does not take into consideration the effects of contingency actions. Therefore, the use of purely physical simulation is somewhat limited in a strategy which cycles between evaluation and planning.

Hence, a more suitable computational environment to test the emergency management would be one that could combine simulation of physical phenomena with others processes such as those related to *Planning "P"*.

3. Process Simulation

3.1 Motivation

In the long-lived field of simulation, many different formalisms were developed to model processes in time [Eker et al. 2003, Vangheluwe 2000, Zeigler et al. 2000]. Each formalism suits better some specific class of processes. For instance, the dispersion of some leaked product into the environment can be modeled as a cellular automata (CA), while the corresponding contingency action plan can be modeled as a workflow. These two types of processes are depicted respectively in Fig. 1 (a) and (b). The first process defines a sequence of states over time while the second represents a set of actions from a workflow being executed in time, where each individual action has a defined start and end time instants. The problem of simulating these two processes independently of each other is obvious. Each of them only represents a partial view of the problem. The physical dispersion simulation would not take into consideration the effects of contingency actions. Likewise, the elaboration of plans as mere workflows does not guarantee that the actions will have their preconditions for execution met and, even if executed, that they will produce the desired results. Again, one illustrative example would be a plan that contains one action for placing contention barriers for oil leaked into the sea without considering whether the teams will have the necessary time to do so. These processes represent partial views precisely because they only consider a subset of all the elements involved.

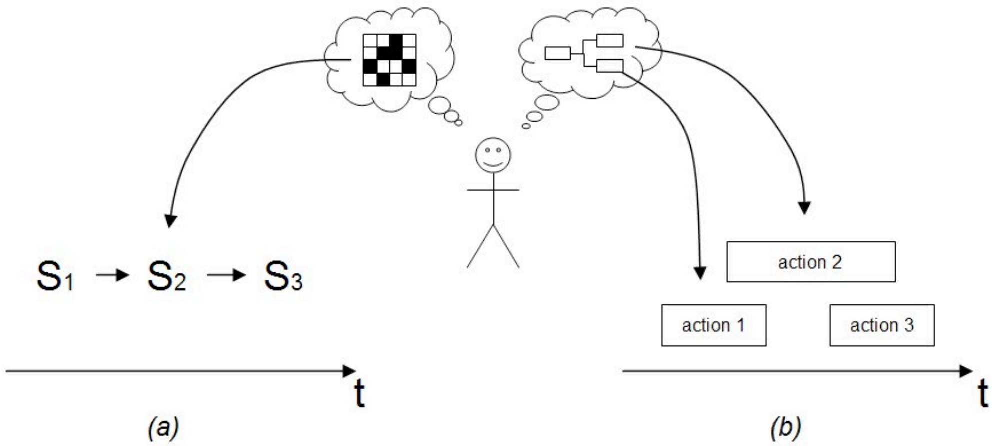


Figure 1. (a) A cellular automaton process represented by a succession of states over time. (b) A **workflow process represented by a set of actions in time**.

Our approach for dealing with this problem is to integrate these isolated processes into a single simulation execution environment, even if they are expressed in different formalisms. Naturally, the simulation must consider all the interferences between them. In order to achieve this, it is necessary to provide a common high-level formalism capable of representing different kinds of processes, such as plans and physical simulations. Also, the simulation environment must be capable of executing interfering processes in parallel.

The most straightforward way to integrate a number of different processes in a simulation environment probably is to run all of them in parallel, propagating causality every time a process generates an event that will possibly affect another process. Fig. 2 depicts two interfering processes being simulated in parallel. The sinuous arrows indicate causality being propagated between those processes.

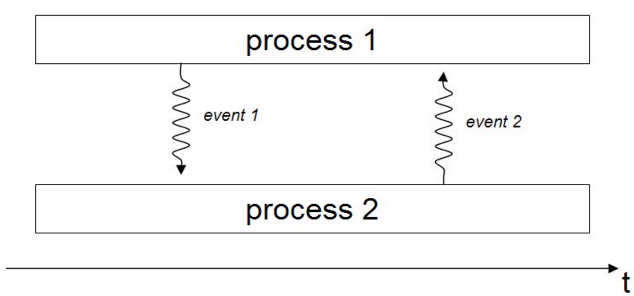


Figure 2. Two interfering processes being executed in parallel.

Three problems must be addressed to integrate processes of different natures. First, we need a common formalism which must be general enough to express the behavior of all individual processes as well as the interferences between them. Second, we need to map the formalisms in which the individual processes are described to this general formalism. Lastly, we need a simulation environment to execute all processes together.

3.2. On the Process Definition Formalism

Abstractly speaking, a process represents change in time. In the simulation field, a wide variety of formalisms to model change in time have been proposed. These formalisms may differ from each other even in the most fundamental aspects such as time representation. Some of them use a continuous numeric scale while others are restricted to discrete time values. They also differ with respect to how they represent changes in the world state with time. Again, some represent changes as discrete instantaneous events while others are able to represent continuous changes by means of differential equations. A complete discussion however is out of the scope of this work. Instead, we shall just point out that generality and the ability to integrate different formalisms is what is important to achieve our goals. Previous research about simulation formalisms have suggested that the *discrete event* approach has good potential for serving as the basis for integrating different simulation formalisms [Vangheluwe 2000, Zeigler et al. 2000]. Therefore, the principles of the discrete event approach were used as basis for the formalism designed to describe processes in a general way.

Each process evolves in a continuous timeline in which instantaneous events cause changes to the internal state of the process. Additionally, a process can generate events that are sent to other processes that are coupled to it. The coupling structure is defined separately from the definition of the processes behavior, thus increasing modularity and reuse. The coupling structure is also dynamic and can be changed during the execution of the simulation. Changes made to the coupling structure are also instantaneous in time, so it can be treated in the same way as a regular event.

Another aspect that was included in the approach is the lifetime of a process. During the course of a simulation, new processes can be started and existing ones can finish their executions. Each process must have definite start and end times. A process can only alter the state of the world within its lifetime. In consonance with other types of changes in the simulation, the start of a new process and the finish of an active one are also treated in the same way as other instantaneous events.

In order to allow GIS-based simulation, a notion of a *GIS environment* is created as a spatio-temporal data repository which represents the physical world. It has a state that changes instantaneously at some points in time, as the simulation advances and new events alter some geographical feature. Fig. 3 shows two processes and their interaction with the GIS environment, as well as with each other.

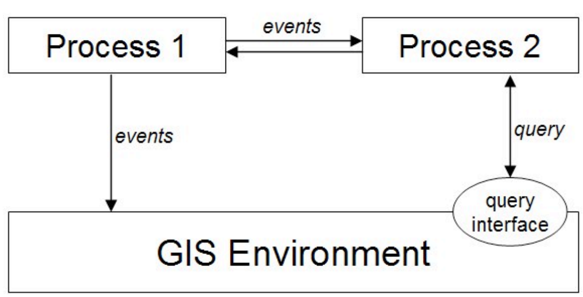


Figure 3. Two processes interacting with the GIS environment.

Because of the generality of the discrete event approach to simulation, it is usually possible to map other dynamic models into it [Zeigler 2000]. Discrete time models such as cellular automata, continuous models such as system dynamics and other hybrid approaches can be implemented on top of the discrete event approach [Vangheluwe 2000]. This makes it possible to reuse most models developed in the GIS field such as physical phenomena, social and economical behaviors and transportation, among others. The idea is to integrate all the relevant dynamic models into one synchronized set of processes and, on top of that, add some extra processes representing the business processes of an organization in that scenario. Then, it will be possible to the organization to foresee the effects of any given course of action.

4. Mapping Workflows to Discrete Event Simulation Processes

In this section, we describe in detail how to map workflow processes into a discrete-event process formalism. This is relevant since workflows are the most common way to represent business processes, as mentioned before, and since we are particularly interested in how organizations may create action plans. It is out of the scope of this paper to discuss how other types of processes are mapped into the discrete event process formalism.

4.1. Workflow Definition

Since there are many different languages and representations for workflows, this section starts by describing the workflow definition used here.

A workflow defines a set of actions and a control flow which imposes restrictions on the order at which the actions should be executed. With respect to their control flow, workflow representations may differ from each other in some of the patterns they use. For the sake of simplicity, we shall consider only the basic patterns defined in [van der Aalst et al. 2003]. Those are *sequence*, *parallel split*, *synchronization*, *exclusive choice* and *simple merge*. Fig. 4 depicts these patterns.

Each basic pattern connects some *preceding actions* to some *following actions*. In Fig. 4, all preceding actions are represented as rectangles on the left side of each pattern and all following actions are placed on the right side. In some cases it makes sense to connect a pattern directly to another pattern instead of an action. However, considering a direct connection between two patterns, placing a dummy action between them will not alter the behavior of the process described by the workflow. In fact, the two will be equivalent. Therefore, it is enough to describe the behavior of the patterns considering that they are only connected to actions. Table 1 lists some of the characteristics of each basic pattern.

The number of preceding and following actions indicates the number of action connections each pattern may have. The trigger time indicates the condition upon which the pattern is triggered. Finally, the following activity describes which of the following actions should be started when the pattern is triggered.

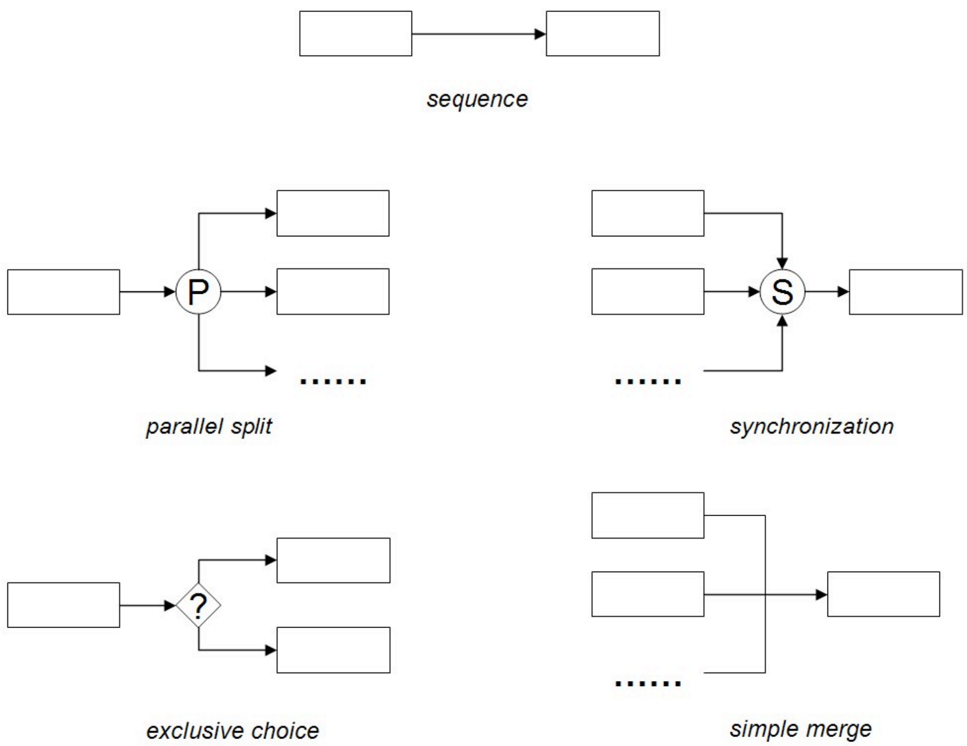


Figure 4. Basic workflow patterns.

Table 1. Properties and behaviors of basic patterns.

pattern	number of preceding actions	number of following actions	trigger time	following activity
sequence	1	1	completion of preceding action	start following action
parallel split	1	n	completion of preceding action	start all following actions in parallel
synchronization	n	1	completion of all preceding actions	start following action
exclusive choice	1	2 (or n)	completion of preceding action	start one of the following actions
simple merge	n	1	completion of any preceding action	start following action

Optionally, the parameters or inputs of the individual actions are also represented. Likewise, an action may also produce some data as output. That data could be consumed either by another action executed after it or by some conditional split operator in the control flow. Therefore, in order to represent action input and output, a data flow must also be represented. It is important to notice that the data flow does not follow the same paths as the control flow. However, it should obviously obey the ordering restrictions imposed by it since an action cannot consume output data from another action that has not yet been executed.

The simplest way to model the data flow is to define a set of variables that will compose the internal state of the workflow process. These variables are accessible by any action. Every time an action or a control operator needs some input data, it can get it from one or more variables. Whenever an action produces some data, it should store it in variables so that later actions can read it. Information stored in variables can also be used by exclusive choice operators to evaluate their conditions when they are triggered. Fig. 5 illustrates a workflow with the basic patterns and some variables.

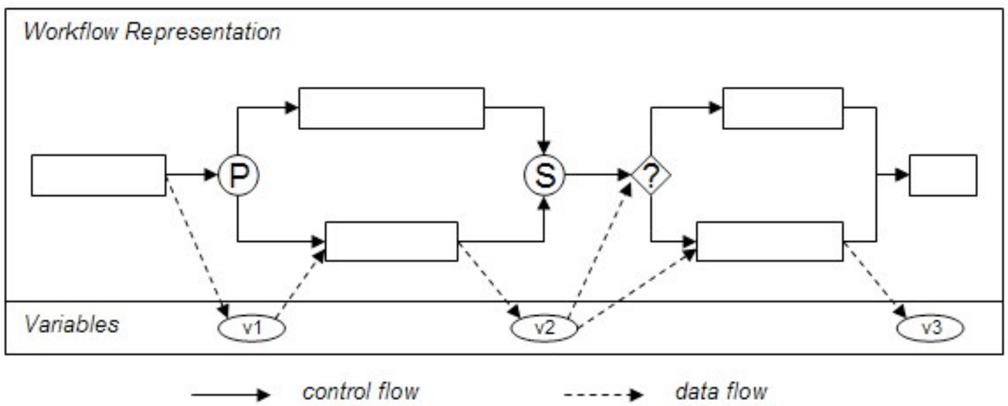


Figure 5. Workflow example with control and data flows.

4.2. Workflow as a Simulation Process

In our approach, we introduce a simulation process to represent the workflow, which we simply call the *workflow process*. This process will fork one individual *action process* every time it executes one action, as illustrated in Fig. 6. For that reason, the workflow process must be able to indicate the start time for each of its actions. Likewise, it needs to receive back the exact finish time of each action process so that the workflow can continue executing the next steps.

For any of the basic patterns, the start time of its following actions is given by a function that takes as input the trigger time of the pattern and, optionally, some values from variables. Actions that are not following actions of any pattern are started at the beginning of the workflow execution. When an action finishes its execution, it must inform the workflow process that it has finished by sending an event to it. That event may cause another pattern to trigger and so on until there are no more actions to execute. The finish time of the workflow process is the same as the finish time of its last executed action. The flow of all this timing information is illustrated in Fig. 6 as dotted arrows. As in the previous section, continuous arrows indicate the control flow and the dashed arrows indicate the data flow.

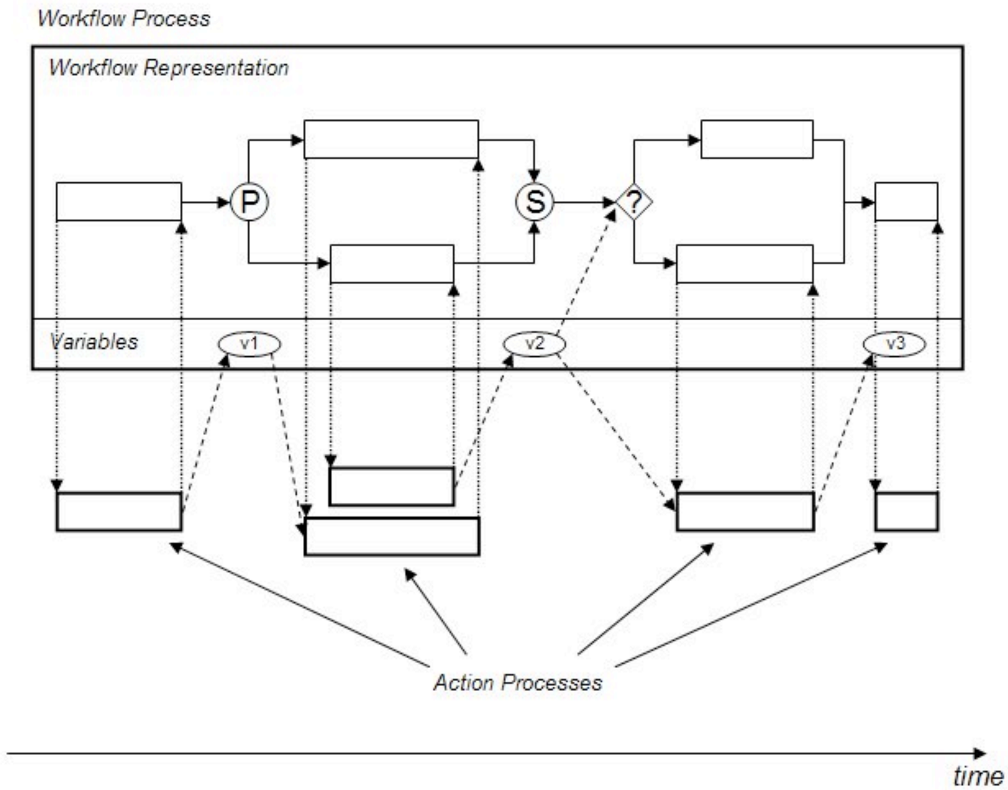


Figure 6. Processes involved in the simulation of a workflow.

Now that the processes to simulate a workflow are defined, we can combine them with processes of a different nature and consider how they interfere with each other. It is worth noting that the workflow variables could be updated by events sent by some process other than the action processes. This is actually a way for a process of a different nature to interfere with a workflow process. Likewise, an action process may send events to processes other than the workflow process, therefore interfering with them.

With the definitions presented in this section, it is possible to identify all the requirements of a workflow representation necessary to simulate a workflow together with other interfering processes. The requirements are summarized in the following list:

- **Provide all actions with typing information so that they can be associated with simulation processes.** Some human-readable workflow representations define the information about what an action does textually. In order to allow automatic simulation, actions must be more strongly typed. Besides, some additional information may be needed providing more detail for enabling simulation. For example, describing the weather conditions as “low tide with south wind” may not be precise enough for a physical simulation.
- **Provide functions to specify the start time of actions and condition evaluation.** As seen before, this is essential to simulate the situation correctly.

- **Provide a way to receive the end times of the actions.** Specific events may be created for providing the workflow process with information about when its executed actions have finished their execution.
- **Provide the workflow process with all necessary information so that it can keep its variables up to date.** By sending events to it, other processes may provide the workflow process with information it needs for correct simulation.

5. Conclusion and Future Work

The paper discusses aspects regarding the integration of different types of processes into a common simulation engine and, in particular, how to handle business processes described as workflows. In this sense, it presents the simulation process, describes a method for mapping a workflow into the simulation process and it identifies which information needs to be added to the workflow description to make it possible to simulate it.

In order to illustrate the ideas proposed, the aspects regarding management of an emergency situation are discussed. The discussion considered the differents processes involved with it and the use of computational simulation to help manage the emergency situation and . Section 3 discussed the use of discrete event approach to simulation, which proved to be a good way of integrating different processes and Section 4 showed how to map workflows into the simulation processes. Finally, it is discussed what a workflow representation must provide in order to enable the engine to simulate it.

As for future work, one possibility is to consider other wofkflow patterns. Indeed, Section 4.1 considered only five patterns, whereas [van der Aalst 2003] defines another fifteen patterns.

Another interesting challenge is that some workflow representations allow the definition of sub-workflows in a hierarchical composition. The approach could be extended to simulate a workflow hierarchy.

The requirements listed in Section 4.2 provide a roadmap to add some simulation functionality to an existing BPM system. For example, the InfoPAE system features a workflow language called XPAE [Casanova et al. 2002] that facilitates defining action plans. However, the current version of the language is not able to consider the dynamic aspect of a real emergency. Extend the language with the requirements presented would be extremely valuable.

References

Bigley, G.A., Roberts, K.H. (2001) ”The Incident Command System: High-Reliability Organizing for Complex and Volatile Task Environments“, The Academy of Management Journal, 44(6), pp. 1281-1299.

Carvalho, M.T., Freire, J., Casanova, M.A. (2001) “The Architecture of an Emergency Plan Deployment System”, In: Proc. III Brazilian Symposium on Geoinformatics, Rio de Janeiro, Brazil.

Casanova, M.A., Coelho, T.A.S., Carvalho, M.T.M., Corseuil, E.T.L., Nóbrega, H., Dias, F.M., Levy, C.H. (2002) “The Design of XPAE – An Emergency Plan

Definition Language", In: Proc. IV Brazilian Symposium on GeoInformatics, Caxambu, MG, Brazil.

Eker J., Janneck J.W., Lee E.A., Liu J., Liu X., Ludvig J., Neuendorffer S., Sachs S., Xiong Y. (2003) "Taming heterogeneity - the Ptolemy approach", In Proceedings of the IEEE, 91(2).

Forrester, J. (1961) "Industrial Dynamics", Cambridge, MA: MIT Press

Metello, M., Casanova, M.A., Carvalho, M.T.M (2008) "Using Serious Game Techniques to Simulate Emergency Situations", In: Proc. X Brazilian Symposium on GeoInformatics, Rio de Janeiro, Brazil.

Michel, F., Ferber, J., Drogoul, A. (2009) "Multi-Agent Systems and Simulation: A Survey from the Agent Community's Perspective", In A. Uhrmacher and D. Weyns (Eds.), Multi-Agent Systems: Simulation and Applications, Taylor and Francis, pp 3-51

van der Aalst, W.M.P., der Hofstede. A.H.M., Kiepuszewski, B., Barros, A.P. (2003) "Workflow Patterns", Distributed and Parallel Databases, 14(1): 5–51(47).

Vangheluwe, H.L. (2000) "DEVS as a common denominator for multi-formalism hybrid systems modeling", IEEE International Symposium on Computer-Aided Control System Design, pp 129-134, Anchorage, Alaska.

von Neumann, J. (1966) "Theory of self-reproducing automata", Illinois: A.W. Burks.

Weske, M. (2007) "Business Process Management - Concepts, Languages, Architectures", Springer.

Zeigler, B.P., Praehofer, H., Kim, T.G. (2000) "Theory of Modeling and Simulation", Academic Press, San Diego, CA.

G-GCS – A Geographical Aware Graph Coupling Structure for GIS Water Flow Representation

Sergio Rosim[1], Antônio Miguel V. Monteiro[1], Camilo Daleles Rennó[1], João Ricardo de F. Oliveira[1]

[1]Instituto Nacional de Pesquisas Espaciais (INPE) Caixa Postal 515 – 12.227-010 – São José dos Campos – SP – Brazil

{sergio,miguel,camilo,joao}@dpi.inpe.br

Abstract. *A new computational representation for dealing with water flow in GIS applications is proposed. Different computer structures have been studied proposed and implemented and have gained a place in most of the commercial systems. Each particular local water flow data structure needs its own set of operators in order to provide support for Distributed Hydrologic Modeling. On this paper a new data structure based on Graph theory and preserving geographical awareness called G-GCS – Geographical Aware-Graph Coupling Structure (G-GCS) is presented. It is as a basis for a unified computer local water flow representation independent of the data structures used for representing the terrain topography.*

1. Introduction

The local flow distribution in a water basin is the most important item to develop distributed hydrologic modeling oriented to hydrological resources. The underlying premise is that terrain main landscape contributor in settling these local flows [Kiss, 2004; Soille, 1994]. The basis for terrain representation in GIS is to partition the region extent. A *Cell* is the unit of this partition set and the *local flow* is the water flow for each cell considering the status of its neighbor cells according to a specific neighborhood rule chosen.

The most common data structures found in GIS libraries and systems for terrain representation dedicated to hydrologic modelling are the DEM (Digital Elevation Models- DEM) [Burrough, 1998] with regular grids, Irregular Triangular Networks – TIN [Chew, 1989], Contour Lines based representation [Dawes, 1988] and Irregular Polygons Tessellations [Tucker, 2001]. Each surface representation chosen comes with its own local water flow extraction functions and its own local flow data structure attached to it. The local flow representation is dependent on the data structure used to represent the terrain topography. For instance, DEM local flow extraction uses the 8-neighbor idea, creating a local flow representation, called *Local Drain Direction* (LDD) [Burrough, 1998].

Distributed hydrologic modelling environments normally assume a unique data structure for terrain representation. This may simplify software development but it can't make use of the properties of the other terrain data structures. The model proposed here helps to simplify software development and at the same time it allows the use of diverse terrain data structure. In this way, decoupling terrain data structure and local flow data structure eliminates the need to codify a given operator for each terrain data structure used.

The main idea is to decouple the local flow representation from the set of functions needed for its generation. The solution proposed consists in having the local flow mapped into a graph-based structure, the G-GCS, which is then coupled to the particular terrain data structure on which its generation has been based. Figure 1 presents the mapping local flow got from DEM, TIN, Contour Lines or from other possible surface data structure representation into the proposed G-GCS.

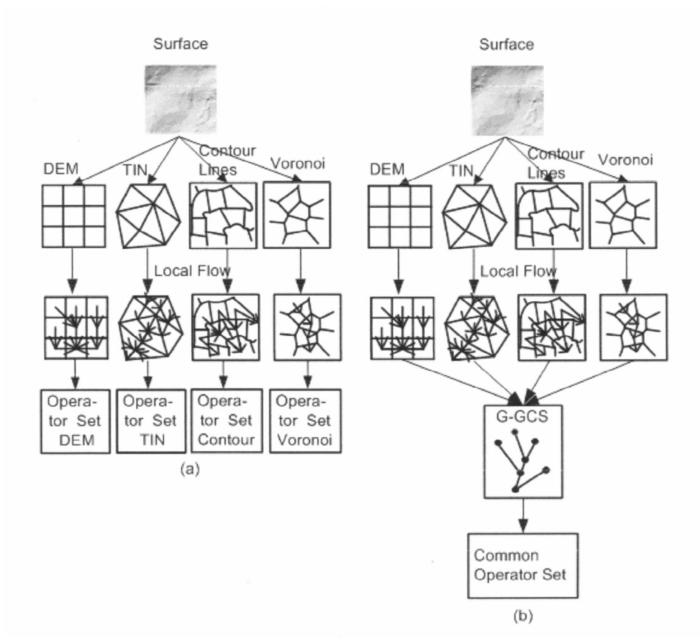


Figure 1. (a) **Present state in GIS Hydrologic Modelling;** (b) **New approach for building GIS Distributed Hydrologic Modelling Applications using the G-GCS idea.**

A feature of this new approach is that any new data structure for terrain representation can be incorporated into the framework with no impact at all on existing running models. The set of basic operators and functions can be extended and more complex operation can be built on top of the basic operator set. In order to demonstrate the G-GCS proposal at work we have defined a set of basic operations and have embedded them into a Distributed Hydrologic Modelling Toolkit prototyping environment based in Haskell and the TerraLib GIS free library [Costa, 2006]. Haskell was used because IT is a language that allows easy and fast implementations of prototypes. Same model was implemented using the PCRaster system that has its local flow representation bound to a grid structure. A comparative analysis based on the outcomes from these two implementations of the same model was conducted and the results are discussed over the sections (5) and (6) of this paper.

2. Previous Works

Methods to extract local water flow from the surface representation structures were developed to DEM, as the D8 unidirectional algorithm where the flow follows to the steepest descent 8-neighbour [O'Callaghan, 1984]. The Rho8 is a stochastic version algorithm of the D8 algorithm [Fairfield, 1991] FD8 and FRho8 are the changes of the previous algorithms, allowing dispersion flowing [Freeman, 1991; Quinn, 1991]. Some improvement has been achieved with methods that remove false pits and plane areas [Soille, 1994; Jenson, 1988].

Hydrological models using these methods of local flow extraction have been embedded in GIS systems as a computerized add-on for GIS hydrological modelling. The ArcGis Hydro Module [Maidment, 2002], The Grass GIS [GRASS, 1993], the Topographic Parameterization (TOPAZ) [Garbrecht, 1997], the Topography based Hydrological Model (TopModel), the MIKE SHE [DHI, 1998] and the PCRaster [Deursen, 1995] are examples of systems using DEM terrain representation based on a regular grid data structure for hydrologic modelling. The Watershed Modelling System (WMS) [Nelson, 1994] uses DEM and TIN data structures. TIN-based Real-Time Integrated Basin Simulator (tRIBS) uses TIN data structure and the TOPOG [Dawes, 1988] e SASHI [Rennó, 2003] that use Contour Lines based representations.

3. Local Flow Extractions from a Set of Different Terrain Data Structures Representation

The extraction of local flow is entirely dependent on the data structure chosen for terrain representation. The consequence of this is that each representation requires its own specific extraction algorithms, as well as different formats to store its associated local flow.

3.1. Extracting local flow from DEM

DEM is a rectangular regular elevation grid formed by cells. The use of some algorithm that approximates the local flow direction by the direction of steepest downhill slope for each cell, results in a new grid overlay called the set of Local Drain Directions – LDD. Each cell in this new grid contains an integer code representing the local flow direction at that cell. Figure 2(a) shows possible local flows in a given cell (center); Figure 2(b) shows a codification mask that supports unidirectional or multidirectional local flows; Figure 2(c) shows the resulting code of the cell representing the sum of the corresponding directional codes.

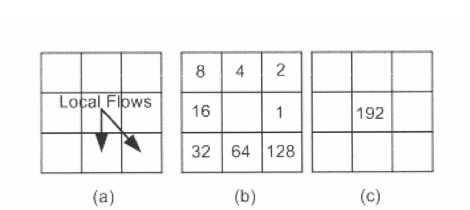


Figure 2. LDD grid creation, (a) **Local flows;** (b) **Codification mask;** (c) **LDD grid.**

3.2. Creating local flow from TIN

TIN structure normally created by Delaunay triangulation, stores information about these triangles and its sides and vertices, preserving the topological neighborhood of the triangle set. Triangles store information about its vertices and neighbor triangles. Sides store information about the two common triangles and about its two defining vertices. Vertices have three coordinates defining their position in space. TIN local flow has two kinds of propagation geometry: either the local flow crosses a triangle, flowing from a side to another, or it flows down along a side. Figure 3 presents an example of these two types of TIN local flow.

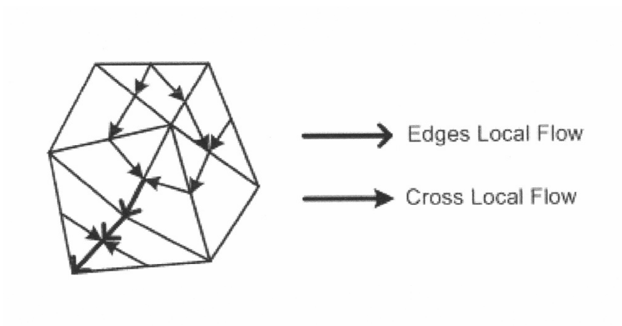


Figure 3. Cross flows and edges local flows

3.3. Creating local flow from Contour Lines

It is necessary to define an uniform flow unit between two neighbor contour lines forming a four-sided irregular polygon where two flow lines link two neighbor contour lines. In this case, local flow inside each flow unit is uniform, and its direction goes from the contour line with higher elevation value to the contour line with lower elevation value. Figure 4 presents local flow from contour lines.

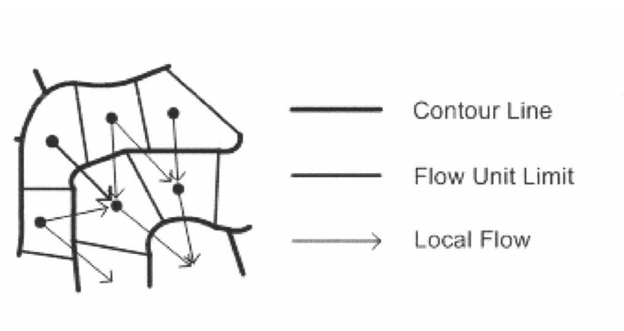


Figure 4. Contour Lines used to create local flows

3.4. Creating local flow from Voronoi diagram

The Voronoi diagram, a TIN dual structure, is a partition of space into cells, each one consists of points closer to a Delaunay triangle vertex than any other vertex. The local flow goes from a Voronoi cell to another through the triangles sides, beginning and ending at the center of neighbor cells, following the steepest downslope criterion [Tucker, 2001]. Figure 5 shows a Voronoi diagram with the local flow triangle sides.

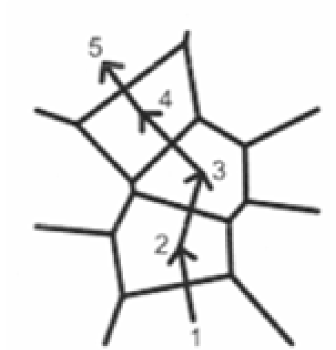


Figure 5. Voronoi used to generate local flows

4. Graph-Based Coupling Structure (GCS)

Labeled acyclic directed graph is a natural way to represent connected local flows. A graph G(V,E) is a vertex set V and an edge set E, where $e \in E, e = (v_i, v_j) \mid v_i \in V, v_j \in V \text{ and } i \neq j$ [Ore, 1962]. A directed graph is a graph with oriented edges. An acyclic graph is a graph without loops. A labeled graph is a graph where nodes and edges can store values. The main advantage of a graph is that it can store data in a structured way. This leads to very useful graph properties comparing to the others structures used to store local flow. Routines accessing graph data usually are more efficient than routines accessing local flow data structures.

Graph is a structure containing ready-to-use information about node links. Applications using graphs use this property to travel through the graph; local flow data structures do not have this property. So routines accessing graph data are usually more efficient than routines accessing local flow data structures.

4.1. LDD to Graph Map

Each LDD grid cell represents a graph node and the flow from a given cell to a neighbor cell defines a graph edge that links these two cells. Figure 6 shows the local flow from cell ‘4’ to cell ‘8’ and its graph representation.

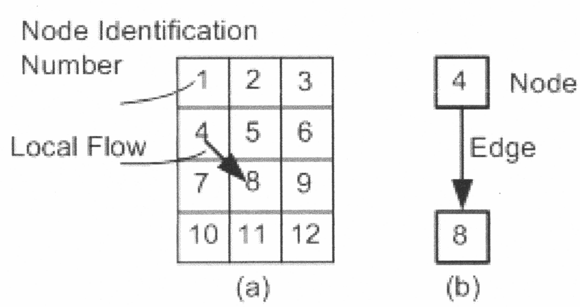


Figure 6. (a) Local flow; (b) Graph representation

4.2. LDD to Graph Map

As TIN local flow has two types of propagation geometry, each type needs a different approach to map TIN local flow to graph.

4.2.1. Triangle crosses local flow to graph

Each triangle side starting or ending as a local flow represents a graph node. Local flow goes from one side to another side of a triangle, passing through their middle points. The graph nodes identifiers are the same associates to the triangles sides during the triangulation process. This enables the remap from graph to triangulation, if desired. Figure 7 shows this local flow type.

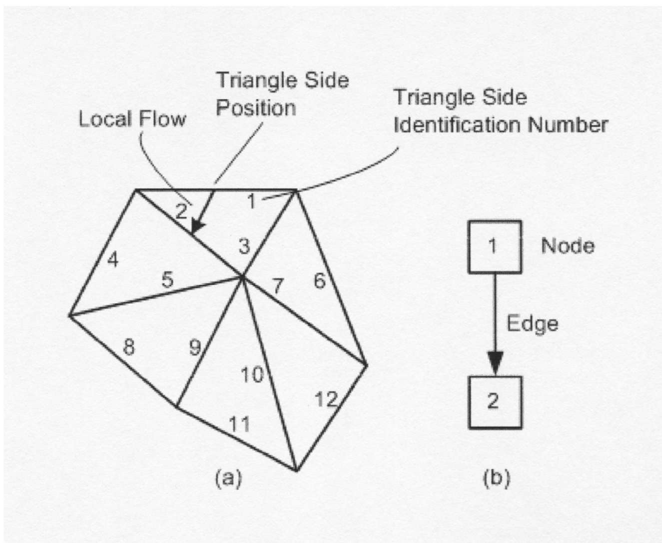


Figure 7. (a) Triangle crosses local flow; (b) Graph representation

4.2.2. Common triangles edge local flow map to graph

When a local flow goes along a triangle side, the vertices of that side represent graph nodes. The graph node identifier corresponding to triangle vertex is computed adding the total number of triangles sides to the vertex identifier from triangulation. Figure 8 shows this graph edge type and the corresponding detailed graph structure.

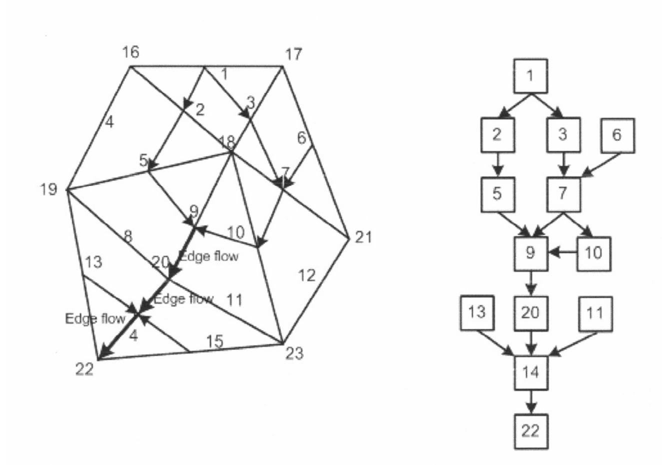


Figure 8. Mapping triangle local flows with edge flows to graph edges

4.2.3. Contour line to graph map

The local flow goes from each cell to one or more neighbors, passing through their centers. The graph node stores the cell identification number and a graph edge is a link between two cells. Multi flow issue is intrinsic in the contour lines data model. Figure 9 presents an example of the mapping from contour lines to graph.

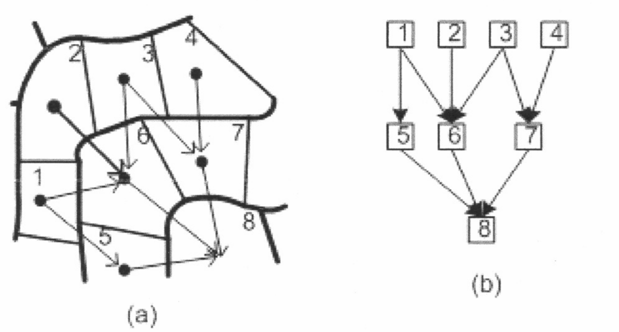


Figure 9. (a) Uniform flow unit local flows; (b) Graph representation

4.2.4. Voronoi to graph map

Each Voronoi polygon is a graph node and each graph edge represents a link between two neighbor polygons. Graph nodes store the Voronoi identification numbers existing in the Voronoi data structure and this approach is similar to grid cell approach. Figure 10 shows Voronoi to graph map.

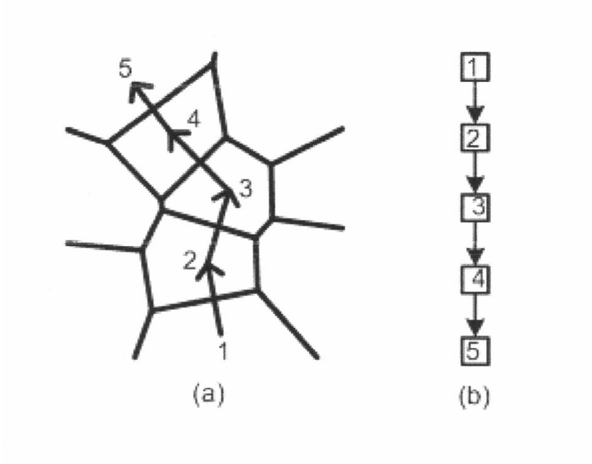


Figure 10. (a) Voronoi local flows; (b) Graph representation

5. A Water Flow implementation Problem with G-GCS

The computational implementation of G-GCS is not dependent of any programming language. Here, the implementation has used the Haskell functional language [Peyton Jones, 2002; Hudak, 2007] and the Functional Graph Library (FGL) [Erwig, 2001], which has several functions to create and manipulate graphs. Haskell language allows quick codification, depuration and prototype testes. Other decisive factor in choosing Haskell was the ease of use of many graph functions provided by FGL.

The example using PCRaster system "Simplified Hydrological Runoff Model" was chosen [Karssenberg, 1997] to compare with the same implementation with the G-GCS proposed. This application uses a LDD grid, a soil infiltration grid and a rainfall temporal series with same grids obtained from three rainfall stations to calculate flow accumulation for each grid cell.

This application aim is to determine the flow accumulation for each grid cell that has an infiltration capacity value. If the sum of the water arriving in a cell, coming from its neighbor cells, plus the rainfall at the same cell, exceeds its infiltration capacity, the water excess will flow out of this cell. Otherwise, the whole water quantity will be retained in the considered cell. Flow accumulation can be calculated using the water balance equation [Rennó, 2003].

P is the rainfall, E_{int} is the part of precipitation intercepted by canopy and evaporated afterwards, E_s is the soil evaporation, E_p is the water evaporation from canopy, Q_{out} is the runoff, Q_{in} is the water flow in the system and I is the quantity of infiltration water. The example of the PCRaster system doesn't consider the evaporation, then $E_{\text{int}} = 0,\ E_s = 0,\ E_p = 0$ and it assumes that the system infiltration capacity as infinite. It means that each time step is independent of the previous time steps. Figure 11 shows all components considered by this modelling.

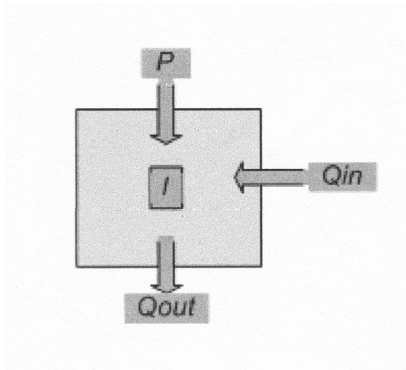


Figure 11. Simplified water flow balance diagram

PCRaster uses the accuthresholdflux and accuthresholdstate functions to accomplish this example. The accuthresholdflux function assigns the water that is transported out of the cell and the accuthresholdstate function calculates the amount of water infiltrated in the cell. Figure 12 presents the grids used in this example that were extracted using the NutShell, a PCRaster graphical user interface.

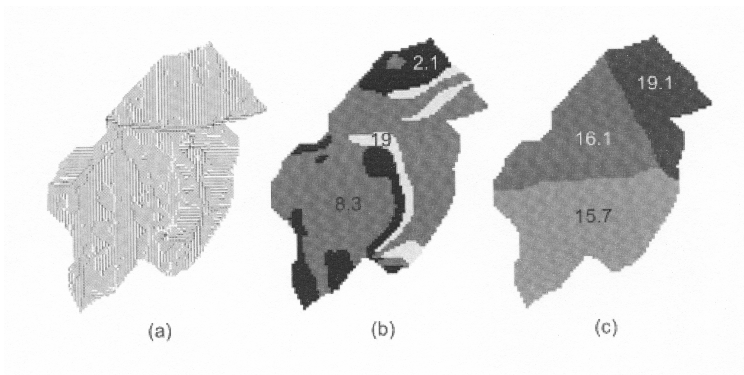


Figure 12. Grids used in this application. Numbers at the Figure are infiltration and rainfall values; (a) **LDD grid;** (b) **Soil infiltration grid;** (c) **Rainfall grid.**

The implementation here devised uses data stored in TerraLib [Câmara, 2000] open source geographical library implemented in C++ language. Terraview, a TerraLib based software, is used to visualize and to manipulate vector and raster data preserved in geodatabases to read, write and visualize the grids. The grids used in this application are in TerraLib format and they can't be directly read in Haskell. A binding in C language called Terra-HS [Costa, 2006] was used to access the TerraLib grids.

Any developing application needs to execute the same steps independent of the particular application. Initially, it is necessary to have in separated files the grids that will be used. The computer execution sequence is: (1) read the LDD, infiltration and rainfall grids of TerraLib; (2) create a graph using LDD grid, assigning the infiltration and rainfall values at the respective graph nodes; (3) execute the accumulate flow function; (4) convert accumulation nodes values to a grid; (5) write this grid in Terralib format. Figure 13 presents the general steps and the particular implementation used.

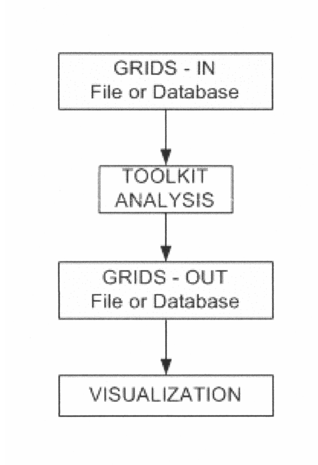


Figure 13. Developing steps to implement an application with G-GCS.

The Haskell accumulation flow function result is converted from graph to grid (4) and is saved in the TerraLib format (5) where it can be seen using TerraView software. Figure 14 shows G-GCS result.

Figure 14. Result obtained with G-GCS approach

The results obtained using the model proposed here is exactly the same obtained using PCRaster, validating the proposed model. This was expected because in both cases the local flow was expected from the same LDD grid. The next step in this work will be the creation of the graph and the flow accumulation calculus for local flows extracted from TIN. The Very Important Points (VIP) [Chen, 1987] algorithm will be used to reduce DEM data points to generate TIN sample points.

To illustrate the graph usage in the drainage network implementation, the drainage definition operator (DefDrain) will be presented. This operator defines a subset of the accumulated area network using a threshold parameter. All network nodes with values greater than the threshold will belong to the network drainage. This operator is easily done by using the graph structure.

6. Conclusions

It is possible to use a single coupling structure to represent local flows generated from different surface representation structures. This was illustrated for the accumulation flow function using DEM structure and the PCRaster system, comparing their results, showing they are similar.

The surface representation structure is decoupled from the operator set linked to the G-GCS structure. This guarantees that the operator set does not need to be duplicated and, at the same time it can be utilized the best characteristics of each surface representation structure.

References

Burrough PA, McDonnell RA (1998) Principles of Geographical Information Systems. Oxford University Press, New York

Câmara G, Souza RCM, Pedrosa B, Vinhas L, Monteiro AMV, Paiva JA, Carvalho MT, Gattass M (2000) TerraLib: technology in support of GIS innovation. II Brazilian Symposium on GeoInformatics, GeoInfo2000, São Paulo.

Chen ZT, Guevara JA (1987) Systematic selection of very important points (VIP) from digital terrain model for constructing triangular irregular networks. In N. Chrisman, editor, Proc. of Auto-Carto 8 (Eighth Intl. Symp. on Computer-Assisted Cartography), pages 50–56, Baltimore, MD. American Congress of Surveying and Mapping.

Chew LP (1989) Constrained Delaunay triangulations. Algorithmica 4(1):98-108

Costa SS, Câmara G, Palomo P (2006) TerraHS: integration of functional programming and spatial databases for GIS application development. VIII Brazilian Symposium on GeoInformatics, GeoInfo2006, Campos do Jordão, São Paulo, pp 109-127

DHI-Danish Hydraulic Institute (1998) MIKE SHE water movement - user guide and technical reference manual. Edition 1.1.

Dawes WR, Short DL (1988) TOPOG series topographic analysis and catchment drainage modelling package: user manual. Canberra, CSICO.

Deursen WPA (1995) Geographical Information Systems and dynamic models development and application of a prototype spatial modelling language. PhD Thesis, Faculty of Spatial Sciences, University of Utrecht

Erwig M (2001) Inductive graphs and functional graph algorithms. Journal of Functional Programming, 11(5):467-492

Fairfield J, Leymarie P (1991) Drainage networks from grid digital elevation models. Water Resource Research, 30(6):1681–1692

Freeman GT (1991) Calculating catchment area with divergent flow based on a regular grid. Computers and Geosciences, 17:413-422

Garbrecht J, Martz LW (1997) Automated channel ordering and node indexing for raster channel networks. Computers and Geosciences, pp 96–146

Garbrecht J, Ogden FL, DeBarry PA, Maidment DR (2001) GIS and distributed watershed models. I:Data Coverages and Sources, Journal of Hydrologic Engineering, 6(6):506-514

GRASS (1993) 4.1 User's reference manual. U.S. Army Corps of Engineers Construction Engineering Research Laboratory, Champaign

Hudak P, Hughes J, Jones, SP, Wadler P. (2007) A history of Haskell: being lazy with class. Third ACM SIGPLAN History of Programming Languages Conference (HOPL-III) San Diego, California, June 9-10, pp 55

Jenson SK, Dominque JO (1988) Extracting topographic structure from digital elevation data for geographic information systems. Photogrammetric Engineering and Remote Sensing 54(11):1593–1600

Karssenberg D, Wesseling CG, Burrough PA, van Deursen WPA (1997) A simplified hydrological runoff model. Department of Physical Geography, Ultrecht University, http://pcraster.geo.uu.nl/demos.html

Kiss R (2004) Determination of drainage network in digital elevation models, utilities and limitations. Journal of Hungarian Geomathematics, vol 2

Maidment DR (2002) Arc Hydro: GIS for water resources. ESRI Press, Redlands CA, 2002, pp 220

Nelson EJ, Jones NL, Miller AW (1994) Algorithm for precise drainage-basin delineation. Journal of Hydraulic Engineering, 120(3):298-312

O'Callaghan JF, Mark DM (1984) The extraction of drainage networks from digital elevation data. Computer Vision, Graphics, and Image Processing, 28:323–344

Ore O (1962) Theory of graph. American Mathematical Society, Michigan

Peyton Jones S (2002) Haskell 98 language and libraries the revised report

Quinn P, Beven K, Chevalier P, Planchon O (1991) The prediction of hillslope flow paths for distributed hydrological modelling using digital terrain models. Hydrological Processes, 5:59-79

Rennó C D (2003) Construção de um sistema de análise e simulação hidrológica: aplicação a bacias hidrográficas (in portughese). Ph.D. thesis, National Institute for Space Research, São José dos Campos

Soille P, Gratin C (1994) As efficient algorithm for drainage network extraction on DEMs. Journal of Visual Communication and Image Representation, 5(2):181-189

Tucker GE, Lancaster ST, Gasparini NM, Brás RF, Rybarczyk SM (2001) An object-oriented framework for distributed hydrologic and geomorphic modelling using triangulated irregular networks, Computers & Geosciences 27:959–973

Heuristics to Site Observers in a Terrain Represented by a Digital Elevation Matrix

Salles V. G. de Magalhães[1], Marcus V. A. Andrade[1], Chaulio Ferreira [1]

[1]Departamento de Informática – Universidade Federal de Viçosa (UFV)
36.570-000 – Viçosa – MG – Brazil

{smagalhaes,marcus,chaulio.ferreira}@dpi.ufv.br

Abstract. *This work deals with the observer siting problem in terrains represented by digital elevation matrices. This problem consists in minimize the number of observers needed to see (or cover) a given percentage of a terrain. Two heuristics are proposed to solve this problem and these heuristics are compared with a method described in literature. Tests show that the proposed heuristics generate solutions that use an average of 7% less observers than the heuristic described in literature.*

1. Introduction

Advances in remote sensing have produced large quantities of high resolution geographic data that require new Geographic Information Systems [Laurini and Thompson 1992] (GISs) techniques to process.

The Earth's surface elevation (terrain) data is usually approximately represented by a digital elevation matrix (DEM) that stores the elevations of regularly sampled terrain points. Elevations of intermediate points are approximated using some interpolation process [Li et al. 2005].

An important group of GIS applications on terrain concerns visibility, i.e., determining the set of points on the terrain that are visible from some particular observer. The observer can be located at some height above the terrain. Applications include telecommunications, environmental planning, autonomous vehicle navigation, and military monitoring [Franklin and Ray 1994, Li et al. 2005, Nagy 1994, Andrade et al. 2009]. One important problem is the positioning of a given number of facilities in order to optimally "cover the terrain". These facilities may be radio, TV, internet or mobile phone towers, and monitoring towers [Ben-Moshe 2005, Ben-Shimol et al. 2007].

This paper presents approximated methods based on heuristics to solve the multiple observer siting problem in terrains represented by digital elevation matrices. More precisely, this work deals with the observer siting problem where the objective is to site the "minimum" amount of observers needed to achieve, at least, a minimum desired coverage. As stated later, this is a NP-Complete problem and, thus, the proposed solution is approximated. This work extends Franklin et al. [Franklin 2002, Franklin and Vogt 2006] technique, that uses a greedy approach to solve the problem. Two other heuristics to solve the problem are proposed: a variation of GRASP method and a greedy heuristic enhanced with a local search.

2. Terrain Visibility

A *terrain* represents a region of the Earth's surface. In the context of this paper, it is a scalar field over a square (in the relevant coordinate system) domain. The terrain's value at any point is the elevation of the corresponding point of the Earth's surface above some reference ellipsoid called the *geoid* that represents sea-level. For this paper, terrain is represented by a matrix of elevation posts on a square grid, whose vertical and horizontal spacing is uniform either in distance, e.g., 30m, or in angle, e.g. 1 arc-second.

Other representations such as triangular splines, or Triangulated Irregular Networks (TINs), are also common. Their implementation is more complicated, especially for operations such as line-of-sight. When only a coarse approximation of the terrain is needed, a TIN is compact. However, in this case, a DEM may be lossily compressed to an extremely small size. Indeed, by separating the abstract data structure, the grid, from the concrete implementation, the compression algorithm, the DEM is conceptually a better designed representation, while easier to implement, and apparently equally compact.

An *observer* is a point in space that "wants" to see or communicate with other points in space, called *targets*. The usual notation for observer and target is O and T. The *base points* of O and T are the points on the terrain directly below O and T, respectively. They are denoted as O_b and T_b. O and T are each at height h above O_b and T_b.

The possible generalization of having different heights for the observer and target, while adding another degree of freedom to the experiments, does not seem to add anything new to the science of the problem. That is, earlier tests lead us to believe that our conclusions are general.

The *radius of interest*, R, of O is the distance out to which it can see, in the absence of obstructions. E.g., if O is a radio transmitter, R is a function of the transmitter power and receiver sensitivity. For convenience, R is measured between O_b and T_b rather than between O and T, which is equivalent when h is much smaller than the radius of the Earth.

T is *visible* from O iff $|T_b - O_b| \leq R$ and there is no terrain point between O and T blocking the line segment, called the *Line of Sight (LOS)*, between them; see Figure 1. In this Figure, T_1 is visible from O but T_2 is not.

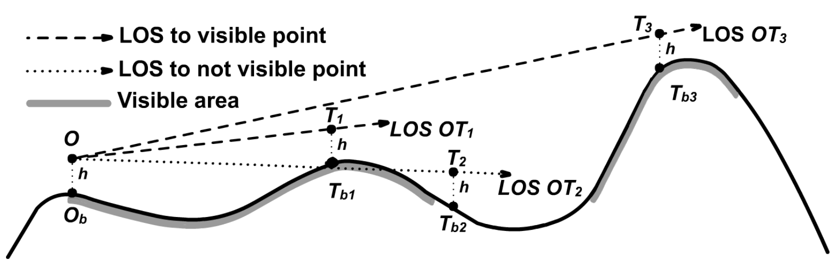


Figure 1. Determining the visibility using a *LOS* in a terrain vertical section.

The *viewshed*, V, of O is the set of base points whose corresponding targets are visible from O. V is stored as a bit matrix with the bits packed into the longest words that can be processed in one machine cycle.

The *visibility index*, $VIX(O)$ is the number of targets with base points within the circle C of radius R centered at O_b that are visible from O.

Points with a large VIX are good candidate places to site observers in order to maximize the area of the terrain that is seen by at least one observer. VIX, which is simply the number of 1-bits in V, is commonly estimated by counting how many of a random sample of targets inside C are visible.

The *joint viewshed*, V, of a set of observers $\mathcal{O} = \{O_i\}$ is the union of the individual viewsheds V_i, i.e., the bitwise-*or* of their bit matrices. The *joint visibility index*, $VIX(\mathcal{O})$ is the number of targets in the terrain that are visible from at least one observer in $\mathcal{O}$. $VIX(\mathcal{O})$ may be normalized to be a percent of the terrain area.

Given a set $\mathcal{O}$ of observers, $VIX(\mathcal{O})$ can be obtained by counting the 1-bits in $V(\mathcal{O})$. Notice that this operation is computationally expensive since it is necessary to join all viewsheds from $\mathcal{O}$ in order to get $V(\mathcal{O})$ and, then, to count the number of 1-bits in $V(\mathcal{O})$.

The *Multi-observer siting* problem means optimizing the locations of a set of observers, called *siting*, so that these observer´s VIX is as large as possible [hoon Kim et al. 2004]. This has important practical facilities-location applications, such as siting mobile phone towers, fire monitoring towers, and radar systems.

This problem is NP-Complete [Nagy 1994], and can be reduced to classical set coverage optimization [Cormen et al. 2001] wich can be defined as: Given a set $S = \{s_i\}$ of sets, choose $C = \{c_i\} \subset S$ such that $\cup c_i = \cup s_i$ and $|C|$ is minimized [Black 2004]. Informally, find the smallest number of sets s_i whose union is equal to the union of all the s_i.

This work will present and compare different heuristics to approximate the following variation of multi-observer siting problem: minimize the number of observers needed to cover a percentage of the terrain area. More specifically, given a set $P = \{P_i\}$ of candidate points to receive observers, the problem consists in finding the smallest subset $S = \{s_i\}$ of P whose visibility index is greater or equal than VIX_{min} (a user-defined value). Thus, the solution to this problem will be represented by a subset of P.

3. Heuristics to solve the observer siting problem

The optimization problems whose exact solutions are hard to be found usually are solved approximately using combinatorial optimization heuristics. Examples of these heuristics include local search techniques [Resende 1999], GRASP (*Greedy Randomized Adaptative Search Procedure*) [Resende 1999], genetic algorithms [Mitchell 1996], simulated annealing [Davis 1987], etc.

These heuristics generally use several attempts (iterations) to get a solution that maximizes (or minimizes) the value of an objective function while satisfying some constraints. As some authors have shown, these heuristics can be combined to get better solutions. For example, Resende et al [Resende and Werneck 2004] proposed a combination of heuristics to solve the p-medians problem (a variation of the facility location problem). This idea consists in using a local search heuristic to improve the solution obtained by the genetic algorithm.

As stated in section 2, this work presents and compares different heuristics to the observer siting problem. More specifically, the greedy heuristic proposed by Franklin et al [Franklin and Ray 1994] was combined with two new heuristics proposed in this work to solve the problem: a greedy heuristic enhanced with a local search and an heuristic based on GRASP. Furthermore, two local search heuristics were proposed to improve the covered area of a given solution (these heuristics were used in the enhanced greedy heuristic and in the GRASP heuristic). One of these heuristics is based on all neighbors local search and the other one uses the spatial location of the solution´s elements to do the search.

3.1. Greedy method

A greedy strategy to solve the observer siting problem named Site, proposed by Franklin et al. [Franklin 2002], consists in generating iteratively the solution (represented by the set S). Given the set of candidates $P = \{P_i\}$, at each step, the point P_i that would most increase the joint visibility index of S is inserted into S. This iteration is executed while $VIX(S)$ is smaller than VIX_{min} See the Algorithm 1.

Algorithm 1 Greedy algorithm used on the observer siting problem

```
S ← {}
while VIX(S) < VIX_min do
  id ← 1 //Suppose that the first element of S is the "best" point.
  for j = 1 to |P| do //Find the point of P that would most increase the joint visibility
  index of S.
    if VIX(S + P_j) > VIX(S + P_id) then
      id ← j
    end if
  end for
  S ← S + {P_id}
end while
return S
```

3.2. Improving the visible area using a local search

Given an initial solution S, a local search method defines a neighborhood to S with solutions that are "near" to S and, then, moves through this neighborhood in order to find a "better" solution than S. Thus, in each step, the local search moves toward a neighbor S' better than S.

This work uses local search techniques to improve the visibility index of a solution with a fixed number of observers. More specifically, given an initial solution S, the local search tries to improve S by keeping the number of elements in S while increasing $VIX(S)$. The number of observers was kept by considering that a neighbor of S is a solution S' with exactly one element different of S. See Figure 2. Notice that this local search doesn´t change the value of the objective function wich considers only the number of observers used by the solution. But, on the other hand, it improves the solution visibility index and, thus, it can be used to make solutions with less observers feasible (see sections 3.3 and 3.4).

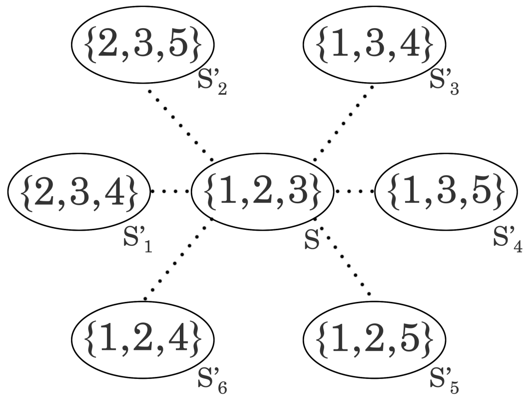


Figure 2. Neighbors of solution $S = \{1, 2, 3\}$ wich is a subset of $P = \{1, 2, 3, 4, 5\}$

The neighbors generation is done using two strategies: generating all neighbors (called *all neighbor local search*) and generating neighbors considering points spatial locations (called *spatial local search*).

In the *all neighbors local search*, all neighbors of a current solution S are generated and, then, the visibility index of each neighbor S' is evaluated. When a neighbor S' whose visibility index is greater than $VIX(S)$ is found, S' is attributed to S (that is, S' becomes the current solution). This process is repeated until a solution $S*$ that doesn´t have a better neighbor is found. $S*$ is called a local optima.

Since the evaluation of a solution is a very expensive operation (it is necessary to join several viewsheds), it is important to reduce the number of neighbors in the local search. So, given a solution S, a solution S' will be considered neighbor to S if and only if there is an observer P_i in S' with P_i not in S and another observer P_j in S with P_j not in S' such that the distance between P_i and P_j is at most s_r, named swap radius. In other words, when the neighbors of S are generated, a point P_i that belongs to S is replaced by another point P_j (that doesn´t belong to S) only if the distance between P_i and P_j is smaller or equal than s_r. See the example of solution in Figure 3: points represented by triangles or pentagons belongs to the current solution while points represented by squares or circles do not belong. In this example, the point represented by a pentagon will be removed and it will be replaced by points represented by squares. Thus, only 4 different neighbors (each one with a square replacing the pentagon) are evaluated. Notice that this figure shows the neighborhood related to the remotion of just one point (neighbors related to the remotion of other points, represented by triangles, will also be evaluated).

The *spatial local search* works similarly to the *all neighbors local search* except that it evaluates just "near" neighbors while the previous one evaluates all the neighborhood. Note that, if s_r is big (greater or equal than the terrain diagonal) the *spatial local search* will be equal to the *all neighbors local search*.

3.3. Greedy method with local search (GreedyLS)

Since the local search methods described in section 3.2 improves the covered area of a solution S not changing the number of observers used by S, a strategy (called GreedyLS)

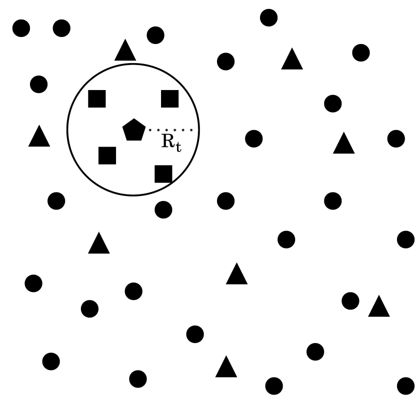


Figure 3. Example of neighborhood considering the spatial location of elements: points represented by triangles and by the pentagon belong to the current solution while points represented by squares and circles do not belong.

was proposed to improve the objective function of solutions obtained by a greedy method by adding, in each iteration, a local search that maximizes the current solution's visible index. Thus, the desired visibility index might be achieved using less iterations and, so, less observers.

Moreover, an additional iteration was included to the end of the method to remove, in each step, the observer that would least contribute to the solution's visibility index (a greedy removal) and, then, a local search is used to improve the current solution's VIX. This iteration is executed while the current solution is feasible (or, in other words, while $VIX(S) \geq VIX_{min}$). See Algorithm 2.

3.4. GRASP

The GRASP method, originally proposed by Resende et al. [Resende 1999], consists in creating an initial solution using a constructive method and, then, improving this solution with a local search. The constructive method is randomized and, thus, in each execution it gives a different solution. This process of construction and improvement is repeated several times and, at the end, the best solution obtained is returned by the method. The stopping criteria can be defined using a time limit, iteration limit, etc. See Algorithm 3.

The GRASP method was adapted to solve the observer siting problem as follow: the constructive step uses a greedy randominzed method and the local search was replaced by an heuristic that improves the current solution using a greedy remotion combined with a local search (similar to the heuristic described in section 3.3). The greedy randomized solutions are created using the following variation of the greedy method described in section 3.1: in each iteration, the k-th best observer is added to the current solution where k is a value randomly chosen between 1 and $RandFactor$. The $RandFactor$ is a randomization factor defined by the user (if $RandFactor$ is equal to 1 the method behaves like a traditional greedy method).

Algorithm 2 GreedyLS method

$S \leftarrow \{\}$
while $VIX(S) < VIX_{min}$ **do**
  $id \leftarrow 1$ //Suppose that the first element of S is the "best" point.
  for $j = 1$ to $|P|$ **do** //Find the point of P that would most increase the joint visibility index of S.
    if $VIX(S + P_j) > VIX(S + P_{id})$ **then**
      $id \leftarrow j$
    end if
  end for
  $S \leftarrow S + \{P_{id}\}$
  $S \leftarrow localSearch(S)$ //Maximize $VIX(S)$ using a local search.
end while
while $VIX(S) \geq VIX_{min}$ **do** //While S is feasible
  $S* \leftarrow S$ //$S*$ will store the best feasible solution found
  $removeWorstObserver(S)$ //Remove the "worst" point from S.
  $S \leftarrow localSearch(S)$ //Maximize $VIX(S)$ using a local search.
end while
return $S*$

Algorithm 3 Original GRASP Method

//Let: randConstructSol(): function that creates a randomized constructive solution
//localSearch(S): local search function
$S* \leftarrow NULL$ //$S*$ will store the best solution found
while Stopping criteria isn´t reached **do**
  $S \leftarrow randConstructSol()$ //Creates a randomized constructive solution.
  $S \leftarrow localSearch(S)$
  if $S*$ = NULL **or** $f(S) > f(S*)$ **then**
    $S* \leftarrow S$
  end if
end while
return $S*$

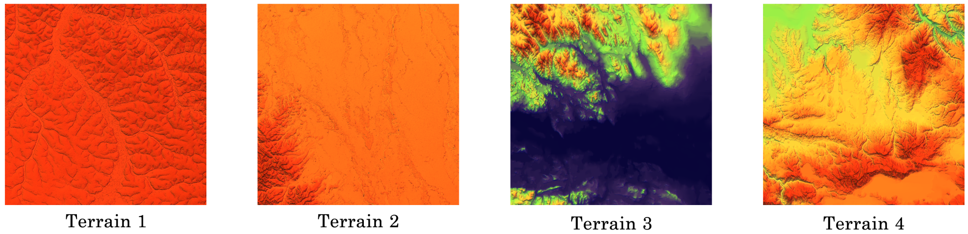


Figure 4. Images of terrains used in the tests

4. Results

The greedy method (proposed by Franklin [Franklin 2002]) and the methods proposed in this work were implemented in C++ and tested in a Core 2 Duo computer with $2,8$GHz and 4GB of RAM memory using terrains (from Brazil and from USA) with different characteristics (lakes, mountains and plain regions) – See Figure 4. These terrains were obtained from NASA SRTM project [Laboratory 2009]. Figure 4 shows some images generated from these terrains. All terrains have a resolution of 1201×1201 points (SRTM3 default resolution) and they are described below:

- Terrain 1: A terrain from Brazilian Midwest region (with mountains).
- Terrain 2: Another terrain from Brazilian Midwest (with plain regions).
- Terrain 3: Terrain from lake Champlain (border between USA and Canada) west region. This terrain has mountain and plain areas.
- Terrain 4: Mountain terrain from Mount Baker (USA) east region.

Observers and targets elevations were defined as 30-meter-tall and the radius of interest was fixed as 2700 meters (this is equivalent to 30 SRTM3 cells). These values are usual, for example, for cell phone towers. Candidate points generation was done using *vix* and *findmax* methods [Franklin 2002] that consist in using the points visibility index in order to choose "good" candidates. The viewsheds were computated using *viewshed* method [Franklin 2002]

GRASP method was implemented using the processing time and number of iterations with no improvement as stopping criteria (the method will end when the time limit or maximum number of iterations are reached). GRASP and GreedyLS were both implemented using spatial local search with swap radius equal to three times the interest radius value.

Each test case was executed only once since in two heuristics (based on the greedy strategy) the solutions are always the same. On the other hand, the method based on *GRASP* could give different solutions when executed more than once, but all the preliminary tests have shown a very small deviation in the obtained solutions.

Table 1 shows the number of observers needed to cover different percentages of the terrains using the methods described in this paper. The tests used 1 hour as time limit and 1000 iterations as iterations limit. Columns *GRASP Improv.* and *GreedyLS Improv.* show the percentage difference between the number of observers used by the greedy method and methods GRASP and GreedyLS, respectively. Charts 5 and 6 show the number of observers needed by each method to reach the different desired visibility

Terrain	Desired *VIX*	Greedy # Obs	GreedyLS # Obs	GRASP # Obs	GreedyLS Improv.	GRASP Improv.
1	75%	411	395	393	4%	4%
	85%	508	471	468	7%	8%
	95%	675	594	619	12%	8%
2	75%	413	395	394	4%	5%
	85%	506	469	467	7%	8%
	95%	670	595	598	11%	11%
3	75%	423	405	405	4%	4%
	85%	525	486	486	7%	7%
	95%	705	630	629	11%	11%
4	75%	465	450	450	3%	3%
	85%	589	561	558	5%	5%
	95%	836	769	786	8%	6%

Table 1. Comparison of number of observers used by methods GRASP, GreedyLS and Greedy to reach the desired visibility index using observers and targets sited 30 meters above the terrains. The methods used 6000 points as candidates to receive observers.

indexes in Terrains 1 and 4, respectively (considering the same tests conditions used in results of Table 1).

Notice that GRASP and GreedyLS methods obtained better results than Greedy method in all tests, with an improvement of, respectively, 6.7% and 6.9% on average.

Table 2 shows the processing time used by each method in the tests described previously. Notice that, as it is simpler, the Greedy method was much more efficient then the proposed methods. Furthermore, notice that GRASP method spent 1 hour of processing time in all tests where the desired visibility index was 95%. This happened because the method ended due to the time limit stopping criteria (in the other tests, GRASP stopped due to the iteration limit stopping criteria).

Figures 7(a), 7(b) and 7(c) shows solutions obtained, respectively, by Greedy, GreedyLS and GRASP methods considering a visibility index equal to 70% and 30-meter-tall observers sited in Terrain 3. The black regions represents the covered areas. The examples use observers with big interest radius (8100 meters) in order to facilitate the solution visualization.

The methods were also tested with 15-meter-tall observers and targets. As the results were similar to the results from Tables 1 and 2, they were not shown in the paper due to space limitations.

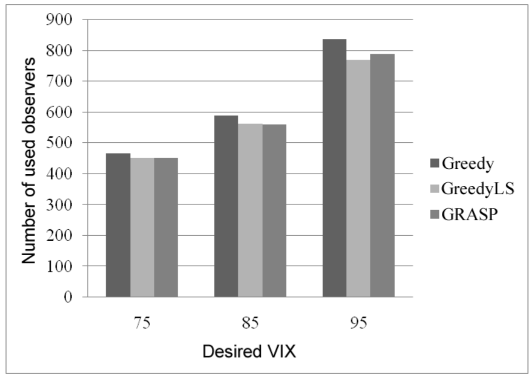


Figure 5. Number of observers needed to reach different visibility indexes in Terrain 1 (considering 30-meter-tall towers).

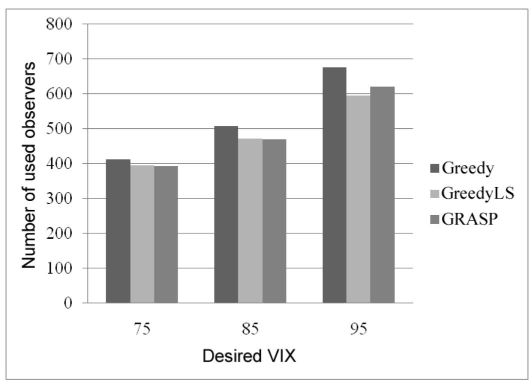
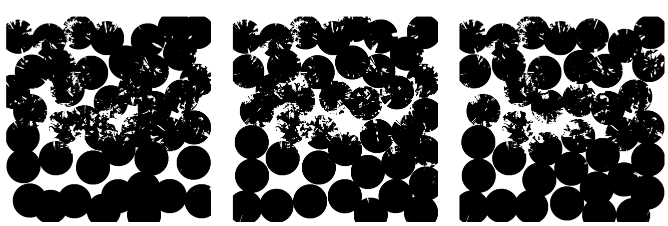


Figure 6. Number of observers needed to reach different visibility indexes in Terrain 4 (considering 30-meter-tall towers).



(a) Greedy method (b) GreedyLS method (c) GRASP method

Figure 7. Solutions obtained using different methods: Greedy method (a) (it used 43 observers), GreedyLS method (c) (it used 41 observers) and GRASP method (c) (it used 40 observers).

Terrain	Desired *VIX*	Time(s) Greedy	Time(s) GreedyLS	Time(s) GRASP
1	75%	7	912	850
	85%	9	1674	1618
	95%	13	3224	3600
2	75%	7	934	1580
	85%	9	1609	1726
	95%	13	3327	3600
3	75%	6	955	730
	85%	9	1687	1518
	95%	13	3314	3600
4	75%	6	992	1097
	85%	9	1761	1299
	95%	16	4303	3600

Table 2. Comparison of processing time (in seconds) used by the three methods to reach the desired visibility index using observers sited 30 meters above the terrains.

5. Conclusions and future work

This work presented the description and implementation of two heuristics to solve the observer siting problem in terrains represented by digital elevation matrices. The results allow to conclude that the proposed methods are able to generate feasible solutions using less observers (an average of 6.8%) than the other method described in literature (the greedy method). This observers saving can represent a considerable saving, mainly when the sited resources involve a large amount of money (cell phone towers, for example).

As a disadvantage, the proposed heuristics demands more processing time then the greedy method.

As a future work, a next step consists in develop graphical user interfaces that facilitate the use of these heuristics and, also, allow the user to modify (using graphical tools) the solution suggested by the heuristics.

The developed methods' source code is available at [de Magalhães et al. 2010].

Acknowledgment

This work was partially supported by FAPEMIG – the State of Minas Gerais Research Foundation and by CNPq – the Brazilian Research Council.

References

Andrade, M. V. A., Magalhães, S. V. G., Magalhães, M. A., Franklin, W. R., and Cutler, B. M. (2009). Efficient viewshed computation on terrain in external memory. *GeoInformatica*. `http://www.springerlink.com/content/p1783648185g1252/`, [accessed on Aug 10].

Ben-Moshe, B. (2005). *Geometric Facility Location Optimization*. Phd thesis, Ben-Gurion University, Israel, Department of Computer Science.

Ben-Shimol, Y., Ben-Moshe, B., Ben-Yehezkel, Y., Dvir, A., and Segal, M. (2007). Automated antenna positioning algorithms for wireless fixed-access networks. *Journal of Heuristics*, 13(3):243–263.

Black, P. E. (2004). Set cover. In Black, P. E., editor, *Dictionary of Algorithms and Data Structures [online]*. U.S. National Institute of Standards and Technology. `http://www.itl.nist.gov/div897/sqg/dads/HTML/setcover.html`, [accessed on Aug 10].

Cormen, T. H., Leiserson, C. E., Rivest, R. L., and Stein, C. (2001). *Introduction to Algorithms*. MIT Press.

Davis, L. (1987). *Genetic Algorithms and Simulated Annealing*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA.

de Magalhães, S. V. G., Andrade, M. V. A., and Franklin, W. R. (2010). Código fonte do site heurísticas. `http://www.dpi.ufv.br/~marcus/projects/SiteHeuristicas.htm`, [accessed on Aug 10].

Franklin, W. R. (2002). Siting observers on terrain. In Richardson, D. and van Oosterom, P., editors, *Advances in Spatial Data Handling: 10th International Symposium on Spatial Data Handling*, pages 109–120. Springer.

Franklin, W. R. and Ray, C. (1994). Higher isn't necessarily better - visibility algorithms and experiments. In *6th Symposium on Spatial Data Handling*, Edinburgh, Scotland. ISBN 2-540-43802-5.

Franklin, W. R. and Vogt, C. (2006). Tradeoffs when multiple observer siting on large terrain cells. In Riedl, A., Kainz, W., and Elmes, G., editors, *Progress in spatial data handling: 12th international symposium on spatial data handling*, pages 845–861. Springer, Vienna. ISBN 978-3-540-35588-5.

hoon Kim, Y., Rana, S., and Wise, S. (2004). Exploring multiple viewshed analysis using terrain features and optimisation techniques. *Computers & Geosciences*, 30:1019–1032.

Laboratory, J. P. (2009). Shuttle radar topography mission. `http://www2.jpl.nasa.gov/srtm`, [accessed on Aug 10].

Laurini, R. and Thompson, D. (1992). *Fundamentals of Spatial Information Systems*. Academic Press.

Li, Z., Zhu, Q., and Gold, C. (2005). *Digital Terrain Modeling - principles and methodology*. CRC Press.

Mitchell, M. (1996). *An Introduction to Genetic Algorithms*. MIT Press, Cambridge, MA.

Nagy, G. (1994). Terrain visibility. *Computers and Graphics*, 18:763–773.

Resende, M. G. C. (1999). Greedy randomized adaptive search procedures (grasp). *Journal of Global Optimization*, 6:109–133.

Resende, M. G. C. and Werneck, R. F. (2004). A hybrid heuristic for the p-median problem. *Journal of Heuristics*, 10(1):59–88.

Interacting with Spatial Data Warehouses through Semantic Descriptions*

Renato Deggau[1,3], **Renato Fileto**[1], **Danilo Pereira**[2,3], **Eugênio Merino**[2]

[1] Post-Graduate Program in Computer Science (PPGCC)
[2]Post-Graduate Program in Knowledge Engineering (EGC)
Federal University of Santa Catarina (UFSC)
PO BOX 476, 88040-900, Florianópolis-SC, BRAZIL

fileto@inf.ufsc.br, merino@deps.ufsc.br

[3]Agricultural Research and Rural Extension Institute of Santa Catarina (EPAGRI)
PO BOX 502, 88034-901, Florianópolis-SC, BRAZIL

{rdeggau,danilopereira}@epagri.sc.gov.br

Abstract. *Users with limited GIS and BI skills usually have troubles to interact with spatial data warehouses. This paper proposes a knowledge-based GUI to help these users to find data marts related to domain specific keywords, understand their contents, and do spatial information analysis. The proposed GUI presents ontological descriptions of the structure, contents, and information analysis capabilities of the data marts. It supports the selection of spatial operators and aggregation functions to compose SOLAP queries. The tables, graphs and maps presented in response to these queries allow further user interaction to gradually refine the analysis. An empirical usability test has been applied to the proposed GUI, in a case study for the agriculture domain, and the results suggest that it meets the user's needs.*

1. Introduction

Decision support in areas such as management of natural resources, agriculture, and disaster prevention often requires the analysis of large volumes of spatial data. A spatial data warehouse (SDW) can help to resolve this kind of demand [Rao et al. 2003, Malinowski and Zimányi 2007, Bimonte et al. 2008]. It extends the capabilities of a traditional data warehouse (DW) for handling large volumes of numeric and categorized data in the dimensional model, such as on-line analytical processing (OLAP), with abilities of typical geographic information systems (GIS) for handling spatial data. The goal is to facilitate the analysis of different kinds of data, and present results in maps whenever it is considered appropriate.

Spatial objects can appear in spatial data as members of dimensions (e.g., multipolygons representing states and cities) and as measures of fact tables (e.g., geographic locations of cases of poisoning by pesticides) [Malinowski and Zimányi 2007]. In addition to the dimensional model and the traditional OLAP operators and aggregation functions, a SDW must support spatial data types integrated to the dimensional model [Fidalgo et al. 2004], and a variety of methods for handling spatial objects [Silva et al. 2008].

*Fapesc (grant 12552-2007-0), CNPq (grant 48139212007-6), and Epagri supported this work.

This work focus on the development and usability test of the graphical user interface (GUI) of an ontology-based system for interacting with SDWs called S^2DW (Semantic and Spatial Data Warehouses). It allows ordinary users to search for spatial data marts, using domain specific keywords, and do spatial information analysis by interacting with the recovered semantic descriptions of spatial data marts. These semantic descriptions are built in accordance with a spatial dimensional model, described in the SDW ontology. This ontology defines the possible structural components, operators, and aggregation functions available in the data marts. It describes the possible connections of the structural components of spatial data marts and the possible uses of operators and functions. These concepts are used in conjunction with terms taken from a domain ontology, for semantically describing spatial data marts according to domain specific knowledge.

A spatial data mart is presented in the proposed knowledge-based GUI as a graph. The GUI allows the user to explore the structure and contents of the data mart. It also assists the user to correctly assemble operators, compatible with the types of the selected members of the dimensions and measures of the fact table, to compose spatial OLAP (SOLAP) queries. The query results can be viewed as tables, graphs and maps. Then the user can interact with these results in order to refine his information analysis.

1.1. Related Work

The research problems being currently considered in spatial data warehouses include: (i) SDW modeling [Fidalgo et al. 2004, Malinowski and Zimányi 2007]; (ii) extracting, transforming and loading (ETL) spatial data into a SDW [Skoutas and Simitsis 2006, Martino et al. 2009]; and (iii) human-computer interfaces to support user interactions with spatial data marts for doing information analysis [Rao et al. 2003, Sell et al. 2008, Xie et al. 2007]. This paper addresses the third problem.

Some proposals from the literature use ontologies and semantic annotations to help users to explore a DW and make information analysis [Sell et al. 2008, Xie et al. 2007]. However, these approaches do not address spatial data. The correct use of the variety of spatial operators and spatial aggregation functions available for information analysis in a SDW requires knowledge about their functioning, inputs and outputs.

[Sell et al. 2008] use ontologies and inference to integrate business semantics with the dimensional data, in order to support information analysis services. In [Xie et al. 2007] IT specialists use an extension of OWL to build mappings between the schema of a DW and domain specific terms used in information analysis, in order to allow domain users to specify their analysis needs and automatically generate data marts to meet them. However, none of these studies present visual interfaces for the user to perform dimensional queries involving manipulation of spatial objects.

The rest of this paper is organized as follows. Section 2 presents a SDW about farms and agricultural production that serves as a case study. Section 3 describes the S^2DW's architecture, the SDW ontology, and an example of domain ontology for agriculture. Section 4 presents the S^2DW's knowledge-based GUI and shows how the user can interact with this GUI. Section 5 describes some experiments that were performed to assess the usability of the proposed GUI, and analyzes the results. Finally, Section 6 presents the conclusions, and some topics for future work.

2. Case Study: A SDW of the Agriculture Domain

Figure 1 shows a data mart of a SDW about farms activities and agricultural production in Santa Catarina State, Brazil. This SDW has been produced with data from Epagri (Agricultural Research and Rural Extension Institute of Santa Catarina). The fact table includes the scalar measures *PlantedArea, HarvestedArea, AmountProduced*, and *FarmsCount*, along with the spatial measure *FarmHeadquarter* (geographic coordinates of farms' headquarters). Such measures are grouped according to levels and members of the dimensions *Place* (with geographic objects representing the *State*, its *Regions*, *Cities*, and *WaterBodies*), *Time*, *Agricultural Product*, and *Farm Properties*.

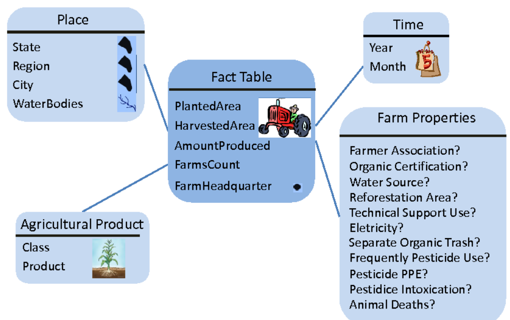


Figure 1. SDW about farms activities and agricultural production

This data mart supports SOLAP with conventional and spatial data. The multi-polygons of the spatial dimension *Place*, can be used in spatial operations to select, group and aggregate data. The results can be presented as tables, graphics or thematic maps. Aggregation of the spatial measure *FarmHeadquarter* can generate subsets of the respective geographic points (e.g., a subset of points that are inside each county) or other geometries (e.g., a Voronoi diagram), depending on the aggregation function chosen among those available for this particular type of spatial data. Different aggregations of spatial measures can reference the same spatial object representation, in order to avoid copies of potentially large representations of spatial objects.

Ordinary users have additional difficulties to interact with SDWs due to the great variety of methods used to manipulate spatial objects. In the following section we present our proposal to facilitate this interaction.

3. The S^2DW System

S^2DW (Semantic and Spatial Data Warehouses) is a system to support information analysis in SDWs [Deggau and Fileto 2009]. It aims to enable the ordinary user to identify spatial data marts related to a topic of interest and pose SOLAP queries on this data mart through a knowledge-based GUI. Additional interactions can be performed on the results of the queries in the form of tables, graphs and maps, in order to refine the analysis.

Figure 2 illustrates the architecture of the S^2DW system and highlights the function of its knowledge-based GUI. The semantic descriptions of spatial data marts refer to the SDW ontology and the domain ontology. The former ontology is fixed and includes conceptual definitions of structural components of spatial data marts (dimensions, levels, measures) and their relationships, in addition to the description of operators and aggregation functions for data manipulation in SDWs. The domain ontology, which varies with the application domain, allows the description of the contents of data marts by using domain-specific knowledge and vocabulary. The S^2DW's GUI provides a graphical view of the knowledge contained in the domain ontology and the semantic descriptions of data marts. It enables the users to explore and use the information analysis capabilities of these data marts by themselves.

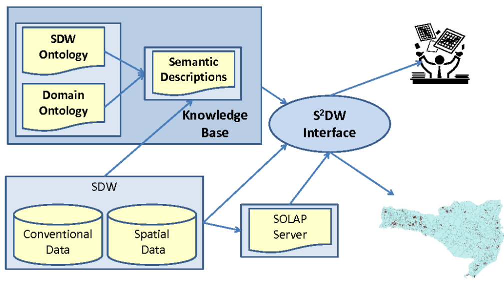


Figure 2. The S^2DW' s general architecture and knowledge-based GUI

3.1. The SDW Ontology

The SDW ontology describes the possible structural compositions and data manipulation facilities in a spatial data mart accessed via S^2DW, according to the dimensional model with spatial extensions from [Fidalgo et al. 2004]. Figure 3 gives a high level view of the hierarchy of concepts of the SDW ontology[1]. It is divided into four portions: (A) the structural components of spatial data marts; (B) the types of spatial entities; (C) the data handling operators used to filter and group information according to the descriptive members of the dimensions; and (D) functions to aggregate measures of fact tables.

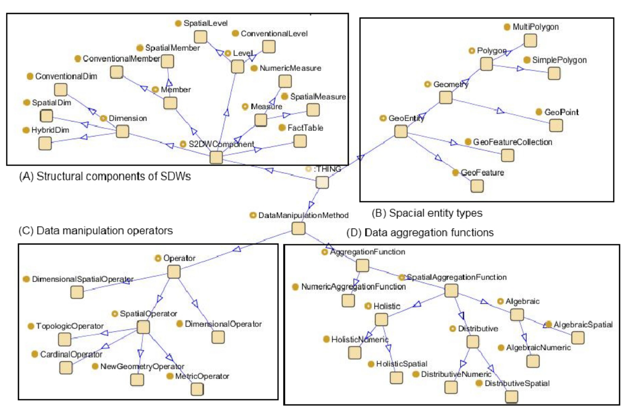


Figure 3. A general view of the SDW ontology

This version of the SDW ontology uses the classification of spatial operators and spatial aggregation functions from [Silva et al. 2008], but the SDW ontology can be adapted to describe other collections of operators and functions.

[1]For simplicity we omit several details of the proposed ontologies. We only present ISA relationships (subsumption hierarchies) and the concepts that are close to the top of the hierarchies.

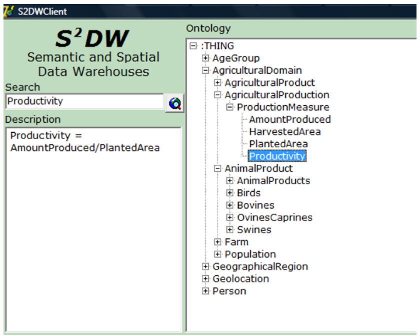


Figure 4. Keyword-based search and navigation on a domain ontology

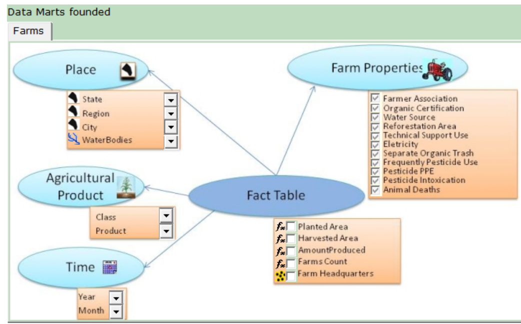


Figure 5. Graphical rendering of the semantic description of a spatial data mart

3.2. A Domain Ontology for Agriculture

S^2DW is domain independent. It can be adapted to different domains by replacing the domain ontology. As such ontology formalizes specific domain knowledge, it enables the system to present data marts, receive and process queries, and present results according to specific domain vocabulary. It facilitates the interactions of users from that particular domain with the SDW. In this work we used an agriculture ontology, which has a small portion of it illustrated on the right side of Figure 4, to describe and facilitate interaction with the SDW about farms and agriculture production.

4. The S^2DW's Knowledge-Based GUI

Figure 4 shows the initial screen of S^2DW. This screen includes an input field for entering keywords on the top left corner, and a hierarchical view of the domain ontology on the right side. The user starts using the system by entering one or more keywords or browsing the view of the domain ontology to choose keywords referring to a subject of interest. The text field on the left bottom side of the window presented in Figure 4 provides information about the posed keyword, taken from the knowledge-base. When the user has finished the indication of subjects, he asks the system to perform a semantic search for data marts with information related to that subject, by clicking on the *Search* button or pressing *Enter*. The user's interactions with the S^2DW's knowledge-based GUI, follows the sequence of steps described below.

1. search for data marts related to a subject;
2. pose one or more SOLAP queries to the recovered data marts:
   (a) select measures, members of dimensions to filter and aggregate these measures, sequences of operators to filter data, and aggregate functions to consolidate data;
   (b) request the execution of the query, by pressing the button *View Results*;
   (c) visualize the results in tables, graphs and/or maps;
   (d) if necessary, refine the information analysis by interacting directly with the presented results (back to the beginning of step 2).

In the following we illustrate this interactive process by means of examples from a case study for the agriculture domain.

4.1. Searching for spatial data marts related to a subject

Suppose the user is interested in analyzing the use of pesticides by farms and its possible impacts on water supplies. If he has any number of simple or composite keywords in his/her mind to express the subject of his concern (e.g. "agrotoxin"), he can provide them, otherwise he can browse the domain ontology view provided by S^2DW, in other to find and select one or more keywords related to his interests. When he asks for the execution of the search with the provided keywords, the system executes a semantic search [Mangold 2007] on the semantic network formed by the ontologies and semantic descriptions of spatial data marts, using a spreading activation algorithm [Crestani 1997] that takes into account the user's preferences with respect to particular denotations of keywords [D'Agostini and Fileto 2009]. After processing this search, S^2DW presents the semantic description of each spatial data mart.

Figure 5 shows the graph representing the semantic description of the data mart from Figure 1 in the S^2DW knowledge-based GUI. This graph is returned in response to a search for the keyword "agrotoxin". This keyword is not explicitly mentioned in the semantic description of that data mart. However, the references to the domain concept "pesticide" in the description of the dimension *Farm Properties* made it possible for the semantic search process to recover this data mart description as a response to the search, because the words "agrotoxin" and "pesticide" are related in the domain ontology.

Next, the user can do information analysis with a spatial data mart using the S^2DW's knowledge-based GUI as described in the following.

4.2. Performing Spatial Information Analysis

The user must select members, measures, operators and aggregation functions in order to specify a SOLAP query using the S^2DW's knowledge-based GUI . In the following we describe the user interactions for posing and analyzing the results of two example queries.

Query 1: Show the geographic distribution of the cases of poisoning by pesticides among the farms of Santa Catarina State in 2003

The user can specify this query by first selecting the spatial measure *Farm Headquarters* and the numeric measure *Number of Establishments*. Then, he could select the value *TRUE* for the feature *Frequent Pesticide Use*, in the dimension *Farm Properties*, in order to indicate the filter condition in this dimension. He could also select the year *2003* in the *Time* dimension, and the state called "Santa Catarina" in the *Space* dimension, in order to indicate the filter conditions in the respective dimensions. The specification of the value *ALL* for the level *City* indicates that the measures must be aggregated by city. The operator *GeoIntersects* and the aggregation function *Collection* are taken as default when the user selects a spatial members and a spatial measure, respectively. After finishing the desired selections, the user press the button *View Results*. Figure 6 illustrates the selected items (the graph in the top left side) and the results returned for this query (the pivot table and the map in the bottom).

Query 2: In the city with the highest number of properties that use them frequently, list the headquarters of farms that are close to the major river.

The municipality with the highest incidence of poisoning by pesticides, *Alfredo Wagner*, can be easily identified in Figure 6. The user can select it, by clicking on the top

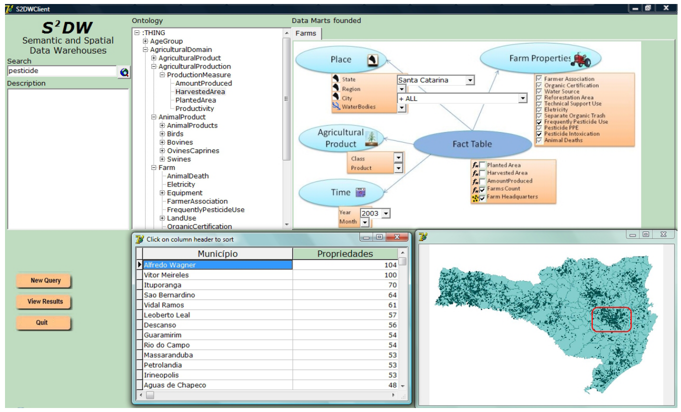


Figure 6. Geographic distribution of the farms with cases of poisoning by pesticides and that make frequent use of pesticides in Santa Catarina State in 2003

city name in the pivot table, or on the representation of the municipality highlighted by a bounding rectangle in Figure 6. Then, the user can further refine his analysis, in order to assess the risk of contamination of *Water Bodies* close to the farms of this city. For doing this, he can zoom-in and reposition the map to focus in the surrounding of the *Itajaí do Sul River*, which is a major water body that crosses the municipality of Alfredo Wagner. The river can be selected by clicking on its representation on the map or by choosing it in the *Water Body* level of the *Space* dimension. Then, the user can click on the icon of the corresponding dimension level, in order to verify the existing operators that can be used to filter information from the fact table, using the spatial objects of the type *Water Body*. This is determined by an inference in the ontology and the semantic descriptions of the data mart. The allowed operators are presented in the dialog box of Figure 7. The explanations accessed by clicking on the respective operator name help the user to find out that he can apply the *Buffer* operator, providing a value for its distance parameter, in order to create an expanded geometry that can be used with the *GeoIntersects* operator to find the farm headquarters that up to 5 kilometers from the *Itajaí do Sul River*. The algorithm below presents the generated SOLAP query in a spatial extension of the MDX[2] language and Figure 8 shows the final results of the analysis in a map on the S^2DW GUI.

```
SELECT Count ([Measures].[Farm Headquarters]),
       Collection ([Measures].[Farm Headquarters])
  ON ROWS {[Local].[City]}
FROM FarmsProdSC
WHERE (
  [Farm Properties].[Frequent Pesticide Use].[True] AND
  [Time].[Year].[2003] AND
  buffer( [Local].[Water Bodies].[Itajaí do Sul River], 5 )
     INTERSECTS
  Farm Headquarters )
```

5. Usability test of the S^2DW's GUI

We developed a prototype of the S^2DW system to analyze the usability of the proposed knowledge-based GUI. It was implemented as a stand-alone application running on Win-

[2]http://msdn.microsoft.com/en-us/library/ms145506.aspx

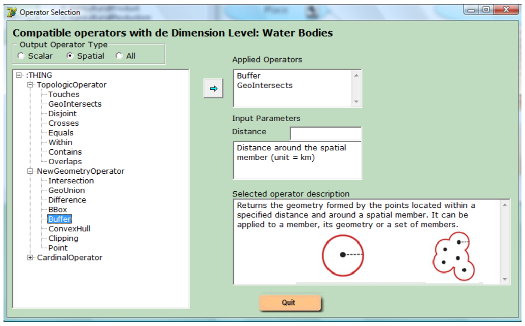


Figure 7. Available operators for filtering information using the spatial member WaterBodies

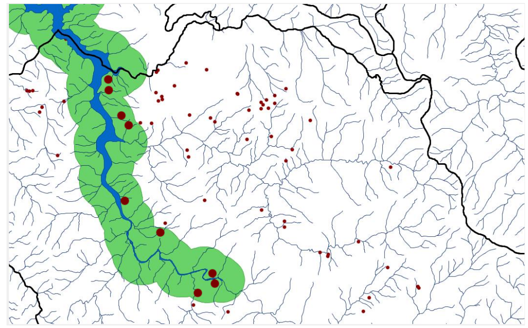


Figure 8. Resulting map after the application of Buffer and the Intersects operators

dows. We used this prototype to perform some experiments with users from the agriculture domain to assess whether the S^2DW's knowledge-based GUI provides an appropriate means for these users to do information analysis. In these experiments we asked the users to do some data analysis tasks on the spatial data mart presented in Figure 1, and assessed their performance on doing these tasks with S^2DW. We also applied a semi-structured questionnaire to the participants, asking them to answer some questions, in order to assess their opinion about proposed system.

5.1. The experimental setting

The experiment used inspection methods in accordance with the norm ISO 9241-11 [ISO/IEC 1991]. The empirical test of usability, also known as observation technique, was chosen as the method of inspection. We used the software Morae [TechSmith 2010] to observe the actions of each user, capture data, and apply the questionnaire. According to [Nielsen 1993], experiments with 6 to 12 users are enough to discover 85% of the usability problems. The experiment was administered to 17 users. The execution of the experiment was divided into three steps, as described below:

Step 1 - Presentation of the S^2DW system We first provided an overview of the information analysis process using S^2DW.

Step 2 - Execution of Information Analysis Tasks: The users were asked to specify some SOLAP queries on the S^2DW knowledge-based GUI. The sequence of tasks was previously stored as a script in Morae. The user read the instructions before performing each task and indicated when he was able to start that task. At the end of each task, he indicated its conclusion, and the system presented him the next task.

Step 3 - Application of the Questionnaire: Upon completion of all the information analysis tasks, Morae prompted the user to answer the questionnaire and give his opinion about doing the proposed tasks with S^2DW.

Morae collected information about each user's interactions with S^2DW, including each manipulation of the GUI components, the time spent per task, and the number of mouse clicks per task. We have also recorded the user's behavior on video. The information analysis tasks requested to the users were the following:

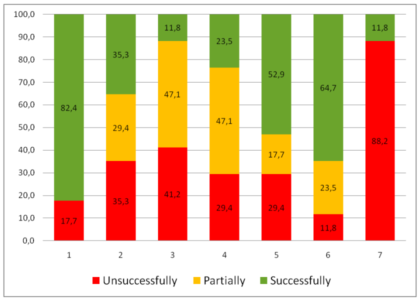


Figure 9. Success per task

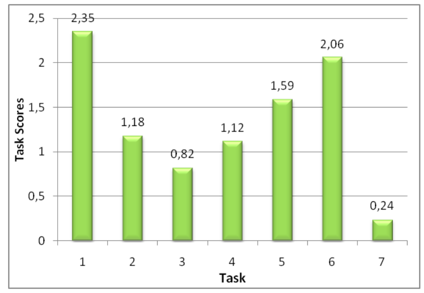


Figure 10. Average score per task

1. Provide keywords or browse the view of the domain ontology in order to find a data mart with information about *agricultural production*.
2. Show the distribution of the *production* of *onions* among the *counties* of *Santa Catarina* state in *2005*.
3. Show the distribution of the *productivity* ($production/plantedarea$) of *onions* among the *counties* of *Santa Catarina* in *2005*.
4. Identify the *city* with the highest yield of *onions* in *2005*, in the table or map resulting from the execution of the previous query.
5. Determine the distribution of *farm headquarters* in the *counties* of *Santa Catarina* that were certified *organic producers* in *2003*.
6. Identify the city with the largest number of *farms* certified as *organic producers* in *2003*.
7. Identify, in this city, the *farm headquarters* that are less than 5 km from a given geographic point.

5.2. Experimental Results

After the execution of the experiment with all users, we manually analyzed the data and video recordings. We assigned a score for the level of achievement of each task for each user, according to the following criteria:

- **Unsuccessful (score = 0):** the user was unable to perform any interaction necessary to perform the task;
- **Partially successful accomplished (score = 1):** the user was able to perform some of the interactions necessary to perform a task;
- **Successful with difficulty (score = 2):** the user was able to perform all the interactions necessary to perform the task, but spending more than 4 minutes; and
- **Successful (score = 3):** the user was able to perform all the interactions necessary to perform the task, spending up to 4 minutes.

Figure 9 presents the percentage of the users with each level of success to perform each task, Figure 10 presents the average score of the users when performing each task, and Figure 11 presents the average time spent by the users to perform each task.

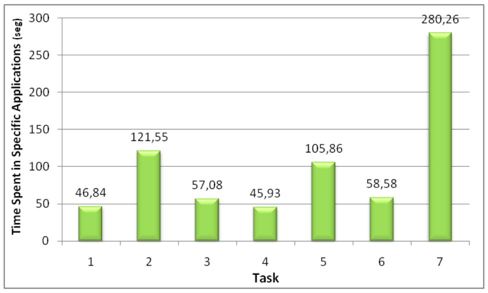


Figure 11. Average time per task

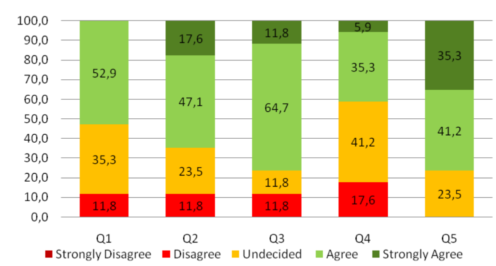


Figure 12. Responses per question

The questionnaire completed by users was composed of the following statements:

1. The proposed interface is clear and easy to understand.
2. The construction of queries is simple and easy to learn.
3. The terms used in the interface are clear.
4. The final results of the queries have been achieved.
5. The interface takes into account the needs of the users.

For each statement above, the user had to choose one of the following scores: (1) Strongly Disagree; (2) Disagree; (3) Undecided; (4) Agree; (5) Strongly Agree. Figure 12 shows the distribution of the users responses for each one of the five questions.

5.3. Analysis of the Results

Most users considered tasks 1 and 2 relatively easy. The percentage of users who were "Unsuccessful" to perform the tasks increases for tasks 2 and 3, as shown in Figure 9. Then this percentage decreases, indicating that a learning process has occurred. Task 7 was considered the most difficulty.

The percentage of successful executions of task 3 was one of the lowest. It involves the measure *productivity*, which is not explicit in the fact table, but has to be calculated from existing measures ($AmountProduced/PlantedArea$). This suggests that some support is necessary to define or recover definitions of composite measures.

The average score of task 7 was the lowest one (Figure 10). There were users who did not find the button to check the spatial operators that could be applied. The high quantity of time spent to execute this task suggests that users knew what they wanted. They tried many actions looking for the button to select the spatial operation to be used.

Figure 11 shows a decrease of the time spent with subsequent tasks of the same analysis, except for task 7. Within the refinement of an information analysis, the interface keeps the previous selections, speeding up the procedures to specify subsequent queries. In task 1, which was supposed to be very fast, many users initially posed some quite arbitrary keywords, just to explore the system.

Analyzing the responses provided by users for the questionnaire, one can observe that there is a total or partial agreement of the users with all statements. The only exception was that the users said that some of the desired results were not achieved (negation of statement 4). However, there was a general agreement that the proposed system meets the information analysis needs of users (statement 5).

6. Conclusions

S^2DW employs ontologies to describe the structure, information contents, and information analysis capabilities of SDWs. Spatial data marts are semantically described in S^2DW by correlating concepts from the SDW ontology with those of a domain ontology. The former, which is domain independent, describes the possible structural components, operators and aggregation functions of a SDW. The latter can be changed in the S^2DW knowledge base, in order to allow adaptation of the system to different application domains, and support user interactions using domain specific vocabularies.

This paper presented the S^2DW's knowledge-base GUI. It provides abstractions to enable the user with limited computational skills to search for spatial data marts containing information about a topic of interest, and perform information analysis on them. It has the following capabilities: (i) semantic search for spatial data marts related to a topic of interest; (ii) present semantic descriptions of the structure and the information analysis capabilities of the spatial data marts of interest, in a knowledge-based GUI; (iii) enable interactions with the data mart through this GUI, assisting the user to correctly use the available information analysis capabilities for posing SOLAP queries on the spatial data mart; and (iv) allow further user interactions with the semantic description of the data mart, and with the results presented for the posed queries in order to support gradual refinement of an information analysis. The benefits of the proposed system apply to conventional DWs, but are magnified in SDWs, due to the variety and the complexity of spatial objects and spatial data handling capabilities.

The results of usability tests of the proposed GUI suggest that it needs some improvements, but meets the user's needs. The users consider the GUI clear and easy to understand. The SOLAP query specification using this GUI is easy to learn and feasible for the majority of the users. Thus, these experiments gave an indication that the development of the proposed system should continue.

The next step of our research is to test and tune alternative formal data structures and algorithms necessary to completely implement the proposed system. Among the facilities that can be included on top of the proposed architecture are natural language interpretation capabilities, rules, and more sophisticated inferences to better advise the users to properly choose and compose the operators for information analysis.

References

Bimonte, R., Tchounikine, A., and Miquel, M. (2008). Spatial olap: Open issues and a web based prototype.

Crestani, F. (1997). Application of spreading activation techniques in information retrieval. *Artificial Intelligence Review*, 11:453–482.

D'Agostini, C. S. and Fileto, R. (2009). Capturing users' preferences and intentions in a semantic search system. In *Proceedings of the 21st International Conference on Software Engineering & Knowledge Engineering (SEKE), Boston, Massachusetts, USA, July 1-3*, pages 587–591. Knowledge Systems Institute Graduate School.

Deggau, R. and Fileto, R. (2009). Enriching Spatial Data Warehouses with Semantic Descriptions. In *WTDBD (Workshop de Teses e Dissertações em Banco de Dados)*, pages 61–66(in Portuguese).

Fidalgo, R., Times, V. C., da Silva, J., and da Fonseca de Souza, F. (2004). Geodwframe: A framework for guiding the design of geographical dimensional schemas. In *DaWaK*, pages 26–37.

ISO/IEC (1991). *Quality characteristics and guidelines for use. - ISO/IEC 9126*. ISO, Geneva, Switzerland.

Malinowski, E. and Zimányi, E. (2007). Logical representation of a conceptual model for spatial data warehouses. *GeoInformatica*, 11(4):431–457.

Mangold, C. (2007). A survey and classification of semantic search approaches. *International Journal of Metadata, Semantics and Ontologies*, 2(1):23–34.

Martino, S. D., Bimonte, S., Bertolotto, M., and Ferrucci, F. (2009). Integrating google earth within olap tools for multidimensional exploration and analysis of spatial data. In *ICEIS*, pages 940–951.

Nielsen, J. (1993). *Usability Engineering*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA.

Rao, F., Zhang, L., Yu, X., Li, Y., and Chen, Y. (2003). Spatial hierarchy and olap-favored search in spatial data warehouse. In *DOLAP*, pages 48–55.

Sell, D., da Silva, D. C., Beppler, F. D., Napoli, M., Ghisi, F. B., Pacheco, R. C. S., and Todesco, J. L. (2008). Sbi: a semantic framework to support business intelligence. In *OBI '08: Proceedings of the first international workshop on Ontology-supported business intelligence*, pages 1–11, New York, NY, USA. ACM.

Silva, J., Times, V. C., Salgado, A. C., Souza, C., Fidalgo, R. d. N., and de Oliveira, A. G. (2008). A set of aggregation functions for spatial measures. In *DOLAP '08: Proceeding of the ACM 11th international workshop on Data warehousing and OLAP*, pages 25–32, New York, NY, USA. ACM.

Skoutas, D. and Simitsis, A. (2006). Designing etl processes using semantic web technologies. In *DOLAP '06: Proceedings of the 9th ACM international workshop on Data warehousing and OLAP*, pages 67–74, New York, NY, USA. ACM.

TechSmith (2010). Morae. `http://www.techsmith.com/morae.asp/`.

Xie, G. T., Yang, Y., Liu, S., Qiu, Z., Pan, Y., and Zhou, X. (2007). Eiaw: Towards a business-friendly data warehouse using semantic web technologies. In *ISWC/ASWC*, pages 857–870.

Towards Intelligent Analysis of Complex Networks in Spatial Data Warehouses

Sergio Izack Vieira Filho[1,4], **Renato Fileto**[1,2],
Andre Salvaro Furtado[2,5], **Ricardo Haus Guembarovski**[3,4]

[1]Post-Graduate Program in Computer Science (PPGCC)
[2]Informatics and Statistics Department (INE)
[3]Post-Graduate Program in Knowledge Engineering (EGC)
Federal University of Santa Catarina (UFSC)
PO BOX 476, 88040-900 - Florianópolis-SC, BRAZIL

[4]Santa Catarina Electrical Central (CELESC)
Av Itamarati, 160 - 88034-900 - Florianópolis-SC, BRAZIL

[5]Geography Department (DG), Santa Catarina State University (UDESC)
Av. Madre Benvenuta, 2007 - Itacorubi - 88.035-001 - Florianópolis-SC, BRAZIL

{sergioivf,fileto,asalvaro}@inf.ufsc.br, ricardohg@celesc.com.br

Abstract. *This paper proposes a knowledge-based model to support spatial and temporal analysis of complex networks in data warehouses. The proposed model, inspired from geography, represents the dynamics of spatial elements in the network. Ontologies, describing element classes from a domain, space partitions, and time periods, respectively, help to define dimensions of data marts customized for specific analysis needs. This model has been implemented in a prototypical framework, with a GUI supporting spatial OLAP and visualization of the varying state of network portions in graphs and maps. It supports the investigation of spatial and temporal patterns and tendencies, as has been observed in an electricity distribution case study.*

1. Introduction

On-Line Analytical Processing (OLAP) [Kimball and Ross 2002] in traditional data warehouses (DW) organized according to the conventional multidimensional model can only handle categoric dimensions and numeric measures [Inmon 2005]. On the other hand, some studies show that around 80% of data are related to some spatial information [Malinowski and Zimányi 2008]. GIS [Chang 2007] are also important to provide information for decision making purposes. The composition of these technologies, DW and GIS, can be useful for several applications, such as the spatial analysis of bird migration, traffic behavior, and power grids [Malinowski and Zimányi 2008]. Spatial data warehouses (SDW) [Bimonte et al. 2006, Escribano et al. 2007] try to reconcile these two successful technologies. In SDW's multi-dimensional and spatial data models [Damiani and Spaccapietra 2006, Malinowski and Zimányi 2008], spatial objects can arise as members of dimensions (e.g., polygons representing states and cities) and as measures of fact tables (e.g., geographic points where equipments are installed). Spatial On-Line Analytical Processing (SOLAP) [Rivest et al. 2005] handle spatial elements integrated to the dimensional model, using a wide variety of operators and aggregation functions for handling spatial objects [Ruiz and Times 2009], besides the traditional OLAP operators and aggregation functions used to handle conventional data. It enables information analysis using geographic features, where topological relations are used to filter and aggregate conventional and spatial measures [Malinowski and Zimányi 2008].

However, current SDW technologies have limited means to handle complex networks of spatial elements (objects located in the space). Complex networks have been used to model and analyze a wide variety of systems and phenomena, ranging from the Internet and the social Web to traffic and power grids [Strogatz 2001, Newman 2003]. This paper proposes a model for handling complex network data in a SDW. This model represents spatial elements, their functions, connections and dynamics as proposed by [Santos 2008]. The forms of these elements are represented as geographic extensions. Their functions are defined by concepts from domain ontologies. These concepts also influence the measures that can be taken from the spatial elements taking part in a process. The domain ontology, along with an ontology about spatial partitions and another ontology about time periods, help to define the dimensions of data marts. We have built a framework based on this model which has a knowledge-based graphical user interface (GUI) for domain experts with basic TI skills to use SOLAP to analyze large amounts of data from vast complex networks, and visualize the results in graphs and maps. The facilities of this GUI include sliders for the user to track the temporal changes of measures of the state of spatial elements in the complex network (such as the load on energy distribution equipments), according to absolute and cyclic time periods. These facilities have been tested using a case study in the area of electricity distribution.

1.1. Related Work

Several SDW models have been proposed in the literature [Damiani and Spaccapietra 2006, Malinowski and Zimányi 2008]. In addition to conventional and spatial data, many applications involve the handling of historical series. Representations of time have to take into account the granularity of the periods of time and cycles of time [Malinowski and Zimányi 2008]. Thus, some multidimensional models have been enriched with temporal [Mendelzon and Vaisman 2000, Moreno et al. 2009] and spatio-temporal [Bertino et al. 2009, Golfarelli and Rizzi 2009] features, giving rise to spatio-temporal data warehouses (STDW) [Savary et al. 2004]. However these models do not consider spatial elements connected in a network.

Works about visualization of the state of complex networks, such as those proposed for the analysis of power grids [Overbye et al. 2007, Moreno-Munoz et al. 2009], on the other hand, do not provide systematic means like SOLAP for the analysis of large volumes of spatial and temporal data from these networks. Other works address the analysis and visualization of trajectories, in applications like the analysis of the traffic [Braz et al. 2007, Leonardi et al. 2010], but they do not incorporate the topology of the spatial network and precise measures of the state of network elements in the data model.

Dynamic networks of elements on earth have been addressed by geography [Santos 2006, Santos 2008, Câmara et al. 2001]. In order to give a better understanding of the complex relationships of geographical elements over time [Santos 2008] proposes to model such elements with respect to four aspects: form, function, structure (connections), and the processes that the connected elements participate. [Câmara et al. 2001] suggests that future geographic information systems would use such ideas from the geography, along with means for knowledge representation, such as ontologies, to cope with dynamic systems in the space.

The remainder of this paper starts with a review of some foundations from geography in Section 2. Section 3 describes the proposed model for handling data from complex networks of spatial elements in data warehouses. Section 4 presents some information analysis using the proposed knowledge-based model and GUI in a case study. Finally, Section 5 enumerates the conclusions and future work.

2. The Geographic Space

The geographic space has such a complexity that it requires an analytical model that fragments the totality of the objects and phenomena taking place in the space [Santos 2006]. A model for spatial analysis of networks of spatial elements connected or interacting with each other is proposed [Santos 2008]. This model is based on descriptions of complex spatial elements, their interrelations, and measures characterizing their state, according to four aspects: form, function, structure, and process.

The *form* is the visible aspect of a spatial element in a given instant. It can be influenced, though in an imperfect way, by the role of the spatial element in an information analysis from a particular viewpoint. The form carries an idea of the purpose and relevance of the spatial element, for performing one or more functions in a system.

The *function* is the action performed by or with the use of the spatial element. It is what is expected to be executed by the spatial element with a given form. The same form of a spatial element can have one or more functions, according to the context, which includes the space and time where the element is inserted. This variation is the consequence of constant changes in the physical environment and the society. Though some forms can last for long periods, they can have new functions assigned to them.

The relationships between spatial elements, i.e., the components of the spatial whole, results in the *structure*. The form-function pairs of the spatial elements can be organized according to flows of relationships among elementary units and their significant combinations [Perroux 1969].

The *process*, by its turn, is a continuous action developing towards some result, which can only be apprehended in a spatio-temporal view [Santos 2008]. By analyzing the process it can be possible to understand the patterns and trends in the evolution of the structure, and the interactions among spatial elements to reach a given result.

The conceptualization of these aspects of the description of spatial elements and their connections is necessary for a complete description of what happens in networks of elements in the space as time goes on. It allows a concrete and precise interpretation of the evolving processes in the geographical space [Santos 2008].

In the real world, the concepts for describing spatial elements are highly dependent on the application domain. For example, the evolution of the traffic can be analyzed in a mesh composed of highways, roads, streets, and other elements of transport networks. On the other hand electricity distribution and consumption occur in networks composed of power generation stations, wires, switches, voltage converters, and consumption meters, among other equipments. However, the partitions of the space and the time periods used for information analysis have many concepts that apply to a variety of domains. Thus, the time and space conceptualizations necessary for information analysis can sometimes be reused in different domains with some customization.

3. The Proposed Model

Our dimensional model for handling information from complex networks in spatial data warehouses extends the model proposed in [Malinowski and Zimányi 2008]. It also relies on ideas of [Santos 2008] and ontologies. Due to space limitations we abstract details of the dimensional model in this paper. We show how our model organizes information of a complex network of spatial elements and how it supports analysis of this information.

3.1. Complex Networks of Spatial Elements

A complex network in our model is represented as a directed graph $G(V, E)$. Each vertex $v \in V$ represents a spatial element. Each directed edge $e \in E$ represents a connection from an element v to an element v' ($v, v' \in V$). Each vertex $v \in V$ has the form:

$$v = (\Gamma, \Phi, P(v,t), N(v,t), S(v,t))$$

where:

- Γ is a set of forms of v (including spatial location in some coordinate system)[1],
- Φ is a set of functions (roles) of v (functions can be defined by concepts in the *domain ontology* and be associated with forms),
- $P(v,t) : V \times T \rightarrow 2^{\{V-v\}}$ is a function that gives the preceding elements of v in the complex network in a given time period $t \in T$,
- $N(v,t) : V \times T \rightarrow 2^{\{V-v\}}$ is a function that gives the following elements of v in the complex network in a given time period $t \in T$, and
- $S(v,t) : V \times T \rightarrow \mathbb{R}^n, (n \in \mathbb{N}, n > 0)$ is a function that gives the values of a set of n state variables of the spatial element v (each spatial element can be a complex non-linear system with n state variables [Strogatz 2001]).

The time T is represented a sequence of periods of some duration (e.g., hour). The complex network G has a connection $e = (v, v') \in E$ from vertex $v \in V$ to vertex $v' \in V$ in a given time period $t \in T$ if and only if $v' \in N(v,t)$ and $v \in P(v',t)$. For simplicity we only consider in this paper the dynamics of the internal states ($S(v,t)$ varying in time) for each $v \in V$, but not the dynamics of the network topology. In other words, we consider the set of network links E fixed, i.e., $\forall v \in V, t, t' \in T : P(v,t) = P(v,t') \wedge N(v,t) = N(v,t')$.

3.2. Linked Spatial Elements

Figure 1 illustrates the *Complex GeoObjects* representing spatial elements of complex networks in our model. The methods for accessing these objects cover the four aspects of the geographic reality (form, function, structure and process) introduced by [Santos 2008]. The method *getForm* returns the form(s) of the object valid for the context provided as argument. The context can include a geographic region (defined by a concept from the *Space Partitions Ontology*), a time period (defined by a concept from the *Time Periods Ontology*) and a scale. The set of returned forms varies with the context and the state of

[1] We intentionally avoid commitment to any standard for representing forms.

the object in the time period specified in the context. The method *getFunction* returns one or more concepts from the domain ontology describing the function(s) of the object in the context provided as argument. The methods *previousObjects* and *nextObjects* return the sets of objects representing previous and next elements in the complex network, respectively. With these methods one can trace the structure of connected components of the complex network. Finally, *getValue*, returns the value of an aggregate measure (e.g., maximum, minimum, average) of a variable of the internal state of the spatial object in the context provided as argument. It enables the analysis of the processes occurring in the complex network, by giving access to information about the time varying state of each spatial element in the complex network.

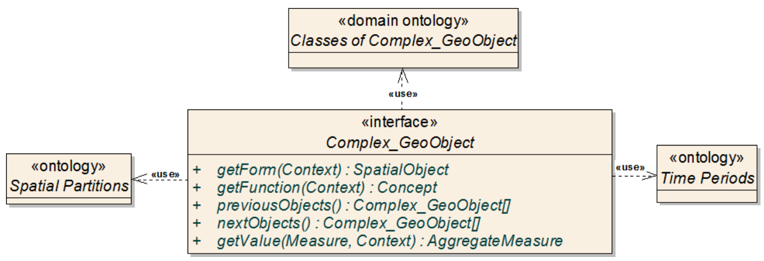


Figure 1. Objects representing spatial elements of complex networks

3.3. Ontologies

The three ontologies referenced in Figure 1 describe the spatial elements and support consolidated analysis of the state of different parts of the complex network along time. Figures 2-a, 2-b and 3 present the top conceptualizations of these ontologies[2]. Analytical dimensions of data marts for addressing specific analysis needs can be specified as views of these ontologies, i.e., extracts with selected concepts, instances and relationships. These subsets constrain the data mart to some focus, avoiding a large, cumbersome and inefficient analytical cube.

The *Spatial Partitions Ontology* and the *Time Periods Ontology* are more stable in our model than the *Domain Ontology*, which describes and classifies specific domain

[2]These are just illustrative portions of the complete ontologies. Many concepts, semantic relationships and all instances are omitted for simplicity and due to space limitation.

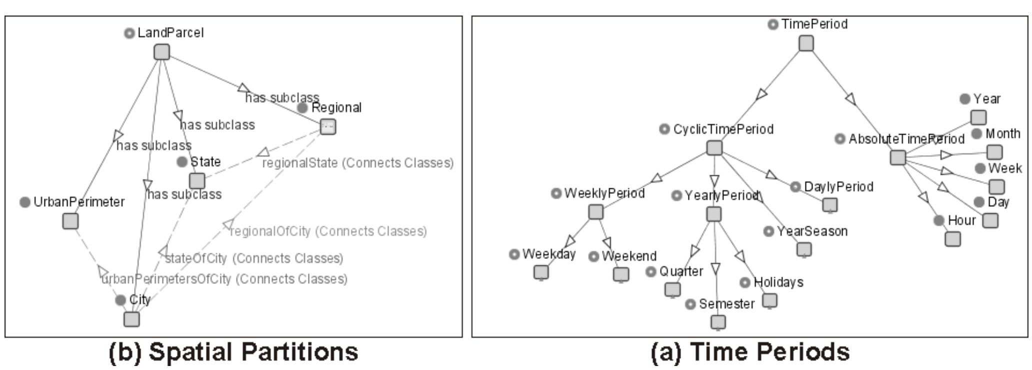


Figure 2. Fragments of ontologies used in the proposed model

concepts. Thus, the latter needs to be changed in order to support information analysis for different application domains. The major concepts of the former ontologies, require minor customizations for different domain. For example, the analysis of electric power grids can use different regions and time periods (more related to social life) than those used for the analysis of and environmental issues (more related to nature).

3.3.1. Spatial Partitions Ontology

Analysis of spatial information requires spatial knowledge, including:

- relationships between spatial partitions and between components of these partitions (e.g., states, which composed of counties);
- homonyms, i.e., different entities referenced by the same name (e.g., *Santa Catarina* referring to a Brazilian State or the island with the same name in that state);
- synonyms, i.e., alternative names referring to the same entity (e.g., *Florianópolis*, *Floripa* or *The Island of Magic* referring to the capital of Santa Catarina State) .

The *Spatial Partitions Ontology* addresses all these issues. It describes hierarchies of partitions of the space in land parcels that can be relevant for information analysis. Land parcels can be countries, states, cities and other regions defined by some criteria (e.g., relief, vegetation, economic and demographic issues). Figure 2-a shows that both *State* and *City* are specializations of *LandParcel*, and that an instance of *State* is composed of instances of *City* or, alternatively, by instances of *Regional*, another specialization of *LandParcel* to divide the territory for energy distribution purposes. The *Spatial Partitions Ontology* also includes instances of these concepts, containment relationships between instances (e.g., *Santa Catarina*, the state, contains the city of *Florianópolis*) and synonyms. Thus, it can be used as a gazetteer to identify and help to solve ambiguities.

3.3.2. Time Periods Ontology

Analysis of a complex network dynamics can refer to two major kinds of time intervals:

- **linear time periods**, which succeed each other in the calendar, such as specific days, weeks, months and years;
- **cyclic time periods**, which repeat in cycles, such as daily periods (e.g., morning, afternoon, evening and night) and yearly periods (e.g., the year seasons).

The *Time Periods Ontology* classifies and describes different kinds of time periods used for information analysis. One user can be interested in analyzing the changing state of portions of a complex network according to linear time (e.g., the evolution of a measure as time goes on) or some cyclic time periods (e.g., the variation pattern of some measure along the seasons or daily periods, considering all the recorded years or days). Figure 2-b shows an extract of the ontology of time periods. This ontology addresses, among other details, the classification on cyclic and absolute time periods, their relationships, and spatial customization of certain time periods according to the geographic location (e.g., the seasons of the year occur in different months in the northern and southern hemispheres).

3.3.3. Electricity Distribution Equipment Ontology

The *Domain Ontology* describes the hierarchy of classes of spatial elements. Concepts of this ontology are used in our model to define the functions of the spatial elements, according to domain specific knowledge. Figure 3 illustrates the high level class hierarchy on domain ontology about components of an electrical distribution grid, used in our case study. This ontology classifies the electricity distribution equipments according to their function and voltage range.

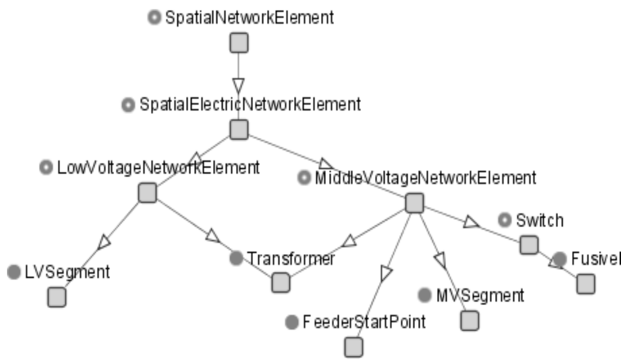


Figure 3. Electrical grid equipments ontology

4. Case Study

The electricity distribution companies provide services for different categories of consumers: residential, business, industrial and rural. The Brazilian Electricity Regulatory Agency (ANEEL) imposes certain rules relative to the quality levels that must be observed for these services. These rules force the companies to develop solutions to optimize their operation and reach quality standards. The control of their activities, the planning of investments and the understanding of technical problems and limitations require the spatial and temporal analysis of the performance of their network of electricity distribution assets. The domain expert needs systematic ways to assess the temporal evolution of the topology and the operation (level of load, number of faults) of the electricity network equipments installed in the geographic space. The analysis of the behavior of these variables is fundamental for right and fast decision making.

4.1. Data Mart for Analyzing an Electricity Distribution Grid

Figure 4 presents the star schema of a data mart developed for the load analysis of the power grid of an electricity distributor: Santa Catarina Electrical Central (CELESC). The three analytical dimensions of this data mart, *Space*, *Time*, and *Equipment* are views of the ontologies presented in section 3.3 specified for addressing some specific analysis needs in this data mart. The levels of the dimension *Space* are specializations of the concept *Land Parcel*: *State*, *Regional* (set of cities from a state under the same electricity distribution manager), *City* and *UrbanArea*. The members of these levels have geographical extents to represent the respective land parcels in maps. Some members are associated to each other according to containment relationships (e.g., *Florianópolis* is a city of *Santa Catarina State*). The level *equipmentType* of the dimensional *Equipment* has as members the types of electricity distribution equipments described in the domain ontology of

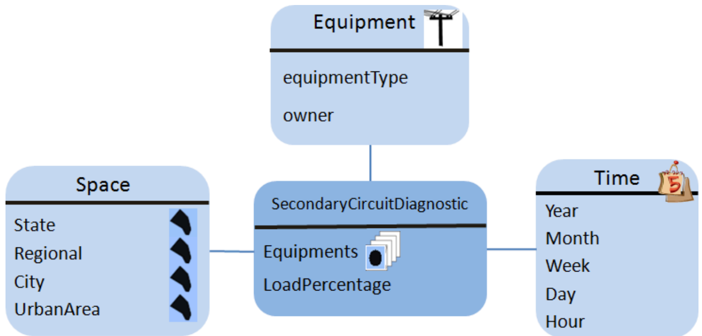


Figure 4. DM schema for an electricity distribution grid

figure 3 and are associated according to the conceptualization defined in that ontology. These equipments can be owned by CELESC or third parties, including consumers in some cases. The fact table maintains the numeric measure *LoadPercentage* (the maximum, minimum or average load percentage for all the equipments installed in a certain region during a time period) and the spatial measure *Equipments* (complex objects representing the equipments and measures of their state along time, as described in section 3).

4.2. Information Analysis

The model proposed in section 3 supports analysis of information from complex networks on the geographic space using OLAP and tracing of the state of the interconnected components of the network in different time periods. In the following we present some spatial information analysis that can be done on this model, using the data mart illustrated in figure 4 as a case study.

4.2.1. Spatial OLAP

The user can initiate his analysis of an extensive complex network by using SOLAP to have an overview of the situation and investigate the consolidated state of the network in different regions and time periods. SOLAP provide systematic means for doing information analysis for strategic and tactic purposes, before going into operational details of the network. Figure 5 illustrates the results of a sequence of *drill-down* operations applied to the data mart illustrated in figure 4 to investigate the distribution of the maximum load percentage of the capacity of the power grid across different regions. As the user identifies the regions with high load he can *drill-down* into these regions to see zoomed-in maps of the distribution of the load in their sub regions. Figure 5(a) shows the distribution of the maximum load percentage of electricity distribution equipments across the cities in the Florianópolis metropolitan area, Figure 5(b) shows details of the load in Tijucas, a particularly loaded city, and Figure 5(c) shows the load on specific equipments installed its urban area. These maximum loads are taken from the SDW measures $S(v, t)$ of each spatial element v contained in the corresponding region, with $t = April$ 2010.

The calculus of the measure *LoadPercentual* presented in figure 5 for each spatial element is based on the relation between the nominal power (kVA) and the energy demand. The nominal power is obtained by summing the power supplied by the medium to

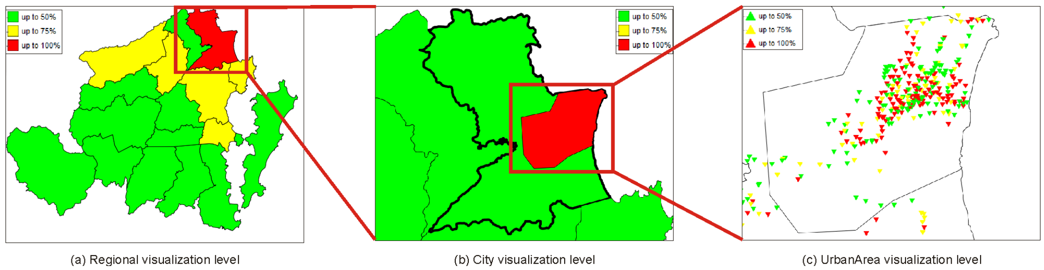


Figure 5. *Drill-Down* on regions to analyze the distribution of the load

low voltage transformers located in the each land parcel. The demand is derived from the total energy consumption (kWh) measured in the consumption units. This estimation can be done by using a statistic approximation function. The calculus of the measures follows a sequence of steps for each land parcel[3]:

1. search for the lvsegments connected in each transformers and consumption units located in the land parcel;
2. sum the nominal power of the transformers in the parcel;
3. sum the power consumed by the consumption units in the parcel, and derive the maximum demand;
4. calculate the relation between the nominal power and the maximum demand.

4.2.2. Trace Analysis of the Network

The trace operation analyses the state of the connected spatial elements of portions of the complex network in different time periods. In the case study of the electric energy distribution it is possible to assess, for example, the energy flow across the equipments installed in the geographic space. Figure 6 shows the evolution of the maximum load of a major electrical energy feeder located in Florianópolis downtown. The user can move the time slides under the graphic showing the variation of the overall load on the feeder in different months (in the right side) in order to visualize the temporal evolution of the distribution of the load in the map (in the left side), month by month. Figure 6(a) shows the load in May 2009 and Figure 6(b) in December 2009. These graphs plotted by following the function $N(v, t)$ (next spatial element) for each equipment v, starting in the $Feeder\ Start\ Point$ for the respective month.

The maps presented in Figure 6 show an estimation of the maximum electrical current flowing on each distribution equipment. This measure is calculated from an estimation of the maximum demand on each equipment, using the formula:

$$I = \frac{DEM}{V \times \sqrt{3}}$$

Where:

[3]We consider the time period fix in order to simplify the explanations. The measures can be aggregated for different time periods as well.

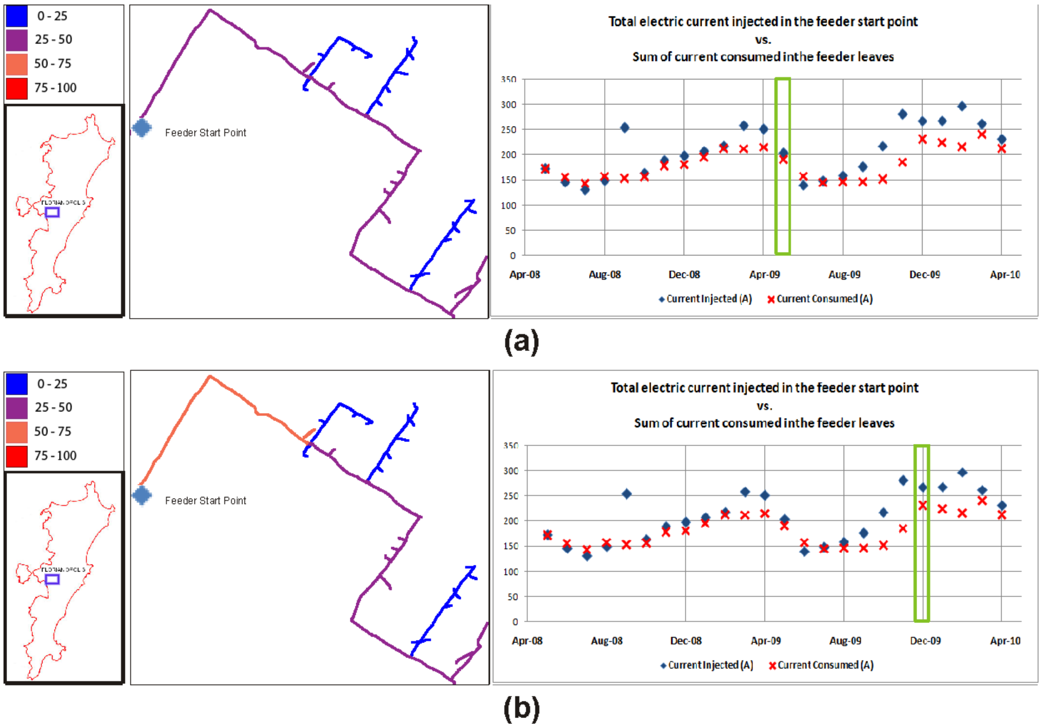


Figure 6. Spatio-temporal analysis of the load in a power grid portion

I is the maximum current consumed (A);
DEM is the maximum estimated demand (kVA); and
V is the operation voltage (V).

The estimation of the demand on the equipments begins in the feeder start point, where it is estimated using the total current injected in the circuit and the total energy consumption. The demand for each other equipment in the same feeder is estimated using the demand in the previous equipments (previous elements in the complex network) and the consumption in the connected consumption units (leaves of the transitive closure of following elements in the network).

Finally, some issues observed in Figure 6 deserve explanations:

- Outliers are due to error in the collection of basic measures such as the energy consumption (from the consumer's meters) and commercial losses due to factors like the joule effect, and energy thieving.
- Some measures of energy consumption are higher than the energy injected in the corresponding portion of the circuit, due to measurement faults and, sometimes, due to modification in the topology of the power grid which are not communicated to electricity supply company.
- The temporal analysis of the load enables the identification of some seasonal patterns. The graphic in Figure 6 shows that the energy consumption is higher during the summer (due to the high proportion of tourists in Florianópolis at this time and the use of air conditioning) and that there has been a vegetative growth in the energy consumption during the last two years.

5. Conclusions and Future Work

The knowledge-based model presented in this paper enables the representation and analysis of historical series of data about the state of interconnected elements of complex networks in a spatial data warehouse. The measures about the state of the network components are estimated and aggregated according to the network topology and a dimensional schema. It supports the analysis of portions of the network using SOLAP, as well as tracing the state of network portions. These analyses are oriented by analytical dimensions like element types, space, and time, which can be derived from ontologies describing the respective conceptualizations. Information visualization in graphs and maps facilitates the identification of spatial and temporal patterns and tendencies.

Currently the proposed model is being tested using a case study in the area of electricity distribution. Future work include: (i) improve the representation of the network topology and dynamic states of the network elements, in order to efficiently store data and process queries for information analysis; (ii) use data sampling from sensors installed in different kinds of network elements to calibrate methods to estimate the distribution of measures of the state of network elements in different time periods; (iii) develop and test knowledge-based human-computer interfaces to help the users to specify the data marts to address specific analysis needs; (iv) exploit inference techniques to help in the specification of data marts and information analysis; and (v) test the proposed model for analyzing information of complex networks in other domains, like the traffic.

References

Bertino, E., Camossi, E., and Bertolotto, M. (2009). Multi-granular spatio-temporal object models: concepts and research directions. In *16th SIGSPATIAL Conf. (ACM-GIS)*.

Bimonte, S., Wehrle, P., Tchounikine, A., and Miquel, M. (2006). GeWOLAP: A web based spatial OLAP proposal. In *OTM Workshops (2)*, pages 1596–1605.

Braz, F., Orlando, S., Orsini, R., Raffaela, A., Roncato, A., and Silvestri, C. (2007). Approximate aggregations in trajectory data warehouses. In *IEEE 23rd Intl. Conf. on Data Engineering (ICDE)*, pages 536–545.

Câmara, G., Monteiro, A., Medeiros, J., and Davis, C. (2001). Fundamentos epistemológicos da ciência da geoinformação. *Introdução à ciência da geoinformação. URL: http://www. dpi. inpe. br/gilberto/livro/introd (access May 2010)*.

Chang, K. (2007). *Introduction to Geographic Information Systems*. McGraw-Hill.

Damiani, M. L. and Spaccapietra, S. (2006). Spatial data warehouse modelling. *Processing and managing complex data for decision support*, page 27.

Escribano, A., Gómez, L. I., Kuijpers, B., and Vaisman, A. A. (2007). Piet: a GIS-OLAP implementation. In *DOLAP*, pages 73–80.

Golfarelli, M. and Rizzi, S. (2009). A Survey on Temporal Data Warehousing. *Database Technologies: Concepts, Methodologies, Tools, and Applications*, page 221.

Inmon, B. (2005). *Building the Data Warehouse*. Wiley, 4th edition.

Kimball, R. and Ross, M. (2002). *The Data Warehouse Toolkit: The Complete Guide to Dimensional Modeling*. Wiley, 2nd edition.

Leonardi, L., Marketos, G., Frentzos, E., Giatrakos, N., Orlando, S., Pelekis, N., Raffaeta, A., Roncato, A., Silvestri, C., and Theodoridis, Y. (2010). T-warehouse: Visual OLAP analysis on trajectory data. pages 1141 –1144.

Malinowski, E. and Zimányi, E. (2008). *Advanced Data Warehouse Design: From Conventional to Spatial and Temporal Applications*. Springer.

Mendelzon, A. O. and Vaisman, A. A. (2000). Temporal queries in OLAP. In *VLDB*, pages 242–253.

Moreno, F., Arango, F., and Fileto, R. (2009). A multigranular temporal multidimensional model. In *MiproBIS Conf. on Business Intelligence Systems*, Opatija, Croatia.

Moreno-Munoz, A., Gil-de Castro, A., Flores, J., and de la Rosa, J. (2009). Hypermedia user interface integration in distribution management systems. In *4th IEEE Conf. on Industrial Electronics and Applications (ICIEA)*, pages 3482–3487.

Newman, M. E. J. (2003). The structure and function of complex networks. *SIAM Review*, 45:167–256.

Overbye, T., Rantanen, E., and Judd, S. (2007). Electric power control center visualization using geographic data views. In *Bulk Power System Dynamics and Control, iREP Symposium*, pages 1–8.

Perroux, F. (1969). *L'Économie du XX Siécle*. Press Universitaries de France, Paris.

Rivest, S., Bedard, Y., Proulx, M., Nadeau, M., Hubert, F., and Pastor, J. (2005). SOLAP technology: Merging business intelligence with geospatial technology for interactive spatio-temporal exploration and analysis of data. *ISPRS Journal of Photogrammetry and Remote Sensing*, 60(1):17–33.

Ruiz, C. V. and Times, V. C. (2009). A taxonomy of SOLAP operators. In *Brazilian Symposium on Databases (SBBD)*, pages 151–165.

Santos, M. (2006). *The Nature of Space: Technique and Time, Reason and Emotion (in Portuguese)*. Edusp, São Paulo, 4 edition.

Santos, M. (2008). *Space and Method (in Portuguese)*. Edusp, São Paulo, 5 edition.

Savary, L., Wan, T., and Zeitouni, K. (2004). Spatio-temporal data warehouse design for human activity pattern analysis. In *15th Conf. on Database and Expert Systems Applications (DEXA)*, pages 814–818, Washington, DC, USA. IEEE Computer Society.

Strogatz, S. H. (2001). Exploring complex networks. *Nature*, 410:268–276.

Using MDA and a UML Profile integrated with international standards to model geographic databases

Filipe Ribeiro Nalon[2], Jugurta Lisboa-Filho[1], Karla A. V. Borges[2], José Luís Braga[1], Marcus V. A. Andrade[1]

[1]Departamento de Informática – Universidade Federal de Viçosa (UFV)
36.570-000 – Viçosa – MG – Brasil

[2]Prodabel – Empresa de Informática e Informação do Município de Belo Horizonte
Av. Pres. Carlos Luz, 1275 – 31230-000 – Belo Horizonte – MG – Brasil

{frnalon,marcus,zeluis}@dpi.ufv.br,jugurta@ufv.br,karla@pbh.gov.br

Abstract. *In the last 20 years, several conceptual data models specific for Geographic Information Systems (GIS) have been proposed. However, so far there isn't a consensus model, which has generated several problems for the GIS area, such as the lack of interoperability among CASE tools that give support to these models. A UML profile, called GeoProfile, was proposed to standardize the task of geographical data modeling. This article shows the integration of GeoProfile with the international standards of ISO 19100 series, which are addressed to geographical information. This integration is presented through the different abstraction levels of the approach Model Driven Architecture (MDA).*

1. Introduction

The activity of software development is a task that requires more and more the use of standardized methodologies and techniques that are widely known. Currently, the main concern of the designer is to understand well the problem domain in order to generate solutions that suit the real necessities of the users.

In order to help in this task of understanding the problem and reducing the system complexity to be developed, the main technique that is used is modeling. A model is a reality simplification [Erikson et al. 2004]. In database design, the construction of models helps to describe the data without having to worry about the implementation details. The *Model Driven Architecture* (MDA) approach [OMG 2003] enables the development of systems using models in different levels of abstraction.

In the classic databases design [Elmasri and Navathe 2003], the most abstract model level is called conceptual model. The data conceptual modeling is done with languages which syntax and semantics have its focus turned to the conceptual and physical representation of a system [Fuentes and Vallecillo 2004]. Currently, one of the most used and accepted languages is the *Unified Modeling Language* (UML), which is extensible, in order to attend to some specific domains, using for that, a mechanism called profile.

An application domain that has been calling attention at present is the geographic domain, due to the current availability of the spatial data sets. In the last 20 years, several researches have been done aiming to create or adapt conceptual data models for geographic applications.

The existence of several models has brought a problem for the area, which is the lack of a modeling standard. Tools have been created for the several models and there is difficult to obtain interoperability among the created solutions, thus making it impossible the reuse of solutions in other projects. Besides that, there are certain modeling requirements of geographic applications that some models support while others do not.

For the standardization of these models, a UML profile, called GeoProfile [Lisboa Filho et al. 2010], was proposed. The GeoProfile is a UML profile for conceptual modeling of geographic database (geoDB), which puts together the characteristics of the main existing conceptual models. As an effort for the geographic information standardization, some organizations, such as the *International Organization for Standardization* (ISO) and the *Open Geospatial Consortium* (OGC), have published international standards to help in the construction of standardized geographic applications.

This article describes the integration of the GeoProfile with the standards of the ISO 19100 series, using the different abstraction levels of the MDA approach. Section 2 describes UML profiles. Section 3 describes the MDA approach. Section 4 describes some requirements for geoDB modeling and the standards from the ISO 19100 series which can be used in the geoDB modeling. Section 5 describes the GeoProfile. Section 6 compares the abstraction levels between the GeoProfile and the ISO standards, presenting the correspondence between both the concepts. In Section 7 some final considerations as well as some future works are presented.

2. UML profile

UML is a modeling language which can be used in several application domains [OMG 2007]. However, there are situations in which the UML designers are not able to express all the concepts of certain domains. Thus, as it is mentioned by Erikson et al. (2004), in order to avoid that UML became too complex, its creators made them extensible, that is, it is possible to adapt it to a domain or specific platform, through its extension mechanisms, which are *stereotypes*, *tagged values* and *constraints*.

The set of these extension mechanisms can be grouped in a UML profile. The intent of the UML profile mechanism is to supply a direct path to adapt an existing metamodel with the constructors that are specific to a particular domain, platform or method. The profile mechanism is consistent with the *Meta Object Facility* (MOF) specification [OMG 2007].

A well-specified UML Profile would have direct support of CASE tools. In other words, once the profile is defined there is no need to implement new CASE tools. *Enterprise Architect* [EA 2010] and *Rational Software Modeler* [RSM 2010] are examples of CASE tools that support UML profiles.

Hence, the development of a UML profile has proven an excellent method to standardize modeling of specific domains, as it uses the language's popularity and tools compatible with UML2, favoring standard acceptance and reducing time for training in new languages.

3. Model Driven Architecture (MDA)

With the promise of improving the software development, the *Object Management Group* (OMG) has adopted the MDA approach, whose main characteristic is the emphasis given to the models. In this approach, the software development process is directed by the modeling activity of the system. A system model is a description using a specific notation. The artifacts produced in MDA are formal models, that is, models that can be understood by computers [Erikson et al. 2004].

In MDA, the system requirements are modeled using a *Computation Independent Model* (CIM). This model is called domain model or business model and it uses a familiar vocabulary to the domain experts. A CIM does not show details of the systems structure, but the environment in which the system will operate, being useful to understand the problem [OMG 2003].

In the second level of abstraction we find the *Platform Independent Model* (PIM). This is a model with an abstraction level relatively high and independent from any implementation technology [OMG 2003].

Later, a PIM should be transformed into a *Platform Specific Model* (PSM). A PSM is customized in order to specify the system in terms of implementation constructors which are available in a specific implementation technology. For instance, a relational database PSM should include terms such as “table”, “column”, “foreign key”, among others. A PIM can be transformed into one or more PSMs. For each specific technology platform a separated PSM is generated. Next step is the transformation of each PSM to source code. This transformation is relatively direct since the PSM is adjusted to a specific technology.

A MDA key element is that the transformations should be automatically executed. Traditionally, the transformations from model to model or from model to code are manually made. In the MDA approach, in the other hand, transformations are executed preferably by tools [Kleppe et al. 2003].

The OMG also provides some ways of transforming models into MDA, one of them is the transformation using UML profiles. A CIM and a PIM can be prepared using a UML profile independent from platform. This model can be then transformed into a PSM using a second profile, of specific platform [OMG 2003].

4. Geographic database modeling

Geographic databases (geoDB) belong to the category of the non conventional databases. Geographic data have, besides the descriptive attributes, a geometric representation in the geographic space; these data are known as geo-spatial or geo-referenced data.

The geoDB modeling holds some particularities that cause the development of specific solutions for this domain. Friis-Christensen et al. (2001) describe a survey of the geographic data modeling requirements, which were classified into five groups: space-temporal properties, roles, associations, constraints and data quality.

Another requirement list is exhibited in Lisboa Filho and Iochpe (1999). In this study, eight groups of requirements are mentioned, five of them equivalent to the ones presented by Friis-Christensen et al. (2001): modeling possibility of the phenomena in the field and object views, spatial aspects, spatial relationships, temporal aspects and quality aspects. The other requirements, which are not explicitly mentioned in the previous paper, are: possibility of differentiation between geographic phenomena and objects without spatial reference, necessity of organizing the phenomena by theme and possibility of modeling of phenomena with more than one spatial representation (multiple representations).

Currently there are several proposals of modeling for geographic data, among the most known are: GeoOOA [Köster et al. 1997], OMT-G [Borges et al. 2001], MADS [Parent et al. 2008], UML-GeoFrame [Lisboa Filho and Iochpe 2008] and *Perceptory's model* [Bédard and Larrivée 2008]. Each of these models presents particular characteristics and try to implement the requirements of geographic applications modeling.

4.1. International standards for geographic information

The efforts for the international standardization in the area of geographic information have been taking place since the last decade through organizations such as ISO and OGC. The Technical Committee ISO/TC 211 is the one responsible for the preparation of the ISO 19100 series, which define the international standards regarding the geographic information field. These standards aim to promote the usage of geographic information in an efficient, effective and economical way, thus contributing to the solution of global problems, such as the humanitarian and ecological problems.

These standards can contribute in several levels of abstraction, from abstract modeling through implementation aspects. In this article some standards related to data models for geographic information, more specifically the ISO 19107 *Spatial Schema* [ISO/TC211 2003], ISO 19108 *Temporal Schema* [ISO/TC211 2002] and ISO 19123 *Schema for Coverage Geometry and Functions* [ISO/TC211 2005] standards, are analyzed.

The ISO 19107 *Spatial Schema* standard specifies schema to describe and manipulate the spatial characteristics of the geographic features. A feature is an abstraction of a real world phenomenon. This abstraction is a geographic feature if it is associated to a relative localization in the Earth [ISO/TC211 2003]. The standard consists of class diagrams that can be used in a application schema, profiles and implementation specifications. It also defines spatial operations, standards for use in the access, query, management, processing, and data exchange of geographic objects. The ISO 19107 standard defines in details the geometric and topological characteristics that are necessary to describe the geographic features.

The ISO 19108 *Temporal Schema* standard defines the concepts regarding the temporal characteristics of geographic information, showing how these characteristics are abstracted from the real world. Jensen (1994) considers two kinds of time: the *valid time* and the *transaction time*. The first one is the time when a fact is true in the observed reality and it is generated by the user. The second one is the time when a fact is stored in a database from which it can be recovered. This international standard emphasizes the valid time instead of the transaction time. The standard consists of a class hierarchy that considers the geometric and topological aspects of the temporal characteristics [ISO/TC211 2002].

The ISO 19123 *Schema for Coverage Geometry and Function* standard, on the other hand, defines a scheme for the spatial characteristics of coverage. Coverage is a feature that has multiple values for each type of attribute and can represent a simple feature or a set of features. They integrate discreet and continuous geographic phenomena [ISO/TC211 2005]. Examples of coverage include raster, TIN, point coverage and polygon coverage. They are used in several specific areas such as, for instance, remote sensing, meteorology, soils and vegetation.

5. GeoProfile – a UML profile for geoDB conceptual modeling

The UML profile proposed for geoDB conceptual modeling, called GeoProfile [Lisboa Filho et al. 2010], puts together the characteristics of the main conceptual data models of the area, previously mentioned, thus seeking to achieve the requirements of geographic applications modeling.

The GeoProfile was specified following the guidelines for specification of UML profiles discussed in Fuentes and Vallecillo (2004) and Selic (2007). The first step was the construction of the domain metamodel [Lisboa Filho et al. 2010], in which the concepts present in the geoDB modeling and the basic requirements were approached.

The way each considered conceptual model in this proposal (GeoOOA, MADS, UML-GeoFrame, OMT-G and Perceptory's model) meets requirements was examined. The inclusion of the main mechanisms present in each of these models into the GeoProfile allows it to meet most requirements of a geoDB. Table 1 summarizes the results obtained in the comparative analysis between requirements and conceptual models, but also displays in its last column the models that most influenced GeoProfile construction in each requirement.

Table 1. Comparison between requirements and models presented, and major contributions to the GeoProfile

Models X Requirements	GeoOOA	MADS	OMT-G	Perceptory	UML-GeoFrame	Contribuition for GeoProfile
Geographical phenomena and conventional objects	Yes	Yes	Yes	Yes	Yes	Perceptory
Field visions and objects	Partial	Partial	Yes	No	Yes	OMT-G
Spatial aspects	Partial	Yes	Yes	Yes	Yes	OMT-G, UML-GeoFrame
Thematic aspects	No	No	Yes	Yes	Yes	UML-GeoFrame
Multiple representations	Partial	Yes	Yes	Yes	Yes	UML-GeoFrame
Spatial relationships	Partial	Yes	Yes	Partial	Partial	MADS, OMT-G
Temporal aspects	Partial	Yes	No	Yes	Partial	MADS, Perceptory

The second step was to extend the UML metaclasses to create the profile itself. In this step the *stereotypes*, *tagged values* as well as the *constraints* were defined. The GeoProfile stereotypes are shown in Figure 1.

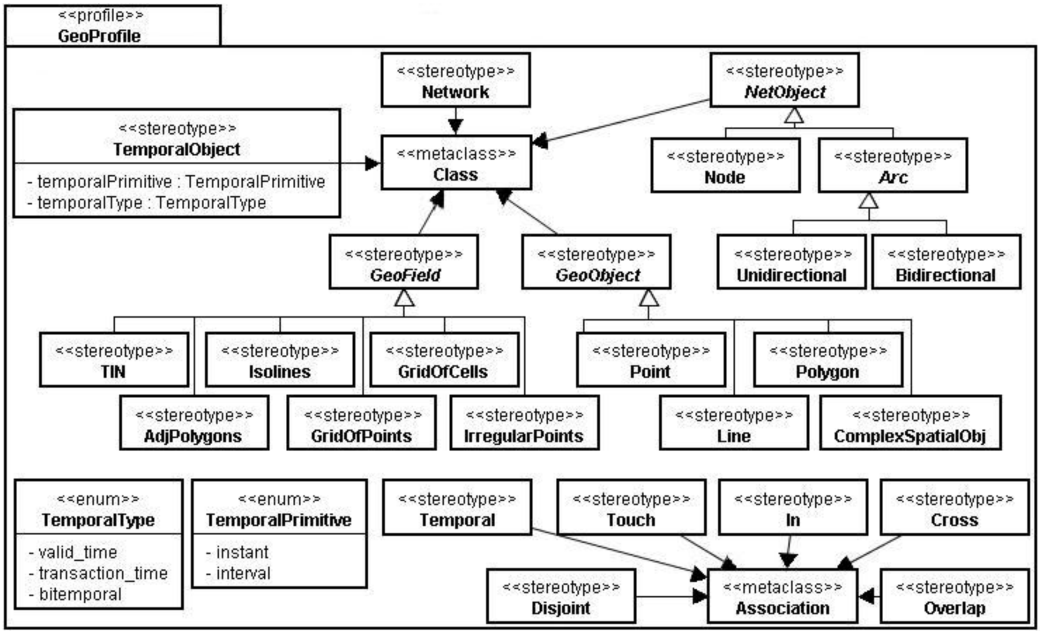


Figure 1. GeoProfile's Stereotypes.

Most of the GeoProfile stereotypes extend the metaclass *Class*. Both the *GeoObject* and *GeoField* stereotypes represent the geographic phenomena perceived in the objects and fields views, respectively. Since these stereotypes were defined as abstracts, as well as the *NetworkObj* and *Arc* stereotypes, they will not be included in the schema classes during the usage of the GeoProfile, but their corresponding subclasses will.

To deal with temporal aspects, the *TemporalObject* stereotype, that also extends the metaclass *Class*, was included. The two enumerations that were included (*TemporalPrimitive* and *TemporalType*) are used to list the possible values that the meta-attributes *(tagged values*) *temporalPrimitive* and *temporalType* may assume, which are: *instant* and *interval*.

Besides the extensions to the metaclass *Class*, extensions to the metaclass *Association* were included. These extensions had the aim of creating stereotypes to serve the topological relationships, which are: *Touch*, *In*, *Cross*, *Overlap* and *Disjoint*. In addition, designers are allowed to indicate that an association between two objects is only valid for one period and this history should be kept in the database. This is done by simply assigning the stereotype *Temporal*.

Besides the stereotypes, some *constraints* were also added, which are useful for the conceptual schema validation. The constraints were defined using the *Object Constraint Language* (OCL) and they have as context the created stereotypes. Details about the constraints specification can be obtained in Lisboa Filho et al. (2010).

6. GeoProfile adequacy to ISO standards using MDA

GeoProfile was designed having in mind higher abstraction levels, helping the designers in the first steps of a geoDB project. This abstraction level, in the classical approach of database design, is called conceptual level, in which only the aspects related to the problem domain are taken into account, without dealing with implementation details. In the MDA approach, this more abstract level is the CIM. According to OMG (2003), such model uses vocabulary that is familiar to the domain experts. The GeoProfile also acts as a CIM, since it represents the geoDB in a more abstract way.

These models of higher abstraction levels should be transformed into models of lower levels, enriched with elements of a more technical order until they achieve implementation details. In the classical approach, this transformation is called logical-conceptual mapping. It is what happens, for example, in the transformation of a schema made in the Entity-Relationship Model for the Relational Model. In the MDA approach, on the other hand, a CIM is transformed into a PIM.

The international standards of the ISO 19100 series, which were analyzed in the previous section, act in a lower level of abstraction as a PIM due to the fact that they present some technical details. Despite the fact they are still in a conceptual level and do not present implementation details, these standards are not in the same GeoProfile abstraction level. Table 2 shows the correspondence between the elements of GeoProfile and elements of the ISO 19100 series standards.

Table 2. Variables to be considered on the evaluation of interaction techniques

Requirements of BDGeo modeling	GeoProfile	Classes in the international standards	Standard
Geographical objects in the object view	*Point*	*GM_Point*	ISO 19107
	Line	*GM_Curve*	ISO 19107
	Polygon	*GM_Surface*	ISO 19107
	ComplexSpatialObj	*GM_Complex*	ISO 19107
Geographical objects in the field view	*TIN*	*CV_TINCoverage*	ISO 19123
	Isolines	*CV_SegmentedCurveCoverage*	ISO 19123
	AdjPolygons	*CV_DiscreteSurfaceCoverage*	ISO 19123
	GridOfPoints	*CV_DiscreteGridPointCoverage*	ISO 19123
	GridOfCells	*CV_GridCell*	ISO 19123
	IrregularPoint	*CV_DiscretePointCoverage*	ISO 19123
Network elements	*Node*	*TP_Node*	ISO 19107
	Arc	*TP_Edge*	ISO 19107
	UnidirectionalArc	*TP_DirectedEdge*	ISO 19107
	BidirectionalArc	*TP_DirectedEdge*	ISO 19107
Temporal objects	*Temporal Object*	*TM_Object*	ISO 19108
	Instant	*TM_Instant*	ISO 19108
	Interval	*TM_Period*	ISO 19108

The execution of these correspondences can be made as a transformation between a CIM, that is a schema using the GeoProfile, and a PIM, that is a schema enriched with elements from the ISO 19100 series standards. For instance, the phenomena perceived in the objects views modeled with the GeoProfile will be mapped to a PIM enriched with the ISO standards in the following way: the classes that were stereotyped as *Point* will be mapped to a class that will have an attribute called *geometry* of *GM_Point* type. In the ISO 19107 standard, *GM_Point* is a kind of basic data for objects with 0-dimension. The same will be done with the other three classes, *Line*, *Polygon* and *ComplexSpatialObj*, which will be mapped to a class with *geometry* attribute of *GM_Curve*, *GM_Surface* and *GM_Complex* types, respectively.

It is important to highlight the fact that these standards offer several ways to model the same geographic information. The correspondence made here was the closest possible to the GeoProfile concepts. The ISO 19100 series standards used above are the ones that come closer to the requirements for geoDB conceptual modeling. For example, the ISO 19107 standard was used to build the correspondence with the GeoProfile stereotypes which represent the geographic objects perceived in the objects view and also with the network elements. The standard is divided into two parts. In the first, which deals with the geometric aspects of geographic information, the correspondence with the geographic objects perceived in the objects view was done. In the second, that deals with the topological aspects, the correspondence was done with the GeoProfile network elements. The ISO 19108 standard was used to build the correspondence with the elements that represent the geographic objects temporal aspects and the ISO 19123 standard was used to make the correspondence with the GeoProfile stereotypes which represent the geographic objects perceived in the fields view.

Regarding the topological relationships, in which the GeoProfile are represented by the *Touch*, *In*, *Cross*, *Overlap* and *Disjoint* stereotypes and extend the *Association* metaclass, in the standard they are dealt with as operations. The ISO 19107 standard is what specifies these operations, which are inherited by all the geometric classes defined in the standard. Therefore, the correspondence with the GeoProfile will not be made, since these operations may be accessed by all the geometric classes, from which the correspondences were made.

6.1. Application example

Figure 2 shows an example of conceptual schema modeled with the GeoProfile. The schema uses a visual notation for the GeoProfile stereotypes. This is a possibility that is suggested by the OMG for UML profiles. In the geoDB modeling visual notation to represent the geographic objects spatial characteristics is used in several models. Some models use other denominations, such as the "pictograms" developed by Bédard and Larrivée (2008). In these schemas a visual notation for the stereotypes `<<Polygon>>` and `<<Point>>` is used.

The schema shows four classes, three of them with spatial characteristics, which were represented by the GeoProfile stereotypes. In this level of abstraction only the "which" geographic representations and not "how" they were implemented were considered, as well as some basic attributes. Therefore, the schema is a CIM, which uses concepts that are the closest to the end users.

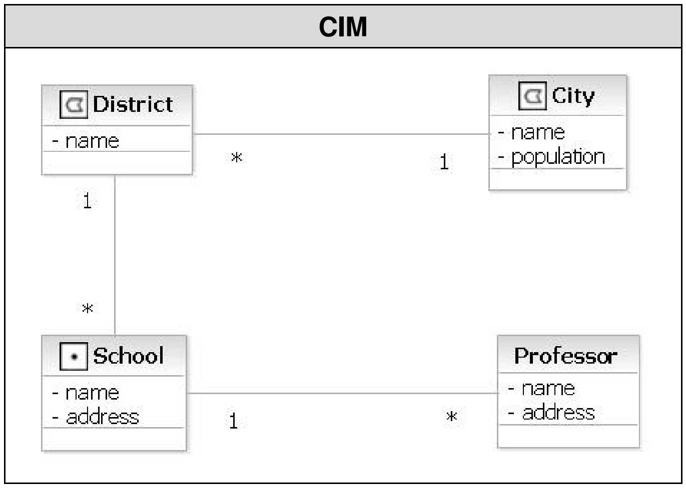


Figure 2. A GeoProfile data conceptual schema (CIM level).

After the construction of the CIM using the GeoProfile, it should be transformed into a PIM, which will take into account some technical details. Figure 3 shows the PIM resulting from this transformation.

The spatial characteristics were transformed into attributes with the types according to the correspondence with the ISO 19100 series standards. For example, the class *City*, which was modeled using the stereotype `<<Polygon>>`, in this level of abstraction has a *geometry* attribute of the *GM_Polygon* type. The same thing was done with the other classes that possess spatial characteristics.

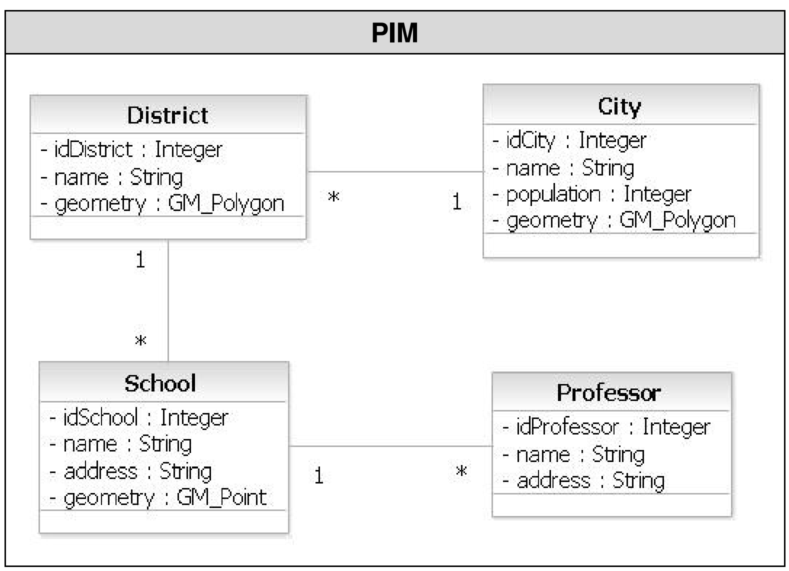


Figure 3. A conceptual data schema enriched with the ISO 19100 series standards (PIM level).

The next step is transforming the PIM into a PSM, that could be, for example, an object-relational data model. However, this level won't be shown in this article. One of the main benefits of the MDA approach is the gain in productivity in the development of software systems through the emphasis given to modeling and to the transformation of models from higher abstraction levels into models of lower abstraction levels in an automated way [Kleppe et al. 2003]. The geoDB project can follow these steps. For example, using tools that support model transformation languages, it will enable the generation of models of lower abstraction levels and, later on, models for specific platforms. An example of model transformation language is *Atlas Tansformation Language* (ATL) [Jouault and Kurtev 2005].

In order to illustrate, a small part of transformation code from the CIM, shown in Figure 2, to the PIM presented in Figure 3, will be shown, using the ATL models transformation language.

The definition of transformations using ATL starts with the transformation module statement as well as the source and target models. The module is defined using the keyword “module” followed by the module name. The keyword “create” indicates the source and target models [Jouault and Kurtev 2005].

After this step, the transformation rules are then defined. Those rules are written using ATL syntax, are saved in files with the extension *.atl* and can use either the declarative or the imperative style. The code presented in Figure 4 shows one of the transformation rules for the CIM and the PIM previously shown. This rule is responsible for creating the classes that have geographic information, that in this case are represented by the GeoProfile stereotypes and for creating the elements that were not contained in the CIM such as, for example, the *geometry* attribute, which type will be the correspondent to the ISO standard.

```
rule stereotypedClass{
   from
        input : GeoProfile!Class(
           not thisModule.emptyGeometry(input.stereotype))
   to
        output : ISO!Class(
          name <- input.name,
                reference <- input.reference ->
           collect( e | thisModule.getReferences(e) ).asSet(),
                 attribute <- input.attribute ->
           collect( e | thisModule.getAttributes(e) ).asSet(),
                 attribute <- id,
                 attribute <- geometry
      ),
      id : ISO!Attribute(
          name <- 'id' + input.name,
          type <- thisModule.integerDataType()
      ),
          geometry : ISO!Attribute(
          name <- input.name + 'Geometry',
          type <- if( thisModule.isPolygon( input.stereotype )) then
                    thisModule.polygonDataType()
                 else
                    thisModule.pointDataType()
                 endif
      )
}
```

Figure 4. Example of an ATL transformation rule.

With the transformation application, the output model is generated in the *XML Metadata Interchange* (XMI) format, which is a standard format for the models exchange and that can be imported by most of the CASE tools with support to the UML2.

7. Final considerations

The GeoProfile development had as main motivation the fact that it can use the UML, together with all its available resources, for example the CASE tools, to conceptually modeling a geographic database. GeoProfile gathers in its definition the main requirements for geographic applications and it uses characteristics of the main existing conceptual data models.

This article has shown the correspondence between the GeoProfile and the ISO 19100 series international standards. The use of standards is very important for the acceptance of the GeoProfile by the scientific community as well as by the geoDB designers.

By using the MDA approach, it was possible to show the difference of abstraction levels between the GeoProfile and the international standards and a possible model transformation, using the ATL. This automation of the transformations constitutes one of the main benefits of the MDA approach.

As future works, we can mention the definition of transformation rules for specific platform models, which have not been dealt with in this article, and also the source code generation, for example, the script for logical specification using SQL.

Acknowledgments

This project was partially funded by CAPES, FAPEMIG and CNPq / MCT / CT-Info.

References

Bédard, Y. and Larrivée, S. (2008). Modeling with Pictogrammic Languages. In *Encyclopedia of GIS*, pages 716–725. Springer-Verlag.

Borges, K.A.V., Davis Jr.C.A. and Laender, A.H.F. (2001) "OMT-G: an object-oriented data model for geographic applications", In: GeoInformatica, v. 5, n. 3, p. 221-260.

Brodeur, L. and Badard, B. (2008). Modeling with ISO 191xx standard. In *Encyclopedia of GIS*, pages 705–716. Springer-Verlag.

EA. (2010). Enterprise Architect. Disponível em: http://www.sparxsystems.com/products/ea/. Acesso em: 01 mar. 2010.

Elmasri, R. and Navathe, S.B. (2003), Fundamentals of database systems, Addison-Wesley Publishing Co., 4th edition.

Erikson, H., Penker, M., Lyons, B and Fado, D. (2004), UML 2 Toolkit, OMG Press, 1th edition.

Friis-Christensen, A., Tryfona, N. and Jensen, C.S. (2001), "Requirements and research issues in geographic data modeling", In: 9th ACM International Symposium on Advances in GIS, p. 2-8.

Fuentes, L. and Vallecillo, A. (2004) "An introduction to UML Profiles" In: The European Journal for the Informatics Professional, v. 5, n. 4, p. 5-13.

ISO/TC211. (2003). ISO 19107: 2003 Geographic Information – Spatial Schema. ISO, Geneva, Switzerland.

ISO/TC211. (2002). ISO 19108: 2002 Geographic Information – Temporal Schema. ISO, Geneva, Switzerland.

ISO/TC211. (2005). ISO 19123: 2005 Geographic Information – Schema for Coverage Geometry and Functions. ISO, Geneva, Switzerland.

Jensen, C.S. (1994). A consensus glossary of temporal database concepts. In ACM SIGMOD, v. 23, n. 1, p. 52-64.

Jouault, F. and Kurtev, I. (2005), "Transforming models with ATL", In: Model Transformations in Practice Workshop (MoDELS 2005), p. 128-138.

Köster, G., Pagel, B. and Six, H. (1997) "GIS-Application development with GeoOOA", In International Journal of Geographical Information Science, v. 11, n. 4, p. 307-335.

Kleppe, A., Warmer, J. and Bast, W. (2003), MDA Explained: The Model Driven Architecture: Practice and Promise, Addison-Wesley Publishing Co., 2nd edition.

Kresse, W. and Fadaie, K. (2004), ISO Standards for Geographic Information, Springer-Verlag, 1th edition.

Lisboa Filho, J., Sampaio, G.B, Nalon, F.R. and Borges, K.A.V. (2010), "A UML Profile for Conceptual Modeling in GIS Domain" In: International Workshop on Domain Engineering (DE@CAiSE'2010), p.18-31.

Lisboa Filho, J. and Iochpe, C. (2008). Modeling with a UML Profile. In *Encyclopedia of GIS*, pages 691–700. Springer-Verlag.

Lisboa Filho, J. and Iochpe, C. (1999) "A study about data conceptual models for geographic database design" In Informática Pública, v. 1, n. 2, p. 67-90 (In Portuguese).

Object Management Group. (2007). Unified Modeling Language: Superstructure, v.2.1.2. OMG, Needham, MA, USA.

Object Management Group. (2003). MDA Guide, v.1.0.1. OMG, Needham, MA, USA.

Parent, C., Spaccapietra, S. and Zimányi, E. (2008). Modeling and multiple perceptions. In *Encyclopedia of GIS*, pages 682–690. Springer-Verlag.

RSM. (2010). Rational Software Modeler. Disponível em: http://www-01.ibm.com /software/awdtools/modeler/. Acesso em: 01 mar. 2010.

Selic, B. (2007), "A systematic approach to domain-specific language design using UML", In: 10th IEEE International Symposium on Object and Component-Oriented Real-Time Distributed Computing (ISORC'07), p. 2-9.

Sendall, S. and Kozaczynski, W. (2003) "Model transformation: the heart and soul of model-driven software development". In IEEE Software, v. 20, n. 5, p. 42-45.

Um algoritmo para identificar padrões comportamentais do tipo *avoidance* em trajetórias de objetos móveis

Alisson Moscato Loy[1], Vania Bogorny[2], Chiara Renso[3], Luis Otavio Alvares[1]

[1]II – Universidade Federal do Rio Grande do Sul, Porto Alegre, Brasil
{amloy, alvares}@inf.ufrgs.br

[2]INE – Universidade Federal de Santa Catarina, Florianópolis, Brasil
vania@inf.ufsc.br

[3]KDD Laboratory – ISTI CNR, Pisa, Italia
chiara.renso@isti.cnr.it

Resumo. *O foco do estudo de padrões comportamentais em trajetórias tem sido a busca por padrões de aglomeração ou semelhança no deslocamento de entidades no espaço e no tempo. Este artigo propõe um algoritmo para a detecção de um novo padrão comportamental que indica quando um objeto móvel está evitando determinadas regiões espaciais. Tal comportamento foi denominado neste artigo de* avoidance. *O algoritmo é avaliado através de experimentos com dados reais coletados na cidade de Porto Alegre.*

1. Introdução

O estudo de padrões comportamentais em trajetórias de objetos móveis está sendo desenvolvido principalmente sob duas óticas de pesquisa, a geométrica [Laube et al 2005, Gudmundsson e Kreveld 2006 , Cao et al 2007, Giannotti et al 2007] e a semântica [Alvares et al 2007, Palma et al 2008, Bogorny et al 2009].

Dentro do estudo de comportamentos de trajetórias sob o ponto de vista geométrico, Laube [Laube et al. 2005] definiu um conjunto de padrões espaço-temporais baseados na direção do movimento e localização do objeto móvel, dentre os quais podemos citar: *flock* - representa um grupo de trajetórias de objetos móveis que se desloca de forma conjunta dentro de um dado raio no espaço; *leadership* - define um deslocamento com padrão de liderança, onde os seguidores devem estar dentro de uma determinada distância de tolerância; *convergence* - representa grupos de trajetórias que se encontram em uma determinada localização em uma dada janela temporal; e *encounter* - semelhante ao padrão *convergence,* porém as trajetórias devem estar juntas num dado raio no mesmo período de tempo. Este e outros trabalhos identificam padrões com base na variação da direção no tempo, e analisam grupos de trajetórias.

Observou-se, porém, a inexistência de estudos que procurem identificar o comportamento de objetos móveis que, ao se deslocar no espaço, evitam determinadas regiões ou evitam outras trajetórias como ocorre quando, por exemplo, pessoas evitam cruzar com determinadas pessoas enquanto caminham em um parque, veículos que mudam de rota frente a situações de tráfego lento ou ainda indivíduos que se deslocam em um ambiente de forma suspeita, evitando câmeras de vigilância ou postos de segurança.

Este artigo tem por objetivo propor um novo algoritmo, capaz de identificar este padrão comportamental *avoidance*, mais especificamente aquele que ocorre quando um objeto móvel evita determinada região espacial.

O artigo está organizado da seguinte forma: a seção 2 trata das heurísticas que norteiam a detecção da ocorrência do padrão comportamental *avoidance*. Na seção 3, um algoritmo construído com base nas heurísticas propostas é apresentado. Experimentos preliminares são apresentados na seção 4 e a conclusão na seção 5.

2. Heurísticas para detecção do padrão *avoidance*

O padrão comportamental *avoidance* ocorre, em termos gerais, quando um objeto móvel está se deslocando em direção a um objeto de interesse ou alvo, como uma câmera de vigilância, por exemplo, se desvia para evitar passar na região de alcance da câmera, e depois volta ao seu trajeto original. O problema é tentar diferenciar o que é realmente um desvio para evitar a câmera de uma mudança de rota por outro motivo qualquer.

Alguns aspectos a serem considerados são:

O objeto móvel deve estar indo em direção ao objeto alvo (o objeto a ser evitado) e se desviar do alvo relativamente perto deste para ser considerado um *avoidance*. Um contra-exemplo seria uma pessoa caminhando e que a um quilômetro de distância se desvia de uma câmera; ela provavelmente mudou de direção por outro motivo e não para escapar da câmera, não caracterizando então o comportamento *avoidance*.

O objeto móvel não deve cruzar (interceptar) a região do objeto alvo (a região de alcance da câmera, por exemplo), pois se o objeto muda de direção mas mesmo assim cruza a região coberta pela câmera, não a evitou, não caracterizando assim um *avoidance*.

A Figura 1 explicita estas idéias. A trajetória t_1 vinha em direção ao objeto alvo, se desvia e depois continua no rumo original, caracterizando um caso de *avoidance*. A trajetória t_2 vai em direção e intercepta o objeto alvo, não o evitando, não caracterizando por conseguinte, um caso de *avoidance*. Já a trajetória t_3, vinha em direção ao objeto alvo mas se desvia deste muito longe (pelo fato do desvio ocorrer fora da *região de interesse* – caracterizada por uma distância d do objeto alvo), não caracterizando portanto o comportamento de *avoidance*.

Para reduzir a incidência de falsos positivos, que podem ocorrer em casos de mudanças de direção dentro da área de interesse ou imprecisão do aparelho de captura dos dados, exige-se que a porção da trajetória que vai na direção do objeto alvo dentro da região de interesse, denominada *subtrajetória direcionada ao alvo*, tenha um comprimento mínimo.

Um objeto móvel que evita a região coberta pela câmera e depois volta a seu trajeto original tem maior probabilidade de estar evitando a câmera do que um objeto que se desviou mas depois não retornou ao rumo anterior. A verificação deste comportamento é realizada testando-se a interseção da trajetória com uma região denominada *região de incremento de confiança*, que é definida como sendo uma região situada entre o objeto alvo e a borda da região de interesse no lado oposto à interseção da trajetória com a borda da região de interesse, e de largura igual ao objeto alvo. A região de incremento de confiança é única para cada trajetória considerada.

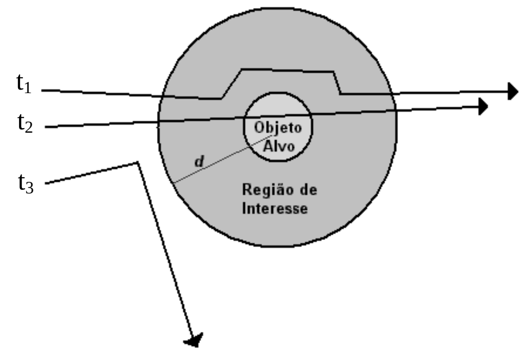


Figura 1. Exemplo de região de interesse, objeto alvo e comportamento de trajetórias.

Nos exemplos da Figura 2, a trajetória t1 possui uma subtrajetória direcionada ao alvo de tamanho válido e intercepta a região de incremento de confiança, apresentando um valor de confiança alto para o *avoidance*. A trajetória t2, da mesma forma, possui uma subtrajetória direcionada ao alvo válida, porém não intercepta a região de incremento de confiança, de modo que apresenta um valor fraco para confiança no *avoidance*. A trajetória t3 não apresenta uma subtrajetória direcionada ao alvo com tamanho válido e a trajetória t4 intercepta o objeto alvo, assim nenhuma destas duas trajetórias apresenta o comportamento *avoidance* em relação ao objeto alvo.

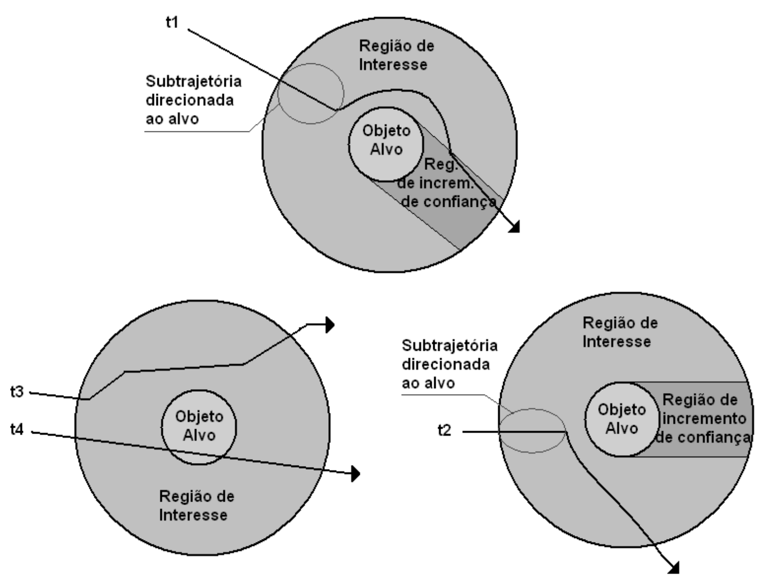


Figura 2. Exemplos de região de incremento de confiança e de subtrajetória direcionada ao alvo.

Uma última heurística utilizada é que a certeza de existir efetivamente um *avoidance* aumenta se o mesmo objeto móvel se desvia de vários objetos alvo.

3. Algoritmo para detecção de *avoidance*

Com base nas heurísticas definidas na seção 2, foi criado um algoritmo para detectar possíveis ocorrências do padrão comportamental *avoidance*, cujo pseudocódigo é apresentado na Figura 3 e explicado a seguir.

```
Entrada:  T // Conjunto de trajetórias
          O // Conjunto de objetos alvo
          d // tamanho do buffer da região de interesse em torno do objeto alvo
          subt // Tamanho mínimo da subtrajetória direcionada ao alvo

Saída:    Avt // Conjunto de graus de confiança de avoidance por trajetória

1.  Início
2.  Para t_i ∈ T | intersects (t_i, buffer(O, d))  faça  // intercepta região de interesse
3.      Para o_k ∈ O  faça
4.          Se intersects (t_i, o_k) // Testa interseção com objeto alvo
5.              av_ik ← none // não é avoidance
6.          Senão
7.              Se Subtrajetoria(t_i,o_k) >= Subt
8.                  RegIncr ← Calcula região de incremento de confiança
9.                  Se intersects (t_i, RegIncr)
10.                     av_ik ← strong  // avoidance forte
11.                 Senão
12.                     av_ik ← weak    // avoidance fraco
13.                 Fim Se
14.             Senão
15.                 av_ik ← none // não é avoidance
16.             Fim Se
17.         Fim Se
18.     Fim Para
19.     Calcula Avt_i
20. Fim Para
21. Retorna Avt
22. Fim
```

Figura 3. Pseudocódigo do algoritmo para detecção de *avoidance*.

Se a trajetória intercepta a região de interesse então não tem *avoidance* (linhas 4 e 5). A função Subtrajetória (linha 7) retorna a maior subtrajetória que vai em direção ao alvo dentro da região de interesse e é detalhada no próximo parágrafo. Se a trajetória intercepta a região de incremento de confiança, então a confiança no *avoidance* é alta (linhas 9 e 10); caso contrário, a confiança no *avoidance* é mais baixa (linhas 11 e 12). O cálculo da confiança total do padrão *avoidance* de uma trajetória *i* é realizado na linha 19, utilizando a fórmula

$$Avt_i = \frac{\sum_{k=1}^{n} Av_{ik}}{n}$$

onde, n corresponde ao número de objetos alvo cuja região de interesse foi interceptada pela trajetória i, e Av_{ik} é a medida de *avoidance* da trajetória i em relação ao objeto alvo k e pode ter os valores 1 (*strong*), 0,5 (*weak*) e 0 (*none*).

A função Subtrajetória (linha 7) considera o primeiro ponto da trajetória dentro da região de interesse e vai pegando os próximos pontos, um a um, enquanto a direção da semireta que vai do ponto inicial ao ponto considerado intersecta o objeto alvo. Repete esta operação com os pontos seguintes, de forma a obter a maior subtrajetória em direção ao objeto alvo.

4. Experimentos preliminares

Para verificar a eficácia do algoritmo foram realizados experimentos preliminares com trajetórias obtidas a pé em uma praça e outras obtidas de carro em ruas e avenidas, todas na cidade de Porto Alegre. Estas trajetórias foram capturadas ora sem qualquer restrição, ora com a determinação de evitarem algumas regiões específicas, mapeadas como objetos alvo.

Para os experimentos de trajetórias de carro, foram coletadas 26 trajetórias e foram demarcadas seis regiões como objetos alvo, simulando locais de monitoramento por câmeras de segurança. Para este grupo de trajetórias, o experimento buscou identificar quais veículos apresentaram um padrão de comportamento do tipo *avoidance*, evitando as regiões monitoradas.

Levando-se em consideração que na região observada a largura máxima das avenidas é de aproximadamente 20m, usou-se esse valor como raio do objeto alvo. Já como buffer da região de interesse, utilizou-se 100m. O tamanho mínimo da subtrajetória direcionada ao alvo foi arbitrado em 8m.

O experimento resultou, como esperado, em seis ocorrências de *avoidance* detectadas, sendo cinco com valor *weak* para confiança local e uma com confiança local *strong*.

Na Figura 4 é possível visualizar parte das trajetórias 4, 17 e 22 e como elas se relacionam com o objeto alvo 3. Neste trecho, a trajetória 22 intercepta o objeto alvo não sendo, portanto, identificada como um possível *avoidance*. Já as trajetórias 4 e 17 apresentam o padrão comportamental procurado, pois evitam o objeto alvo. Nesta imagem também é possível visualizar a região de incremento de confiança gerada para a trajetória 4.

Os experimentos preliminares realizados com trajetórias de pedestres também tiveram bom resultado, tendo o algoritmo detectado corretamente os casos de *avoidance*. Este experimento não está detalhado por falta de espaço.

5. Conclusão

Este artigo apresenta um algoritmo para detecção de um novo padrão comportamental denominado *avoidance*, que ocorre quando objetos móveis evitam determinadas regiões espaciais. Para testar sua eficácia, foram realizados experimentos iniciais com dados reais obtidos com aparelhos GPS por veículos e por pessoas a pé. O algoritmo desenvolvido identificou com sucesso as trajetórias que evitaram as regiões estáticas informadas como objeto alvo, tanto nas trajetórias a pé quanto nas de carro, retornando o resultado esperado. Testes em maior escala estão sendo realizados para avaliar o comportamento e o desempenho do algoritmo com um volume de dados maior.

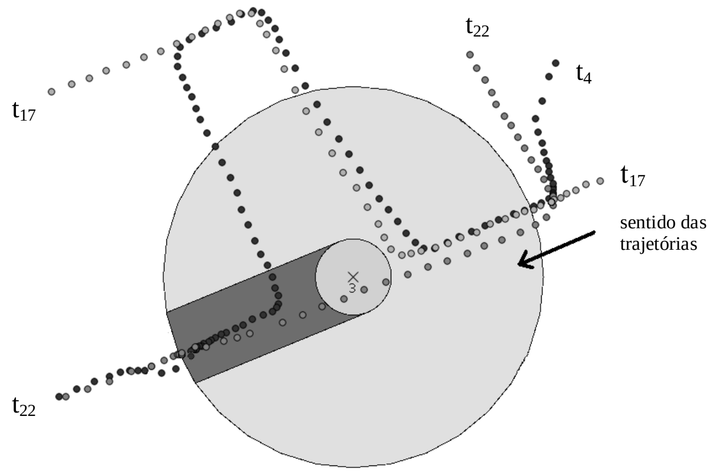


Figura 4. Visualização de parte das trajetórias obtidas de carro em relação ao objeto alvo 3.

6. Agradecimentos

Os autores agradecem ao CNPq (projeto 307588/2008-4), à FAPESC (CP005/2009), ao Projeto Europeu MODAP e ao CNR - Short Term Mobility Program, Itália, pelo financiamento parcial desta pesquisa.

Referências

Alvares, L. O.; Bogorny, V.; Kuijpers, B.; Macedo, J. A. F.; Moelans, B.; Vaisman, A. (2007). A Model for Enriching Trajectories with Semantic Geographical Information.. In: Proc. of the ACM 15th International Symposium on Advances in Geographic Information Systems (ACM-GIS'07), Seattle, Washingthon, 7-9 November, p. 162-169.

Bogorny, V.; Kuijpers, B.; Alvares, L.O. (2009). ST-DMQL: A Semantic Trajectory Data Mining Query Language. In: International Journal of Geographical Information Science. Taylor and Francis, p.1245-1276.

Cao, H., Mamoulis, N., and Cheung, D. W. (2007). Discovery of periodic patterns in spatiotemporal sequences. IEEE Transactions on Knowledge and Data Engineering, 19(4):453–467.

Giannotti, F., Nanni, M., Pinelli, F., and Pedreschi, D. (2007). Trajectory pattern mining. In Berkhin, P., Caruana, R., and Wu, X., editors, KDD, ACM, p. 330–339.

Gudmundsson, J.; Kreveld, M.(2006). Computing longest duration flocks in trajectory data. In GIS'06: Proceedings of the 14th annual ACM international symposium on Advances in Geographic Information Systems, New York, NY, USA, ACM Press, p. 35–42,

Laube, P., Kreveld, M., Imfeld, S. (2005). Finding remo - detecting relative motion patterns in geospatial lifelines. In Fisher, P.F., ed.: Developments in Spatial Data Handling. Proceedings of the 11th International Symposium on Spatial Data Handling. Springer, Berlin Heidelberg, p. 201–214.

Palma, A. T; Bogorny, V.; Kuijpers, B.; Alvares, L.O. (2008). A Clustering-based Approach for Discovering Interesting Places in Trajectories. In: 23rd Annual Symposium on Applied Computing, (ACM-SAC'08), Fortaleza, Ceara, 16-20 March Brazil. pp. 863-868.

Definição de uma gramática dos conceitos relacionados a Movimento e Mudança em objetos espaço-temporais

Leila Weitzel[1], Cirano Iochpe[2]

[1]Faculdade de Computação – Universidade Federal do Pará (UFPA)
Folha 31, Quadra 7 Lote especial - Cx Postal 101 - 68.507-590 - Marabá - PA - Brasil

[2]Instituto de Informática – Universidade Federal do Rio Grande do Sul (UFRGS)
Caixa Postal 15.064 - 91.501-970 - Porto Alegre - RS - Brazil

lmartins@ufpa.br, ciochpe@inf.ufrgs.br

Abstract. *The space-time representation has been one of the greatest demands on research and development in geographic information systems. The goal of this paper is to provide a grammar of movement and change in space-time domain. The grammar allowed an unambiguous definition and offers a better semantic analysis. In fact, the effort to establish the terminology demonstrated the lack of standardization of concepts in this area.*

Resumo. *A representação espaço-temporal tem sido uma das maiores demandas no desenvolvimento e pesquisa nos últimos anos em sistemas de informação geográfica. O objetivo deste artigo é descrever uma gramática que mapeie os conceitos de movimento e mudança no domínio espaço-temporal. A gramática aqui descrita permitiu uma definição não ambígua facilitando a análise semântica, e pode servir de base para descrever outros modelos. O esforço no sentido de estabelecer uma terminologia demonstrou na realidade a falta de padronização no emprego de conceitos relacionados neste domínio.*

1. Introdução

Em Weitzel (2009) foi realizado um levantamento das diferentes formas de representação dos aspectos de mudanças e movimento (aspectos dinâmicos) no contexto de objetos e fenômenos espaço-temporais. Serviram de base os modelos, metamodelos e ontologias geográficas encontrado em: Frank et al (2001) e Cheylan et al (2001). Com este estudo foi possível verificar que existem diferentes níveis de abstração e soluções de modelagem para um mesmo conjunto de conceitos. Acredita-se que uma análise mais profunda destes conceitos e de suas relações poderá revelar similaridades, complementaridades, conflitos e outras dependências entre eles. Dando prosseguimento a esta pesquisa o presente artigo procura definir, de forma não ambígua os conceitos pesquisados em Weitzel (2009). Busca-se assim estabelecer uma compatibilização terminológica através de um conjunto de regras sintáticas e semânticas, aqui definido, como uma gramática normativa.

2. Definição do Objeto Espaço-Temporal – (OET)

O Geoframe-Temporal (Rocha, 2001) fornece uma hierarquia de classes de objetos geográficos que podem ser modelados a partir do mundo real. O OET é uma especialização da classe Fenômeno Geográfico na visão de objeto proposto por (Lisboa Filho & Iochpe, 1999).

Os OET apresentam características (a) Espaciais: são: (i) Localização ou posição que é dada por uma referência em coordenadas geográficas na superfície do globo terrestre; (ii) Geometria refere-se ao formato do objeto (polígono, linha, ponto ou combinações destas geometrias) e (iii) Relações Espaciais entre pares de objetos. (b) A Espaço-Temporal é a característica espacial que pode varia ao longo do tempo. (c) A Descritiva é aquela que apenas qualifica um objeto e independe da localização do objeto no espaço. (d) A Temporal é a característica Descritiva que varia ao longo do tempo. Salienta-se que a alteração no valor das características Espaciais se dá por meio das operações descritas em Hornsby & Egenhofer (1997).

Os OET possuem também um Identificador (Oid) que individualiza cada instância deste objeto, permanecendo invariável durante toda a sua Vida (sua existência). Assim um OET pode ser representado por: (Oid, (Cj)t), onde: Oid é o identificador único do objeto, (Cj) é um conjunto de características j = 1...n , sendo n é o total de características presentes no tempo t.

O parâmetro Cj é expandido em (N(Cj), Cc(Cj), Dom(Cj)), onde: N(Cj) é a função unária que fornece o nome da j-ésima Característica; Cc(Cj) é a função unária que fornece o nome da Categoria da j-ésima Característica e Dom(Cj) é o domínio da j-ésima Característica que identifica o conjunto de valores válidos para uma característica. Por exemplo, (1, (Localização, Espacial, Coordenada UTM))

3. Descrição dos conceitos

Esta seção apresenta um conjunto de definições gramaticais que serão utilizadas para descrever os conceitos levantados em (Weitzel, 2009). Cabe salientar que a descrição suporta a abordagem orientada a objetos de acordo com o proposto em (Worboys, 1994). Considera-se assim, que uma visão do mundo baseada em objetos é adequada para representar fenômenos geográficos tanto na visão de campo quanto na visão de objeto.

3.1. Estado (ST) de um OET

O ST em um tempo qualquer t é dado pelo seu Identificador e por um conjunto específico valores associados às suas características $ST(OET)_t$ = (Oid , (C_j, Val_{jt})), onde OET apresenta um conjunto de valores válidos para as j-ésimas Características no tempo t. Então: $ST(OET)_{t1}$ = (1, (Localização, Espacial, Coordenada UTM, 26°64'50.80"N; 83°16'43.87"W)).

3.2. Evento

Evento é uma ocorrência[1]. Seja o Evento de Queimada então, $ST(OET)_{t1}$ = (1,(Cobertura Vegetal, Nativa)) e em t_2 $ST(OET)_{t2}$ = (1,(Cobertura Vegetal, Desflorestada)).

3.3. Transição de Estado (TE) de um OET

Mostra os diferentes Estados que um objeto passou ao longo do tempo. TE = (($Event_j$)$_t$, $(ST)_t$, (op) , $(ST)_{(t+1)}$), onde: ($Event_j$)$_t$ é um conjunto de Eventos que determinam a Transição de Estado; $(ST)_t$ é o Estado inicial do OET, anterior à mudança; (op) é a operação e $(ST)_{(t+1)}$ é o Estado alcançado após $Evento_j$ no tempo t+1. Exemplo, Mudança na Capital do País TE = (Transferência da Capital, (Localização, Espacial, Coordenada UTM, 26°64'50.80"N;

[1]Ou um fator ou fenômeno da natureza ou artificial produzido por ações antrópicas em uma visão de Campo.

83°16'43.87"W)$_t$, move, (Localização, Espacial, Coordenada UTM, 20°34'5.8"N; 23°6'3.87"W)$_t$).

3.4. Vida OET

Ao longo de sua existência um OET pode sofrer sucessivas mudanças, passando por ***n*** Estados diferentes. O símbolo "≤" vai denotar qualquer relação de ordem parcial. Assim, um conjunto A, junto com seu Ordenamento R é chamado de Conjunto Parcialmente Ordenado denotado por (A,R). Representa-se o conceito de Vida como um ordenamento parcial de Estados: Vida (OET) = (ST , ≤), onde ST é o conjunto de Estados e ≤ é o ordenamento parcial destes estados.

4. Definição das Operações Válidas sobre Objetos

Operações modificam características dos OET. A seguir são descritas os tipos de operações em função da mudança que ela provoca nestes objetos de acordo com Hornsby & Egenhofer (1997). **Grow e Shrink**: são validas para objetos OET. Considera-se que estas operações alteram o valor da característica Espacial *geometria* destes objetos, dando a impressão de *movimento de fronteira* destes objetos. **Merge e Split**: são operações válidas para objetos OET. Considera-se que estas operações alteram a característica Espacial *geometria* (a forma) destes objetos

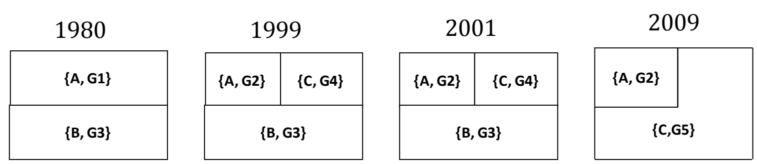


Figura 1: Exemplo de operação de split e merge em um objeto

Sejam os intervalos de tempo seqüenciais (**Ts** = Time Stamping) **Ts** = {1980, 1999, 2001 e 2009}. Estados que cada objeto apresenta ao longo do tempo, supondo que $\mathbf{t_0}$ é o tempo inicial e $\mathbf{t_n}$ o tempo final que delimita dois estados. Decidiu-se apenas ilustrar as características G_i (geometria) e o tipo de cobertura do solo. Ressalta-se que as operações de **merge** e **split** seguem o padrão proposto por Hornsby & Egenhofer (1997) e por isso omitidas aqui. Em seu trabalho, Hornsby & Egenhofer (1997) descrevem estas operações usando, o que os autores chamam de CHANGE DESCRIPTION LANGUAGE – CDL, um framework baseado em uma representação icônica para expressar a semântica da mudança.

Tabela 1: Tabela dos Estados dos Objetos A, B e C

O_{id}	Geometria	Cob. solo	t_0	t_n
A	G1	Urban	1980	1999
A	G2	Urban	1999	2001
A	G2	Corn crop	2001	2009
A	G2	Corn crop	2009	-
B	G3	Urban	1980	1999
B	G3	Urban	1999	2001
B	G3	Urban	2001	kill
C	G4	Corn crop	1999	2001
C	G5	Corn crop	2001	2009

Neste caso o objeto A sofreu uma operação de split em **Ts** = 1980 dando origem ao objeto C, já em **Ts** = 2009 o objeto B e C sofreram uma operação de merge dando origem um novo objeto C, e como conseqüência o objeto B sofreu uma operação de Kill. Em Hornsby & Egenhofer (1997) tem um estudo detalhado destas operações. **Kill, Create e Reincarnate**: são operações válidas objetos OET. Na Figura1, em **Ts** = 1999, o objeto A sofreu split originando dois objetos, o objeto A com geometria G2 e o objeto C com geometria G4. O objeto C nessa figura foi criado por meio de uma operação create. No **Ts** = 2009 os objetos C e B sofreram uma operação de merge, e assim o objeto B ainda nessa figura desapareceu por meio da operação de kill. **Move**: Esta operação resulta em mudanças que podem ser percebidas pela trajetória que o objeto percorre, relacionado ao conceito de movimento. Representam as mudanças na posição (localização espacial) de um objeto.

5. Descrição dos Modelos Estudados

5.1. Modelo de Frank (2001)

Em Frank (2001) objetos sócio-econômicos OSE são unidades utilizadas para descrever fenômenos econômicos (zona urbana, rural, plantação etc.) ou sociais (demografia, etnia etc). Estes objetos são espaços subdivididos pelo homem para fins sociais ou econômicos. O OSE tem o mesmo comportamento de um fenômeno geográfico com visão de campo. Conceito *Mudança* neste modelo ocorre quando se altera o valor de uma característica Descritiva, ou seja, $\{\exists c_j \in (C_j) \mid Val(C_j)_t <> Val(C_j)_{(t+1)} \wedge Cc(C_j)=Descritiva\}$. Então o conceito *Mudança* pode ser descrito como uma Transição de Estado, TE = $((Event)_t, (OET)_t, (op), (OET)_{(t+1)})$**.**

O conceito de *Movimento* é dado quando o valor de uma ou mais Característica Espacial (posição ou geometria) varia. Então: MOV_{Frank} ocorre quando $\exists\ c_j \in (C_j) \mid Val(C_j)_t <> Val(C_j)_{(t+1)} \wedge Cc(C_j)=Espacial$. Por analogia, o conceito de *Movimento* é dado por: TE = $((Event_j)_t, (OET)_t, (op)_t, (OET)_{(t+1)})$ onde $\exists\ c_j \in (C_j) \mid Val(C_j)_t <> Val(C_j)_{t+1} \wedge Cc(C_j) = Espacial$. Frank (2001) admite o *Movimento de Fronteira* $Mov_{boundary}$ que ocorre quando existe a alteração do valor da característica Espacial *geometria* de um objeto. Por exemplo, o desmatamento de uma parcela de floresta, pode-se então pensar em um *Movimento de Fronteira*. O conceito *Movimento de Fronteira* é um tipo de *Mudança* onde apenas a característica Espacial *geometria* se altera, por meio de operações *shrink* ou *grow*. Outros tipos de *Movimento de Fronteira* podem ser provenientes de operações *kill/reincarnate*. Exemplificando, os rios da Região Amazônica, em determinada estação do ano têm elevação temporária do seu nível pelo aumento da vazão proveniente as chuvas intensas. No inverno, o processo se dá de forma inversa, os rios se esvaziam e desaparecem pela escassez de chuva. Assim, na visão de campo, um objeto apresenta *Movimento de Fronteira* $Mov_{boundary}$ quando: $\exists\ c_j \in (C_j) \mid Val(C_j)_t <> Val(C_j)_{(t+1)} \wedge Cc(C_j) = Espacial \wedge N(C_j) = geometria$. Por analogia, o conceito de *Movimento de Fronteira* é dado por: TE = $((Event_j)_t, (OET)_t, (op), (OET)_{(t+1)})$ onde $\exists\ c_j \in (C_j) \mid Val(C_j)_t <> Val(C_j)_{(t+1)} \wedge Cc(C_j) = Espacial \wedge N(C_j) = geometria$. Esse tipo de movimento cria um ou mais eventos que irão forçar movimentos de fronteiras em objetos da vizinhança. Na visão de objeto, um objeto apresenta movimento de fronteira $Mov_{boundary}$ quando: $\exists\ c_j \in (C_j) \mid Val(C_j)_t <> Val(C_j)_{(t+1)} \wedge Cc(C_j) = Espacial \wedge N(C_j) = geometria$. Por analogia, o conceito de *Movimento de Fronteira* é dado por: TE = $((Event_j)_t, (OET)_t, (op)_t, (OET)_{(t+1)})$ onde $\exists\ c_j \in (C_j) \mid Val(C_j)_t <> Val(C_j)_{(t+1)} \wedge Cc(C_j) = Espacial \wedge N(C_j) = geometria$.

5.2 Modelo de Cheylan (2001)

O conceito de *Evento* pode ser mapeado para a gramática por uma tupla que contém o identificador deste objeto, um conjunto de características e de operadores que alteram o *Estado* de objetos da realidade. Para o autor o *Estado* de um objeto é uma ocorrência temporal que se intervala a dois *Eventos*. Para Cheylan (2001) assim como para Frank (2001) *Movimento* é percebido quando uma ou mais características espaciais se alteram ao longo do tempo. A noção de *Movimento* é dada pela alteração da sua posição ou da sua geometria. Assim uma mudança no sentido de movimento ocorre quando: $\exists c_j \in (C_j) \mid Val(C_j)_t <> Val(C_j)_{(t+1)} \wedge Cc(C_j)=Espacial$. Por analogia, o conceito de Movimento de Cheylan (2001) -

$\mathbf{MOV_{Cheylan}}$ é dado por: TE=((Event$_j$)$_t$, (OET)$_t$, (op)$_t$, (OTE)$_{(t+1)}$) onde $\exists\ c_j \in (C_j) \mid Val(C_j)_t <> Val(C_j)_{(t+1)} \wedge Cc(C_j)$=Espacial. Ainda para o autor um objeto pode sofrer mudanças ao longo de sua *Vida* (Life) alterando o valor de uma ou mais características. Assim por analogia, o conceito *Vida* pode ser representado por um ordenamento parcial de *Estados*: Life (OTE$_i$) = (ST , $\leq$); Onde: ST é o conjunto de *Estados* do objeto i ; $\leq$ é o ordenamento parcial destes *Estados*.

Tabela 3: quadro Resumo dos modelos e gramática

Modelo/autor	Nome do conceito	Relação de equivalência na gramática	Gramática do termo	Restrições	Semântica associada ao termo
Frank (2001)	Mudança	Transição de Estado (TE)	((Event$_j$)$_t$, (OTE)$_t$, (op)$_{kt}$, (OTE)$_{(t+1)}$)	$\exists\ c_j \in (C_j) \mid Val(C_j)_t <> Val(C_j)_{(t+1)} \wedge Cc(C_j)$ = Descritiva	Mudanças ocorrem quando uma ou mais característica descritiva se altera ao longo do tempo
Frank (2001)	Movimento	Transição de Estado (TE)	((Event$_j$)$_t$, (OTE)$_t$, (op)$_{kt}$, (OTE)$_{(t+1)}$)	$\exists\ c_j \in (C_j) \mid Val(C_j)_t <> Val(C_j)_{t+1} \wedge Cc(C_j)$ = Espacial	Movimento ocorre quando uma ou mais característica espacial se altera ao longo do tempo
Frank (2001)	Movimento de Fronteira	Transição de Estado (TE)	((Event$_j$)$_t$, (OTE)$_t$, (op)$_{kt}$, (OTE)$_{(t+1)k}$)	$\exists\ c_j \in (C_j) \mid Val(C_j)_t <> Val(C_j)_{(t+1)} \wedge Cc(C_j)$ = Espacial $\wedge$ N(C$_j$) = geometria	É alteração do valor da característica Espacial *geometria*
Cheylan (2001)	Movimento	Transição de Estado (TE)	((Event$_j$)$_t$, (OTE)$_t$, (op)$_t$, (OTE)$_{(t+1)}$)	$\exists\ c_j \in (C_j) \mid Val(C_j)_t <> Val(C_j)_{t+1} \wedge Cc(C_j)$ = Espacial	Movimento ocorre quando uma ou mais característica espacial se altera ao longo do tempo
Cheylan (2001)	Vida	Life (OTE)	(ST , $\leq$)	-	Representa um ordenamento parcial de Estados

4 Discussão

O mundo real é um ambiente altamente complexo, não estático e de difícil reconhecimento e verificação. Devido a esta complexidade, o processo de captura da realidade para efeito de modelagem envolve abstrações, generalizações e aproximações. A representação espaço-temporal tem sido uma das maiores demandas no desenvolvimento e pesquisa nos últimos anos em sistemas de informação geográfica, pois, o tempo é uma característica intrínseca e, presente em, ou associada aos fenômenos do mundo real. Fenômenos (eventos) acontecem em um dado momento alterando características dos objetos da realidade. Estas mudanças alteram o Estado de um objeto ao longo de sua Vida. A gramática aqui descrita permitiu uma definição não ambígua facilitando a análise semântica. Esta gramática pode servir de base para descrever outros modelos encontrados na literatura. O esforço no sentido de estabelecer uma terminologia demonstrou na realidade à falta de padronização no emprego de conceitos relacionados a movimento e mudança neste domínio.

References

Cheylan, J.. Time, actuality, novelty and history. (2001). In: FRANK, A. U.; RAPER, J. ; Cheylan, J. (eds.) Life and motion of socio-economic units, London: Taylor & Francis.

Frank, A. U.; Raper, J.; Cheylan, J. (2001). Life and Motion of Socio-Economic Units. ESF Series, London, Taylor & Francis, p. 353.

Hornsby, K. and M. Egenhofer. (1997), "Qualitative Representation of Change," In Spatial Information Theory - A Theoretical Basis for GIS, International Conference COSIT '97, Laurel Highlands, PA, Vol. 1329, Lecture Notes in Computer Science, S. Hirtle and A. Frank, Eds. Berlin: Springer-Verlag, 1997, pp. 15-33.

Lisboa F. J.; Costa, A.C.; Iochpe, C. (1999), "Projeto de Banco de Dados Geográficos: mapeando esquemas GeoFrame para o SIG Spring," in I BRAZILIAN WORKSHOP ON GEOINFORMATICS - GEO-INFO'99, Anais... São José Dos Campos:INPE, Unicamp, Campinas., pp. 2-5.

Rocha, L. V. (2001). GeoFrame-T: um framework conceitual temporal para aplicações de Sistemas de Informação Geográfica. Porto Alegre: PGCC da UFRGS. Dissertação de Mestrado.

Weitzel, L. (2009). Estudo e documentação de características espaço-temporais em modelos de dados e em ontologias, Trabalho Individual I, Programa de Pós-Graduação em Computação- PPGC, UFRGS.

Worboys, M. F. (1994). Object-Oriented Approaches to Geo-Referenced Information. International Journal of Geographical Information Systems 8(4): 385-399.

Dicionário de tipos de feições geoespaciais para o território brasileiro

Ivanildo Barbosa[1,2], Marco Antonio Casanova[2]

[1]Seção de Ensino de Engenharia Cartográfica – Instituto Militar de Engenharia (IME)
Praça General Tibúrcio, 80 – CEP 22290-270 – Rio de Janeiro – Brasil

[2]Departamento de Informática – Pontifícia Universidade Católica do Rio de Janeiro (PUC – Rio)
Rua Marquês de São Vicente, 225 - CEP 22451-900 – Rio de Janeiro – Brasil

{ibarbosa,casanova}@inf.puc-rio.br

Resumo. *Este trabalho apresenta uma análise das especificações vigentes no Brasil para a elaboração de produtos geoespaciais, visando à verificação da completude das mesmas no que diz respeito à diversidade de feições geoespaciais representáveis. A relação de termos de cada uma das especificações é comparada com o dicionário de feições geoespaciais da Alexandria Digital Library (ADL), verificando combinação exata entre termos, sinônimos, ou omissões em qualquer das partes. Foram detectadas diferenças relevantes nas abordagens para classificação das feições, bem como a omissão de termos classificando elementos que não ocorrem em território brasileiro.*

Abstract. *This paper presents an analysis of the current Brazilian specifications for creating geospatial products, aiming at checking their completeness concerning the diversity of representable geospatial features. The list of terms of each specification is compared with the Alexandria Digital Library (ADL) Feature Type Thesaurus, checking for the exact match for terms, for synonyms and for missing terms. The analysis detected important differences on the approaches for feature classification, as well the omission of certain terms in the specifications, mostly corresponding to types of geographic features not occurring in Brazil.*

1. Introdução

Existe uma crescente demanda por dados espaciais e uma consequente demanda pela sua disponibilização. Um dos objetivos da Infraestrutura Nacional de Dados Espaciais (INDE) é o compartilhamento dos dados geoespaciais de origem federal, estadual, distrital e municipal [Brasil 2008].

Os dados geoespaciais abrangem bases em mídia analógica, como as cartas topográficas impressas, ou dados digitais em formato matricial ou vetorial. No caso das cartas do mapeamento sistemático, produzidas pelo Instituto Brasileiro de Geografia e Estatística (IBGE) e pela Diretoria de Serviço Geográfico (DSG) do Exército Brasileiro, as feições são classificadas em categorias de informação conforme especificado no manual T34-700 [DSG 1998], destinado a "*normatizar a representação dos acidentes* *naturais e artificiais em cartas topográficas e similares nas escalas de 1:25.000, 1:50.000, 1:100.000 e 1:250.000*". A atribuição de simbologia às feições depende da pertinência dessas feições a uma das categorias de informação pré-definidas, respeitadas as condições de representatividade nas escalas mencionadas, muito embora nem todos os dados geoespaciais de interesse geral não estejam restritos a essas escalas, nem ao conjunto de dados representáveis em mapas topográficos.

Com o processo de informatização da cartografia, a DSG e o IBGE realizaram esforços para especificar a estruturação dos dados espaciais vetoriais, surgindo as versões da Mapoteca Topográfica Digital [IBGE 1996] e da Tabela da Base Cartográfica Digital [DSG 1997]. Contudo, esses padrões eram incompatíveis, demandando mecanismos de conversão entre arquivos produzidos por aqueles órgãos.

Atualmente, está em vigor a segunda versão das Especificações Técnicas para Estruturação de Dados Geoespaciais Digitais Vetoriais (ET-EDGV) [CONCAR 2007]. Nelas estão definidos os esquemas aplicáveis às diferentes categorias de informação disponíveis em formato vetorial. Em cada categoria são descritas as classes e respectivos atributos, seus tipos e domínios. A recuperação das informações contidas nos produtos gerados com base nessas especificações está condicionada à perfeita associação do significado das categorias às representações das feições, permitindo a usuários de diversas especialidades conhecer de maneira inequívoca o conteúdo disponibilizado.

Os termos empregados na caracterização das feições geoespaciais representáveis conforme as especificações apresentadas formam um vocabulário adotado pelas instituições citadas em Brasil [2008], que pode ser usado como base para a construção de um dicionário (*thesaurus*) – *vocabulário de uma linguagem de indexação controlada, estruturada de modo que os relacionamentos entre conceitos são explicitados* [ISO, 1986 *apud* Breitman *et al*, 2007]. Uma vez delimitado o domínio de atuação, são estabelecidos os critérios de categorização dos termos que, neste trabalho, serão baseados naqueles adotados em cada especificação, aplicando alterações pontuais visando à busca por diferenças de conteúdo entre o ADL FTT e o dicionário de feições geoespaciais baseado nas especificações brasileiras.

Este trabalho tem por objetivo avaliar a abrangência do vocabulário criado a partir das especificações nacionais para produção de dados geoespaciais (verificando ausência ou excesso de termos) comparando-o com o dicionário de tipos de feições (*Feature Type Thesaurus – FTT*) da Biblioteca Digital Alexandria (*Alexandria Digital Library – ADL*), desenvolvida pela Universidade da Califórnia. Essa comparação será realizada pelo alinhamento sintático e semântico dos vocabulários, respeitando as relações entre os termos (sinônimos, generalização e especialização, *etc*) e a manutenção do significado durante a tradução dos termos.

A correlação entre termos em diferentes idiomas facilitará a associação das feições a ontologias adequadas, agregando semântica aos dados, assim como fornecerá subsídios para o desenvolvimento de mediadores e mecanismos de busca na web para descoberta de produtos e serviços.

2. Análise dos termos relativos ao domínio

Cada uma das especificações nacionais apresenta um critério primário de classificação das feições em categorias, de acordo com as suas particularidades. Neste tópico, serão comparadas as definições dadas para as categorias especificadas no manual T34-700, na MTD e na ET-EDGV. A seguir, encontram-se listadas as categorias empregadas em cada especificação.

a) *T34-700* (9 categorias): Hidrografia, Vegetação, Limites, Pontos de Referência, Localidades, Sistema de Transporte, Altimetria, Edificações e Infraestrutura;

b) *MTD* (8 categorias): Hidrografia, Vegetação, Limite, Ponto de Referência, Localidade, Sistema Viário, Hipsografia, Obra e Edificação;

c) *ET-EDGV* (13 categorias): Hidrografia, Vegetação, Limites, Pontos de Referência, Localidade, Sistema de Transportes, Relevo, Energia e Comunicações, Abastecimento de Água e Saneamento Básico, Educação e Cultura, Estrutura Econômica, Administração Pública, Saúde e Serviço Social.

Uma análise imediata permite correlacionar, pela semelhança sintática, nomes de categorias comuns às 3 especificações: *Hidrografia*, *Localidade(s)*, *Limite(s)*, *Ponto(s) de Referência* e *Vegetação*. O termo *Sistema de Transportes* aparece explícito em duas especificações, e na terceira ocorre apenas o termo *Sistema* acompanhado do qualificador *Viário*. Contudo, a correlação entre *Sistema Viário* e *Sistema de Transportes* precisa ser avaliada com base nas descrições, da mesma forma que os demais itens sem semelhança sintática. O termo *Edificação* aparece no plural ou associado ao termo *Obra*, o que sugere a correlação ente duas categorias de especificações diferentes.

Analisando outras descrições, é possível estabelecer alguns vínculos entre alguns dos termos (grifando palavras relevantes em comum):

- *Altimetria*: elementos ***hipsográficos***, que representam o ***relevo*** da superfície terrestre, relativamente ao *datum* vertical de referência. Este relevo é representado por meio de curvas de nível e pontos de altitude;
- *Hipsografia*: não foi encontrada descrição dessa categoria, mas uma menção ao uso de cores hipsométricas como forma de representação do relevo; e
- *Relevo*: Categoria que representa a forma da superfície da Terra e do fundo das águas tratando, também, os materiais expostos, com exceção da cobertura vegetal.

A correlação do termo *Edificação* leva a associar as categorias *Edificações* (TBCD) e *Obra e Edificação* (MTD). Dentre as categorias ainda não analisadas nas ET-EDGV, esse termo aparece na descrição de *Educação e Cultura*, *Estrutura Econômica*, *Administração Pública* e *Saúde e Serviço Social*. Pode-se interpretar que houve especialização da classe *Edificações* (TBCD), condicionando a pertinência da feição à categoria à sua finalidade.

Prosseguindo com a análise das categorias especificadas no T34-700, observa-se que o termo *Infraestrutura* ocorre parcialmente (na forma de estrutura) nas categorias *Energia e Comunicações* e *Abastecimento de Água e Saneamento Básico*, ambas das ET-EDGV. A correlação semântica acarretada pelo termo *Estrutura* ainda é fraca, uma vez que o mesmo aparece na descrição de *Sistema de Transportes* (ET-EDGV). Contudo a maior afinidade da categoria com o termo *Transporte* e a descrição de *Infraestrutura* (T34-700) empregando um conjunto de feições relacionadas a atividades de infraestrutura torna a primeira associação mais apropriada. Algumas camadas de informação contidas na MTD correlacionam elementos referentes a *Energia e Comunicações* e *Abastecimento de Água e Saneamento Básico* como *Obra e Edificação*, pela própria definição de Infraestrutura (T34-700) como "edificações de".

Outros termos são introduzidos na criação de classes especializadas ou como elementos de listas controladas, domínio de um atributo de uma classe, devido às suas semelhanças segundo o critério de classificação adotado.

3. Hierarquização de termos

Para verificar a abrangência da classificação empregada nas especificações nacionais para elaboração de documentos cartográficos, foi criada uma estrutura de dicionário conforme o padrão internacional contendo os termos referentes a feições e classes geoespaciais. Com isso, é possível identificar o contexto geoespacial de cada termo.

Foram desconsiderados na composição do vocabulário (consequentemente, do dicionário) termos referentes a classes não instanciáveis das ET-EDGV e outras que, reunidas, compõem uma feição, não representando, isoladamente, elementos geoespaciais. Neste caso, foram considerados os termos relativos às feições integrais (exemplo, considerar *rio* e descartar *trecho de massa d'água*).

Outro fator a ser considerado nas especificações é que o nome das classes são escritos visando à implementação (não havendo espaços, acentuação ou caracteres especiais, ou contendo abreviaturas). Para fins de composição do dicionário, foram selecionados os termos mais próximos já existentes nas outras especificações ou aqueles empregados na descrição da classe.

Com base nas definições contidas em ISO (1986) *apud* Breitman *et al* (2007), as feições geoespaciais podem ser classificadas como *Top Terms* (TT), *Broader Terms* (BT) e *Narrowed Terms* (NT). Essa classificação se baseia nos relacionamentos entre as feições representadas pelos termos, retratando principalmente *pertinência* e *similaridade*.

Estabelecer critérios para descrever o relacionamento entre termos é uma tarefa delicada, uma vez que estão sendo analisadas feições reais no terreno. Porém, alguns fatos podem auxiliar a escolha de termos relacionados (como os relacionamentos existentes entre as classes, no caso específico de termos das ET-EDGV).

A escolha de *top terms* exige um grau de abstração que deve levar em consideração os objetivos propostos na elaboração das especificações. Com o objetivo de facilitar a comparação entre os dicionários, os termos contidos nas especificações nacionais será associado à estrutura do FTT da ADL. Desta forma, a primeira categorização separa os elementos geoespaciais em Elementos Naturais e Elementos Artificiais (*man-made*). Seguindo esse critério, elementos de ***Relevo***, ***Hidrografia*** e ***Vegetação*** integrariam o primeiro grupo, enquanto o segundo agrega as categorias *Sistema de Transporte*, *Edificações*, *Infraestrutura*, *Limites* e *Pontos de Referência*, que são ***Elementos Construídos.*** Nesta mesma categoria se encontram os elementos de

Relevo, Hidrografia (*Estruturas Hidrográficas*) e Vegetação produzidos por ação humana. *Localidades*, que são definidas e agrupadas segundo critérios político-administrativos, e outros tipos de particionamento do espaço com finalidades administrativas (setores censitários, por exemplo) são abrangidas pelo termo ***Áreas Administrativas***, existente na ADL.

Todo TT já é um BT de um conjunto de termos e não possui um BT para si, por estarem no topo da classificação das feições. Dentro de cada TT, pode-se observar uma primeira divisão, com base nas categorias originais. Isso significa que ***Sistema de Transporte***, ***Limites***, ***Edificações***, ***Infraestrutura***, ***Pontos de Referência, Localidades*** e ***Estruturas Hidrográficas*** formam o primeiro conjunto de NT.

Dentro de cada uma das categorias das ET-EDGV (e em alguns casos no T34-700), podem ocorrer especializações que indicam correlação entre os termos envolvidos (outras classes). Os termos especializados tornam-se NT dos termos generalizados, que passam a ser os BT daqueles. Por outro lado, há classes que abrangem diversos tipos de feições sem que haja especialização (diferentes instâncias de uma classe).

Foram considerados sinônimos os termos *Hipsografia* (com *Altimetria*), *Sistema de Transportes* (com *Sistema Viário*) e aqueles qualificados como *temporário* (com o correspondente *intermitente*). Em geral, os termos possuem *descrição excludente*, mesmo em casos que a diferença entre os termos seja apenas na extensão. Nesses casos, o termo pode ser considerado quase sinônimo (UF).

4. Conclusões

Reunidos todos os termos relacionados a feições geoespaciais elencados nas especificações analisadas, recorreu-se à tradução dos mesmos de modo que esse conjunto pudesse ser comparado aos vocábulos e expressões constantes no ADL FTT. Foram empregados os recursos de tradução do *Google*, do Glossário de Gestão Costeira Integrada e, no caso de ambiguidade ou dúvida sobre a corretude da tradução, o termo foi consultado no glossário *WordNet*. A tabela 1 apresenta algumas das comparações realizadas.

Tabela 1 – comparação de termos descritos na ET-EDGV e no ADL FTT

Termo EDGV	**Tradução**	**Termo ADL FTT**	**Significado combina?**
rio	*river*	*rivers*	Sim
cachoeira	*waterfall*	*Waterfalls (Used for: cascades, cataracts, falls)*	Em parte
fenda	*gap*	*gap*	Não

Foi observado que a diferença no critério de categorização dos termos dificultou o trabalho, principalmente separando elementos naturais de elementos artificiais (construídos, antrópicos ou *man-made*). Portanto, a classificação em TT, BT e NT pode ser feita de maneira mais eficiente, sob o ponto de vista funcional das especificações analisadas, adotando o critério de categorização das ET-EDGV por ser mais uniforme e objetivo.

Por se tratar de especificações para representação de feições presentes no Brasil, não são encontrados elementos relacionados aos ambientes polares (tundras, fiordes, geleiras, etc.) de regiões específicas do globo (*savanas*, *desertos*, *vulcões*, *áreas de atividade tectônica*, etc). Também foi observado que alguns termos usados no Brasil, quando traduzidos, tornam-se um único termo, assim como foi detectada a ausência de termos como *córrego*, *cânion* e *arquipélago*. Em contrapartida, algumas categorias apresentam termos de aplicação muito específica, algumas delas não presentes no FTT da ADL, encontradas apenas em glossários de domínios específicos. Isso reflete uma disparidade na abordagem das categorias, com grande profundidade em alguns casos e superficialmente em outros.

Como sugestões de trabalhos futuros, pode-se destacar a busca por ontologias de domínio que possam abranger as feições relacionadas nas especificações com suas propriedades e predicados textuais e espaciais, assim como a implementação de um dicionário de termos geoespaciais retratando seus relacionamentos e associando-os a URI (*Uniform Resource Identificator*), cuja codificação também é objeto de estudo.

Referências

BRASIL, Decreto n. 6.666, de 27 de novembro de 2008. "Institui, no âmbito do Poder Executivo federal, a Infra-Estrutura Nacional de Dados Espaciais - INDE, e dá outras providências".

BREITMAN, Karin K., CASANOVA, Marco A., Truszkowski, Walter, "Semantic Web: Concepts, Technologies and Applications", London: Springer-Verlag London Limited. 2007. 327 p.

COMISSÃO NACIONAL DE CARTOGRAFIA, "Especificações Técnicas para Estruturação de Dados Geoespaciais Digitais Vetoriais". http://www.concar.ibge.gov.br/arquivo/94@ESPECIFICACOES%20TECNICAS%20PARA%20ESTRUTURACAO%20DE%20DADOS%20GEOESPACIAIS%20VETORIAIS_V2.0_10_10_2007.pdf. Brasília, 2007.

DIRETORIA DE SERVIÇO GEOGRÁFICO, "Manual Técnico T34-700: Convenções Cartográficas", 2ª Ed., Brasília: 1998.

DIRETORIA DE SERVIÇO GEOGRÁFICO, "Tabela da Base Cartográfica Digital", Brasília: 1997.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA, "Mapoteca Topográfica Digital: Documento Geral – Versão 3.0." IBGE, Diretoria de Geociências, Departamento de Cartografia. Rio de Janeiro: 1996.

University of California – Santa Barbara, "Alexandria Digital Library Feature Type Thesaurus", http://www.alexandria.ucsb.edu/~lhill/FeatureTypes/ver070302/index.htm. Abril.

Revista de Gestão Costeira Integrada, "Glossário das Zonas Costeiras", http://www.aprh.pt/rgci/glossario/index.html. Maio 2010.

Modelos Baseados em Agentes em Mudanças de Uso e Cobertura da Terra: o Caso da Moratória da Soja em Santarém

Talita Oliveira Assis, Ana Paula Dutra de Aguiar, Gilberto Câmara, Pedro Ribeiro de Andrade

Instituto Nacional de Pesquisas Espaciais (INPE)
Av. dos Astronautas, 1758 - 12227-001 - São José dos Campos, SP, Brasil

talitaoliveiraassis@gmail.com, anapaula@dpi.inpe.br, {gilberto.camara, pedro.andrade}@inpe.br

Abstract. *Agent-based models allow treats individuals, organizations and institutions representing the interaction of actors among themselves, and among the actors with the environment. This paper conceptualizes such relations and explores the example of moratorium soybean pact that has marked out the soybean in the Amazon region applying the discussed concepts.*

Resumo. *Os modelos baseados em agentes possibilitam tratar indivíduos, organizações e instituições representando a interação dos atores entre si, e entre os atores com o ambiente. Este trabalho conceitua tais relações e explora o exemplo da moratória da soja, pacto que tem balizado o mercado da soja na região amazônica, aplicando os conceitos discutidos.*

Palavras-chave: *Modelagem dinâmica espacial, Relações, Células, Agentes.*

Introdução

Mudanças de uso e cobertura da terra são transformações na superfície terrestre resultantes um sistema complexo de interações humanas e ambientais [Turner et al. 1995] que podem causar impactos em escalas locais e globais. Estas mudanças referem-se tanto a conversões entre classes (por exemplo, floresta para agricultura) quanto as alterações dentro de uma mesma classe, como a intensificação da agricultura e a degradação de uma floresta [Lambin et al. 2006]. Em contextos socioeconômicos, biofísicos e políticos específicos, múltiplos atores influenciam as diferentes trajetórias de mudança do uso da terra. Para capturar e melhor compreender estas mudanças é necessário estudar as interações sociais subjacentes nas diferentes escalas.

Na área de modelagem e cenários de mudanças de uso da terra, a construção de modelos que sejam capazes de representar a interação entre os diversos atores envolvidos e suas relações com trajetórias de mudanças de uso da terra são de grande interesse. Os modelos computacionais auxiliam na capacidade mental de modelagem, de forma a permitir tomadas de decisão mais informadas [Costanza e Ruth 1998]. *Modelos baseados em agentes* são uma alternativa interessante para representar a heterogeneidade dos atores [Gilbert 2007], possibilitando tratar tanto indivíduos quanto organizações. Estes modelos devem representar a interação dos atores entre si, e entre os atores com o ambiente. Exemplos de modelos que utilizam este tipo de abordagem são o modelo de dinâmicas urbanas nas cidades da América Latina de Barros (2004) e o modelo de segregação de Schelling (1971).

Baseado em Andrade et al. (2008), o presente trabalho conceitua as quatro relações envolvendo agentes e células em um ambiente de modelagem de uso e cobertura da terra. Estes conceitos são utilizados na construção de um modelo hipotético para explorar as relações na dinâmica de uso da terra na Amazônia brasileira.

Modelagem baseada em agentes

Um modelo baseado em agentes busca compreender um determinado processo social através da descrição das entidades mínimas que compõem este sistema, chamadas *agentes*. Segundo Wooldridge e Jennings (1995), agentes possuem quatro características que os definem: a) autonomia: não há um controle global do que o agente faz; b) habilidade social: é possível interagir com outros agentes; c) reatividade: é possível reagir apropriadamente a estímulos vindos do seu ambiente e d) proatividade: existem objetivos que norteiam as decisões, mas os agentes tomam suas próprias iniciativas. A partir da percepção do ambiente os agentes escolhem e executam uma ação podendo, desta forma, modificar o ambiente e influenciar outros agentes [Vidal et al. 2001].

Uma das grandes vantagens dessa abordagem é a possibilidade de se modelar atores com características e habilidades heterogêneas, tornando possível trabalhar diretamente com as conseqüências das suas interações [Gilbert, 2007]. Adicionalmente, modelos baseados em agentes estão centrados no conceito de “emergência”, onde as ações e interações entre cada agente resultam em um padrão macro que não foi descrito nas regras de comportamento dos agentes, propiciando o entendimento dos processos e suas conseqüências [Matthews et al. 2007, Gilbert 2007].

No contexto de mudanças de uso da terra os modelos baseados em agentes consistem de entidades autônomas (atores), um ambiente onde os agentes interagem (normalmente um espaço celular) e regras que definem estas interações. A principal função do ambiente é prover um contexto espacial para os agentes [Huigen and Fischer, 2003] que podem representar atores como produtores e instituições (políticas ou particulares). Desta forma, existem dois tipos de entidades espaciais básicas neste tipo de modelo: *células*, que representam o espaço geográfico, e *agentes*, que representam entidades autônomas capazes de tomar decisões.

Conceituação das relações entre as entidades espaciais

Uma relação é um mapeamento que descreve as conexões entre duas entidades. Segundo Andrade et al. (2008), dado que existem dois tipos de entidades espaciais nos modelos baseados em agentes, quatro tipos de relações são formadas através da combinação de duas classes de entidades. Elas são:

- Célula→célula: relações de vizinhança entre células, representando a proximidade espacial entre elas. Podem ser baseadas na distância euclideana ou em métricas que geram distâncias relativas, como por exemplo as redes de conectividade.
- Agente→célula: trata da relação entre um agente e o espaço celular, descrevendo sobre quais células um determinado agente tem domínio.

- Célula→agente: a relação entre o espaço celular e o agente define os agentes que pertencem ou que têm influência sobre a célula. Vários agentes podem pertencer a uma mesma célula, assim como um conjunto de células pode ser influenciado pelo mesmo agente, dependendo de fatores como a resolução celular.
- Agente→agente: é a relação entre agentes do mesmo tipo ou de tipos diferentes. Pode ser baseada em fatores externos, como relações de mercado, ou em relações de adjacência das células pertencentes aos respectivos agentes, como no caso de produtores que possuem propriedades vizinhas.

Neste trabalho, estas relações são usadas como base para a construção de um modelo computacional para estudar a dinâmica de uso da terra em Santarém, na Amazônia brasileira. Este estudo de caso é apresentado a seguir.

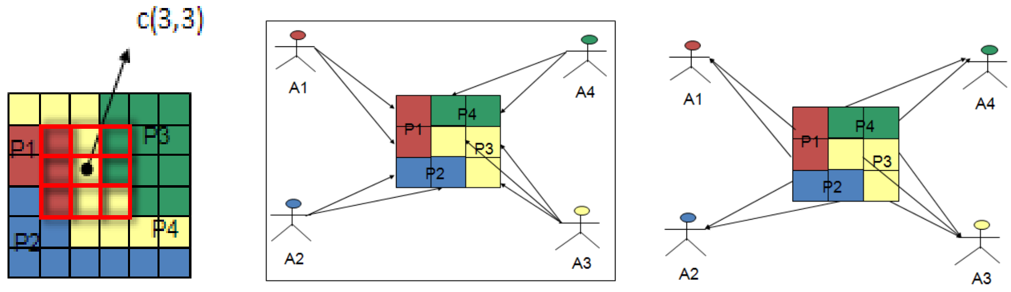
Estudo de caso

O município de Santarém, no oeste do Pará, teve sua dinâmica de uso da terra nos últimos anos baseada principalmente na substituição de áreas de agropecuária familiar e pastagem em áreas de cultivo de grãos. Isto estimulou o desmatamento e contribuiu para um re-ordenamento territorial com a substituição de um grande número de pequenas propriedades por um pequeno número de grandes propriedades.

O avanço da agricultura mecanizada na região, bem como seus impactos ambientais e sociais, ocasionaram a reação de diversos setores da sociedade civil, que resultou no pacto conhecido como "Moratória da Soja". Este firma o compromisso de não comercializar soja plantada após outubro de 2006, proveniente de áreas do Bioma Amazônico que foram desflorestadas a partir da data da assinatura do compromisso (GTS, 2008). Este arranjo tem balizado o mercado da soja na região surgindo, então, a necessidade de representação desta relação o que ajudará a entender de que forma ela influencia no comportamento dos atores, diferenciando os produtores e compradores de grãos que aderem a este acordo daqueles que não fazem esta adesão, e suas implicações nas dinâmicas do uso da terra na Amazônia.

Neste trabalho criamos um modelo que considera as classes floresta e desflorestamento onde, além do produtor de grãos, foi considerado o agente do tipo cerialista associado ao acordo da moratória da soja. No contexto deste modelo, as relações discutidas são aplicadas para representar as interações entre os agentes conforme enumerado abaixo:

- Célula→célula: vizinhança entre as propriedades, baseada na adjacência das células que as compõe. Este fato pode ser verificado considerando a vizinhança de Moore de cada uma das células. Na figura 1(a), por exemplo, c(3,3) é vizinho de c(2,3) que pertence a propriedade P1 e de c(4,3) que pertence à propriedade P3. Logo, as propriedades P1, P2 e P3 são vizinhas.



(a) célula→célula (b) agente→célula (c) célula→agente

Figura 1. Relação envolvendo células

- Agente→célula: relação que conecta o produtor às células que compõem a sua propriedade, como mostra a figura 1(b).
- Célula→agente: permite que cada célula saiba quem é seu proprietário, podendo assim efetuar transições de acordo com as decisões de um agente específico, como exemplificado na figura 1(c).
- Agente→agente: há dois tipos de relação entre agentes neste caso. No primeiro, a relação de vizinhança entre os produtores com base na vizinhança das células que compõem as propriedades que lhes pertencem. No segundo, uma relação de mercado baseado na venda de produtos do produtor para o cerealista associado a moratória da soja(figura 2).

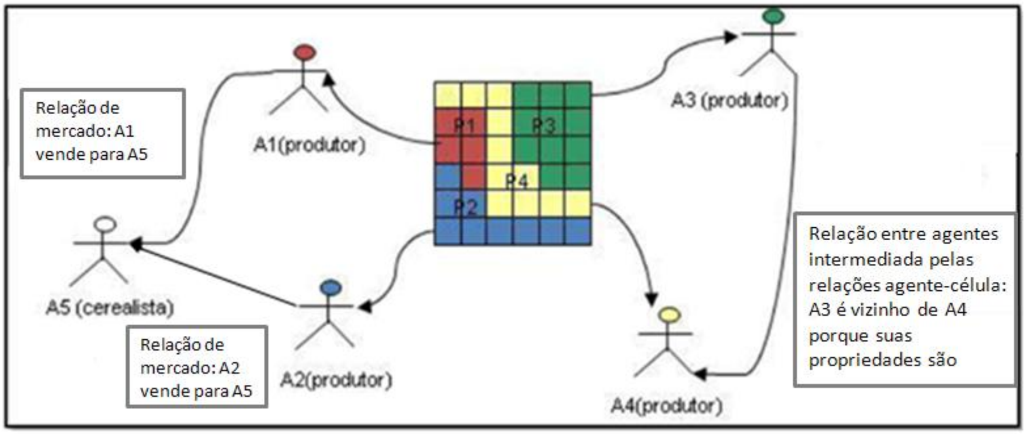


Figura 2. Relações agente→agente

O modelo descrito considera que produtores que se relacionam com o cerealista ficam impedidos de desmatar**,** por constar esta claúsula no pacto. Os demais produtores desmatam segundo variáveis de capital e custo de desmatamento por célula e limite permito de desmatamento ná área da propriedade.

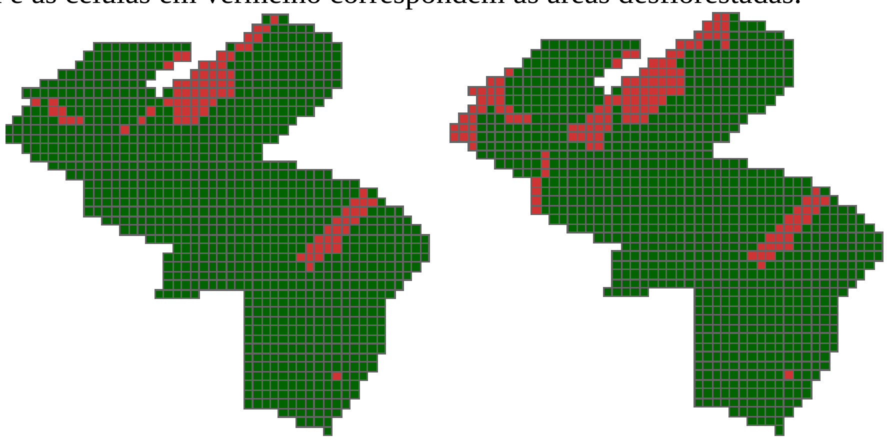
Experimentos

O modelo construído simula a dinâmica da terra entre os anos de 2007 e 2020 e utiliza uma área dividida em onze propriedades, sendo cada uma delas pertencente a um produtor de grãos diferente, conforme mostrado na figura 3.

Figura 3. Divisão das propriedades

Os produtores das propriedades de 7 a 11 se relacionam com o cerealista. A existência desta relação não permite o desmatamento dentro de suas propriedades.

Os resultados obtidos são apresentados a seguir. O 4(a) mostra mapa de uso para 2007, 4(b) mostra a simulação para 2020. As células em verde correspondem as áreas de floresta e as células em vermelho correspondem as áreas desflorestadas.



(a) Mapa 2007 **(b) Simulação 2020**

Figura 4. Mapas de uso para os anos de 2007 e 2020

É visualmente perceptivel que a área onde os agentes tem relação com a moratória permanecem inalteradas a partir de 2007.

Considerações Finais

As relações exploradas neste trabalho são o ponto de partida para que este modelo possa evoluir para modelos mais complexos considerando outras entidades envolvidas na dinâmica de uso da terra da região. Desta forma, novas classes de uso e tipos de agente poderão ser encorporados ao modelo.

O comportamento dos agentes também pode ser refinado, com novos parâmetros para decisão, novas relações de mercado e com instituições regulamentadoras que acarretam em novas restrições as quais podem ou não ser integralmente respeitadas. Além disso, a influência dos vizinhos nas suas tomadas de decisão também pode ser considerada. Este conjunto de melhorias será utilizado futuramente buscando a aproximação do modelo com a realidade observada.

Referências

Andrade, P. R., Monteiro, A. M. V., Camara, G. (2008) "Entities and Relations for Agent-Based Modelling of Complex Spatial Systems". In. *I Brazilian Workshop on Social Simulation* (BWSS/SBIA).

Barros, J. X. (2004) "Urban growth in Latin American cities. Exploring urban dynamics through agent-based simulations". Tese (Doutorado), University of London, Inglaterra.

Coelho, A. (2009) "Modelagem de dinâmica de uso da Terra e cobertura vegetal em área de expansão de grãos na regiao do oeste paraense". Dissertação (mestrado), Universidade Federal do Pará.

Costanza, R., Ruth, M. (1998) "Using Dynamic Modeling to Scope Environmental Problems and Build Consensus", In. *Environmental Management*, p. 183-195.

Gilbert, N. (2007) Agent-based models, Sage Publications Inc., London, Inglaterra.

GTS - Grupo de Trabalho da Soja (2008) "Moratória da Soja: Avanços e Próximos Passos".

Huigen, M., Fischer, G. (2003) Agent Based Modelling in Land Use and Land Cover Change Studies. In: *Interim Repor*.

Lambin, E. F., Geist, H., Rindfuss, R. R. (2006) "Introduction: Local processes with Global impacts", In. Lambin, E., Geist, H. (Ed.). *Landuse and land-cover change: local processes and global impacts*. The Berlin: Springer, p. 1-8.

Matthews, R. B., Gilbert, N. G., Roach, A., Polhill, J. G., & Gotts, N. M. 2007, "Agent-based land-use models: a review of applications", In. *Landscape Ecology*, vol. 22, pp. 1447-1459.

Schelling, T. (1971) "Dynamic models of segregation", In. *Journal of Mathematical Sociology*, vol. 1, p. 143-186.

Turner, B. Skole D., Sanderson S., Fischer G., Fresco L., Leemans R. (1995) Land-Use and Land-Cover Change (LUCC): Science/Research Plan, HDP Report No. 7. IGBP Secretariat. Stockholm.

Vidal, J. M., Buhler, P. A., Huhns, M. N. (2001) "Inside an agent", In. *IEEE Internet Computing*, vol. 5, p. 82–86.

Wooldridge, M., Jennings, N. (1995) Intelligent Agents: Theory and Practice. v. 10. 1995. p. 115-152.

Proceedings XI GEOINFO, Campos do Jordão, Brazil, Nov. 28 to Dec. 1, 2010, INPE, p. 182–187.

Heurística para o posicionamento de reservatórios d'àgua em terrenos representados por matrizes de elevação

J. R. Sousa, K. Nogueira, M. V. A. Andrade, J. Lisboa Filho, T. L. Gomes

Departamento de Informática – Universidade Federal de Viçosa (UFV)
36.570-000 – Viçosa – MG – Brazil

joserui.ufv@gmail.com, keiller.nogueira@ufv.br, marcus@dpi.ufv.br, jugurta@ufv.br, thiago.luange@ufv.br

Abstract. *This paper presents a method to determine the 'best place' in a terrain for the construction of a dam that allow to store a given volume of water. The main contributions of the proposed method are: a heuristic to select the 'candidate points' where to build the dam, the dam's direction and extension is automatically determined and the storage capacity is calculated more accurately.*

Resumo. *Este trabalho apresenta um método para determinação do "melhor local" em um terreno para a construção de uma barragem capaz de armazenar um certo volume de água. As principais contribuições são: uma heurística que seleciona os "pontos candidatos" onde construir a barragem, a direção e a extensão da barragem são calculadas automaticamente e a capacidade do reservatório gerado é determinada de maneira mais precisa.*

1. Introdução

A água é um elemento essencial à vida humana, sendo necessária em muitas atividades como abastecimento urbano, geração de energia elétrica, irrigação, navegação, pesca e, até mesmo, na condução de resíduos residenciais.

Portanto, torna-se imperioso adotar mecanismos de suporte ao gerenciamento de recursos hídricos, adotando como ferramenta o uso de sistemas de informação geográfica (SIG). Dentre as diversas aplicações envolvendo hidrologia, uma questão importante é a construção de reservatórios para armazenamento de água, que tem por objetivo maior suprir as demandas hídricas de regiões e comunidades que se encontram em situação de intermitência de água (Lopes e Freitas 2003).

O objetivo deste trabalho é determinar qual a "melhor" região de um terreno, representado por uma matriz digital de elevação (MDE), para se construir uma barragem que gere um reservatório capaz de armazenar um dado volume de água.

2. Trabalhos Relacionados

Este trabalho supõe que a região alagada será formada exclusivamente pela água proveniente dos rios da região, isto é, não serão considerados os possíveis alagamentos produzidos por precipitações chuvosas. Assim, a determinação da região a ser alagada será realizada utilizando a rede de drenagem do terreno que, de um modo geral, é composta pela direção de fluxo de escoamento e pelo fluxo acumulado que podem ser determinados por um dos métodos descritos em (Moore et al, 1991).

Segundo o DNOCS (Departamento Nacional de Obras Contra as Secas) (DNOCS, 1981) a fundação de uma barragem é a região que suporta o maciço, incluindo as ombreiras e o fundo do vale. A sua preparação deve ser cuidadosa, porque no subsolo de fundação e no maciço é que se encontram os pontos fracos de uma barragem. Por isso é suma importância uma posterior observação da região onde e como será posicionada a barreira.

Geralmente, o volume do reservatório pode ser obtido traçando várias curvas de nível, tomando um espaçamento uniforme entre elas, e calculando o volume entre duas curvas (Matos, 2003). Assim, o volume total corresponde à soma desses volumes. Para obter o volume de cada seção é necessário encontrar a área das superfícies geradas pelas curvas de nível. De forma geral, esta área pode ser obtida pelo método de Gauss (Beyer,1987) que utiliza os pontos conhecidos do polígono para alimentar sua função e obter uma aproximação da área geométrica.

Vale dizer que não foi encontrado nenhum trabalho na literatura que descreva um outro método para obter o local mais adequado para a construção de uma barragem de modo que o método proposto neste trabalho pudesse ser avaliado comparativamente.

3 Metodologia

Como já foi dito, dado um terreno *T* e um valor *k,* o objetivo é determinar em que ponto do terreno (na verdade, de um rio em *T*) deve-se construir uma barragem de modo que esta seja capaz de armazenar um volume de água igual ou maior do que *k*.

De um modo geral, a solução proposta se baseia em utilizar uma heurística para avaliar a condição de cada ponto do rio receber uma barragem (seção 3.1). Caso o ponto seja apto, o passo seguinte é determinar a orientação, a altura e a extensão do reservatório, utilizando o algoritmo da seção 3.2.

Uma vez definida a barragem, o método da seção 3.3 calcula a área alagada e, desta forma, é possível determinar o volume do reservatório, utilizando o método da seção 3.4. Caso o volume for maior que *k* então o "custo" da construção da barragem naquele ponto é avaliado sendo que para isso, é utilizada uma função de custo[1] que envolve algumas variáveis (restrições) e seus respectivos pesos. O objetivo final é obter o ponto onde este "custo" seja mínimo. Se o volume do reservatório for menor que *k*, a altura corrente da barragem é acrescida de 1 metro e o processo retorna na etapa de determinação da represa. Isso ocorre até que o volume seja atendido ou até que a altura seja a máxima possível (o valor mínimo dos máximos locais à "direita" e à "esquerda" do ponto analisado (seção 3.1).

3.1 Heurística para determinar a aptidão de um ponto para receber uma barragem

Em áreas planas, a construção de uma barragem produz reservatórios rasos, o que pode aumentar consideravelmente o custo de construção, pois para se armazenar um grande volume de água nessas áreas é necessário construir uma barragem grande e/ou alta. Por

[1] No momento, o método utiliza uma versão preliminar da função de custo que considera apenas a área da lâmina d´água do reservatório. Na versão final, esta função deverá levar em conta se o reservatório irá alagar regiões com certas características como existência de estradas de rodagem, de ferrovias, áreas indígenas, de preservação, etc.

outro lado, em áreas de relevo irregular (vales), uma barragem relativamente pequena gera um reservatório com maior capacidade de armazenamento e com um menor espelho d'água. Regiões com vales são mais indicadas para construção da represa.

Para determinar se um ponto pertence a uma região plana ou não, isto é, se ele deve ser descartado ou não, a heurística utiliza a direção do fluxo no ponto para avaliar o terreno na direção em que a barragem será construída. Mais precisamente, dado um ponto *p* e dada a direção da barragem, este processo determina a elevação máxima dos pontos da barragem que estão "à direita" e "à esquerda" de *p.*

Após a obtenção dos máximos locais, o próximo passo é verificar se o ponto pertence a uma região plana ou íngreme. Este processo é realizado analisando a tangente do ângulo formado pela linha horizontal e pela "linha" que liga o ponto com o máximo local. Caso essa tangente seja maior que um certo valor pré-estabelecido, então o terreno é classificado como íngreme e o ponto em questão é candidato para construção da represa. Caso contrário, o terreno é considerado plano, e o ponto é descartado. Veja Figura 1a. Na verdade, um ponto só é considerado candidato, caso, em ambos os sentidos, o terreno seja considerado íngreme.

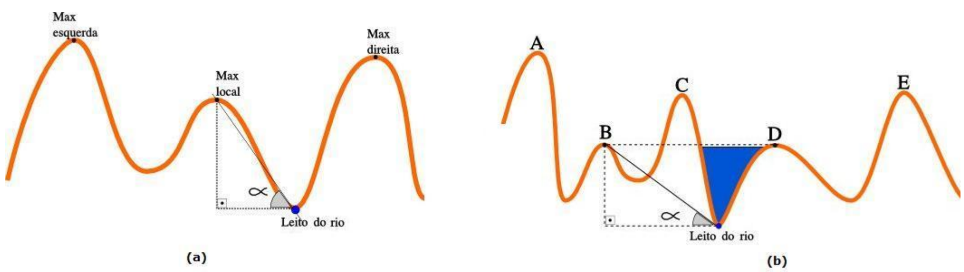


Figura 1. (a) A seleção do ponto baseado na sua inclinação (b) O máximo local C irá impossibilitar a construção da barragem ligando os pontos B e D. Logo, não é possível selecionar o B para teste em relação à inclinação.

Testes realizados mostraram que esta heurística obtém, em média, 70% das "melhores soluções", isto é, das soluções ótimas encontradas pelo método "força bruta".

3.2 Orientação e extensão da barragem

Dado um terreno *T* representado por uma matriz de elevação *M*, suponha que se deseja construir uma barragem de altura *h* num determinado local do terreno. Com um ponto do rio selecionado a orientação da barragem é determinada de modo que a barragem fique perpendicular ao fluxo do rio. Sua extensão é definida como sendo todos os pontos do terreno, ao longo da "linha" perpendicular ao rio, cuja elevação é menor do que a altura *h.* Partindo de *p* é realizado um percurso, em ambas as direções, sobre a "linha" perpendicular ao rio, até que uma posição do terreno com altura maior do que *h* seja alcançada (Figura 2).

3.3 Região alagada

Uma vez definida a barragem (sua posição, altura e extensão), o próximo passo é identificar a(s) região(ões) do terreno que serão alagadas devido o represamento do rio.

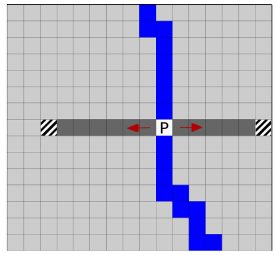


Figura 2. No desenho, a barragem de 12 pontos parte do ponto *P* e se estende até atingir os pontos de altura maior que *h* (pontos hachurados).

Definição: Dada uma matriz M e uma posição $p=(p_i, p_j)$ nesta matriz, as posições vizinhas a p são (p_i+1, p_j), (p_i+1, p_j+1), (p_i, p_j+1), (p_i-1, p_j+1), (p_i-1, p_j), (p_i-1, p_j-1), (p_i, p_j-1) e (p_i+1, p_j-1). Dois pontos p_1 e p_2 são conexos se existe uma seqüência de pontos $q_1, q_2, ..., q_n$ com $q_1=p_1$ e $q_n=p_2$ tal que q_i é vizinho a q_{i+1} para todo $i=1,2,..., n-1$. Uma componente conexa na matriz M é um conjunto de pontos conexos que possuem uma determinada característica em comum. Em particular, uma $cc(h)$ é uma componente conexa formada pelos pontos conexos cuja altura é no máximo h.

Assim, a região a ser alagada após a construção de uma barragem de altura h será formada pelos pontos conexos à barragem cuja elevação é inferior ou igual à altura desta barragem. Mais precisamente, a região alagada será formada pela componente conexa $cc(h)$ que possui pelo menos um ponto de um rio.

Para determinar esta (variação da) componente conexa, usa-se uma adaptação do método *Connect*, proposto por (Franklin, 2006), que dada uma matriz de bits, isto é, contendo valores 0 e 1, obtém as componentes conexas formada por 0.

Assim, o método foi adaptado da seguinte forma: dada uma matriz de elevação M de dimensão $n \times n$ e dado o valor h, o método gera uma matriz M' de bits, também de dimensão $n \times n$, tal que a posição (i,j) na matriz M' irá conter o valor zero caso o valor na posição (i,j) da matriz M seja menor ou igual a h; caso contrário, se na posição (i,j) da matriz M houver um valor maior do que h então será armazenado o valor 1 na posição (i,j) da matriz M'. Assim, o método *Connect* é aplicado à matriz M' e as componentes conexas obtidas correspondem às componentes conexas na matriz M cujos pontos possuem elevação no máximo h. Daí, para determinar a região alagada basta verificar qual(is) componente(s) possui(em) pelo menos um ponto num rio da rede de drenagem sendo que o respectivo ponto do rio deve estar a montante[2] da barragem.

3.4 Volume do reservatório gerado pela barragem

Suponha uma barragem de altura h construída numa MDE com resolução de r metros, isto é, onde cada célula corresponde a uma região com área $r \times r$. Assim, cada célula p com elevação e_p pertencente ao reservatório formado pela barragem é capaz de armazenar um volume de água $v(p)$ dado por (veja figura 3):

---

2 Montante é todo um ponto de referência ou secção de um rio que se situa antes de um ponto referencial qualquer de um curso de água. Sendo assim a foz de um rio é o ponto mais a jusante deste rio, assim como a nascente é o seu ponto mais a montante.

$$v(p) = r^2 \; x \; (h - e_p)$$

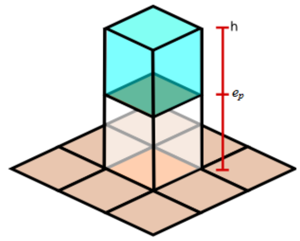


Figura 3. Volume de uma célula da MDE. Calculado pela multiplicação da "área da célula" pela diferença entre a altura da lâmina dágua e a altura da célula.

O volume (máximo) total do reservatório será dado pelo somatório do volume de todas as células que pertencem à região alagada.

5 Resultados

Foram realizados vários testes utilizando dados captados pela *Shuttle Radar Topography Mission* (SRTM) (Sousa et al, 2006). Dentre eles, serão apresentados os resultados considerando uma região abrangendo a cidade de Viçosa, MG, Brasil, cujas coordenadas de latitude e longitude são 20°45'14"S, 42°52'55"O, respectivamente. A região utilizada inclui parte da bacia do ribeirão São Bartolomeu e é representada por uma MDE de 120 x 122 ou seja 14.640 células. Cada célula corresponde a uma região do terreno de dimensão 90 x 90 metros.

No primeiro teste, o objetivo foi obter um reservatório com, pelo menos, um milhão de metros cúbicos de água. O peso da variável que representa a lâmina d'água foi 3 e o peso da variável que representa a extensão da barragem foi 2. O resultado obtido pela heurística proposta foi o mesmo que o obtido pelo método da força bruta em que todos os pontos são testados. O resultado é mostrado na figura 4a. Este resultou numa barragem com extensão de 180 metros (2 pixels) e altura de 8 metros. A área alagada foi de 26 pixels ou 210.600m². O valor da função objetivo foi 82.

Em um segundo teste, o volume desejado foi de, no mínimo, 5 milhões de metros cúbicos de água. As variáveis que representam a lâmina d'água e extensão do reservatório receberam, respectivamente, os valores 2 e 8. O resultado obtido usando a heurística e usando força bruta foram os mesmos. O resultado é mostrado na figura 4b. Neste caso, foi gerada uma barragem com extensão de 270 metros (3 pixels) e altura de 19 metros. A área alagada totalizou 84 pixels ou 680.400m³. O valor da função objetivo nesse ponto foi de 192.

6 Conclusão e trabalhos futuros

Neste trabalho foi apresentado um método para a determinação da "melhor posição" para a construção de uma barragem que produza um reservatório com capacidade mínima igual a um valor pré-determinado. As principais contribuições são: a direção e a extensão da barragem, assim como a região alagada, são determinadas automaticamente considerando a altura da barragem; uma heurística seleciona os pontos candidatos e é apresentado um método alternativo (mais simples e mais preciso) para a determinação do volume do reservatório correspondente à região alagada.

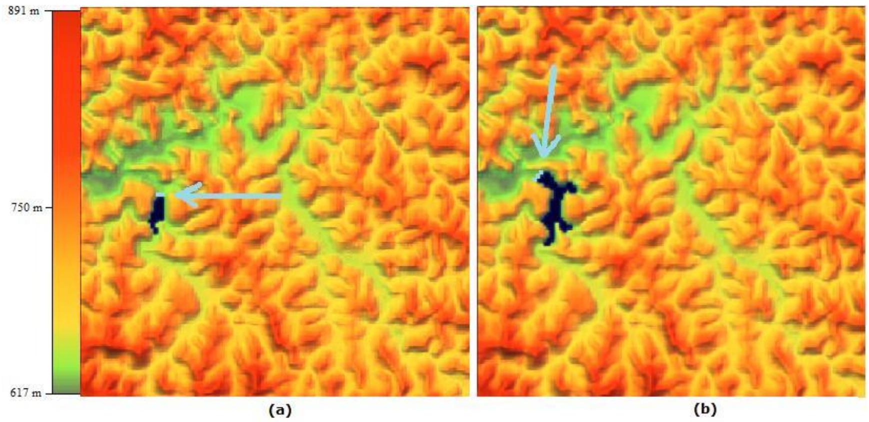


Figura 4. A parte mais escura da figura representa a área alagada devido a construção de uma barragem. (a) O volume do reservatório gerado é 1.012.500 m^3. (b) O volume do reservatório gerado é 5.078.700 m^3.

Este trabalho se encontra em andamento e o próximo passo será englobar a análise do terreno onde a barragem será construída para evitar que sejam alagadas áreas urbanas, indígenas e de proteção ambiental, além de ferrovias e rodovias.

Agradecimentos

Este trabalho foi parcialmente financiado pela FAPEMIG – Fundação de Amparo à Pesquisa do Estado de Minas Gerais e pelo CNPq – Conselho Nacional de Desenvolvimento Científico e Tecnológico.

Referências

Beyer, W. H. (Ed.). CRC Standard Mathematical Tables, 28th ed. Boca Raton, FL: CRC Press, pp. 123-124, 1987.

DNOCS. Departamento Nacional de Obras Contra as Secas (DNOCS). Intruções gerais a serem observadas na construção das barragens de terra. 2ª. Edição revista e ampliada. Fortaleza. 1981. (Preparação das Fundações, 30-31).

Franklin, W. R and Landis, E. Connected Components on 1000x1000x1000 Datasets. *In 16th Fall Workshop in Computational Geometry,* Smith College, Northampton, MA, 10-11 Nov 2006.

Lopes, A.V. Freitas, M. A. S. Avaliação das demandas e ofertas hídricas na bacia do rio São Francisco usando modelo de rede de fluxo.

Moore, I. D. Grayson, R. B. and Ladson. A. R. Digital terrain modelling: a review of hydrological, geomorphological and biological aplications. *Hydrological Processes*, 5:3-30, 1991.

Matos, A, T., Silva, D. D., Pruski, F. F., 2003. *Barragens de Terra de Pequeno Porte*, 2ed, Viçosa: UFV, 124p. (caderno didático 73)

Sousa, L. Nery, F. Matos, J. Metodologias de processamento de dados SRTM para a produção de modelos digitais de direções de escoamento. VIII Congresso da Água. Figueira da Foz, Portugal, 2006.

EpiGTA/TerraView: Plug-in para Simulação de Enfermidades Infecciosas em Redes Sociais com Fluxo de Populações

Carolina S. Pena[1], Marcelo A. Costa[1], André de Souza Gomes[2]

[1]Departamento de Estatística – Universidade Federal de Minas Gerias (UFMG)
CEP 31.270 - 901 – Belo Horizonte, MG – Brasil

[2]Programa de Pós Graduação em Engenharia Elétrica – Universidade Federal de Minas Gerais (UFMG)

carolinasend@gmail.com, azevedo@est.ufmg.br, asgomes86@gmail.com

Abstract. *This project aims to develop a TerraView plug-in application that simulates spatial and temporal spread of highly contagious diseases based on social networks with flow of populations. Three discrete-time stochastic SIR (Susceptible-Infective-Recovered) models will be implemented. The main application of this project is the study of the spread of foot-and-mouth disease in the Brazilian territory using the electronic records of cattle movements, provided by the Ministry of Agriculture and Livestock.*

Resumo. *Este projeto tem como objetivo implementar uma ferramenta computacional, na forma de um plug-in do aplicativo TerraView, capaz de gerar cenários de disseminação espaço-temporal de doenças altamente contagiosas. Para isso é utilizado um grafo do fluxo de movimentos de populações em redes sociais. Três modelos estocásticos de tempo discreto do tipo SIR (Suscetível-Infectado-Recuperado) serão implementados. A aplicação inicial deste projeto é o estudo da disseminação da Febre Aftosa no território brasileiro a partir do registro da movimentação de bovinos, registradas nas Guias eletrônicas de Trânsito Animal (e-GTA).*

1. Introdução

O desempenho do setor agropecuário brasileiro ao longo das últimas décadas contribuiu substancialmente para atenuar os efeitos locais da crise financeira internacional. Em 2008, o agronegócio brasileiro respondeu por um em cada três reais de renda gerada no país e contribuiu decisivamente para o superávit da balança comercial. O setor abrangeu 28% do PIB, 36% das exportações e 37% dos empregos (MAPA, 2009).

A evolução da condição sanitária têm sido um fator primordial para a manutenção e ampliação dos mercados internacionais. No entanto, alguns importantes mercados, como Estados Unidos, Japão e a União Européia ainda estabelecem restrições à carne bovina in natura, alegando problemas sanitários.

Por outro lado, a movimentação de animais no país é intensa, especialmente para as espécies bovina e suína. Se um surto de uma doença ocorrer, o trânsito de gado poderia ser suspenso temporariamente para ajudar a conter o avanço da doença. No entanto, parar todo esse transporte pode resultar em gastos muito altos para os produtores, sendo que as indenizações pagas pelo governo pela intervenção e pelo sacrifício de animais não cobrem todos os custos.

Vindo de encontro às necessidades apontadas acima, este projeto propõe a implementação de uma ferramenta computacional, calibrada com os bancos de dados de trânsito de animais. Essa ferramenta, denominada *EpiGTA*, permitirá avaliar, mediante simulações, os possíveis cenários de disseminação inicial de doenças de animais altamente contagiosas no país, a partir do fluxo de movimentos de animais. O intuito é que o software possa auxiliar aos órgãos oficiais de defesa sanitária e sanidade animal no desenvolvimento e otimização de potenciais estratégias de prevenção e controle de tais doenças, visando impedir a sua incursão/re-introdução/disseminação no país.

2. Material e Métodos

A movimentação de bovinos no território brasileiro é atualmente documentada eletrônicamente através da emissão de Guias de Trânsito Animal (GTAs). As informações utilizadas no desenvolvimento desse projeto foram retiradas dessas guias, são elas: municípios de origem e destino, finalidade (cria, engorda, reprodução, abate ou exposição), data da emissão da GTA e quantidade de animais transportados. Os dados analisados nesse artigo são referentes às GTAs emitidas em 2006 no estado do Acre.

Uma forma de analisar dados de trânsito de animais entre propriedades é através de uma rede de contatos direcionada, na qual as propriedades rurais são representadas por nós e os movimentos de gado entre propriedades são as arestas. A matriz de adjacência, G, é uma representação matemática adequada para descrever as conexões entre os nós de uma rede (Keeling & Eames, 2005). Considerando nós como indivíduos de uma população, define-se $G_{ij}=1$, se existe uma conexão que permita que a infecção passe do indivíduo i para o indivíduo j, e $G_{ij}=0$, em caso contrário. Nesse estudo considera-se que a infecção pode ocorrer somente em um sentido, dessa forma, a rede é representada por um grafo direcionado (Harary, 1969; Bollobás, 1979).

Representações que utilizam uma única matriz de adjacência para um período de n dias de simulação são chamadas de redes estáticas. Nesse trabalho, a representação dos movimentos será feita por meio de uma rede dinâmica. Na prática, esse tipo de rede consiste em um conjunto de inúmeras redes estáticas, uma para cada unidade de tempo. Dessa forma, se houver uma movimentação de animais da fazenda i para a fazenda j no dia t, então a rede para o dia t deve conter uma aresta $i \rightarrow j$. A rede dinâmica para um período de n dias contados a partir do dia t_0 é dada por:

$$G(t_0, n) = \{G(t_0), ..., G(t_{n-1})\},$$

As redes dinâmicas têm sido consideradas o padrão ouro de representação de movimentos em diversos estudos (Vernon & Keeling, 2008). Em termos de transmissão de doenças, quanto maior o número de conexões de uma rede, mais provável é a disseminação da doença.

Para realizar simulações, serão incorporados três modelos estocásticos de tempo discreto do tipo SIR (Suscetível-Infectado-Recuperado). A suposição básica destes modelos é que um indivíduo pode passar pelos estágios de suscetibilidade, infecção e recuperação, e que a imunidade é permanente após a recuperação. Existe ainda uma simplificação do modelo SIR, no qual o individuo infectado não se recupera, nesse caso considera-se que o tempo de recuperação é infinito (R = ∞), reduzindo o modelo SIR para o modelo SI (Suscetível-Infectado).

Em um dos modelos avaliados, a menor unidade considerada é a fazenda (Vernon & Keeling, 2008). Nas simulações desse modelo todas as fazendas começam no estado susceptível (S), exceto uma delas que é aleatoriamente escolhida para iniciar o processo no estado infectado (I). O modelo é sequencialmente atualizado utilizando passos diários. Durante cada dia de simulação a doença é transmitida, através do contato (aresta), de fazendas infectadas para fazendas susceptíveis (nos), com probabilidade *p*. As fazendas se mantêm no estado infectado (I) por um número inteiro de interações, μ, e então passam para o estado recuperado (R). Nós que atingem o estado *R* se mantém no mesmo até o final da simulação. O parâmetro *p* se mantém constante durante qualquer simulação especifica e o parâmetro μ pode ser substituído por uma variável aleatória como, por exemplo, uma distribuição exponencial com média λ.

Nesse tipo de modelo todas as fazendas são consideradas idênticas, de tal forma que nem mesmo o número de animais transportados afeta a dinâmica da transmissão da doença, conforme é descrito em (1) (Vernon & Keeling, 2008):

$$\begin{cases} p(\,fazenda(i,t+1) = I \mid fazenda(i,t) = S\,) = 1 - \prod_{fazenda(j,t)=I} 1 - pG_{j,i}(t) \\ p(\,fazenda(i,t+1) = R \mid fazenda(i,t-\mu_i) \neq S\,) = 1 \end{cases} \tag{1}$$

onde $fazenda(i,t) \in \{SIR\}$ é o estado do nó (fazenda) *i* no tempo *t*, μ_i é o tempo de recuperação da *fazenda i* e $G_{ji}(t) = 1$ se houve movimentação do nó *j* para o nó *i* no instante *t* e $G_{ji}(t) = 0$, caso contrário. A probabilidade de contaminação da *fazenda i* no tempo *t+1* (primeira linha da equação) depende a estrutura da rede, enquanto o processo de recuperação é independente, operando ao nível da propriedade (nó). Os dois outros modelos avaliados consideram o animal como a menor unidade de trabalho. Nesse caso, todos os animais começam no estado susceptível (S), exceto um deles que é aleatoriamente escolhido para iniciar o processo no estado infectado (I). A fazenda é considerada infectada (I) se possui pelo menos um animal infectado em seu rebanho. Durante cada dia de simulação, animais infectados espalham a doença dentro do rebanho do *j*-ésimo nó com probabilidade $\pi_j(t)$. A doença é transmitida para fazendas susceptíveis através do envio de animais contaminados, com probabilidade $p_{ji}(t)$. Os animais se mantêm no estado infectado (I) por um número inteiro de interações, μ, e então passam para o estado recuperado (R). Animais que atingiram o estado R se mantém no mesmo estado até o final da simulação. Nesse tipo de modelo considera-se ainda outra medida de controle da infecção, o abate de animais. Quando um movimento da GTA é classificado como abate, os animais retirados da fazenda de origem são sacrificados, impossibilitando que esse tipo de movimentação cause contaminação na fazenda de destino. Ao final de cada dia, verifica-se a existência de animais infectados em todas as fazendas presentes na simulação. Caso não haja animais infectados em seu rebanho, a fazenda é considerada susceptível (S).

Seja $S_j^+(t)$ o número de animais susceptíveis na fazenda *j* no tempo *t* após todas as movimentações do dia *t*; $I_j^+(t)$ o número animais infectados na fazenda *j* no tempo *t* após todas as movimentações apuradas no dia *t*. Similarmente, $I_j^-(t)$ representa o número animais infectados na fazenda *j* no tempo *t* no início do dia (ou seja, antes das movimentações) e $R_j^+(t)$ o número animais recuperados na fazenda *j* no tempo *t* após todas as movimentações do dia *t*, então pode-se dizer que:

$$I_j^-(t+1) \mid I_j^+(t), S_j^+(t) \sim Binomial(S_j^+(t), \pi_j(t)),$$

onde $\pi_j(t)$ é a probabilidade de contaminar um novo animal na fazenda *j* no tempo *t* dado que existem $I_j^+(t)$ animais infectados na fazenda *j* no tempo *t*.

No modelo de Greenwood (Greenwood, 1931) $\pi_j(t) = \pi$ é uma constante que não depende do número $I_j^+(t)$ de infectados. No modelo Reed-Frost (Abbey, 1952) $\pi_j(t) = 1-(1-\pi)^{I_j^+(t)}$. Dessa forma, o modelo de Reed-Frost considera que a probabilidade de um novo animal se tornar infectado depende do número de animais infectados atualmente presentes no rebanho. Seja $n_{ji}(t)$ o número de animais transportados da fazenda *j* para a fazenda *i* no tempo *t* e seja $X_{ji}(t)$ o número de animais infectados em $n_{ji}(t)$ animais transportados, então pode-se dizer que:

$$X_{ji}(t) \sim Binomial(n_{ji}(t), \pi_j(t)).$$

A probabilidade $p_{ji}(t)$ de contaminação da fazenda *i* pela fazenda *j* no instante *t* é definida por:

$$p_{ji}(t) = P(X_{ji}(t) \geq 1) = 1-(1-\pi_j(t))^{n_{ji}(t)} \tag{2}$$

Dessa forma, nos modelos de Greenwood e Reed-Frost, a dinâmica de transmissão da doença é descrita por:

$$\begin{cases} p(fazenda(i,t+1) = I \mid fazenda(i,t) = S) = 1 - \prod_{fazenda(j,t)=I} 1 - p_{ji}(t)G_{j,i}(t), \\ p(animal(u,t+1) = R \mid animal(u,t-\mu_u) \neq S) = 1, \end{cases} \tag{3}$$

onde $animal(u,t) \in \{S\,I\,R\}$ é o estado do animal *u* no tempo *t*, $fazenda(i,t) \in \{S\,I\}$ é o estado da fazenda *i* no tempo *t* e $p_{ji}(t)$ é descrito em (2).

Uma das inovações trazidas por esse trabalho foi retirar os modelos de Greenwood e Reed-Frost do contexto de populações confinadas (Tuckwell & Williams, 2007) expandindo a sua utilização para cenários em que populações, ou propriedades, estão interligadas por uma rede dinâmica de movimentos.

3. Descrição dos dados

Como exemplo, foi utilizada a rede de contatos entre as fazendas do estado do Acre construída a partir das GTAs emitidas no ano de 2006. A Figura 1 mostra a distribuição espacial da movimentação de bovinos no estado. O mapa foi gerado a partir do plug-in "*Flow*" do software *TerraView*.

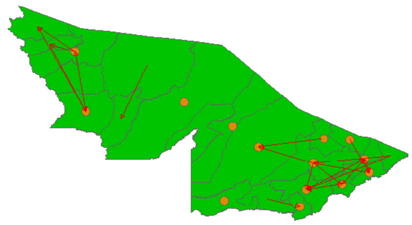


Figura 1. Fluxo de movimentos entre os municípios do Acre no ano de 2006

Os pontos vermelhos na Figura 1 revelam a presença de fluxo interno, ou seja, de GTAs emitidas entre propriedades localizadas no mesmo município. As setas indicam a direção da movimentação (origem → destino).

4. Resultados

Os resultados preliminares mostrados nesse artigo são baseados nas movimentações do período de janeiro a fevereiro, totalizando 2220 GTAs; foram realizadas 10.000 simulações para cada dia do período estudado. As simulações foram direcionadas para o estudo do espalhamento da febre aftosa. Em consonância com características dessa doença, nas simulações dos modelos Greenwood e Reed-Frost, considerou-se que $\pi = 0{,}4$ (ver equação (3)) e no modelo Nível Fazenda foi utilizado $p = 0{,}4$ (ver equação (1)). Para o modelo SIR considerou-se ainda que o parâmetro $\mu \sim Poisson(15)$.

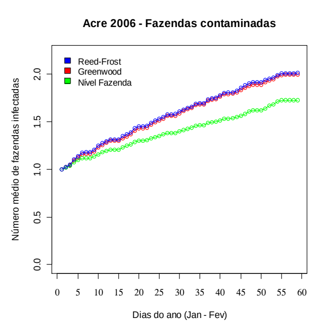
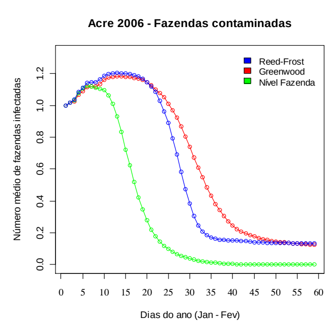
Modelo SI	Modelo SIR
	

Figura 2. Número médio de fazendas infectadas por dia de simulação

Os resultados obtidos a partir do modelo SI mostraram que o número médio de fazendas contaminadas por dia nos modelos de Greenwood e Reed-Frost é praticamente o mesmo. Por outro lado, o modelo Nível Fazenda estima valores relativamente mais baixos, especialmente com o crescimento do tempo desde o início da epidemia.

No modelo SIR, observa-se que após 10 dias do início da contaminação, o modelo Nível Fazenda já atingiu o cenário mais crítico da doença. A partir desse ponto o número médio de fazendas contaminadas possui tendência decrescente, chegando à extinção dos focos da doença no decorrer do período simulado. Para os modelos Greenwood e Reed-Frost, a involução da doença começa a acorrer após 15 dias do aparecimento do primeiro foco, sendo que o decaimento do número médio de fazendas infectadas no Reed-Frost é mais acentuado do que no modelo de Greenwood.

5. O aplicativo *EpiGTA*

O aplicativo *EpiGTA* permitirá ao usuário escolher um dos três modelos apresentados anteriormente. As simulações poderão ser realizadas utilizando apenas informações de GTAs, sem dados do rebanho. Caso esteja disponível o número de animais presentes nas propriedades, o programa será capaz de simular cenários mais reais. O aplicativo *EpiGTA* irá gerar resultados na forma de mapas e curvas epidemiológicas (ver também Coelho *et al*, 2008) . A interface sugerida é apresentada a seguir.

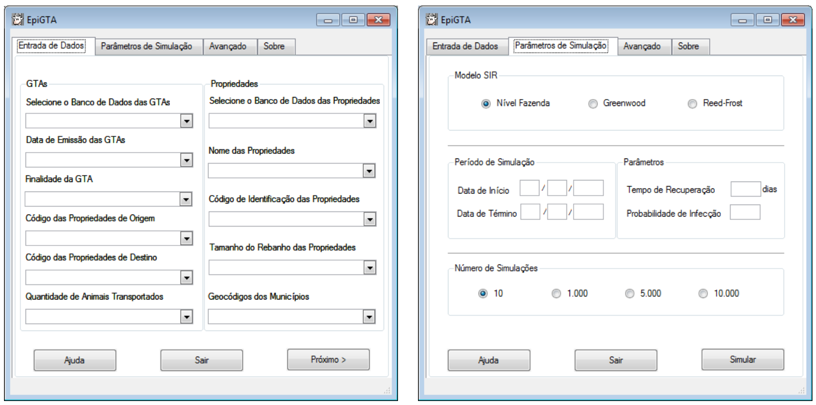


Figura 3. Interface do aplicativo *EpiGTA*

Referências

Abbey, H. (1952). An examination of the Reed-Frost theory of epidemics, *Hum. Biol.* 24, 201.

Bollobás, B. (1979). Graph theory. New York: *Springer*

Coelho, F.C., Cruz, O.G. & Codeço, C.T. (2008). Epigrass: a tool to study disease spread in complex networks. *Source Code for Biology and Medicine*. 3:3.

Greenwood, M. (1931). On the statistical measure of infectiousness, J. *Hyg. Camb*. 31, 336.

Harary, F. (1969). Graph theory. Reading, MA: *Addison – Wesley*.

Keeling, M.J. & Eames, K.T.D. (2005). Networks and epidemic models. *Journal of the Royal Society Interface*, 2, 295-307.

MAPA, Ministério da Agricultura Pecuária e Abastecimento (2009). Relatório de avaliação do plano plurianual 2008-2011, Ano Base 2008. Brasília, D.F, 91p. Disponível em: <http://www.agricultura.gov.br>. Acesso: Março 2010.

TerraView – Aplicativo para visualização, consulta e análise de dados geográficos – Disponível em: http://www.dpi.inpe.br/terraview> Acesso em: Outubro 2010.

Tuckwell, H.C. & Williams, R.J. (2007). Some properties of a simple stochastic epidemic model of SIR type. *Mathematical Biosciences*. 208, 76 – 97.

Vernon, M.C. & Keeling, M.J. (2008). Representing the UK's cattle herd as static and dynamic networks. *Proceedings of the Royal Society B*, Published online, 1-8

Proposta de Planejamento em Etapas para Redes Secundárias de Telecomunicações

Hideson A. Silva[1], Luiz E. S. Oliveira[2], Alceu S. Britto Jr[3]

[1]Companhia Paranaense de Energia (COPEL)
[2]Departamento de Informática – Universidade Federal do Paraná
[3]Departamento de Informática – Pontifícia Universidade Católica do Paraná
Curitiba – Paraná – Brasil

hideson@copel.com, lesoliveira@inf.ufpr.br, alceu@ppgia.pucpr.br

Abstract. *This paper presents an approach to planning the last mile access nodes of telecommunication secondary networks, aiming to find the best position, the amount of equipment and their best linkage. The proposed approach is to use georeferred information as a reference to the construction of telecommunication network. This process is carried out within two stages: the first one selects the amount of equipment and their position and the second one determines the best connection path between the equipment and the customers.*

Resumo Este trabalho apresenta uma abordagem no planejamento dos pontos de acesso da última-milha das redes secundárias de telecomunicações, com o objetivo de encontrar a melhor localização, quantidade de equipamentos e a ligação entre eles. A proposta é utilizar informações georreferenciadas como uma referência na construção da rede secundária. O processo é realizado em duas etapas: a primeira seleciona a quantidade dos equipamentos e a segunda determina o melhor caminho de ligação entre os equipamentos e clientes.

1. Introdução

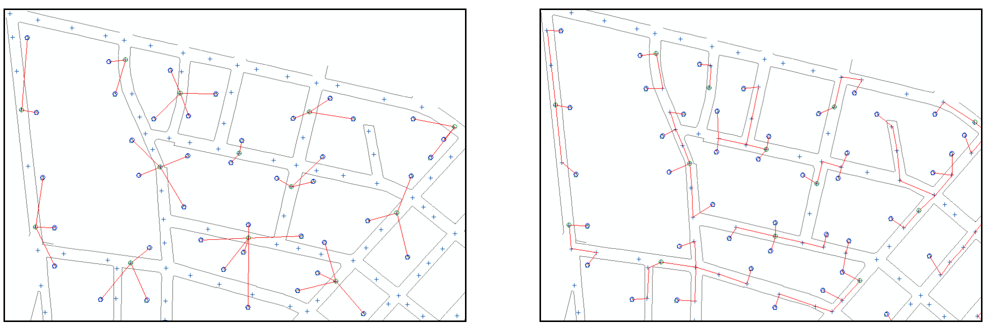
O planejamento de redes é também estudado como *Steiner-tree problem* (STP)[1] e há na computação evolucionária propostas para este problema. Por exemplo, utilizando algoritmos genéticos (AG) Souza et al [2] propõe a avaliação de redes secundárias de energia elétrica apresentando resultados com a simulação de um circuito com grande concentração de unidades consumidoras. Utilizando o algoritmo genético paralelo (uma variação do AG), Huy e Nghia[3] apresentam uma solução em que a *fitness* é avaliada com a heurística DNH (*Distance Network Heuristic*) e as soluções encontradas são comparadas com as metaheurísticas: PGS (*Parallel Genetic Algorithm for Steiner Tree Problem*), GRASP (*Greedy Randomized Adaptive Search Procedure*), EGA (*Enhanced Genetic Algorithm*), PGRASP (*Parallel Greedy Randomized Adaptive Search Procedure*) através de testes realizados com a base OR-Library[4] .

No trabalho dos autores Ding e Ishii[5] é proposta uma solução para o problema *Dynamic Steiner tree* (DST), utilizando o *Online Genetic Algorithm* (OLGA), com o algoritmo *Prim* para a avaliação da *fitness*. Os testes realizados, com a base OR-Library, são comparados com os algoritmos DNH (*Distance Network Heuristics*), SPH (*Sortest Path Heuristic*) e ADH (*Average Distance Heuristic*).

Outro algoritmo utilizado é o *Discrete Particle Swarm Optimization* (DPSO), uma adaptação do enxame de partículas (PSO-*Particle Swarm Optimization*), e é utilizado por Zhong e Huang para uma comparação com o AG através de testes com a base OR-Library. Eles apresentam resultados melhores com o DPSO para solucionar o STP. Outra adaptação do algoritmo PSO é apresentada por Zanh et al [7] com o objetivo de acelerar o processamento com uma técnica em que apenas as soluções promissoras da partícula são avaliadas pela função de *fitness*.

Além do AG e do PSO utilizados nos trabalhos apresentados, outras heurísticas, tal como a Busca Tabu (BT), como é apresentada em Silva e Rodrigues[1] , Ribeiro e Souza[8] e Xu e Glover[9] , são utilizadas para solucionar o STP.

Os trabalhos, em geral, apresentam soluções para o STP considerando apenas a distância *Euclidiana* entre os nós, conforme a Figura 1(a). Mas as restrições físicas da geografia, tais como ruas, edifícios, postes de energia elétrica, etc, devem ser consideradas, como é apresentado na Figura 1(b).



(a) (b)

Figura 1.Construção de redes.

Para resolver esse problema, propomos uma solução baseadas em duas etapas utilizando dados georreferenciados, a qual é demonstrada através de resultados experimentais com dois algoritmos de otimização: AG[10] e PSO[11] .

O artigo está organizado em: a seção dois apresenta o problema de planejamento da rede secundária; na seção três está o modelo matemático; a seção quatro descreve a solução proposta. E os resultados são apresentados e discutidos na seção cinco.

2. Planejamento de Redes de Telecomunicações

Para efeitos de planejamento a rede é dividida em rede primária, que mantém os equipamentos centrais, e a rede secundária que mantém a ligação com os clientes (última-milha). Em resumo um projeto deve fornecer as seguintes informações:

- Quantidade de equipamentos a serem instalados.
- Localização geográfica dos equipamentos.
- Definição da ligação do cliente ao equipamento.

A última-milha é construída com base na ligação de um nó de demanda (cliente) a um nó de acesso (equipamento) e os custos estão diretamente relacionados com a distância entre os nós e com o custo da construção dos nós de acesso. Portanto há uma quantidade ideal de nós de acesso a ser utilizada para um determinado conjunto de demandas e esta quantidade e localização devem ser identificada no planejamento.

Quando é um trabalho manual é complexo para relacionar todas as variáveis do problema, identificar cada possibilidade de rede, selecionar os pontos geográficos e a cada iteração deste processo reavaliar todas essas variáveis.

3. Modelo Matemático

A função objetivo (1) minimiza a somatória do custo de ligação entre os nós de acesso aos de demandas e somando o custo de ativação dos nós de acesso. E esta adição do custo de ativação auxilia encontrar a quantidade ideal de nós.

A restrição (2) representa que todo nó de demanda *i* deve ter uma a conexão com um nó de acesso *j* e devem fazer parte da solução final. A variável x_{ij} tem o valor 1 nos casos em que é válida a ligação por esta conexão e 0 caso contrário.

A restrição (3) restringe a conexão entre o nó de demanda *i* com o nó de acesso *j*, obrigando que cada nó de demanda tenha apenas a ligação com um nó de acesso.

Variáveis do problema:

M: conjunto de nós de demanda.
N: conjunto de nós de acesso.
c_{ij}: custo de conexão da demanda *i* ao nó de acesso *j*.
b_j: custo da instalação do nó de acesso *j*.

Variáveis de decisão:

x_{ij}: identifica conexão da demanda *i* ao nó de acesso *j* quando assume o valor 1.
e_j: identifica se o nó de acesso *j* está ativado.

Função Objetivo:

$$\min \sum_{i \in M} \sum_{j \in N} c_{ij} x_{ij} + \sum_{j \in N} b_j \, e_j \qquad (1)$$

Restrições:

$$\sum_{j \in N} x_{ij} = 1, \quad i \in \mathrm{M} \qquad (2)$$

$$x_{ij} \in \{0,1\}, i \in M, j \in N \qquad (3)$$

4. Metodologia Aplicada

A abordagem proposta é a otimização do planejamento da última-milha da rede secundária de telecomunicações, utilizando como sugestão de locais para instalação dos equipamentos, as informações geográficas dos postes de distribuição de energia elétrica existentes na região. E com estes dados, mais as coordenadas dos clientes a serem atendidos, é proposto o processamento em duas etapas:

1o.) Utilizar um algoritmo de otimização, com o cálculo de *fitness* baseado na distância *Euclidiana*, para selecionar a quantidade e localização dos equipamentos.
2o.) Processar as soluções obtidas na primeira etapa, utilizando um algoritmo de menor caminho, para obter a solução com o melhor caminho entre os nós.

A busca de soluções utilizando apenas o cálculo de *fitness* através da distância *Euclidiana* não retorna soluções reais, por não considerar a geografia da região, mas tem um custo computacional menor, em comparação ao utilizar uma *fitness* apenas com um algoritmo que calcule o melhor caminho a cada iteração. Neste trabalho é utilizado para cálculo do melhor caminho o algoritmo *Dijkstra*[12] .

Os dados são primeiramente processados no algoritmo evolucionário com a função de *fitness* baseada na distância *Euclidiana*, obtendo o conjunto de soluções com a quantidade e localização dos equipamentos de telecomunicações.

Na segunda etapa, processando a solução obtida anteriormente, o algoritmo *Dijkstra* fornece o melhor caminho com a real distância entre os nós de demanda com os nós de acesso. Isso otimiza o processo que no primeiro momento realiza um cálculo mais rápido (distância *Euclidiana*) e em seguida apenas as melhores soluções são processadas na definição do melhor caminho (algoritmo *Dijkstra*). As informações das coordenadas de instalação dos equipamentos, mais o caminho real percorrido desde o cliente até o equipamento selecionado, resolvem a questão de restrições geográficas.

O cromossomo no AG e a partícula no PSO são representados por vetores binários, nos quais cada posição do vetor representa um nó do problema, sendo cada nó uma indicação de equipamento com a coordenada geográfica. No exemplo da Figura 2 é apresentada uma representação binária com quatro nós de acesso selecionados e os respectivos nós de demanda atendidos.

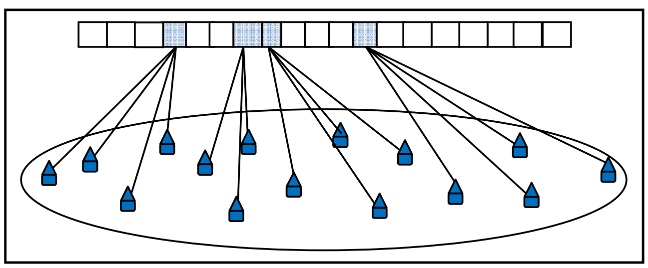


Figura 2.Representação simbólica de uma solução.

6. Testes e Resultados

Os testes foram executados com o AG, na ferramenta Matlab® 7.1 SP3 (*gatool* versão 2.0) e com o PSO, implementado em linguagem C pelos respectivos autores do algoritmo. O equipamento utilizado foi um microcomputador equipado com processador *dual core* 1,83 Mhz, 2 Gb de memória RAM e sistema operacional Windows XP®.

Para exemplificar o planejamento pela proposta descrita é dada a seguinte situação hipotética, com duas bases de dados distintas, para a construção da rede:

Elaborar um projeto para construir uma rede que deverá atender os clientes da região informada os quais são identificados por suas coordenadas geográficas. E para possíveis locais de instalação dos equipamentos de telecomunicações são fornecidas as coordenadas dos postes de distribuição de energia elétrica.

1a.) Com 790 pontos de clientes e 1697 postes na região, totalizando 2487 nós para o problema com 2603 possibilidades de ligação entre eles (arestas).

2a.) Com 104 pontos de clientes e 581 postes na região, totalizando 685 nós para o problema com 796 possibilidades de ligação entre eles (arestas).

Como exemplo observa-se que a solução parcial da Figura 1(a), obtida com os cálculos pela distância *Euclidiana,* não é real, pois as ligações entre os nós não considera a geografia da região. Mas após a segunda etapa é apresentada na Figura 1(b) a solução final com o melhor caminho entre os nós considerando a geografia.

Os resultados são apresentados nas Tabela 1 e Tabela 2, e para cada algoritmo proposto há na coluna 'Solução' o custo encontrado, na coluna 'Iteração' a respectiva iteração em que foi selecionado e a coluna '*Dijkstra*' apresenta as soluções após o processamento da segunda etapa. Observa-se que o custo com o algoritmo *Dijkstra* é maior que o realizado pela distância *Euclidiana*, pois considera a geografia da região.

Tabela 1.Resultados dos dados da base 1.

	A G		Dijkstra	P S O		Dijkstra
	Solução	Iteração		Solução	Iteração	
1	**45.541**	**8.926**	**56.393**	**45.392**	**1.424**	**55.292**
2	**45.425**	**5.154**	**55.495**	**45.515**	**1.278**	**55.556**
3	**45.717**	**9.411**	**56.614**	**45.616**	**1.456**	**56.100**
4	**45.959**	**7.619**	**55.901**	**45.329**	**1.322**	**56.504**
5	**45.583**	**3.893**	**56.061**	**45.473**	**1.411**	**55.008**
6	**45.495**	**6.933**	**55.000**	**45.641**	**1.333**	**55.476**
7	**45.716**	**9.509**	**55.078**	**45.513**	**1.314**	**54.442**
8	**45.595**	**8.097**	**55.531**	**45.533**	**1.372**	**56.022**
9	**45.378**	**8.265**	**54.983**	**45.798**	**1.539**	**55.960**
10	**45.613**	**9.415**	**55.825**	**45.497**	**1.373**	**55.070**

Tabela 2.Resultados dos dados da base 2.

	A G		Dijkstra	P S O		Dijkstra
	Solução	Iteração		Solução	Iteração	
1	10.668	9.909	13.691	10.934	890	14.185
2	10.600	7.162	13.336	10.614	560	13.376
3	10.518	9.560	**13.154**	10.657	702	13.592
4	10.699	8.986	13.377	10.783	955	13.869
5	10.621	9.216	13.811	10.795	718	14.636
6	10.565	6.834	14.178	10.741	734	**13.306**
7	10.597	8.262	14.639	10.640	941	14.024
8	10.570	9.423	13.550	10.805	602	13.788
9	10.521	9.868	13.386	**10.585**	782	13.914
10	**10.448**	9.040	13.835	10.755	780	15.017

*A unidade da solução depende da unidade utilizada na função objetivo (1), se é utilada em valores monetários ou em metros.

Para executar os mesmos projetos propostos nos testes acima foi entrevistado um projetista que considerou em média uma semana de trabalho. Mas com a utilização da presente proposta é possível executar em um dia de trabalho. Sendo os valores das soluções obtidas manualmente similares aos encontrados pela metodologia proposta.

7. Conclusão

Este trabalho propõe um diferencial no desenvolvimento de projetos de redes utilizando ferramentas conhecidas na literatura e assim adicionar inteligência ao processo de planejamento. Com isto obtêm-se um custo menor de mão de obra especializada e proporciona um aumento de qualidade ao projeto.

Os custos encontrados pelos algoritmos de otimização foram semelhantes, apenas divergindo no número de iterações necessárias para convergir na melhor solução, concluindo que ambos os algoritmos são indicados como ferramentas para o planejamento de redes com esta abordagem proposta.

A utilização das ferramentas apresentadas com o processamento dos dados em duas etapas resultou em soluções factíveis de serem construídas, indicando não só os nós envolvidos, mas a sua ligação através dos cabos entre os postes de energia elétrica respeitando a geografia da região.

Como próximo trabalho pode ser melhorado o custo computacional com a implementação de "janelas de atuação", em que o processamento de seleção de um nó de acesso seria restrito a uma região reduzida e próxima a ele, semelhante ao apresentado por Navarro e Rudnick[13] [14] para redes de energia elétrica. Outra proposta de estudo futuro é considerar a capacidade dos equipamentos, a flexibilidade desejada para a rede e a dependência entre as redes primárias e redes secundárias e com isto otimizar simultaneamente a rede primária e secundária.

Agradecimentos

Agradecimento à Companhia Paranaense de Energia (COPEL), aos seus profissionais, em especial o Eng. Alceu Aroldo Schwingel.

Referências Bibliográficas

[1] SILVA, Hideson A., RODRIGUES, Luiz Carlos de A. **Estudo da Otimização de Redes Baseada no Problema de Steiner.** Curitiba: UTFPr, 2004. 172p. Dissertação - Pós-Graduação em Engenhara Elétrica, UTFPr, Curitiba, 2004.

[2] SOUZA, Alexandre A. A. de, NEVES Jr, Flávio, LOPES, Heitor S. **Sistema de avaliação da rede secundária de distribuição utilizando Algoritmos Genéticos.** Espaço Energia, Curitiba-PR, nro 5, p. 34-41. Outubro/2006.

[3] HUY, Nguyen V. NGHIA, Nguyen D. **Solving Graphical Steiner Tree Problem Using Parallel Genetic Algorithm.** IEEE International Conference on Research, Innovation and Vision for the Future, 2008. p. 29-35.

[4] BEASLEY, J. E. OR-LIBRARY: distributing test problems, URL: http://people.brunel.ac.uk/~mastjjb/jeb/info.html, acessado em jan/2010.

[5] DING,Shan. ISHII,Naohiro. **An Online Genetic Algorithm for Dynamic Steiner Tree Problem.** IECON-Industrial Electronics Society,2000. p.812-817 vol.2.

[6] ZHONG, Wen-Liang. HUANG, Jian. ZHANG, Jun. **A Novel Particle Swarm Optimization for the Steiner Tree Problem in Graphs.** IEEE-Congress on Evolutionary Computation, 2008. p. 2460-2467.

[7] ZHAN, Zhi-Hui. ZHANG, Jun. LI, Yun. CHUNG, Henry S. **Adaptative Particle Swarm Optimization.** IEEE Transactions on Systems, Man and Cybernetics. Vol. 39, no. 6, pp. 1362-1381, 2009.

[8] RIBEIRO, Celso C. SOUZA, Maurício C. de. **Tabu Search for the Steiner Problem in Graphs.** Networks 36(2): p.138-146, 2000.

[9] XU, Jiefeng. CHIU, Steve Y., GLOVER, Fred. **Tabu Search Heuristics for Designing a Steiner Tree Based Digital Line Network.** Universidade do Colorado, Colorado, Waltham. Relatório Técnico. 35p. 1995.

[10] GOLDBERG, David. **Genetic Algorithms in Search, Optimization and Machine Learning.** Addison-Wesley, 1989.

[11] KENNEDY, J. EBERHART, R. **Particle swarm optimization.** Procs of the International Conference on Neural Networks, pp. 1942-1948, 1995.

[12] CORMEN, Thomas H., LEISERSON, Charles E., RIVEST, Ronald L., STEIN, Clifford. **Algoritmos: teoria e prática.** Tradução da segunda edição. Rio de Janeiro: Campus, 2002. pp. 470-474.

[13] NAVARRO, Alejandro, RUDNICK, Hugh. **Large-Scale Distribution Planning – Part I: Simultaneous Network and Transformer Optimization.** IEEE Transactions on Power Systems. Vol. 24, no. 2, pp. 741-751, 2009.

[14] ______. **Large-Scale Distribution Planning – Part II: Macro-Optimization With Voronoi´s Diagram And Tabu Search.** IEEE Transactions on Power Systems. Vol. 24, no. 2, pp. 752-758, 2009.

Sistemas de Geoinformação: Tendências em Tecnologias para o Acesso e Interação de Tomadores de Decisão

Carlos Alberto de Carvalho[1], Maria Cecília Calani Baranauskas[2], Evaristo Eduardo de Miranda[1]

[1]Embrapa Monitoramento por Satélite - Av. Soldado Passarinho, 303
Fazenda Chapadão – 13070-115 – Campinas– SP – Brasil

[2]Instituto de Computação – Universidade Estadual de Campinas (UNICAMP)
Caixa Postal 6176 –13.084-971– Campinas– SP – Brasil

{calberto,mir}@cnpm.embrapa.br, cecilia@ic.unicamp.br

Abstract. *This article presents a set of geographical data access technologies and interaction devices used in user interface and geographic information systems and its use in the redesign of a existing prototype of a WebGIS directed to support decision making in the infrastructure of sugar cane agroenergy.*

Resumo. *Este artigo apresenta um conjunto de tecnologias de acesso a dados geográficos e de dispositivos de interação utilizados em interface de usuário e sistemas de informações geográficas (SIG) e sua aplicabilidade no redesign de um protótipo existente de uma aplicação WebGIS voltada para o suporte a tomada de decisão na infraestrutura da agroenergia da cana-de-açúcar.*

1. Introdução

Existe um projeto conjunto da Embrapa Monitoramento por Satélite e do Gabinete de Segurança Institucional da Presidência da República (GSI/PR) que trata sobre o monitoramento da infraestrutura crítica da agroenergia, no setor do etanol da cana-de-açúcar. A energia proveniente de derivados da cana-de-açúcar, como etanol e cogeração de energia elétrica, já supera aquela produzida pelas hidrelétricas na matriz energética brasileira. Em 2009, as hidrelétricas contribuíram na matriz com 15,3% e os derivados da cana-de-açúcar com 18,1% [EPE 2010]. A agroenergia da cana-de-açúcar tem uma forte componente de territorialidade e dinâmica temporal, pois as áreas de plantio estão em crescente expansão [MAPA 2010]. Além disso, o etanol depende de uma série de infraestruturas de transporte e de armazenamento para ser levado desde as usinas de cana-de-açúcar até os postos de combustíveis. Dessa forma, a informação espacial é um elemento importante no suporte à tomada de decisão em diversas questões da agroenergia.

Como uma forma de visualização dessa informação espacial foi desenvolvido um protótipo inicial de WebGIS com as seguintes características: a) uma interface de usuário para ser utilizada em navegador web em *desktop*; b) disponibilização de funções de geoprocessamento como geração de entornos (*buffers*), relatórios e cálculos de relações espaciais (usinas a uma dada distância de rodovias); c) a disponibilização dos dados geográficos utiliza os padrões específicos do servidor de mapas [Carvalho &

Miranda 2009], sem a capacidade de compartilhar ou inserir planos de informação geográfica de outras bases de dados.

Os tomadores de decisão na infraestrutura crítica da agroenergia são especialistas nas áreas de cana-de-açúcar, produção de etanol, segurança de infraestruturas, agricultura, economia e planejamento, e não necessariamente especialistas em geoprocessamento. Eles utilizam conhecimentos prévios em suas respectivas áreas de atuação em conjunto com as informações geográficas disponibilizadas neste protótipo inicial para suas análises e planejamentos. Porém, eles apresentaram dificuldades em utilizar o WebGIS e suas ferramentas disponíveis (cálculos de distâncias, entornos e relatórios), comuns para um usuário especialista em geoprocessamento.

O objetivo deste artigo é apresentar um *redesign* de um protótipo inicial de WebGIS, a partir do levantamento de tecnologias de acesso a dados geográficos e de dispositivos de interação utilizados em interface de usuário e sistemas de informações geográficas (SIG), para facilitar a sua utilização por um usuário tomador de decisão que não é necessariamente especialista em geoprocessamento.

2. Tecnologias para Acesso e Interação

A partir de uma revisão da literatura sobre interfaces de usuário e sistemas de informações geográficas, foi levantado um conjunto de tecnologias de acesso a informações geográficas e dispositivos de interação. Entre as tecnologias de acesso estão: a integração de bases de dados espaciais utilizando padrões abertos de interconexão OGC, a democratização da informação geográfica proporcionada por *geo-mashups* [Boulos et. al. 2008] e os globos virtuais como o Google Earth [Schöning et. al. 2008]. Entre os dispositivos de interação estão as telas grandes utilizadas para geocolaboração [MacEachren et. al. 2003], os monitores 3D e as tecnologias hápticas [Faeth et. al. 2008], as telas *multi-touch* [Buxton 2007] e os mapas em dispositivos móveis [Sankar and Bouchard 2009].

Outras tecnologias e dispositivos apresentam um futuro promissor, mas dependem de *hardware* específico e da queda dos custos para se tornarem opções viáveis, tais como imersão na geoinformação através de mundos virtuais [Boulos and Burden 2007], estereoscopia 3D [Boulos and Robinson 2009], dispositivos de *feedback* [Faeth et. al. 2008], múltiplas interações [Schöning et. al. 2009] e holografia com toque [Iwamoto et. al. 2008].

3. *Redesign* do protótipo

Para efetuar o *redesign* do protótipo inicial, estas tecnologias e dispositivos foram avaliados para definir sua viabilidade de implementação, baseada nos critérios: a) diferentes formas de interação que estimulem a percepção sobre a informação geográfica; b) a capacidade de conexão com diferentes bases de dados espaciais; c) custo; d) disponibilidade de equipamentos. A Tabela 1 mostra a viabilidade de aplicação ao protótipo inicial:

Tabela 1. aplicabilidade ao protótipo

Tecnologias/Dispositivo	Características principais	Aplicabilidade ao protótipo
Feedback e Visualização 3D [Faeth et. al. 2008]	Imersão do usuário na informação geográfica através da utilização de hardwares específicos.	Não foi considerada por que, apesar de custos em queda do hardware, tomada isoladamente é uma tecnologia mais voltada para o especialista em SIG.
Globos Virtuais (Google Earth) [Schöning et. al. 2008]	**Possibilidade de importar a informação geográfica do mapa plano para o formato KML**	**É aplicável pelo seu baixo custo e pelo grande número de usuários que conhecem e operam esta aplicação.**
Interfaces *multitouch* [Buxton 2007]	Múltiplos cliques em uma tela, permitindo interação mais natural.	Apesar do crescente uso de equipamentos que suportam *multitouch*, ainda existe o desafio para as aplicações de SIG diferenciarem cliques de pessoas diferentes.
Múltiplas interações [Schöning et. al. 2009]	*Hardware* específico para utilizar pés e mãos na navegação da geoinformação	Ainda são propostas em desenvolvimento.
Estereoscopia [Boulos and Robinson 2009]	Imersão do usuário na visualização 3D.	É uma área mais consolidada para os especialistas em SIG e depende de óculos e monitores especiais.
Holografia tátil [Iwamoto et. al. 2008]	Imersão do usuário na manipulação de elementos, adicionando componentes de percepção tátil.	Uma tecnologia promissora para o futuro, mas ainda em estágio de pesquisas e desenvolvimento.
Visualização [Rhyne and MacEachren 2004]	**Formas diversas de visualizar a mesma geoinformação, através de mapas, gráficos e tabelas.**	**É aplicável por ser consolidada no campo da geoinformação e seu baixo custo, permitindo percepções diferenciadas sobre a informação geográfica.**
Web 2.0 e *mashups* [Boulos et. al. 2008]	**Utilização de padrões abertos de interconexão de geoinfomação (OGC WMS/WFS) e construção de *geo-mashups***	**É aplicável pelo seu baixo custo, capacidade de utilizar mais de uma interface de usuário (Por exemplo: Google Earth e OpenLayers), capacidade de conexão com bases de dados remotas e gratuitas.**
Geocolaboração [MacEachren et. al. 2003]	**Uso de telas grandes para trabalho colaborativo.**	**É aplicável, pois pode utilizar lousas digitais *singletouch* no lugar de telas grandes *multitouch*.**
Mundos virtuais [Boulos and Burden 2007]	Imersão total do usuário em um mundo 3D de informação geográfica.	Tecnologia promissora para abordar a imersão na informação geográfica, mas ainda está em estágio de propostas e desenvolvimento.
Mobilidade [Sankar and Bouchard 2009]	Uso de dispositivos móveis com capacidade de navegar em mapas e perceber sua própria localização geográfica através de GPS embutidos.	É aplicável pelo custo em queda de dispositivos móveis com suporte a GPS, mas ainda enfrenta dificuldades de acomodar a informação geográfica em telas pequenas.

Um dispositivo disponível no projeto da infraestrutura crítica da agroenergia é uma lousa digital interativa *Hitachi Starboard FX-duo-77*, operada por uma caneta apontadora e por toque, onde um *datashow* projeta a tela do computador sobre uma lousa com sensores que são utilizados para movimentar o ponteiro do mouse. Este dispositivo foi utilizado como uma tela grande para o *redesign* do protótipo.

Após a seleção das tecnologias e dispositivos aplicáveis, um segundo protótipo foi construído. O servidor de mapas foi adaptado para gerar conexões WMS/WFS do padrão OGC a partir dos dados espaciais e gerar os planos de informação geográfica (rodovias, ferrovias, usinas), o que permitiu a utilização de uma interface de usuário para WebGIS conhecida como *OpenLayers* para implementar a versão dos mapas em 2D. Estes planos de informação também foram exportados para o globo virtual utilizando um conjunto de scripts na linguagem PHP que convertem o formato gerado na conexão WMS/WFS para o formato KML.

O segundo protótipo foi construído para ser utilizado tanto na tela grande da lousa digital quanto no *desktop*, utilizando as versões de mapas em 2D da interface *OpenLayers* quanto o globo virtual *Google Earth*.

A figura 1 ilustra este *redesign*:

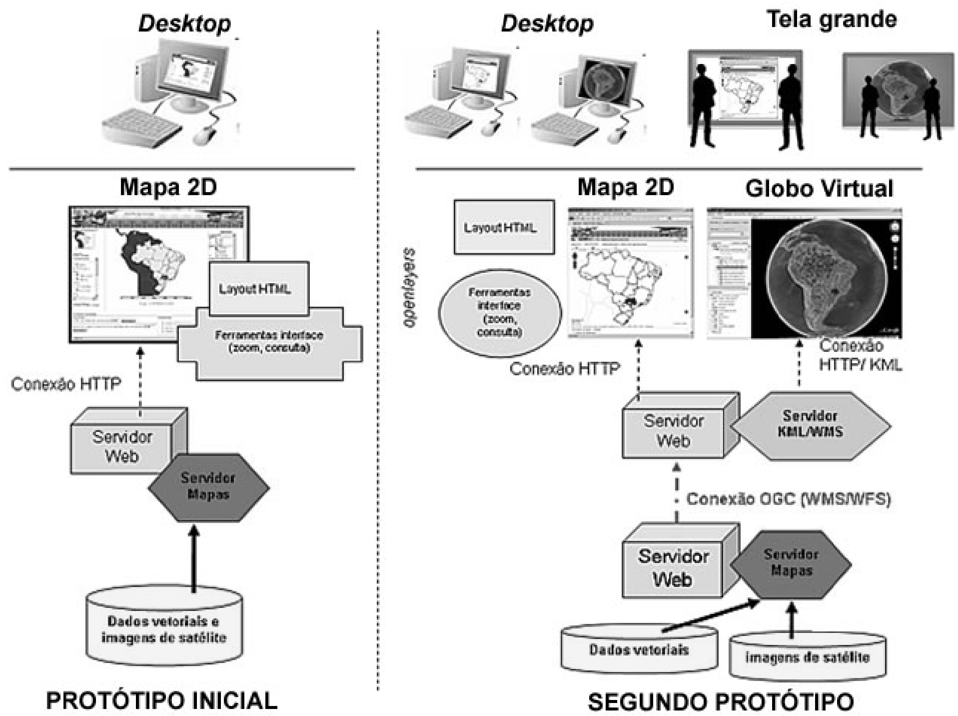


Figura 1. *Redesign* do protótipo

Este segundo protótipo foi submetido a um teste com sete usuários tomadores de decisão da infraestrutura da agroenergia. Em um cenário simplificado e fictício, a produção de etanol de uma região A deveria ser levada por ferrovia ou rodovia até um depósito na cidade B, mas as rodovias X e Y estavam impedidas. O usuário deveria escolher uma rota alternativa. Esta tarefa foi executada em 4 combinações de tecnologias e dispositivos (*desktop* e mapa plano, *desktop* e globo virtual, tela grande e mapa plano, tela grande e globo virtual). O objetivo foi observar se estas diferentes combinações facilitaram a percepção sobre os elementos do mapa e globo virtual, dando suporte à tomada de decisão.

Os testes foram filmados e os comentários dos usuários gravados. Em uma análise preliminar desse material são destacados alguns itens: a) todos os usuários conseguiram cumprir a tarefa nas 4 combinações; b) todos os usuários tiveram dificuldades de interagir com a tela grande. Ainda assim, a maioria dos usuários relatou que apesar da dificuldade, esta forma de interação é interessante e estimula a análise da informação geográfica; c) entre as dificuldades encontradas na interação com a tela grande estão a execução do equivalente do "duplo clique", que é um "duplo toque" com a caneta apontadora ou com o dedo. O uso do dedo ocorreu para ações mais precisas, como selecionar uma *checkbox*. Porém, como a imagem é projetada sobre a tela grande, ao utilizar o dedo para apontar, o braço causa sombra sobre a tela, ocultando parte do mapa. Com o uso da caneta apontadora este efeito é minimizado e aumenta o alcance sobre a tela grande; d) Nas combinações *desktop/*mapa plano e *desktop/*globo virtual não foram encontradas dificuldades de interação, concentrando-se as dúvidas e sugestões dos usuários nas simbologias, legendas e inclusão de planos de informação.

4. Conclusão

Este trabalho apresentou uma proposta de efetuar o *redesign* de um protótipo de WebGIS de apoio a tomada de decisão para usuários que não são necessariamente especialistas em geoprocessamento, a partir de um levantamento de um conjunto de tecnologias de acesso e de dispositivos de interação. Os resultados preliminares dos testes com os usuários sugerem que a disponibilização de combinações diferentes de tecnologias e dispositivos facilitou a execução de uma tarefa de tomada de decisão. As dificuldades de utilização da tela grande justificam-se devido a ser uma forma de interação nova para os usuários e por não ter sido feito um treinamento inicial de manipulação desse dispositivo. O material gravado dos testes necessita ainda de uma análise mais detalhada para resgatar as sugestões dos usuários a respeito de simbologias, tipos de ferramentas de zoom e legendas a serem utilizadas nos mapas e a inclusão de novos planos de informação.

6. Agradecimentos

O trabalho descrito neste artigo foi financiado pela Embrapa, dentro de seu programa de pós-graduação e o seu desenvolvimento foi realizado com o apoio da Embrapa Monitoramento por Satélite, do Escritório de Análise e Monitoramento de Imagens de Satélite da Secretaria de Acompanhamento e Estudos Institucionais do GSI-PR e do Instituto de Computação da Universidade Estadual de Campinas.

Referências

Boulos, M., Robinson, L. (2009). "Web GIS in practice VII: stereoscopic 3-D solutions for online maps and virtual globes". International Journal of Health Geographics, 2009. http://www.ij-healthgeographics.com/content/8/1/59, fev. 2010.

Boulos, M.; Scotch, M.; Cheung, K.; Burden, D. (2008). "Web GIS in practice VI: a demo playlist of geo-mashups for public health neogeographers". International Journal of Health Geographics, http://www.ij-healthgeographics.com/content/7/1/38>. Acesso em: 12 fev. 2010.

Boulos, M. and Burden, D. (2007). “Web GIS in practice V: 3-D interactive and real-time mapping in Second Life”. International Journal of Health Geographics, http://www.ij-healthgeographics.com/content/6/1/51, fev. 2010.

Buxton, B. (2007). “Multi-Touch Systems that I Have Known and Loved”. Microsoft Research, http://www.billbuxton.com/multitouchOverview.html, mar. 2010.

Carvalho, C. A. and Miranda, E. E. de. (2009). “Um sistema de gestão de informações geográficas em WebGIS para o controle do monitoramento por satélite das obras do PAC”. Campinas, SP: Embrapa Monitoramento por Satélite, 43 p.

EPE. (2010). “Balanço Energético Nacional 2010 - Resultados Preliminares - ano base 2009”. Empresa de Pesquisa Energética - Ministério de Minas e Energia. https://ben.epe.gov.br/downloads/Resultados_Pre_BEN_2010.pdf, jul. 2010.

Faeth, A., Oren, M., and Harding, C. (2008). “Combining 3-D geovisualization with force feedback driven user interaction”, In: Proceedings of the 16th ACM SIGSPATIAL international Conference on Advances in Geographic information Systems (Irvine, California, November 05 - 07, 2008). GIS '08. ACM, New York, NY, 1-9.

Iwamoto T, Tatezono M, Hoshi T, Shinoda H. (2008). Airborne Ultrasound Tactile Display. ACM SIGGRAPH 2008: New Tech Demos, Proceedings of the International Conference on Computer Graphics and Interactive Techniques: 11-15 August 2008; Los Angeles, California.

MacEachren, A. M., Brewer, I., Cai, G., & Chen, J. (2003). “Visually-Enabled Geocollaboration to Support Data Exploration and Decision-Making”. International Cartographic Conference, Durban, South Africa, pp. 394-401.

MAPA. (2010). “Plano Nacional de Agroenergia 2006-2011 / Ministério da Agricultura, Pecuária e Abastecimento, Secretaria de Produção e Agroenergia”. 2. ed. rev. - Brasília, DF : Embrapa Informação Tecnológica, 2006. 110 p.

Rhyne, T. M., MacEachren, A. (2004). “Visualizing geospatial data”. In ACM SIGGRAPH 2004 Course Notes (Los Angeles, CA, August 08 - 12, 2004). SIGGRAPH '04. ACM, New York, NY, 31.

Sankar, K. and Bouchard, S. A. (2009). “Chapter 9: Web 2.0 and Mobility. Enterprise Web 2.0”, Cisco Press, http://www.networkworld.com/subnets/cisco/062609-ch9-web2-mobility.html, mar. 2010.

Schöning, J., Daiber, F., Krüger, A., and Rohs, M. (2009). “Using hands and feet to navigate and manipulate spatial data”. In: Proceedings of the 27th international Conference Extended Abstracts on Human Factors in Computing Systems (Boston, MA, USA, April 04 - 09, 2009). CHI EA '09. ACM, New York, NY, 4663-4668.

Schöning, J., Hecht, B., Raubal, M., Krüger, A., Marsh, M., and Rohs, M. (2008). “Improving interaction with virtual globes through spatial thinking: helping users ask 'why?'”. In Proceedings of the 13th international Conference on intelligent User interfaces (Gran Canaria, Spain, January 13 - 16, 2008). IUI '08. ACM, New York, NY, 129-138.

Index of authors