Multidisciplinary Research on Geographical Information in Europe and Beyond
Proceedings of the AGILE'2012 International Conference on Geographic Information Science, Avignon, April, 24-27, 2012
ISBN: 978-90-816960-0-5
Editors: Jérôme Gensel, Didier Josselin and Danny Vandenbroucke

Extending TimeML and SpatialML languages to handle imperfect spatio-temporal information in the context of natural hazards studies

Mouna Snoussi
Laboratory of Informatics of Grenoble,
University Joseph Fourier,
St Martin d'Hères, France
mouna.snoussi@imag.fr

Jérôme Gensel
Laboratory of Informatics of Grenoble,
University Joseph Fourier,
St Martin d'Hères, France
jerome.gensel@imag.fr

Paule-Annick Davoine
Laboratory of Informatics of Grenoble,
University Joseph Fourier,
St Martin d'Hères, France
paule-annick.davoine@imag.fr

Abstract

Natural hazards analysis uses data sets from different sources with different formats (video, audio, books, etc.). The content of these data sets is mainly expressed using natural language describing where and when natural hazards events (eruption, avalanches, flood, etc.) took place. However, natural language contains a lot of vague and imprecise expressions and especially those used to indicate places (around, near to, north of, etc.) and dates (between 1710 and 1711, at the beginning of the century, etc.). For a better exploitation and exchange of this kind of data sets, a spatial and temporal representation based on markup languages (SpatialML and TimeML) can offer a good interoperability and an easy share of such data. However, none of the existing markup languages handles the representation of imprecise and vague spatial/temporal information. For this purpose, we propose extensions of the SpatialML and TimeML mark-up languages that explicitly integrate the representation of imperfect spatio-temporal information.
Keywords: Imperfection, spatio-temporal information, markup language.

1 Introduction

In the context of natural hazards, data sets are collected in different formats (audio, video, images and text). In order to be able to exploit and analyze such kinds of data sets, a first and important step consists in transforming these data sets that contain information expressed using natural language [1] into an exchangeable and interoperable format. In the context of natural hazards, data sets contain most of the time (almost inherently) both spatial and temporal information that describe events by specifying their date(s), location(s), duration(s), etc.

However, when considering time and space, one can observe that expressions in natural language often rely on terms like: around, near to, between, that lead to vague and imprecise interpretation and understanding. The challenge is then to handle such imperfection by identifying its kind and by measuring its degree, producing metadata about the quality of information that can be very helpful for a better understanding and analysis of spatio-temporal phenomena.

On the other hand, it is now widely acknowledged that existing markup languages based on XML facilitates interoperability in data exchange. Then, they are good candidates to exchange data sets in the field of natural hazards too. While SpatialML [2], Geography Markup Language (GML1), or Keyhole Markup Language (KML2) are the main representatives of markup languages dedicated to spatial information, Translingual Information Detection, Extraction, and Summarization (TIDES), and TimeML3 are those dedicated to temporal information.

However, none of the previous markup languages proposals handles the imprecision and vagueness representation over spatial and temporal information. Even if this kind of metadata is required in different domains that deal with spatio-temporal information (text mining, metadata extraction, geographic information retrieval, GIS, etc.) they still may not be well represented in the case of SpatialML and TimeML languages.

In this work, we propose to extend two current markup languages, SpatialML and TimeML, towards the representation of imprecision and vagueness values and degrees for spatio-temporal information.

The remainder of this paper is organized as follows. Section 2 introduces the imperfection terminology. Section 3 presents our approach. Section 4 describes the extensions we propose for SpatialML and TimeML markup languages for the description of the level of imprecision and vagueness of spatio-temporal information. Section 5 defines metadata quality imperfection indicators. Section 6 presents related work. Section 7 concludes and gives future directions for this work.

2 Imperfection terminology

In [3], a complete taxonomy of imperfect information (see figure 1) is proposed. In the context of natural hazards, vagueness and imprecision are the most frequent types that characterize imperfection of spatio-temporal information.

Below, we recall some general definitions of imprecision and vagueness [4] that we also adopt when referring these kinds of imperfection:

- Imprecision: imprecise information is characterized by a sub-set of values of its definition domain while precise information is characterised by a unique value.
  For example: the eruptive crack of the lava is located between 1000 and 1200 meters height, corresponds to an imprecise information, relatively to a context in

[1] http://www.opengeospatial.org/standards/gml
[2] http://www.opengeospatial.org/standards/kml/
[3] http://www.timeml.org/site/index.html

which the height attribute is defined in the [0, 10000] interval.

- Vagueness: this type of imperfection is similar to imprecision. However, vague information refers to a larger (and possibly infinite) set of values.
  For instance, using "in the north of" to express the location of an eruption, knowing that the location attribute is valued by a pair of coordinates, makes in this case the information about the location "vague".

According to these two definitions, we consider that imprecise information is to be represented by a set (finite or not, ordered or not) of elements, while vague information is to be represented by a fuzzy subset of elements [5] whose content cannot be precisely or is subjectively defined. We can also state that, imprecision is a particular case of vagueness.

Examples follow for each imperfection previously defined in the field of natural hazards:

-Spatial imprecision: The earthquake is located at 39.063°N, 119.739°W +/- 1.9km horizontal; +/- 4.5 km depth.

-Temporal imprecision: The eruption happened the 1st or 2sd of January 1899.

-Spatial vagueness: The last eruption took place on the northern part of the crater.

-Temporal vagueness: The eruption occurred during (around) the 18th century.

As we can see in Figure 2, imperfect information is divided into three main classes: Imprecision, Inconsistency and Uncertainty. Each class is composed by different sub classes according to one criterion. For instance, the imprecision class is divided into two sub-classes, the first sub-class which contains vagueness concerns only data without error and the second sub class is related to data tainted with error.

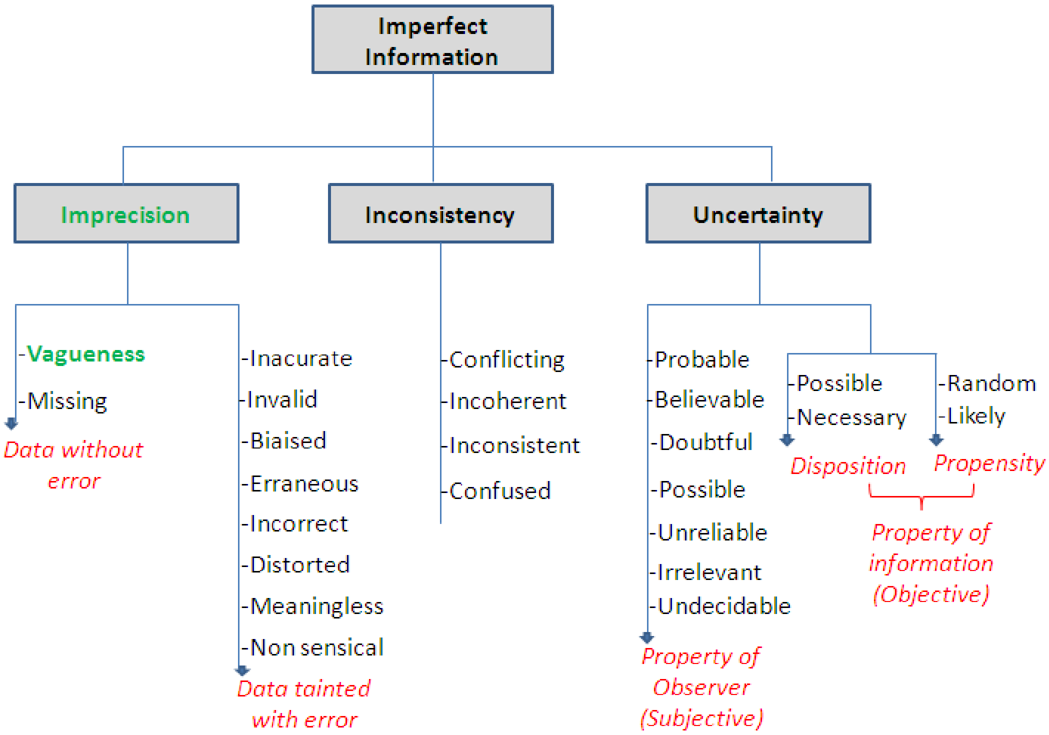
3 Our approach

In the context of natural hazards analysis, experts usually have to deal with large amount of data sets regarding the location and the timing of the events that they have to analyze. Moreover, in order to get as accurate results as possible, the data sets should be as precise as possible or at least should come with metadata describing the imperfection they suffer from. When data are expressed through sentences in natural language, some imprecise terms or expressions are used to describe where and when natural hazards events occur. We here propose to handle imperfect spatio-temporal information (see figure 2), from the acquisition of collections of spatio-temporal information, to the visualization and exchange of such an imperfect spatio-temporal information.

The first step, consist in gathering datasets coming in different formats, then as a second step the entire data sets will be transcribed into textual format. Once the data are presented in textual file, they are parsed towards extended SpatialML and TimeML files containing tags dedicated to the representation of imperfection.

To identify imperfection, the parsers perform a semantic text analysis using a predefined list of natural language terms like: near to, before, after, between, around, etc. Once the imperfections are identified and tagged in the XML file using SpatialML and TimeML tags, a GUI can be generated to visualize the XML file data using spatio-temporal map techniques. To reduce and quantify imperfection already

Figure 1: Taxonomy of imperfect information.



Multidisciplinary Research on Geographical Information in Europe and Beyond
Proceedings of the AGILE'2012 International Conference on Geographic Information Science, Avignon, April, 24-27, 2012
ISBN: 978-90-816960-0-5
Editors: Jérôme Gensel, Didier Josselin and Danny Vandenbroucke

identified by the parser, an expert can interact with the visualization interface to modify the degree of uncertainty of imperfect values. At the end, the expert validates the XML file and publishes the exchangeable data sets.

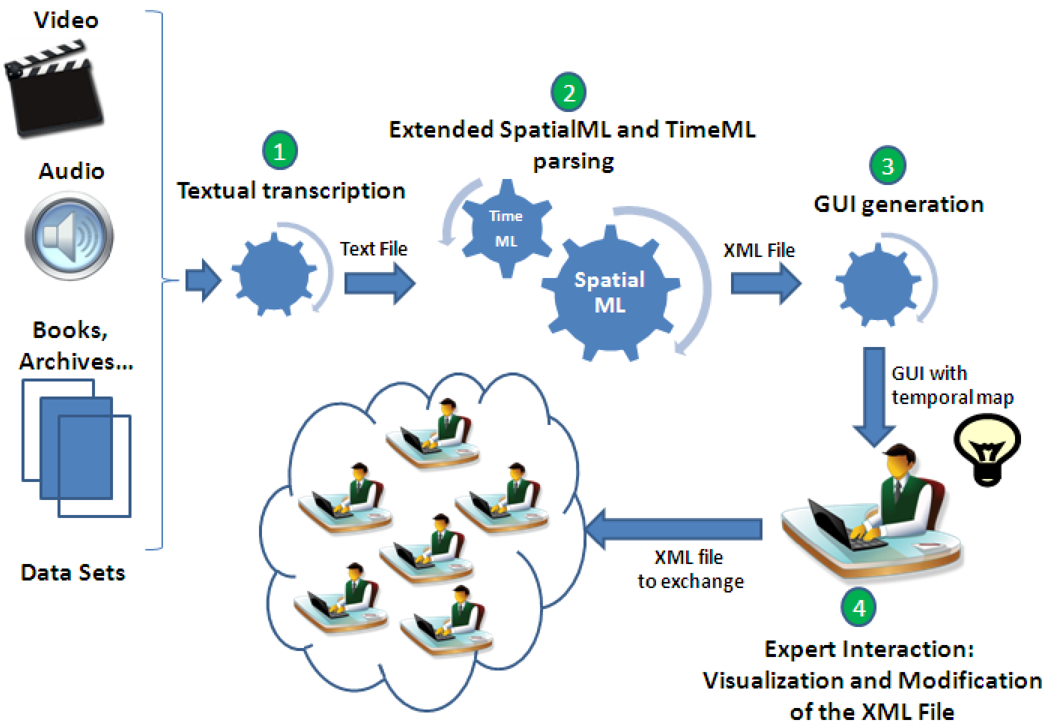
In this paper, we focus on the extensions of SpatialML and TimeML markup languages that allow the representation of vague and imprecise spatial and temporal information. Moreover, we present some examples taken from the natural hazards context and especially from the volcano eruption domain to validate our proposal.

represent the lava's path. The path's source is explicitly described ("*Piton de la Fournaise*") while the end of the path is only described using a direction ("north") and a distance ("1 miles") that implies the creation of an unknown location (<PLACE id=4>) to indicate the end of the path.

In order to improve the representation of imperfection, we propose to help an expert in describing as precisely as possible the imperfection, through a user graphical interface.

For instance, in the example above, the expert could draw on a map some polygons that correspond to possible ends of the path he/she has determined, assigning to each of this polygon a degree of uncertainty.

Figure 2: Our approach



4 Markup language extension

In this section, we present SpatialML and TimeML Markup language extensions to handle imprecision and vagueness imperfection related to spatial and temporal information.

4.1 SpatialML extension

The SpatialML [2] mark-up language proposes numeric representation of places and also defines symbols to express spatial relations between those places based on RCC8 relation. In the example below, we can see how the SpatialML represents spatial information using: the tag <PLACE> for expressing locations, the tag <SIGNAL> for expressing properties of locations, and the tag <RLINK> for expressing existing spatial relations between two locations formerly identified by the <PLACE> tag.

The example on figure 3 shows a concrete case of SpatialML limitation for the representation of a place described using vague expression. In this example the RLINK tag is used to

Figure 3: Example of a SpatialML representation.

The lava flowed 1 miles north from the Piton de la Fournaise

<PLACE id=1 type="PPLC" country="RE" form="NAM">Piton de la Fournaise</PLACE>
<SIGNAL id=2 type="DISTANCE">1 miles</SIGNAL>
<SIGNAL id=3 type="DIRECTION">north</SIGNAL>
<PLACE id=4/>
<RLINK id=5 source=1 target=4 distance=2 direction="N" signals= "2 3">

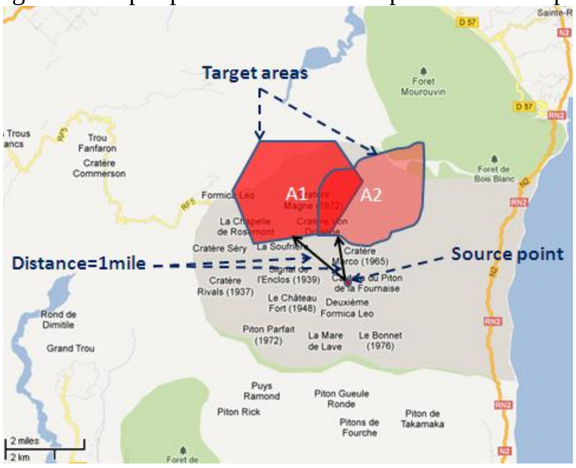
The uncertainty degree representation can be handled by the use of an existing framework, the uncertML framework [6]. It offers probabilistic representation of uncertainty in random quantities. Following the GIS literature, the representation of vague geometries is usually done using the fuzzy sets theory [7,8] that matches with the nature of the geometric vagueness representation. However, the fuzzy set theory has not been handled by the UncertML framework. Moreover, the

Multidisciplinary Research on Geographical Information in Europe and Beyond
Proceedings of the AGILE'2012 International Conference on Geographic Information Science, Avignon, April, 24-27, 2012
ISBN: 978-90-816960-0-5
Editors: Jérôme Gensel, Didier Josselin and Danny Vandenbroucke

definition of the fuzzy set membership function is still a challenge. Therefore, we choose a more simple way to represent the uncertainty during the assignment of geometries by the expert. The uncertainty assignment we use is based on an interval [0..1], where 0 and 1 represent respectively the lowest and highest uncertainty degrees. It can also be more intuitive and easily understandable by other users.

On figure 4 below, the expert has specified two possible areas that represent the areas where the path of lava ends, as first described in figure 3. The areas selected by the expert correspond to imperfect values that can be assigned to the unknown locations to offer more precise information than the textual vague one.

Figure 4: Map representation of the SpatialML example



We propose a new tag named IMPERFECTION that extends the SpatialML markup language to allow the representation of spatial imperfect information (see figure 5 below):

Figure 5: Specification of <IMPERFECTION>.

```
<!ELEMENT IMPERFECTION ( #PCDATA ) >
<!ATTLIST IMPERFECTION id ID #REQUIRED >
<!ATTLIST IMPERFECTION type (VAGUENESS |
IMPRECISION ) #REQUIRED >
<!ATTLIST IMPERFECTION tagtype (PLACE |
RLINK | SIGNAL | LINK) # REQUIRED >
<!ATTLIST IMPERFECTION tagID IDREF #REQUIRED
>
<!ATTLIST IMPERFECTION imperfectionValues
CDATA #REQUIRED >
<!ATTLIST IMPERFECTION
uncertaintyValuesDegrees CDATA #REQUIRED >
<!ATTLIST IMPERFECTION
imperfectionAttributeName CDATA #IMPLIED >
<!ATTLIST IMPERFECTION comment CDATA
#IMPLIED >
```

The imperfection tag we propose has six mandatory attributes:

- `id`: the identifier of this imperfection
- `Type`: the type of imperfection: imprecision, vagueness.
- `tagtype`: the SpatialML element type concerned by this imperfection.
- `tagID`: the SpatialML elements ID handling this imperfection.
- `imperfectionValues`: the set of possible values for the SpatialML elements.
- `uncertaintyValuesDegrees`: the uncertainty degree assigned by the expert. The value lies in the [0..1] interval (0 represents the lower uncertainty degree and 1 the highest).

Using the IMPERFECTION tag, we can now represent the vague imperfection related to the previous example, and we can specify the possible areas representing the final destination of the lava. The selection of the possible imperfection values (geometry of the areas) can be made using a graphical user interface (GUI) composed by a map that shows the different imperfection values (surfaces, line, points). In the previous example, let us consider that the expert has selected two main areas (see figure 4) that represent two possible ends for the lava path, each having a degree of uncertainty. The example of the figure 6, shows the use of the imperfection tag to represent the expert assignment.

Figure 6: SpatialML extension with Imperfection tag.

The lava flowed 1 miles north of the Piton de la Fournaise

```
<PLACE id=1 type=“PPLC” country=“RE”
form=”NAM”>Piton de la Fournaise</PLACE>
<SIGNAL id=2 type=“DISTANCE”>1 miles</SIGNAL>
<SIGNAL id=3 type=“DIRECTION”>north</SIGNAL>
<PLACE id=4/>
<IMPERFECTION id="5" Type= "vagueness"
tagType="PLACE" tagID="4"
ImperfectionValues="Polygon((10,30)(10,25)(30,20) (28,60)
(40,45) (10,45)Polygon((20,40)(30,30)(10,25)(30,20)(28,60)
(30,65)(70,90))" uncertaintyValuesDegrees= "0.2 0.7"/>
```

4.2 TimeML extension

TimeML, is a mark-up language which can be used to annotate temporal expressions and events in textual document. Using TimeML, we can annotate an event and its relative position according to another one using for each annotation a specific tag (for example: <EVENT> to tag an event, <TIMEX3> to tag a temporal expression (time, dates, duration, etc), <TLINK> to express temporal relation among event, etc). Moreover, it is possible to give confidence values to be assigned to any tag and to any attribute of any tag. The confidence value expresses the level of confidence that the metadata provider has in assigning the temporal indication to an event or observation. For example, we can add the confidence annotation to TIMEX3 to indicate the degree of annotation correctness.

In the current TimeML version, the confidence tag (confidence values are defined on [0..1]) can be used as an alternative way to represent uncertainty relative to incomplete or vague knowledge. For example, for a temporal indication of an event either it is represented by a precise value such as “10.50.00 a.m.“, the annotator can assign a confidence value, which is below 1 and express that is a doubt regarding the assessment. In the same way, a vague temporal indication like “after noon” can be annotated with a full confidence

value of 1 if the annotator feel sure about this temporal indication either it is vague.
However, this alternative does not allow the specification of the type of the imperfect information (imprecise or vague). Moreover, only one value can be defined for the vague or the imprecise temporal information and then annotated by the confidence tag to specify its confidence level.
As for SpatialML extension, we propose to introduce the IMPERFECTION tag to represent vague and imprecise temporal information. The IMPERFECTION tag allows us to specify the imperfection type and its possible values within the uncertainty degree of the provider.
To illustrate the use of the imperfection tag for temporal information, we propose to represent the imprecision related to an eruption that might have occurred "between the 12th and the 16th of January 1820":

Figure 7: TimeML extension with Imperfeciton tag.

```
Between
<TIMEX3 tid="t1" type="Date" Value="1820-01-12"> 12 of January 1820</TIMEX3>
<TIMEX3 tid="t2" type="Date" Value="1820-01-16"> 16 of January 1820</TIMEX3>
<IMPERFECTION id="1" Type= "IMPRECISION" tagType="TIMEX3" tagID="t1 t2" ImperfectionValues="1820-01-12 1820-01-13 1820-01-14 1820-01-15 1820-01-16 " uncertaintyValuesDegrees="0.2 0.4 0.6 0.7 0.8"/>
```

As shown in the example above, imprecision is related to a date defined by an interval of possible values {1820-01-12, 1820-01-13, 1820-01-14, 1820-01-15, 1820-01-16}. To each date in the set of imperfection values, the expert has assigned a degree of uncertainty. Here, the less uncertain date is the 12th of January 1820 with the lower uncertainty degree equal to 0.2.

5 Metadata quality: imperfection indicators

The ISO quality standard 19157[4] proposes a quality process to represent the quality of geographic information. However, the use of this kind of standard offers quality metadata, which are difficult to understand due to their complexity and poor semantic. Based on our IMPERFECTION tag we can generate imperfection indicators that can fit with the ISO 19157 standard schema. The imperfection indicators are calculated according to the number of imperfection tags, which is identified for each tagtype (spatial or temporal) and also for each imperfection type (vagueness or imprecision):

- Spatial Vagueness Indicator = Number of IMPERFECTION tag with tagtype=spatial and type=vagueness / total number of spatial tags.
- Spatial Imprecision Indicator = Number of IMPERFECTION tag with tagtype=spatial and type=imprecision / total number of spatial tags.
- Temporal Vagueness Indicator = Number of IMPERFECTION tag with tagtype=temporal and type=vagueness / total number of temporal tags.
- Temporal Imprecision Indicator = Number of IMPERFECTION tag with tagtype=temporal and type=imprecision / total number of temporal tags.

Our metadata quality indicators are simple to calculate and hence offer a significant overview on the spatial and temporal data quality inside the document.

6 Related Work

Imperfection, as defined in [4], includes several concepts such as uncertainty, imprecision, inconsistency, indeterminacy and vagueness. Inconsistency, indeterminacy, imprecision and vagueness are related to information content (its value), while uncertainty is partial knowledge of the true value of the information. Other studies focus on a subset of imperfections [9] and [3], and in particular the spatial uncertainty [10, 11].

In [3], a survey of the various forms of imperfect data is presented. Moreover, the author claims that the most difficult task would be, in this case, the recognition of the imprecision or uncertainty nature encountered in a given data set.

On the other hand, few works on representing imperfect spatial and temporal information were made in the field of textual annotation [12]. [13] propose an extension to allow the representation and management of temporal metadata possibly imperfect, in the context of the INSPIRE directive by adopting TimeML specification language. In this work, the authors present and study the case of temporal queries with imperfect or ill-defined metadata using catalogue services. To represent imperfection, authors opted for the fuzzy set formalism. They also take into account the granularity of the queried data by transforming the user's temporal query granularity into the granularity of the data they have to be matched. However, this proposal does not handle all the aspect of spatio-temporal imperfect data, for example, the combined imperfection that can be propagated from the spatial and temporal information involved in a spatio-temporal relation is still not be envisaged or studied.

In the literature, we find substantial development in the area of temporal information markup language representation (Timex, ISO-TimeML, TimeBank, etc.) and geographic information markup language representation (SpatialML [2], GML, KML, etc.) but there is only few works representing both the imperfect temporal and spatial information [14, 15]

In [14], the authors propose a spatio-temporal framework in XML based on GML and an object oriented integration of temporal models into GIS to describe changes of an object over time. Moreover, they integrate in their temporal structure the notion of determination to represent fuzzy temporal borders. However, the temporal structure that they propose does not represent the temporal relation that can exist among objects neither the complex temporal expression related to the rate of change of an object (for example: frequency).

7 Conclusion and Future Work

In this paper we have proposed extensions to SpatialMl and TimeML markup languages in order to handle vague and imprecise spatial and temporal information. These extensions allow the identification of spatial and temporal imperfections present within a data set by using the IMPERFECTION tag.

[4] http://www.iso.org/iso/fr/catalogue_detail.htm?csnumber=32575

Moreover, we propose an approach aiming to reduce the vagueness and imprecision of spatial and temporal information by integrating expert evaluation and quantification of imperfection within the IMPERFECTION tag representation.

Our research is now directed towards the propagation of the imperfection represented in spatial or temporal attributes to the spatio-temporal relations (Allen's relation [16], RCC8) that link them.

References

[1] A. Peltier, P. Bachèlery, T. Staudacher. Magma transport and storage at Piton de La Fournaise (La Réunion) between 1972 and 2007: A review of geophysical and geochemical data. Journal of Volcanology and Geothermal Research, 93–108, 2009.

[2] MITRE. Spatialml: Annotation scheme for marking spatial expressions in natural language. 2007. online:< http://sourceforge.net/projects/spatialml/>

[3] P. Smets. Imperfect information: Imprecision – Uncertainty. UMIS, 1999.

[4] H. D. Kraft. G. Pasi Disco. G. Bordogna. Vagueness and Uncertainty in Information Retrieval: How can Fuzzy Sets Help?, IWRIDL, India, 2006.

[5] L.A. Zadeh,. Fuzzy sets, Information Control, vol 8, pp.338-353, 1965.

[6] M. W. Williams. Probabilistic uncertainty in an interoperable framework, Thesis, Aston University, 2011.

[7] C. De Runz., E. Desjardin., F. Piantoni,. M. Herbin. Using fuzzy logic to manage uncertain multi-modal data in an archaeological GIS. International Symposium on Spatial Data Quality (ISSDQ), Pays-Bas, Enschede, juin 2007.

[8] P. Fisher,. Sorites paradox and vague geographies. Fuzzy Sets and Systems, 113, p. 7-18, 2000.

[9] E. Zimanyi. Incomplete and uncertain information in relational databases, Université libre de Bruxelles, Thèse. , 1992.

[10] M. Schneider. Uncertainty Management for Spatial Data in Databases: Fuzzy Spatial Data Types, SSD'99, p. 330-351, 1999.

[11] J. Zhang, M. Goodchild. Uncertainty Geographical Information, Taylor & Francis, 2002.

[12] J. Pustejovsky and J. L. Moszkowicz. Integrating Motion Predicate Classes with Spatial and Temporal Annotations. COLING (Posters), p. 95-98, 2008.

[13] G. Bordogna, P. Carrara, M. Pagani, M. Pepe, A. Rampini. Managing imperfect temporal metadata in the Catalog services of Spatial Data Infrastructures compliant with INSPIRE, IFSA-EUSFLAT, 2009.

[14] A. Zipf, S. Kruger. TGML-Extending GML by Temporal Constructs- A Proposal for a Spatiotemporal Framework in XML, ACM-GIS, USA, 2001.

[15] M. Guerrero Nieto, M. J.Garcia Rodriguez, U. A. Zambrana, W. Siabato, M. Bernabé Poveda, Incorporating TimeML into a GIS, IJCLA, pp. 269-283, 2010.

[16] J. F. Allen, Maintaining Knowledge about Temporal Intervals. Communication of the ACM. P. 832-834, 1983.

ISBN: 978-90-816960-0-5
Editors: Jérôme Gensel, Didier Josselin and Danny Vandenbroucke