Basic Notions and Rationale of the Handling of Imperfect Information in Spatio-Temporal Databases

Rita de Caluwe[1], Guy de Tré[1], and Gloria Bordogna[2]

[1] Department of Telecommunications and Information Processing
Ghent University
Sint-Pietersnieuwstraat 41, B-9000 Gent (Belgium)
{rita.decaluwe,guy.detre}@ugent.be
[2] Institute for Environmental Process Dynamic (IDPA)
Dep. of Milan – Lab. of Georesources, National Research Council (CNR)
via Pasubio 5, 24044 – Dalmine (BG), Italy
gloria.bordogna@idpa.cnr.it

Basic Notions

Space and time, both are familiar notions, although they strangely do not let capture themselves in strict definitions. Everything people do, is experienced in space and time. Space and time are, therefore, the logical prerequisites of all human experience. As such they are used as a framework for querying and reasoning. The perspectives in this book are limited to querying and reasoning about information kept in databases and that carries time and space aspects. The underlying time and space models themselves are only presented in as far as they are needed for that purpose. In the same way as models able to deal with imprecision and uncertainty can be conceived, these notions can find their way in the querying and reasoning formalisms too, whether applied to traditional database models or to the more sophisticated database models.

With respect to databases and how to find out and use the information contained in them, it makes sense to distinguish clearly between the data model and the database model, strata on top of which querying and reasoning can be built. Database systems exist by the virtue of database models and thus of the underlying data models; it has to be seen as a term with the flavour of computer applications, hence coloured by implementation aspects and offerings on the software market. To come to a more comprehensive meaning of flexible querying and reasoning as dealt with in this book, it can be helpful to first have a closer look to the meaning of data, data model, database model and database system.

Data are factual information used as a basis for reasoning, discussion and computation. The are what is known in the real world, facts and figures carrying semantics, which is in the end most of the time represented in a very clumsy way in a computer environment. It is noticeable that data can be related among them.

A **database** is any collection of data that is specially structured and organised for rapid search and retrieval by a computer. In a database data are represented according to a data model and is related one another according to a database model.

A **data model** is defined by a **data type**, being a set of legal values, together with a (sufficient) number of (a non-conflicting set of) assorted operators on these values. In traditional systems, only a very limited number of basic types is made available, but sometimes the systems are granted type generators, which converts them into open ended type systems. The most common examples of data types are of the primitive number types integer, real, character type, ..., which can be arranged in arrays, list structures and so on, sometimes in more complicated structures turned into new types by using type generators; appropriate operations are assorted to them as arithmetic operations on numbers. In general, very few of the real world semantics can be formally captured in the models, but database models and database systems built on this kind of strata have been used for years in the absence of more sophisticated data models. As far as there is no need for more powerful systems, there is nothing wrong with using these models: their availability simply is a blessing. However, many times there is a real lack of expressing imprecise and uncertain data, especially of data that only can be described in linguistic terms supported by common understanding; the provided data models are not of sufficient, support, for instance to qualify an item as tall or ephemeral or quick or to indicate the possibility of one event or another, happened or to happen. Even when one is able to cope in an acceptable way with failing semantics at the level of the data model, it can be very helpful to be offered the possibility to leave a number of details to the system, thanks to the use of underpinning more advanced models. A number of such models have been proposed in literature. These models, known in literature as fuzzy models and uncertainty models, rely in essence on fuzzy set theory, on probability theory and on possibility theory and can be easily implemented in the context of open type systems. In general they are seen as generalizations of traditional models, sometimes as extensions of them. A very straightforward way of dealing with them, is to see them as the basic models of which the traditional ones are special cases, for particular parameter values.

The notion of **database model** defined on top of a data model takes into account and incorporates into the model the expression of the relationships between the data, network, hierarchical, relational and object database systems,

being the best known at present. In up to date settings, these relationships are interpreted as constraints, which more and more play a role in database theory.

Databases are therefore instantiations of database models. **Spatio-temporal databases** are databases in which space and time are types of data, at least conceptually, implementation issues sometimes preventing their direct instantiation as basic types.

The spatio-temporal database application field is wide and **GIS (Geographical Information Systems)** is an important part of it. In this context the management of imprecise and /or uncertain spatio-temporal information is particularly important since any observation of geographic phenomena is affected more or less by some kind of imperfection. Many of the contributions in this book are directly centred on GIS, some others are more general or focus on related application fields, presenting theoretical viewpoints and techniques that are inspiring or can be used mutatis mutandis for GIS.

Flexible querying tries to enhance querying expressiveness as offered in traditional database systems in many different ways, always aiming thereby to ease the extraction of data and information from the database relevant to the user. It can be applied to databases based on traditional data models, as well as to databases based on other, more advanced data models, such as the ones offering the possibility to deal with fuzzy and uncertain data. In this book the focus is on flexible querying related to the spatial and temporal components of the information that is represented in a database. Spatio-temporal querying implies the specification of spatial and/or temporal constraints on the entities. The logics used to build the expressions that make up the core of the query can be the traditional binary logic, a ternary logic —as Kleene's logic—, a four-valued logic —as Belnap's logic— or even more general ones. In particular, fuzzy and uncertainty models mostly rely on fuzzy logic and possibilistic logic. Much work has also been done to bridge the gap between the use of stringent formal forms of queries, as imposed by most of the systems at present, and query formulations that allow to express more naturally the information needs of the user, as by natural language sentences. From the foregoing discussion it should be clear that flexible querying constitutes a desirable plus-value in the hands of the user, particularly useful in the cases in which has to be dealt with spatial and temporal information that is affected by imprecision and uncertainty. For historical reasons, the development of the research on flexible querying took off with applications on traditional databases, obviously the easiest part to explore in the long way to go.

Flexible reasoning is not so much heard as a stand-alone term. Most of the time it is used in combination with flexible querying. The techniques for formal reasoning rely on logic as query formulation does, but additionally exploit the deductive capabilities drawn from the inference rules that are valid

in the logic. Reasoning about data and information that resides in a database thus can bring up in a straightforward way more information than simple querying can. Spatio and temporal reasoning can exploit specific logics such as temporal logic. By flexible spatio and temporal reasoning is meant qualitative reasoning about the dynamic changes in the spatial domain, derived by either imprecision or uncertainty (or both), in case of incomplete knowledge of the available information.

Rationale of the Handling of Imperfect Information in Spatio-temporal Databases

This book on flexible querying and reasoning in spatio-temporal databases is divided into two parts. The first part bundles the contributions in which the focus is rather on advances at the theoretical level, thereby discussing examples and opening further perspectives. The second part presents contributions of which the core is a well-developed application.

Though not the focus of the book, representation of the information is an important component in the majority of the chapters, as querying and reasoning are mostly developed on top of one or another specific data and/or database model for the representation of points, lines and objects, the properties associated with them and their intrinsic position, mutual position in aggregations of objects, or their behaviour during the course of time. Most of the authors recognise the need to handle imprecision and uncertainty, ranging from dealing with the static aspects of imprecise or uncertain positions to the dynamic aspects of sequences of changes during time course and possible trajectories. The volume shows definitely how the use of the more advances techniques that enable to deal with imprecision and uncertainty, can add value to the solution of a number of problems related to GIS.

THEORETICAL PART

Chapter 1 by Suzana DRAGICEVIC proposes an approach to represent geographic entities with high spatial variability. Geographic objects are classified according to the characteristics of crispness and vagueness of their body and boundary and are represented in the database by a snapshot model that records distinct states of a dynamic phenomenon at known time instants. The status of the phenomenon at an intermediate spatio-temporal location is derived by a fuzzy temporal interpolation of the known states. To this end probable trajectories of transitions between classes are defined which can model fast, slow or even cyclic changes. An example of fuzzy spatio-temporal interpolation to model land-use changes is presented.

Chapter 2 by Martin ERWIG treats the querying of the spatial dynamic phenomena, by introducing spatio-temporal patterns as a useful means to express repeating or cyclic constraints. The focus is on deriving constraints that allow spatio-temporal patterns to become well-designed composable abstractions that can be smoothly integrated into spatio-temporal query languages. The outline of a language with graphical notation is presented, suited to address problems in GIS and in the related fields of geo- and eco-sciences.

Chapter 3 by Cindy CHEN, Haixun WANG and Carlo ZANIOLO presents a customisable spatio-temporal extension of SQL by enriching a small set of its built-in primitives with User Defined Aggregates that can be programmed directly through the extended SQL itself, hence in a very easy way for the experienced SQL user. The obtained spatio-temporal objects are modelled by counter clockwise directed triangles. A satisfactory overall performance is achieved through the use of the language.

Chapter 4 by Mirco NANNI, Alessandra RAFFAETA, Chiara RENSO and Franco TURINI offer a constraint logical language, STACLP, that provides a set of spatial and temporal primitive operators allowing the user to perform temporal reasoning on spatial data. The chosen formalism can be used as an advanced spatio-temporal query language on geographical data, but as well can be exploited as a deductive rule-based approach to represent domain knowledge on such data. Moreover, it is well suited to represent trajectories of moving objects. STACLP allows to tackle several analysis tasks requiring the integration of deductive and inductive capabilities. A case study in the field of behavioural ecology is added.

Chapter 5 by Gloria BORDOGNA and Sergio CHIESA presents, in an object-based framework, an ontology for imperfect spatio-temporal information, based on fuzzy set theory and possibility theory A dual representation of dynamic spatial phenomena is proposed in which both uncertainty and imprecision can be modelled as orthogonal concepts, affecting either the spatial or the temporal dimensions. Operations going along with the kind of imperfection and methods for the evaluation of spatio-temporal constraints are studied. Abstract spatio-temporal data types are introduced, to guarantee efficient retrieval.

Chapter 6 by Guy DE TRE, Rita DE CALUWE, Jörg VERSTRAETE and Axel HALLEZ discusses the advantage of using generalized constraints as introduced by Lotfi ZADEH to enrich spatio-temporal database models in order to cope with imperfect information. To handle spatio-temporal information, a new data type is introduced. Relying on a many-valued logic based on extended possibilistic truth values, missing information can also be dealt with. It is shown how the generalized constraints enable to specify the semantics of the database and to impose selection criteria for flexible querying. Both data

modelling and querying aspects are addressed in this chapter.

Chapter 7 by Fangju WANG introduces new parsing techniques employing approximate rule matching based on a fuzzy grammar and on possibility theory, for a natural language interface to a GIS, in which can be dealt with uncertainty. This parse strategy aims at minimal computational cost.

Chapter 8 by Adnan YAZICI, Öznur YAVUZ and Roy GEORGE highlights all the static and dynamic aspects of spatio-temporal data objects in the settings of the management of a video database, conform to the MPEG-7 standard. Querying based on the XML Path Language is a well developed part of the chapter.

Chapter 9 by Eliseo CLEMENTINI studies in depth objects with fuzzy boundaries, defined as broad boundaries i.e. delimited by their minimum and maximum extent. New interpretations of the fuzzy boundaries and new application fields for them are investigated. An application is presented in which aggregations of spatial objects are considered, either as tessellations or as networks; it is studied how uncertainty can affect the definition of the position of their common borders and common endpoints.

Chapter 10 by Maria SOMODEVILLA and Fred PETRY too studies vague regions; these are defined by a core and boundary, approximated by their minimum bounding rectangles. It is shown how an expert through an interactive process can achieve a more precise representation of the vague region. Thematic and temporal dimensions have been incorporated in the model to complement spatial reasoning. Both location-based querying and feature-based querying are supported by the model and illustrated in the chapter.

Chapter 11 by Markus SCHNEIDER pays attention to the definition of fuzzy spatial data types and predicates to suit integration into query languages. The 9-intersection model for topological operations is considered as a basis for defining fuzzy topological operators for fuzzy regions. To integrate fuzzy topological predicates into an SQL-like query language, qualitative linguistic descriptions of nuances of topological relationships are introduced, such as "nearly completely inside", "somewhat inside", "quite overlap", etc.

APPLICATION PART

Chapter 12 by Paola CARRARA, Gabriella PASI, Monica PEPE and Anna RAMPINI discusses a possibility-based model to index remote sensing images. This study contributes to the efficient exploitation of image archives, as could be crucial in emergency and disaster management. The capability of spectral

signatures as indexes of images from remote sensing is investigated. The management of uncertainty is faced at distinct levels: in the image classification phase in associating the pixels to one or more classes, in the indexing phase in associating images with spectral signatures and in the retrieval phase to compute the relevance of the images. The proposed method has been evaluated in a database of images representing heterogeneous landscapes.

Chapter 13 by Josef BENEDIKT, Sebastian REINBERG and Leopold RIEDL explores how to evaluate and use linguistic representations for spatial data. Spatial information analysis is carried out with the help of flow-charts based on an easy-to-use programming language. It is shown that sensitivity analysis can be simplified by adopting a fuzzy representation of geographic entities characterized by unsharp boundaries.

Chapter 14 by Ben DERUDDER and Frank WITLOX studies the classification of the major cities in the world city network. This chapter provides evidence that a fuzzy set approach can aid in an exploratory data mining analysis of these inherently vague spatial data. Central to the approach is the generalization of a clustering method using a parameterised fuzzy c-means algorithm. This provides more flexibility in classifying city objects and better supports the analysis of (identified patterns in) world city networks than with a traditional approach.

Chapter 15 by Sidharta GAUTAMA, Johan D'HAEYER and Wilfried PHILIPS addresses image analysis and shows how spatial constraints can help to automatically detect changes in geographic information. The system uses error-tolerant graph matching to find correspondences between the detected image information and the road vector data. Spatial relations between objects play here an important role in finding reliable object-to-object mappings. Flexible querying is realized through the use of relaxation labelling, which in combination with the taking into account of acceptable data inconsistencies, leads to more reliable results for detecting change than the use of global transformation schemes.

Chapter 16 by Tom DE MUER, Andy VERKEYN and Dick BOTTELDOOREN studies noise annoyance and uses a fuzzy rule base to model the vague and complex interactions between its triggers. Large parts of them are related to geographical information, as source location (roads and railway lines), land use, building location and size etc. Many of the input data are imperfect. Imprecision and uncertainty are identified as different flavours of imperfection that need to be treated using different conceptual frameworks. A final issue addressed in this chapter, is the presentation of the vague and uncertain outcome and the aggregation to regions of interest for local policy makers (neighbourhoods, cities, regions).

This book originates from an acknowledgement by the editors of a lack on a basic document in the field of dealing with spatio-temporal databases, offering an overview of the state of the art thus far. So many people have put their efforts in trying to resolve spatio-temporal problems in a computer environment, ranging from modelling information, over querying, mining and reasoning, to say exploiting at best this kind of information, that bringing their efforts together in a book has been felt to be beneficial as well for the further research of those already involved, as for providing a starting platform to those who wish to step into this fast developing, promising, useful field. This volume, companion to the volume Fuzzy Modelling of Spatial Data for Geographic Information Systems, edited by Maria COBB, Fred PETRY and Vincent ROBINSON, focuses on the aspect of exploiting the information which is kept in databases, on querying and reasoning, incorporating the modelling aspects only as far as necessary, but of course highlighting the ultimate purpose of the research, as application is, illustrated by some stringent examples.

Well-known researchers that pioneered the fields of uncertainty and imprecision management in databases and geographical information systems have contributed to the book. Contacts with new researchers have been established made through the process of elaboration of this volume. We are grateful to all of them for the efforts they have put into their contribution and not in the least for their spirit of cooperation, which was utmost encouraging to us and turned our part of the job into a truly joyful and exciting one. For the text editing in Latex and figure layouts, we could account with the valuable help of Jörg VERSTRAETE. While putting this book together, despite the increasing workload never estimated at its real value at the time of the initiative, we not only never came to doubt about the meaningfulness of our enterprise, but rushed towards the end with even greater enthusiasm than ever. We hope the reader will share it in welcoming this book.

Ghent
February 2004

Rita DE CALUWE
Guy DE TRE
Ghent University

Gloria BORDOGNA
National Research Council
Institute for Environmental Process Dynamic