"Time Granularities in Databases, Data Mining, and Temporal Reasoning" (Book Review)

Massimo Franceschet
Department of Sciences, University of Chieti-Pescara, Italy
E-mail: francesc@sci.unich.it
Angelo Montanari
Department of Computer Science, University of Udine, Italy
E-mail: montana@dimi.uniud.it

Since the beginning of the human effort to understand natural phenomena, the problem of measuring time and that of properly relating different time measures occupy a central position. Periodic natural events, such as the earth rotation, the alternation of lunar phases, and the earth revolution around the sun, were commonly identified as suitable time units (granularities), and calendars were devised as a way of expressing the relations between different units (granularities). In computer science, it is widely recognized that the ability of providing and relating temporal representations of the same reality at different 'grain levels' is an important research theme and a major requirement for many applications, including temporal database design, temporal data conversion, temporal database inter-operability, data mining, time management in workflow systems, specification and verification of real-time reactive systems, temporal constraint reasoning, and natural language processing. Any time granularity can be viewed as the partitioning of a temporal domain in groups of elements, where each group is perceived as an indivisible unit (a granule). The description of a fact can use these granules to provide it with a temporal qualification, at the appropriate abstraction level. However, adding the notion of time granularity to a formalism does not merely mean that one can use different time units to represent temporal quantities in a unique flat model, but it involves semantic issues related to the problem of assigning a proper meaning to the association of statements with the different temporal domains of a layered model and of switching from one domain to a coarser/finer one.

The book aims at providing a unifying model to be used for the design, development, and implementation of information system applications that need system support to reason about time granularities. It is organized as follows. In Chapter 1, after a succinct introduction to the subject of the book, the authors define the notion of time granularity as a suitable mapping from the set of integers, called index set, to the power set of the time domain. According the given definition, a granularity maps each index into a subset of the time domain in such a way that the image subsets do not overlap, their order, with respect to the time domain, is the same as the order of their indexes, and the subset of the index set whose elements map to nonempty subsets of the time domain is contiguous. Every nonempty subset of the time domain resulting from the application of the granularity mapping is called a *granule*. Such a definition covers standard granularities, such as `Days` and `Months`, granularities with noncontiguous granules, such as `Business-Days`, and granularities with nonconvex intervals as granules, such as `Business-Months`. Chapter 2 is the main chapter of the book, where such a set-theoretic characterization of time granularity is systematically pursued. The authors elaborate on the given notion of time granularity, identifying the structural properties and relationships between granularities. The key concept is that of *granularity system*. Granularities that come into play in specific contexts or applications are a (usually finite) subset of those satisfying the given definition. A granularity system formally captures this subset by specifying the time domain, the properties of the member granularities, and their relationships. Furthermore, the authors define an algebra of time granularities, called Calendar Algebra, that allow one to generate new granularities through the application of suitable algebraic operations to given granularities. Calendars can be defined as a subset of granularity systems generated from a single bottom granularity by applying the algebraic operations. A special attention is given to the problem of finding the granule (if any) of a specified granularity that covers a given granule of another granularity, e.g., the month that covers a given day. Granule conversion plays a major role in database applications because it allows the user to view temporal information at different granularity levels.

In the other chapters of the book the authors explore the use of the proposed framework in various application domains. Chapters 3 and 4 are respectively concerned with the classical problems of logical design and query processing in the context of temporal databases. As for logical design, the authors show that the addition of a granularity notion makes it possible to express dependencies constraining the evolution of relation instances over time. They introduce a notion of temporal functional dependency to constrain functional dependencies to hold over a specific period of time, e.g., a snapshot, a finite stretch of time, and the whole timeline. For instance, it allows one to constrain the rank of a professor to not vary during a given year and his salary to remain unchanged during a given month. On the basis of this notion, the authors define temporal normal forms and decomposition algorithms that properly extend the traditional ones developed for (atemporal) relational databases. In particular, they provide some operations that allow one to reconstruct the original relation from the component relations of a lossless decomposition. The formalization of semantic assumptions related to time granularity and their use in the query evaluation process are the subject of Chapter 4. The addition of a time granularity dimension to temporal constraint satisfaction problems is investigated in Chapter 5. Exact and approximated algorithms are developed, and their complexity is carefully analyzed. An application of these algorithms to data mining is presented in Chapter 6. In the conclusions, the authors outline a number of open issues and research directions.

The book has several merits. First of all, it is the first attempt to analyze a number of issues related to time granularity in a systematic way. It provides an in-depth illustration of the set-theoretic/algebraic approach to time granularity developed by the authors, and it compares such an approach with alternative ones in the areas of temporal databases and artificial intelligence. The proposed time granularity framework presents two main advantages: naturalness and generality. As for naturalness, it allows one to easily define and manipulate time granularities. For instance, new granularities of practical interest, such as `Business-Weeks`, `Business-Months`, and `Years-since-2000`, can be easily generated by applying user-friendly operations to standard granularities, such `Days`, `Weeks`, and `Months`. As for generality, it subsumes well-known formalisms for time granularity developed in the areas of artificial intelligence and temporal databases, such as the temporal interval collection formalism and the slice formalism. Furthermore, the book effectively supports the claim that time granularity is a major requirements for many systems through the presentation of a number of non-trivial applications.

As it is inevitable, the book suffers from some limitations. First, some basic algebraic problems have not been addressed satisfactorily by the Calendar Algebra. This is the case of the equivalence problem (given two granularity expressions, decide whether they are equivalent), and the minimization problem (given a granularity expression, find the shortest equivalent one). These problems are of obvious theoretical and practical importance, but they have not been systematically investigated. Furthermore, the problem of actually supporting granularity conversion is only partially worked out in a rather complex way.
Second, the book is actually restricted to a single, though important, approach to time granularity that came out in the area of temporal databases. Well-known alternative approaches to time granularity, such as the logical one, and relevant application domains, such as that of reactive system specification and verification, have been almost completely neglected. As an example, a logical framework for time granularity, based on a many-level view of temporal structures, with matching logics and decidability results, has been proposed in [4] and later extended in [3]. Temporal logics for time granularity pair standard temporal logic operators with contextualization and projection operators to select and to move across granularities. They are interpreted over a variety of layered temporal structures, and their expressive power is systematically studied. A comparison between the syntactic flavor of the logical approach and the semantic perspective adopted in the book would have been very useful.
Third, as the authors notice in the conclusions, time granularity is not a long established topic and thus a number of issues and open problems are still the subject of on-going research. Since the publication of the book, the authors themselves deepen some contents of the book, such as the treatment of user-defined granularities [1] and the formalization of the Calendar Algebra [5]. Furthermore, a promising alternative string-based approach to the representation and manipulation of infinite time granularities has been recently proposed in [6]. Infinite

granularities are modeled as infinite strings over a suitable finite alphabet, and, whenever a granularity is (ultimately) periodic, it can be finitely represented as an ordered pair, where the first element is a finite prefix and the second element is a finite repeating pattern. An automata-theoretic solution that revises and extends the string-based one has been later outlined in [2]. It views granularities as strings that can be generated by automata, thus making it possible to (re)use well-known results from automata theory for studying granularities.

We conclude by pointing out some topics that would have deserved further attention in the book. First of all, a systematic complexity analysis of the operations of Calendar Algebra is missing, as well as possible optimizations of given algorithms, e.g. decomposition algorithms in Chapter 3, are often briefly discussed, but they are not fully outlined. Qualitative time granularity is mentioned at the end of Chapter 2, but there is neither a presentation of its distinctive features nor a discussion of its relationships with quantitative time granularity. Finally, the emerging field of spatial granularity and its relationships with that of time granularity are too briefly discussed at the end of Chapter 3.

References

[1] C. Bettini and R. De Sibi. Symbolic representation of user-defined time granularities. *Annals of Mathematics and Artificial Intelligence*, 30(1-4):53–92, 2000.

[2] U. Dal Lago and A. Montanari. Calendars, time granularities, and automata. In *Proceedings of the International Symposium on Spatial and Temporal Databases*, LNCS 2121, pages 279–298, Springer, 2001.

[3] M. Franceschet. *Dividing and Conquering the Layered Land*. PhD Thesis Series CS 2002/2, Department of Mathematics and Computer Science, University of Udine, 2002.

[4] A. Montanari. *Metric and Layered Temporal Logic for Time Granularity*. ILLC Dissertation Series 1996-02, Institute for Logic, Language and Computation, University of Amsterdam, 1996.

[5] P. Ning, S. Jajodia, and X. S. Wang. An algebraic representation of calendars. *Annals of Mathematics and Artificial Intelligence*. To appear.

[6] J. Wijsen. A string-based model for infinite granularities. In *Proceedings of the AAAI Workshop on Spatial and Temporal Granularity*, pages 9–16, AAAI Press, 2000.