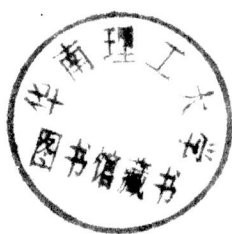# Journal on

Data Semantics IV

Stefano Spaccapietra

Editor-in-Chief

Springer

Stefano Spaccapietra (Ed.)

Journal on Data Semantics IV

 Springer

Volume Editor

Stefano Spaccapietra
EPFL-IC-IIF-LBD, INJ 236 (Bâtiment INJ), Station 14
1015 Lausanne, Switzerland
E-mail: stefano.spaccapietra@epfl.ch

Lecture Notes in Computer Science   3730

The LNCS Journal on Data Semantics

Computerized information handling has changed its focus from centralized data management systems to decentralized data exchange facilities. Modern distribution channels, such as high-speed Internet networks and wireless communication infrastructure, provide reliable technical support for data distribution and data access, materializing the new, popular idea that data may be available to anybody, anywhere, anytime. However, providing huge amounts of data on request often turns into a counterproductive service, making the data useless because of poor relevance or inappropriate level of detail. Semantic knowledge is the essential missing piece that allows the delivery of information that matches user requirements. Semantic agreement, in particular, is essential to meaningful data exchange.

Semantic issues have long been open issues in data and knowledge management. However, the boom in semantically poor technologies, such as the Web and XML, has boosted renewed interest in semantics. Conferences on the Semantic Web, for instance, attract crowds of participants, while ontologies on their own have become a hot and popular topic in the database and artificial intelligence communities.

Springer's *LNCS Journal on Data Semantics* aims at providing a highly visible dissemination channel for the most remarkable work that in one way or another addresses research and development on issues related to the semantics of data. The target domain ranges from theories supporting the formal definition of semantic content to innovative domain-specific application of semantic knowledge. This publication channel should be of the highest interest to researchers and advanced practitioners working on the Semantic Web, interoperability, mobile information services, data warehousing, knowledge representation and reasoning, conceptual database modeling, ontologies, and artificial intelligence.

Topics of relevance to this journal include:

- Semantic interoperability, semantic mediators
- Ontologies
- Ontology, schema and data integration, reconciliation and alignment
- Multiple representations, alternative representations
- Knowledge representation and reasoning
- Conceptualization and representation
- Multimodel and multiparadigm approaches
- Mappings, transformations, reverse engineering
- Metadata
- Conceptual data modeling
- Integrity description and handling
- Evolution and change
- Web semantics and semi-structured data
- Semantic caching
- Data warehousing and semantic data mining
- Spatial, temporal, multimedia and multimodal semantics

- Semantics in data visualization
- Semantic services for mobile users
- Supporting tools
- Applications of semantic-driven approaches

These topics are to be understood as specifically related to semantic issues. Contributions submitted to the journal and dealing with semantics of data will be considered even if they are not within the topics in the list.

While the physical appearance of the journal issues is like the books from the well-known Springer LNCS series, the mode of operation is that of a journal. Contributions can be freely submitted by authors and are reviewed by the Editorial Board. Contributions may also be invited, and nevertheless carefully reviewed, as in the case for issues that contain extended versions of the best papers from major conferences addressing data semantics issues. Special issues, focusing on a specific topic, are coordinated by guest editors once the proposal for a special issue is accepted by the Editorial Board. Finally, it is also possible that a journal issue be devoted to a single text.

The journal published its first volume in 2003 (LNCS 2800), its second volume at the beginning of 2005 (LNCS 3360), and its third volume in Summer 2005 (LNCS 3534). The first two volumes are special issues composed of selected extended versions of the best conference papers. The third volume is a special issue on "Semantic-Based Geographical Information Systems", coordinated by guest editor Prof. Esteban Zimányi. This fourth volume is the first "normal" volume, consisting of spontaneous submissions on any of the topics of interest to the journal. Currently planned volumes include a special issue on emergent semantics.

The Editorial Board comprises one Editor-in-Chief (with overall responsibility) and several members. The Editor-in-Chief has a four-year mandate to run the journal. Members of the board have a three-year mandate. Mandates are renewable. More members may be added to the board as appropriate.

We are happy to welcome both readers and authors, and hope we will share this privileged contact for a long time.

Stefano Spaccapietra
Editor-in-Chief
http://lbdwww.epfl.ch/e/Springer/

JoDS Volume IV - Preface

This fourth JoDS volume is the outcome of the selection of papers spontaneously submitted to the journal, in particular in response to a Call for Papers issued on July 17, 2004. The call invited submissions on any topic that falls within the scope of the journal.

Altogether, 38 submissions were reviewed. After the first round of reviews, 24 submissions were asked to perform a major revision and resubmit. Most of these were actually resubmitted and went through a second round of reviews, with the same reviewers as allocated for the first round. Eventually, 10 papers were accepted for publication, after some last modifications suggested by the reviewers.

Accepted papers cover a wide range of topics, from traditional data semantics (information modeling, data model transformation, knowledge representation, data and schema integration) to the newest trends (multimedia, Semantic Web annotation, information extraction, and knowledge discovery).

A new Call for Papers is open at the moment for a volume to appear in 2006. We are looking forward to your contributions.

Stefano Spaccapietra
Editor-in-Chief

In Memoriam

Prof. Maurizio Panti

On July 3, 2005, at 9:45AM, Prof. Maurizio Panti, Head of the Department of Computer, Management and Automation Engineering at the Polytechnic University of Marche, passed away.

Approximately seven months earlier, he had been diagnosed with a late-stage aggressive cancer. In this brief period of illness, he never stopped his activities, showing devotion to his work and students and a strong will power that will be an example for all of us. He spent his final hours surrounded by family and friends.

Prof. Maurizio Panti was full professor of Information Systems and Data Bases. He promoted the growth of informatics both in academia, contributing to the foundation of the Informatics Institute, which then developed into the present department, and in the regional economic arena, also serving as a member in the Scientific and Technical Board for the Regional Information System.

His loss will be greatly felt by all those who knew and worked with him.

Prof. Hongjun Lu

Prof. Hongjun Lu was a world-renowned researcher who served the database community with dedication and distinction in various capacities over the years. He was a trustee of the VLDB Endowment, a member of the ACM SIGMOD Advisory Board (1998–2002), an associate editor of IEEE Transactions on Knowledge and Data Engineering (TKDE), Chair of the Steering Committee of the International Conference on Web-Age Information Management (WAIM), and Co-chair (1998-2001) and Chair (2001-2003) of the Steering Committee of Pacific-Asia Conference of Knowledge Discovery and Data Mining (PAKDD). In December 2004, the China Computer Federation Database Society honored him with a Contribution Award and, just recently, he was honored with this year's inaugural PAKDD Distinguished Contribution Award.

His passing is a great loss not only to the China Database Society but also to the International Database Society. He has made sustained and outstanding contributions to the international database research community as well as to database research.

Organization

External Reviewers

All Editorial Board members contributed reviews for the selection of the submissions. In addition, a number of colleagues helped us in this reviewing task. We would like to express here our gratitude for their cooperation and our warmest thanks for the job they did.

Bettina Berendt

Sonia Bergamaschi

Claudio Bettini

Mokrane Bouzeghoub

Anita Coleman

Jos De Bruin

Flavio De Paoli

Giovanni Denaro

Ying Ding

Gillian Dobbie

Anastasios D. Doulamis

Oleksandr Drutskyy

Birte Glimm

Sven Hartmann

Georges Hébrail

Martin Hepp

Jane Hunter

Mustafa Jarrar

Roland Kaschek

Evgeny Zolin

Yehuda Koren

Mong Li Lee

Lei Li

Xiaonan Lu

R. Manmatha

Andrea Maurino

Rym Mili

Moira Norrie

Axel Polleres

Helen Purchase

Brian Roark

Ingo Schmitt

Richard Smiraglia

Peter Spyns

Paolo Terenziani

Christelle Vangenot

Michalis Vazirgiannis

James Z. Wang

X. Sean Wang

JoDS Editorial Board

Lecture Notes in Computer Science

For information about Vols. 1–3748

please contact your bookseller or Springer

Vol. 3795: H. Zhuge, G.C. Fox (Eds.), Grid and Cooperative Computing - GCC 2005. XXI, 1203 pages. 2005.

Vol. 3794: X. Jia, J. Wu, Y. He (Eds.), Mobile Ad-hoc and Sensor Networks. XX, 1136 pages. 2005.

Vol. 3793: T. Conte, N. Navarro, W.-m.W. Hwu, M. Valero, T. Ungerer (Eds.), High Performance Embedded Architectures and Compilers. XIII, 317 pages. 2005.

Vol. 3792: I. Richardson, P. Abrahamsson, R. Messnarz (Eds.), Software Process Improvement. VIII, 215 pages. 2005.

Vol. 3791: A. Adi, S. Stoutenburg, S. Tabet (Eds.), Rules and Rule Markup Languages for the Semantic Web. X, 225 pages. 2005.

Vol. 3790: G. Alonso (Ed.), Middleware 2005. XIII, 443 pages. 2005.

Vol. 3789: A. Gelbukh, Á. de Albornoz, H. Terashima-Marín (Eds.), MICAI 2005: Advances in Artificial Intelligence. XXVI, 1198 pages. 2005. (Sublibrary LNAI).

Vol. 3788: B. Roy (Ed.), Advances in Cryptology - ASIACRYPT 2005. XIV, 703 pages. 2005.

Vol. 3787: D. Kratsch (Ed.), Graph-Theoretic Concepts in Computer Science. XIV, 470 pages. 2005.

Vol. 3785: K.-K. Lau, R. Banach (Eds.), Formal Methods and Software Engineering. XIV, 496 pages. 2005.

Vol. 3784: J. Tao, T. Tan, R.W. Picard (Eds.), Affective Computing and Intelligent Interaction. XIX, 1008 pages. 2005.

Vol. 3783: S. Qing, W. Mao, J. Lopez, G. Wang (Eds.), Information and Communications Security. XIV, 492 pages. 2005.

Vol. 3781: S.Z. Li, Z. Sun, T. Tan, S. Pankanti, G. Chollet, D. Zhang (Eds.), Advances in Biometric Person Authentication. XI, 250 pages. 2005.

Vol. 3780: K. Yi (Ed.), Programming Languages and Systems. XI, 435 pages. 2005.

Vol. 3779: H. Jin, D. Reed, W. Jiang (Eds.), Network and Parallel Computing. XV, 513 pages. 2005.

Vol. 3778: C. Atkinson, C. Bunse, H.-G. Gross, C. Peper (Eds.), Component-Based Software Development for Embedded Systems. VIII, 345 pages. 2005.

Vol. 3777: O.B. Lupanov, O.M. Kasim-Zade, A.V. Chaskin, K. Steinhöfel (Eds.), Stochastic Algorithms: Foundations and Applications. VIII, 239 pages. 2005.

Vol. 3776: S.K. Pal, S. Bandyopadhyay, S. Biswas (Eds.), Pattern Recognition and Machine Intelligence. XXIV, 808 pages. 2005.

Vol. 3775: J. Schönwälder, J. Serrat (Eds.), Ambient Networks. XIII, 281 pages. 2005.

Vol. 3774: G. Bierman, C. Koch (Eds.), Database Programming Languages. X, 295 pages. 2005.

Vol. 3773: A. Sanfeliu, M.L. Cortés (Eds.), Progress in Pattern Recognition, Image Analysis and Applications. XX, 1094 pages. 2005.

Vol. 3772: M. Consens, G. Navarro (Eds.), String Processing and Information Retrieval. XIV, 406 pages. 2005.

Vol. 3771: J.M.T. Romijn, G.P. Smith, J. van de Pol (Eds.), Integrated Formal Methods. XI, 407 pages. 2005.

Vol. 3770: J. Akoka, S.W. Liddle, I.-Y. Song, M. Bertolotto, I. Comyn-Wattiau, W.-J. van den Heuvel, M. Kolp, J. Trujillo, C. Kop, H.C. Mayr (Eds.), Perspectives in Conceptual Modeling. XXII, 476 pages. 2005.

Vol. 3769: D.A. Bader, M. Parashar, V. Sridhar, V.K. Prasanna (Eds.), High Performance Computing – HiPC 2005. XXVIII, 550 pages. 2005.

Vol. 3768: Y.-S. Ho, H.J. Kim (Eds.), Advances in Multimedia Information Processing - PCM 2005, Part II. XXVIII, 1088 pages. 2005.

Vol. 3767: Y.-S. Ho, H.J. Kim (Eds.), Advances in Multimedia Information Processing - PCM 2005, Part I. XXVIII, 1022 pages. 2005.

Vol. 3766: N. Sebe, M.S. Lew, T.S. Huang (Eds.), Computer Vision in Human-Computer Interaction. X, 231 pages. 2005.

Vol. 3765: Y. Liu, T. Jiang, C. Zhang (Eds.), Computer Vision for Biomedical Image Applications. X, 563 pages. 2005.

Vol. 3764: S. Tixeuil, T. Herman (Eds.), Self-Stabilizing Systems. VIII, 229 pages. 2005.

Vol. 3762: R. Meersman, Z. Tari, P. Herrero (Eds.), On the Move to Meaningful Internet Systems 2005: OTM 2005 Workshops. XXXI, 1228 pages. 2005.

Vol. 3761: R. Meersman, Z. Tari (Eds.), On the Move to Meaningful Internet Systems 2005: CoopIS, DOA, and ODBASE, Part II. XXVII, 653 pages. 2005.

Vol. 3760: R. Meersman, Z. Tari (Eds.), On the Move to Meaningful Internet Systems 2005: CoopIS, DOA, and ODBASE, Part I. XXVII, 921 pages. 2005.

Vol. 3759: G. Chen, Y. Pan, M. Guo, J. Lu (Eds.), Parallel and Distributed Processing and Applications - ISPA 2005 Workshops. XIII, 669 pages. 2005.

Vol. 3758: Y. Pan, D.-x. Chen, M. Guo, J. Cao, J.J. Dongarra (Eds.), Parallel and Distributed Processing and Applications. XXIII, 1162 pages. 2005.

Vol. 3757: A. Rangarajan, B. Vemuri, A.L. Yuille (Eds.), Energy Minimization Methods in Computer Vision and Pattern Recognition. XII, 666 pages. 2005.

Vol. 3756: J. Cao, W. Nejdl, M. Xu (Eds.), Advanced Parallel Processing Technologies. XIV, 526 pages. 2005.

Vol. 3754: J. Dalmau Royo, G. Hasegawa (Eds.), Management of Multimedia Networks and Services. XII, 384 pages. 2005.

Vol. 3753: O.F. Olsen, L.M.J. Florack, A. Kuijper (Eds.), Deep Structure, Singularities, and Computer Vision. X, 259 pages. 2005.

Vol. 3752: N. Paragios, O. Faugeras, T. Chan, C. Schnörr (Eds.), Variational, Geometric, and Level Set Methods in Computer Vision. XI, 369 pages. 2005.

Vol. 3751: T. Magedanz, E.R.M. Madeira, P. Dini (Eds.), Operations and Management in IP-Based Networks. X, 213 pages. 2005.

Vol. 3750: J.S. Duncan, G. Gerig (Eds.), Medical Image Computing and Computer-Assisted Intervention – MICCAI 2005, Part II. XL, 1018 pages. 2005.

Vol. 3749: J.S. Duncan, G. Gerig (Eds.), Medical Image Computing and Computer-Assisted Intervention – MICCAI 2005, Part I. XXXIX, 942 pages. 2005.

Table of Contents

Generic Relationships in Information Modeling

Mohamed Dahchour[1], Alain Pirotte[2], and Esteban Zimányi[3]

[1] Institut National des Postes et Télécommunications,
Av. Allal Al Fassi, Rabat, Morocco
`dahchour@inpt.ac.ma`
[2] Université catholique de Louvain, IAG School of Management,
1 Place des Doyens, B-1348 Louvain-la-Neuve, Belgium
`pirotte@info.ucl.ac.be`
[3] Université Libre de Bruxelles, Department of Computer and Network Engineering,
CP 165/15, 50 Av. F. Roosevelt, B-1050 Brussels, Belgium
`ezimanyi@ulb.ac.be`

Abstract. Generic relationships are abstraction patterns used for structuring information across application domains. They play a central role in information modeling. However, the state of the art of handling generic relationships leaves open a number of problems, like differences in the definition of some generic relationships in various data models and differences in the importance given to some generic relationships, considered as first-class constructs in some models and as special cases of other relationships in other models. To address those problems, we define a list of dimensions to characterize the semantics of generic relationships in a clear and systematic way. The list aims to offer a uniform and comprehensive analysis grid for generic relationships, drawn from a careful analysis of commonalities and differences among the generic relationships discussed in the literature. The usefulness of those dimensions is illustrated by reviewing significant generic relationships, namely, materialization, role, aggregation, grouping, and ownership. Based on those dimensions, a new metamodel for relationships is proposed.

1 Introduction

Information modeling is the activity of creating abstract representations of some aspects of physical and social systems and their environment. Information models are typically built in the early stages of system development, preceding design and implementation. But information models can also be useful even if no system is contemplated: they then serve to clarify ideas about structure and functions in a perception of the world.

Advances in information modeling involve narrowing the gap between concepts in the real world and their representation in information models by identifying powerful abstractions allowing to better represent application semantics (see, e.g., [1,29,31,34,38,46]).

Generic relationships are such powerful abstraction mechanisms. They are high-level templates for relating real-world entities. Well-known generic relationships include the following.

- *Classification* relates a class with a set of objects sharing the same properties. An object must be an instance of at least one class. It is also known as *is-of.* For example John is an instance of class Person.
- *Association* represents a structural connection among classes. Associations can be *binary* or *n-ary* ($n \geq 3$). An example of a binary association is teaches(Professor,Course). An example of a ternary association is prescription(Doctor,Medicine, Patient).
- *Generalization* relates *superclasses* to their specializations called *subclasses.* Subclasses inherit all properties (attributes, methods, roles, integrity constraints) from their superclasses. Subclasses may define new specific properties. For example, Vehicle is a generalization of Car.
- *Aggregation* associates an *aggregate* (or whole or composite) to its *components* (or parts). It is also known as *part-whole* or *part-of.* For example, Car is an aggregation of Body, Engine, and Wheel.

Generic relationships model patterns abstracting collections of related *specific* relationships. Specific relationships are instances of generic relationships in a particular application. For example, Vehicle◁—Car is a specific generalization with pattern SuperClass◁—SubClass (see Figure 1).
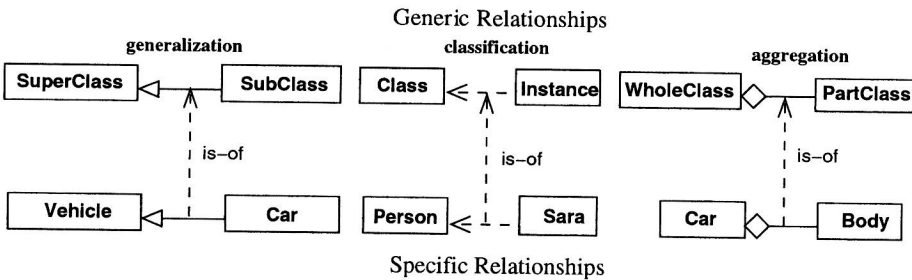
Fig. 1. Generic and specific relationships

Research on information modeling has studied other generic relationships like *materialization* [9], *ownership* [48], *role* [45,10], *grouping* [33], *viewpoint* [32], *generation* [14], *versioning* [3,21], *realization* [25], *transition* [15], and *refinement* [41]. These generic relationships naturally model phenomena typical of complex application domains whose semantics escapes direct representation with classical generic relationships (i.e., association, generalization, classification, and aggregation).

Generic relationships play a central role in information modeling. However, the state of the art of handling them leaves open a number of problems. The semantics of some generic relationships differs among models. Other generic relationships have often been badly understood, underestimated, or merely ignored in some models. In addition, some generic relationships, considered as first-class constructs in some models, are considered as special cases of other relationships in other models. For instance, in UML [41], the aggregation relationship

is considered a special kind of the ordinary association, whereas in most models (e.g., [46,36,19,30,23]) aggregation enjoys a status of its own and comes with more specific features than those defined in UML. Another example concerns the *grouping* relationship relating a collection (e.g., TennisClub) and its members (e.g., TennisPlayer). Grouping was defined in [33] as an independent generic relationship with specific characteristics, while it is just considered as a special case of aggregation, e.g., in [46,19].

Another problem concerns the adequacy of choosing some generic relationships rather than others when modeling an application domain. For example, it can be argued that the relationship between students and employees on the one hand and persons on the other hand is more adequately modeled as a role relationship than as a generalization. Generalization seems more adequate to represent the relationship between males and females on the one hand and persons on the other hand. The idea is that the role relationship captures the temporal and evolutionary aspects of real-world objects (e.g., persons may be students and later become employees), while the usual generalization relationship deals with their more static aspects (e.g., most persons are permanently males or females).

Such difficult questions of adequacy or validity of generic relationships for modeling real-world situations are not directly discussed in this paper, although they are illustrated through a number of examples. Instead, the paper precisely characterizes the structural semantics of generic relationships, to help conceptual database designers precisely evaluate the adequacy of choosing one model rather than another.

We argue that some problems with generic relationships mainly concern the absence of formalizable dimensions or criteria along which the relationships can be characterized in a systematic way. The paper defines such dimensions and illustrates their effectiveness by reviewing some generic relationships. Implementation issues are not presented in this paper. They are discussed in detail in [8].

Notations and Conventions. Table 1 gathers the main notations used in the paper. We use UML [41] notations to specify classes, instances, generalization, instantiation, and aggregation. We add notations to represent concepts that have no equivalent in UML. We prefer to draw associations as boxes with rounded corners rather than using the UML notation. As in UML, instances of relationships are called links.

The rest of the paper is organized as follows. Section 2 presents a preliminary classification of generic relationships. Section 3 presents an overview of our relationship model. Section 4 presents the characteristics of a basic binary relationship (denoted $\mathcal{BBR}$). Section 5 defines a set of dimensions that characterize binary generic relationships. Section 6 reviews several generic relationships in the light of those dimensions. Section 7 gives some guidelines to identify and define new generic relationships. Section 8 presents a new metamodel for generic relationships based on their semantics as presented in Sections 4 and 5. Section 9 summarizes and concludes the paper.

Table 1. Notations and their meanings

Notation	Meaning
$I \dashrightarrow C$ or $I \in C$	I is an instance of class C
$C_1 \!-\!\boxed{\text{R}}\!-\! C_2$	R is a binary association between classes C_1 and C_2
$S \longrightarrow\!\!\triangleright G$	S is a subclass of superclass G
$W \diamond\!\!-\! P$	whole class W is composed of part class P
$A \longrightarrow\!\!* C$	abstract class A materializes as concrete class C
$R \longrightarrow\!\circ O$	R is a role class of object class O
$O \prec\!\cdots P$	owner class O owns property class P
$M \longrightarrow\!\!\rightarrow S$	M is a member class of set S
$\mathcal{BBR}$	the basic binary relationship
$C_1 \overset{R}{\rule{1.2cm}{0.4pt}} C_2$	R is a binary relationship between classes C_1 and C_2
$\rho_C(R)$	the role played by class C in relationship R
$C_1(v_1, v_2) \overset{R}{\rule{1.2cm}{0.4pt}} C_2$	cardinality of role $\rho_{C_1}(R)$ is (v_1, v_2)
$\pi_C(R)$	the set of instances of C participating in R with role $\rho_C(R)$
$R(C_1, C_2)$	R relates C_1 and C_2
$R(c_1, c_2)$	there is an instance of R (a link) relating c_1 and c_2
$R_1 \otimes R_2$	relationships R_1 and R_2 are exclusive
$\rho_C(R_1) \otimes \rho_C(R_2)$	roles $\rho_C(R_1)$ and $\rho_C(R_2)$ are exclusive
$R_1 \subseteq R_2$	relationship R_1 is inclusive in relationship R_2
$\rho_C(R_1) \subseteq \rho_C(R_2)$	role $\rho_C(R_1)$ is inclusive in role $\rho_C(R_2)$
$C_0 \overset{R \mid d}{\rule{1.2cm}{0.4pt}} \{C_1, \ldots, C_n\}$	partition of R in classes C_i according to discriminator d

2   Classification of Generic Relationships

Generic relationships can be classified along the following three orthogonal dimensions, as depicted in Figure 2: (1) degree, (2) structurality and dynamicity, and (3) dependency on application domains.

Degree. It is the number of participating classes in a relationship. A relationship of degree two is said to be *binary*, and one of degree n ($n \geq 3$) is *n-ary*. Examples of binary generic relationships include:

- *classification* (of pattern Class←--Instance) relates an instance (e.g., Sarah) to its class (e.g., person);
- *generalization* (of pattern SuperClass◁—SubClass) relates a superclass (e.g., persons) to its subclasses (e.g., males and females).
- *materialization* (of pattern Abstract—*Concrete) [39,9] relates a class of categories (e.g., models of cars) with a class of more concrete objects (e.g., individual cars);
- *ownership* (of pattern Property···≻Owner) [48] relates an owner class (e.g., persons) and a property (e.g., cars) possessed by their objects;
- *aggregation* (of pattern WholeClass◇—PartClass) [30,46] relates composites (e.g., cars) to their components (e.g., body and engine);