


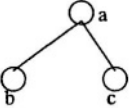
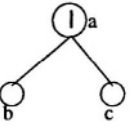


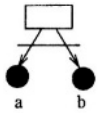
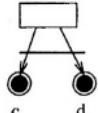



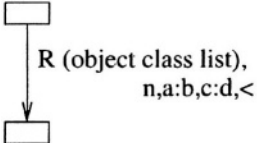
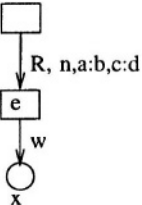


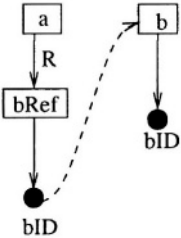
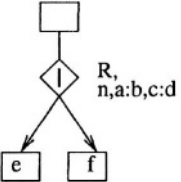
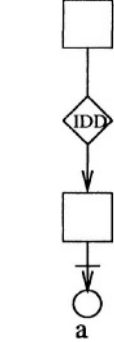

Appendix A

ORA-SS Notation

The following tables summarize the notation of ORA-SS diagrams.

notation	description
	object class with name <i>a</i>
	attribute <i>b</i> , where <i>x</i> represents the cardinality, ? is 0 or 1, + is 1 or more, * is 0 or more, and the default value for <i>x</i> is 1.
	attribute <i>b</i> where the ordering of the value of the attributes is important. <i>x</i> is either + or *, and the default value is *.
	composite attribute <i>a</i> with component attributes <i>b</i> and <i>c</i>
	disjunctive attribute <i>a</i> is either <i>b</i> or <i>c</i>
	identifier/primary key <i>c</i>
	candidate key <i>d</i>

notation	description
	<p>composite identifier/primary key</p>
	<p>composite candidate key</p>
	<p>derived attribute</p>
	<p>attribute with unknown structure or whose structure is heterogeneous</p>
	<p>the ordering on the attributes of object class <i>a</i> is important</p>
	<p>relationship type with name <i>R</i>, with participating objects <i>object class list</i>, of degree <i>n</i>, where the participation of the parent has minimum <i>a</i> and maximum <i>b</i>, and the child has minimum <i>c</i> and maximum <i>d</i>, and the ordering of the object classes is important. The name is optional. The object class list is optional and is included only if the object classes of the relationship type are separated by object class(es) not relevant to the relationship type. The default degree is 2, default parent participation constraint is $0 : m$, default child participation constraint is $1 : n$, and default on ordering is no ordering.</p>
	<p>attribute <i>x</i> belongs to relationship type <i>R</i>. The default (without label <i>R</i> on the edge) shows that attribute <i>x</i> belongs to object class <i>e</i>.</p>

notation	description
	<p>reference object class <i>bRef</i> references referenced object class <i>b</i> with identifier <i>bID</i></p>
	<p>disjunctive relationship type: either object class <i>e</i> or object class <i>f</i></p>
	<p>weak object class: attribute <i>a</i> is a weak identifier</p>
	<p><i>b</i> inherits from <i>a</i> (inheritance diagram)</p>

References

- Abiteboul, S. (1999). On views and XML. In *Proceedings of 18th ACM Symposium on Principles of Database Systems*.
- Abiteboul, S., Amann, B., Cluet, S., Eyal, A., Mignet, L., and Milo, T. (1999a). Active views for electronic commerce. In *Proceedings of 25th International Conference on Very Large Data Bases*.
- Abiteboul, S., Buneman, P., and Suciu, D. (1999b). *Data On the Web-From Relational to Semistructured Data and XML*. Morgan Kaufman Publishers, San Francisco, California.
- Apparao, V. and Byrne, S. (1 October 1998). Document object model (DOM) level 1 specification. W3C Recommendation.
- Arenas, M. and Libkin, L. (2004). A normal form for XML documents. *ACM Trans. Database Syst.*, 29(1):195–232.
- Baru, C. K., Gupta, A., Ludascher, B., Marciano, R., Papakonstantinou, Y., Velikhov, P., and Chu, V. (1999). XML-based information mediation with MIX. In *SIGMOD 1999, Proceedings ACM SIGMOD International Conference on Management of Data*.
- Bray, T., Paoli, J., and Sperberg-McQueen, C. M. (Oct. 2000). Extensible markup language (XML) 1.0. 2nd edition. <http://www.w3.org/TR/REC-xml>.
- Buneman, P., Davidson, S., Fan, W., Hara, C., and Tan, W.C. (2001a). Keys for XML. In *Proceedings of the Tenth International World Wide Web Conference*.
- Buneman, P., Davidson, S., Fan, W., Hara, C., and Tan, W.C. (2001b). Reasoning about keys for XML. In *International Workshop on Database Programming Languages*.
- Carey, M. J., Kiernan, J., Shanmugasundaram, J., Shekita, E. J., and Subramanian, S. N. (2000). Xperanto: Middleware for publishing object-relational data as XML documents. In *Proceedings of 26th International Conference on Very Large Data Bases*, pages 646–648.
- Chen, Y.B., Ling, T.W., and Lee, M.L. (2002). Designing valid XML views. In *Proceedings of 21st International Conference on Conceptual Modeling*.
- Christophides, V., Cluet, S., and Simson, J. (2000). On wrapping query languages and efficient XML integration. In *Proceedings of the 2000 ACM SIGMOD International Conference on Management of Data*, pages 141–152.
- Cluet, S., Veltri, P., and Vodislav, D. (2001). Views in a large scale XML repository. In *Proceedings of 27th International Conference on Very Large Data Bases*, pages 271–280.
- Date, C. J. (1975). *An Introduction to Database Systems*. Addison Wesley, 1st edition.
- Deutsch, A., Fernandez, M., and Suciu, D. (1999). Storing semistructured data with STORED. In *ACM SIGMOD*, pages 431–442.

- Deutsch, A. and Tannen, V. (2003). Mars: A system for publishing xml from mixed and redundant storage. In *VLDB*.
- Dobbie, G., Wu, X., Ling, T.W., and Lee, M.L. (2000). ORA-SS: An object-relationship-attribute model for semi-structured data. Technical Report TR21/00, School of Computing, National University of Singapore.
- Embley, D.W. and Mok, W.Y. (2001). Developing XML documents with guaranteed “good” properties. In *Proc. of 20th International Conference on Conceptual Modeling*.
- Fan, W. and Simon, J. (2000). Integrity Constraints for XML. In *Proceedings of the Nineteenth ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems, Dallas, Texas, USA*, pages 23–34. ACM.
- Fernandez, M., Tan, W., and Suciu, D. (2000). SilkRoute: Trading between relations and XML. In *Proceedings of the 9th International World Wide Web Conference*.
- Florescu, D. and Kossmann, D. (1999). Storing and querying XML data using an RDBMS. *IEEE Data Engineering Bulletin*, 22(3):27–34.
- Goldman, R. and Widom, J. (1997). Dataguides: Enabling query formulation and optimization in semistructured databases. In *Proc. of 23rd International Conference on Very Large Data Bases*, pages 436–445.
- ISO/IEC (2000). Information technology - text and office systems - regular language description for XML (RELAX) - part 1: RELAX core. DTR 22250-1.
- Lee, D. and Chu, W. (2000). Constraints-preserving transformation from XML document type definition to relational schema. In *Proc. 19th International Conference on Conceptual Modeling*, pages 323–338.
- Lee, S.Y., Lee, M.L., Ling, T.W., and Kalinichenko, L.A. (1999). Designing good semi-structured databases. In *Proc. 18th International Conference on Conceptual Modeling*, pages 131–145.
- Ling, T. W. (1989). A normal form for sets of not-necessarily normalized relations. In *Proceedings of the 22nd Hawaii International Conference on System Sciences*, pages 578–586. IEEE Computer Society Press.
- Ling, T. W. and Yan, L. L. (1994). NF-NR: A practical normal form for nested relations. *Journal of Systems Integration*, 4:309–340.
- Ling, T.W. (1985). A normal form for entity-relationship diagrams. In *Proc. 4th International Conference on Entity-Relationship Approach*.
- Ling, T.W., Goh, C.H., and Lee, M.L. (1996). Extending classical functional dependencies for physical database design. *Information and Software Technology*, 38:601–608.
- Ling, T.W. and Teo, P.K. (1993). Inheritance conflicts in object-oriented systems. In *Proc. 4th International Conference on Database and Expert Systems Applications, DEXA '93*, pages 189–200.
- Mani, M., Lee, D., and Muntz, R.R. (2001). Semantic data modeling using XML schemas. In *Proc. of 20th International Conference on Conceptual Modeling*, pages 149–163.
- Manolescu, I., Florescu, D., and Kossmann, D. (2001). Answering XML queries on heterogeneous data sources. In *Proceedings of 27th International Conference on Very Large Data Bases*, pages 241–250.
- McHugh, J., Abiteboul, S., Goldman, R., Quass, D., and Widom, J. (1997). Lore: A database management system for semistructured data. *SIGMOD Record*, 26(3):54–66.
- Mo, Y. and Ling, T.W. (2002). Storing and maintaining semistructured data efficiently in an object-relational database. In *Proc. of 3rd International Conference on Web Information Systems Engineering (WISE 2002)*, pages 247–256.
- Ni, W. and Ling, T.W. (2003). GLASS: A graphical query language for semi-structured data. In *Proc. of Eighth International Conference on Database Systems for Advanced Applications (DASFAA '03)*.

- Ozsoyoglu, Z.M. and Yuan, L.Y. (1987). A new normal form for nested relations. *ACM Transaction on Database Systems*, 12(1).
- Shanmugasundaram, J., Tufte, K., Zhang, C., He, G., DeWitt, D.J., and Naughton, J.F. (1999). Relational databases for querying XML documents: Limitations and opportunities. In *Proc. of 25th International Conference on Very Large Data Bases*, pages 79–90.
- Suciu, D. (1998). Semistructured data and XML. In *Proc. of 5th International Conference on Foundations of Data Organization*.
- Thompson, H.S., Beech, D., Maloney, M., and (Eds), N. Mendelson (May 2001). XML Schema Part 1: Structures. <http://www.w3.org/TR/xmlschema-1>.
- Wang, Q.Y., Xu, J.X., and Wong, K.F. (2000). Approximate graph schema extraction for semistructured data. In *Proc. of 7th International Conference on Extending Database Technology*.
- Widom, J. (1999). Data management for XML: Research directions. *IEEE Data Engineering Bulletin*, 22(3):44–52.
- Wu, X., Ling, T.W., Lee, M.L., and Dobbie, G. (2001a). Designing semistructured databases using ORA-SS model. In *Proc. of 2nd International Conference on Web Information Systems Engineering*.
- Wu, X.Y., Ling, T.W., Lee, S.Y., Lee, M.L., and Dobbie, G. (November 2001b). NF-SS: A normal form for semistructured schemata. In *Proceedings of the International Workshop on Data Semantics in Web Information Systems (DASWIS/ER2001)*. Springer-Verlag.

Index

- ANY, 44
- Active View, 5, 111, 136
- Aggregate, 145
- Agora, 5
- Attribute of object class, 43
- Attribute of relationship type, 44
- Attribute, 43–45
- Binary relationship type, 38–39
- CM Hypergraph, 3, 18–19, 21, 35
- Candidate identifier, 44
- Composite attribute, 44, 60, 64, 85
- Composite candidate identifier, 44
- Composite identifier, 44
- DOM, 3, 12, 14, 34
- DTD, 4, 8–12, 34
- Data model, 3
- Data path expression, 74
- Data redundancy, 143–146
- DataGuide, 3, 15–16, 59, 74
- Default value, 44
- Denormalization, 139
- Dependent object class, 43, 88
- Derived attribute, 44, 145
- Disjunction, 49
- Drop, 116–120, 132
- Element set, 8
- Element, 7
- Entity Relationship, 21, 79
- Extended Entity Relationship, 4, 21–24, 35
- Extended functional dependency, 82
- Extraction Rules, 60
- Fixed value, 44
- Fourth normal form, 79
- Functional dependency diagram, 52–55
- IDREF, 61, 64, 70
- IMS, 47, 141–143
- Identifier dependency relationship type, 43, 88
 - 132, 134
- Identifier, 44–45, 61, 64
- Inheritance diagram, 55–56
- Instance diagram, 49, 51
- Integration, 2
- Invalid view, 114
- Join, 116, 121, 124–125, 132
- Key, 107
- Level of nested relation, 83
- Logical pointer, 142
- MIX, 5, 111, 137
- Many to many relationship type, 143
- Multivalued attribute, 60, 64, 85
- NF ORA–SS schema diagram, 88–91
- NF-NR, 80, 84–85
- NF-SS, 4, 109
- NNF, 80
- Nested relational model, 80
- Normalization, 4, 77, 80
- O-NF, 85, 87
- OEM, 3, 13, 15–16
- ORA-SS, 4, 12, 28, 32, 35
- Object class list, 38
- Object class, 37–38, 64, 85, 88
- Object relational model, 158
- Object, 60
- Ordering, 48
- Participation constraint, 38, 66
- Physical design, 5
- Physical pairing, 142–143
- Project, 117
- R-NF, 86–88
- RELAXNG, 4, 9
- Recursive query, 145
- Redundancy, 77
- References, 46–47
- Relational database, 139
- Relationship type, 38–39, 41–43, 65, 86, 88
- Relationship, 8
- Relatively stable attribute, 147, 154
- Relatively stable relationship type, 147, 154
- Replicated 3NF, 140–141

- Replicated NF ORA-SS schema, 149
- Reversible view, 130
- S3-NF, 4, 108
- S3-graph, 3, 16–18, 29–31, 34
- Schema diagram, 37–39, 41–43, 45–49
- Schema extraction, 4, 59, 62
- Select, 116, 132
- SilkRoute, 5, 111
- Simple attribute, 60
- Single-valued attribute, 60, 64, 85
- Strong functional dependency, 140
- Subelement, 61, 63
- Swap, 116, 125, 127–128, 130, 132
- Symmetric query, 141, 143
- Ternary relationship type, 38–39, 66
- Text segment, 63
- Third normal form, 79
- Update anomaly, 140
- Valid view, 111
- Views, 5
- Weak identifier, 43
- Weak key, 107
- XML Schema, 4, 8
- XML Tree, 4, 24–28, 35
- XML element, 60
- XML, 2, 62
- XNF, 4, 109–110
- XPERANTO, 5, 111
- Xyleme, 5, 111
- YAT, 5

About the Authors

Dr Gillian DOBBIE is currently an Associate Professor in the Department of Computer Science at the University of Auckland, New Zealand, and Deputy Director of the Software Engineering Programme.

[See <http://www.cs.auckland.ac.nz/people/profile.php?id=gdob002>]

She received a Ph.D. from the University of Melbourne, an M.Tech.(Hons) and B.Tech.(Hons) in Computer Science from Massey University. She has lectured at Massey University, the University of Melbourne, and Victoria University of Wellington, and held visiting research positions at Griffith University and the National University of Singapore.

Her research interests include formal foundations for databases, object oriented databases, semistructured databases, logic and databases, data warehousing, data mining, access control, e-commerce and data modeling. She has published 27 international refereed journal and conference papers. Some of the publications are listed in <http://www.informatik.uni-trier.de/ley/db/indices/a-tree/d/Dobbie:Gillian.html>.

She is programme co-chair on ADC05 and ADC06, and has served as programme co-chair on WEBH2001 and WEBH2002. She has served on programme committees for many international conferences including DOOD97, ADC98, DaWaK01, WISE2002, and ACE2003, and has refereed papers for international journals such as TPLP and VLDB.

Dr. Mong Li LEE is currently an Assistant Professor in the School of Computing at the National University of Singapore. She received her Ph.D. degree in Computer Science from the National University of Singapore in 1999. Her thesis examines translation, integration and update issues in a federated database environment. She was awarded the IEEE Singapore Information Technology Gold Medal for being the top student in the Computer Science program in 1989.

Mong Li joined the Department of Computer Science, National University of Singapore, as a Senior Tutor from 1989 to 1999. In 1999, she was appointed Fellow in the School of Computing and lectured Introduction to Programming in JAVA, a Lecture-on-Demand module. She was a Visiting Fellow at the Computer Science Department, University of Wisconsin-Madison and Consultant at QUIQ Incorporated, USA, from September 1999 to August 2000.

Her research interests include the cleaning and integration of heterogeneous and semistructured data, performance database issues in dynamic environments, and medical informatics. Her work has been published in ACM SIGMOD, ACM SIGKDD, VLDB, ICDE and ER conferences. She is a co-Editor for the Proceedings of the 17th International Conference on Conceptual Modeling (ER 1998) and Proceedings of VLDB 2002 Workshop EEXTT and CAiSE 2002 Workshop DIWeb (LNCS #2590, Springer-Verlag). She is a Program Committee member of VLDB (2002, 2003, 2004), DASFAA (2003, 2004) and ER (1998, 1999, 2001, 2003, 2004) and a reviewer for IEEE TKDE and DAMI journals.

Dr. Tok Wang LING is a Professor of the Department of Computer Science, School of Computing at the National University of Singapore, Singapore. He was the Head of Information Technology Division, Deputy Head of the Department of Information Systems and Computer Science, and a Vice Dean of the School. Before joining the University as a lecturer in 1979, he was a scientific staff at Bell Northern Research, Ottawa, Canada. He received his Ph.D. and M.Math., both in Computer Science, from Waterloo University, Canada, and B.Sc.(1 Hons) in Mathematics from Nanyang University, Singapore.

His research interests include Data Modeling, Entity-Relationship Approach, Object-Oriented Data Model, Normalization Theory, Logic and Database, Integrity Constraint Checking, Semistructured Data Model, and Data Warehousing. He has published more than 150 international journal/conference papers and chapters in books, mainly in data modeling. He also co-edited 12 conference and workshop proceedings.

He organized and served as program committee co-chair of DASFAA'95, DOOD'95, ER'98, WISE 2002, and ER 2003. He organized and served/serves as conference co-chair of Human.Society @Internet conference in 2001 and 2003, WAIM 2004, ER 2004, and DASFAA 2005. He served/serves as workshop co-chair of DOOD'95 Post-Conference Workshops, the 8th International Parallel Computing Workshop, and the International Workshop on Conceptual Model-directed Web Information Integration and Mining held in conjunction with ER 2004.

He serves/served on the program committees of more than 100 international database conferences since 1985. He is currently the chair of the steering committee of International Conference on Database Systems for Advanced Applications (DASFAA), a member of the steering committee of International Conference on Conceptual Modeling (ER) and the International Conference on Human.Society@Internet. He was chair and vice chair of the steering committee of ER conference and a member of the steering committee of International Conference on Deductive and Object-Oriented Databases (DOOD).

He is an editor of the journal Data & Knowledge Engineering, International Journal of Cooperative Information Systems, Journal of Database Management, Journal of Data Semantics, and World Wide Web: Internet and Web Information Systems. He is also an advisor of the ACM Transactions on Internet Technology. He is a member of ACM, IEEE, and Singapore Computer Society.