CSE-291 Spring'03: Ontologies in Data Integration

Bertram Ludäscher

April 29, 2003

1 Final Set of Topics

1. (1 T) *Model Management.*

   Present Koch's approach to model management [KOCH, 2001, Chapter 6] and its foundations (query writing). Briefly compare with other current approaches to model management, e.g.

   [MELNIK *et al.*, 2003], [MADHAVAN *et al.*, 2002], [POTTINGER & BERNSTEIN, 2003].

2. (1 T) *Biological Taxonomies: The Prometheus Taxonomic Model and the LITCHI Approach*

   Describe the Prometheus taxonomic model and the approach taken in the LITCHI project.

   [PULLAN *et al.*, 2000], [EMBURY *et al.*, 1999], [EMBURY *et al.*, 2001]

3. (2 T+P) *Reasoning with Ontologies.*

   Give an overview of the FaCT description logic reasoner [HORROCKS, 1999] (what language does it handle, what is the basic underlying calculus used, etc.) and demonstrate examples of using FaCT in class. (If possible use in connection with [OIL, 2002])

4. (4 T) *Comparative Analysis of Semantic Web Languages: OWL, RDF(S), and related formalisms.*

   Study the logic formalisms underlying OWL (description logics, first-order) and RDF(S). Provide a comparative analysis of the features of the languages (e.g., expressiveness, querying and reasoning support)

   [BAADER *et al.*, 2003] [OWL, 2003] [HORROCKS, 2003] [RDF, 2003]

5. (4 T+P) *Formal Concept Analysis: Introduction and Applications.*

   Provide an introduction to FCA and illustrate some applications. Demonstrate an example (e.g., using Toscana)

   [GANTER & WILLER,] [BURMEISTER, 2003] [GANTER & WILLE, 1999] [TOS, 2003]

6. (4 P+T) *Semantic Web Tools.*

   Study OilEd and Protégé and do a modeling exercise (e.g., a part of [STRUIK *et al.*, 2002] or [NOY & MCGUINNESS, 2001]) in each of them; present the main features of the systems, their underlying formalism, and a demo of the modeling exercise in class.

   [BECHHOFER *et al.*, 2001], [OIL, 2002], [PRO, 2003]

7. (4 T) *Ontologies in Data Integration.*

   Present the ontology-based mediation approach by Tzitzikas et al. [TZITZIKAS *et al.*, 2002], [TZITZIKAS *et al.*, 2001] and the one used in TAMBIS [PEIM *et al.*, 2002], [GOBLE *et al.*, 2001].

8. (2 P+T) *The GeoReference Online System and Approach.*

   Present the underlying theory of the model matching approach used by GeoReference Online [LIMITED, 2002], and demonstrate examples in class.

2 Prelimimary Set of Topics

"T" and "P" stand for *theory* (literature) and *practice* studies (hands-on/modeling experiment), respectively. The references given under each topic are preliminary and may be reduced/extended/changed somewhat, depending on the detailed assignment.

- **Ontology Foundations and Formal Ontologies**, (1–4 units T)

  [GRUBER, 1993], [GUARINO & GIARETTA, 1995], [USCHOLD & GRÜNINGER, 1996], [GUARINO, 1997], [GUARINO, 1998], [SOWA, 2000a, Chapter 2], [SOWA, 2000b], [SOWA, 2000c] [SOWA, 2002]

- **Ontologies and Information Integration**, (1–2 units T)

  [WACHE *et al.*, 2001], [STUCKENSCHMIDT, 2003b], [STUCKENSCHMIDT, 2003a], [DA SILVA *et al.*, 2002], [PEIM *et al.*, 2002]

  [ANDRADE & SALTZ, 2000] [BRILHANTE & ROBERTSON, 2001]

- **Conceptual Modeling and Ontologies for Geologic Maps, Geospatial and Geologic Ontologies** (1–4 units T+P)

  [WORBOYS & DUCKHAM, 2002]

  [BRODARIC & GAHEGAN, 2002]

  [VISSER *et al.*, 2002] [FONSECA *et al.*, 2002]

  Rock classification (T+P) [STRUIK *et al.*, 2002]

- **Biological Taxonomies** (1–4 units T)

  [EMBURY *et al.*, 1999], [EMBURY *et al.*, 2001], [PULLAN *et al.*, 2000], ...

  [GRAHAM, 2001]

- **Biological Pathways and other Biological Ontologies and Graph Databases** (1–4 units T+P)

  EcoCyc, BioCyc: [KARP, 1999], [KARP, 2000], [KARP, 2001], [BIO, 2003]

  Gene Ontology: [CONSORTIUM, 2002a]

  Conceptual Modeling of Genomic Information: [PATON *et al.*, 2000]

  UMLS: [UML, 2003b], [UML, 2003a]

  [MCENTIRE *et al.*, 2000], [OLKEN, 2003]

  [KRISHNAMURTHY *et al.*, 2003]

- **Semantic Web Formalisms and Tools** (1–4 units T+P)

  [NOY & MCGUINNESS, 2001], [FALKOVYCH *et al.*, 2003], [OIL, 2002], [BECHHOFER *et al.*, 2001] [PRO, 2003]

  [RDF, 2003], [OWL, 2003] [HORROCKS, 2003]

  [CONSORTIUM, 2002b], [ANGELE & SURE, 2002]

- **Topic Maps** (1 T+P)

  [XTM, 2001], ...

- **Formal Concept Analysis** (1–4 units T+P)

  [GANTER & WILLE, 1999], [TOS, 2003]

- **Reasoning with Ontologies** (1–4 units T+P)

  [HORROCKS, 1999]

  [OIL, 2002]

- **Model Management** (1–4 units T)

  [MADHAVAN *et al.*, 2002], [MELNIK *et al.*, 2003], [POTTINGER & BERNSTEIN, 2003],

  [KOCH, 2001, Chapter 6]

3 Some Mathematical/Logic Background

[SOWA, 2001], [SHAPIRO, 2000], [HODGES, 2001]

References

[ANDRADE & SALTZ, 2000] Henrique Andrade & Joel Saltz. Query Optimization in Kess – An Ontology-Based KBMS. In *XIV Brazilian Symposium on Databases (SBBD)*, 2000. `http://citeseer.nj.nec.com/226856.html`.

[ANGELE & SURE, 2002] Jürgen Angele & York Sure, editors. *EON2002 Evaluation of Ontology-based Tools, EKAW02 Workshop Proceedings*, Sigüenza, Spain, September 2002. `http://sunsite.informatik.rwth-aachen.de/Publications/CEUR-WS/Vol-62/`.

[BAADER *et al.*, 2003] Franz Baader, Ian Horrocks, & Ulrike Sattler. Description Logics as Ontology Languages for the Semantic Web. In Dieter Hutter & Werner Stephan, editors, *Festschrift in honor of Jörg Siekmann*, LNAI. Springer, 2003.

[BAIN & GILES, 1997] K.A. Bain & J.R.A. Giles. A Standard Model for Storage of Geologic Map Data. *Computers and Geosciences*, 23:613–620, 1997.

[BECHHOFER *et al.*, 2001] Sean Bechhofer, Ian Horrocks, Carole Goble, & Robert Stevens. OilEd: a Reason-able Ontology Editor for the Semantic Web. In *Proc. KI2001, Joint German/Austrian conference on Artificial Intelligence*, LNAI 2174, 2001. `http://potato.cs.man.ac.uk/papers/ki2001.pdf`.

[BIO, 2003] BioCyc Knowledge Library, 2003. http://biocyc.org/.

[BRILHANTE & ROBERTSON, 2001] Virgínia Brilhante & David Stuart Robertson. Metadata-Supported Automated Ecological Modelling. In Claus Rautenstrauch & Susanne Patig, editors, *Environmental Information Systems in Industry and Public Administration*. Idea Group Publishing, Hershey PA, 2001. http://www.dai.ed.ac.uk/groups/ssp/psfiles/virginia/EnvIS_chap.aw.ps.

[BRODARIC & GAHEGAN, 2002] Boyan Brodaric & Mark Gahegan. Distinguishing Instances and Evidence of Geographical Concepts for Geospatial Database Design. In Egenhofer & Mark [EGENHOFER & MARK, 2002].

[BRODARIC & HASTINGS, 2002] Boyan Brodaric & Jordan Hastings. An Object Model for Geologic Map Information. In *Proc. Spatial Data Handling Symposium*, Ottawa, Canada, July 2002.

[BURMEISTER, 2003] Peter Burmeister. Formal Concept Analysis with ConImp: Introduction to the Basic Features. http://www.mathematik.tu-darmstadt.de/~burmeister/ConImpIntro.ps, 2003.

[CONSORTIUM, 2002a] Gene Ontology Consortium. GO, 2002. http://www.geneontology.org/.

[CONSORTIUM, 2002b] OntoWeb Consortium. A survey on ontology tools. OntoWeb Deliverable 1.3. Asunción Gómez Pérez editor, 2002. http://citeseer.nj.nec.com/525623.html.

[DA SILVA *et al.*, 2002] Flávio Corrêa da Silva, Wamberto Weber Vasconcelos, David Stuart Robertson, Virginia Brilhante, Ana de Melo, Marcelo Finger, & Jaume Agustí. On the Insufficiency of Ontologies: Problems in Knowledge Sharing and Alternative Solutions. *Knowledge-Based Systems Journal*, 15(3):147–167, 2002. http://citeseer.nj.nec.com/382117.html.

[EGENHOFER & MARK, 2002] Max J. Egenhofer & David M. Mark, editors. *Geographic Information Science, 2nd Intl. Conference (GIScience)*, number 2478 in LNCS, Boulder, CO, September 2002. Springer. http://link.springer-ny.com/link/service/series/0558/tocs/t2478.htm.

[EMBURY *et al.*, 1999] Suzanne M. Embury, Andrew C. Jones, Iain Sutherland, W. A. Gray, Richard J. White, John S. Robinson, Frank A. Bisby, & Sue M. Brandt. Conflict Detection for Integration of Taxonomic Data Sources. In *Intl. Conf. on Scientific and Statistical Database Management, (SSDBM)*, Cleaveland, Ohio, 1999. http://citeseer.nj.nec.com/embury99conflict.html.

[EMBURY *et al.*, 2001] Suzanne M. Embury, Sue M. Brandt, John S. Robinson, Iain Sutherland, Frank A. Bisby, W. Alex Gray, Andrew C. Jones, & Richard J. White. Adapting Integrity Enforcement Techniques for Data Reconciliation. *Information Systems*, 26(8):657–689, 2001. http://dx.doi.org/10.1016/S0306-4379(01)00044-8.

[FALKOVYCH *et al.*, 2003] K. Falkovych, M. Sabou, & H. Stuckenschmidt. UML for the Semantic Web: Transformational Approaches. In B. Omelayenko & M. Klein, editors, *Knowledge Transformation for the Semantic Web*. IOS Press, 2003. http://www.cs.vu.nl/~heiner/public/KTSW.pdf.

[FONSECA *et al.*, 2002] Frederico Fonseca, James Martin, & M. Andrea Rodríguez. From Geo- to Eco-Ontologies. In Egenhofer & Mark [EGENHOFER & MARK, 2002].

[GANTER & WILLE, 1999] Bernhard Ganter & Rudolf Wille. *Formal Concept Analysis – Mathematical Foundations*. Springer, 1999. http://www.springer.de/cgi-bin/search_book.pl?isbn=3-540-62771-5.

[GANTER & WILLER,] Bernhard Ganter & Rudolf Willer. Applied Lattice Theory: Formal Concept Analysis. http://www.math.tu-dresden.de/~ganter/psfiles/concept.ps.

[GOBLE *et al.*, 2001] C. Goble, R. Stevens, G. Ng, S. Bechhofer, N. Paton, P. Baker, M. Peim, & A. Brass. Transparent Access to Multiple Bioinformatics Information Sources. *IBM Systems Journal*, 40(2):534–551, 2001. http://www.research.ibm.com/journal/sj/402/goble.pdf.

[GRAHAM, 2001] Martin James Graham. *Visualising Multiple Overlapping Classification Hierarchies*. PhD thesis, Napier University, UK, 2001. http://citeseer.nj.nec.com/545920.html.

[GRUBER, 1993] T. R. Gruber. A Translation Approach to Portable Ontology Specifications. *Knowledge Acquisition*, 5:199–220, 1993. http://ksl-web.stanford.edu/KSL_Abstracts/KSL-92-71.html.

[GUARINO & GIARETTA, 1995] Nicola Guarino & Pierdaniele Giaretta. Ontologies and Knowledge Bases: Towards a Terminological Clarification. In N.J.I Mars, editor, *Towards Very Large Knowledge Bases*, pages 25–32. IOS Press, Amsterdam, 1995. http://ontology.ip.rm.cnr.it/Papers/KBKS95.pdf.

[GUARINO, 1997] N. Guarino. Understanding, Building, and Using Ontologies: A Commentary to Using Explicit Ontologies in KBS Development, by van Heijst, Schreiber, and Wielinga. *International Journal of Human and Computer Studies*, 46:293–310, 1997. http://citeseer.nj.nec.com/guarino97understanding.html.

[GUARINO, 1998] N. Guarino. Formal Ontology and Information Systems. In N. Guarino, editor, *Proceedings of the 1st International Conference on Formal Ontologies in Information Systems*, pages 3–15. IOS Press, 1998. http://citeseer.nj.nec.com/guarino98formal.html.

[HODGES, 2001] Wilfrid Hodges. Model Theory. Stanford Encyclopedia of Philosophy entry, 2001. http://plato.stanford.edu/entries/model-theory/.

[HORROCKS, 1999] Ian Horrocks. The FaCT System, 1999. http://www.cs.man.ac.uk/~horrocks/FaCT/.

[HORROCKS, 2003] Ian Horrocks. Logical Foundations for the Semantic Web. Seminar given at the University of Glasgow, March 2003. http://www.cs.man.ac.uk/~horrocks/Slides/glasgow.pdf.

[KARP, 1999] Peter D. Karp. EcoCyc: The Resource and the Lessons Learned. In *Bioinformatics Databases and Systems*. Kluwer, 1999. http://www.ai.sri.com/pkarp/pubs/ecocyc-lessons.ps.

[KARP, 2000] Peter D. Karp. An Ontology for Biological Function Based on Molecular Interactions. *Bioinformatics*, 16(3):269–285, 2000. http://www.ai.sri.com/pubs/files/887.ps.

[KARP, 2001] Peter D. Karp. Pathway Databases: A Case Study in Computational Symbolic Theories. *Science*, 293:2040–2044, 2001. http://www.ai.sri.com/pubs/full.php?id=880.

[KOCH, 2001] Christoph Koch. *Data Integration against Multiple Evolving Autonomous Schemata*. PhD thesis, Technische Universität Wien, Austria, 2001. http://www.dbai.tuwien.ac.at/staff/koch/download/thesis_20010516_1500_final.pdf.

[KRISHNAMURTHY *et al.*, 2003] L. Krishnamurthy, J. Nadeau, G. Ozsoyoglu, M. Ozsoyoglu, G. Schaeffer, M. Tasan, & W. Xu. Pathways Database System: An integrated set of tools for biological pathways. In *ACM Symposium on Applied Computing (SAC)*, 2003. http://nashua.cwru.edu/pathways/Sac%202003.pdf.

[KUHN, 2002] Werner Kuhn. Modeling the Semantics of Geographic Categories through Conceptual Integration. In Egenhofer & Mark [EGENHOFER & MARK, 2002].

[LIMITED, 2002] GeoReference Online Limited. Matching in MineMatch and LegendBurster: Theory and Practice, 2002. http://www.georeferenceonline.com/LegendBurster/MatcherTheoryAndPractice.html.

[MADHAVAN *et al.*, 2002] J. Madhavan, P. A. Bernstein, P. Domingos, & A.Y. Halevy. Representing and Reasoning About Mappings between Domain Models. In *18th National Conference on Artificial Intelligence (AAAI)*, 2002. http://www.cs.washington.edu/homes/jayant/Pubs/SemanticsAAAI02.pdf.

[MCENTIRE *et al.*, 2000] R. McEntire, P. Karp, N. Abernethy, D. Benton, G. Helt, M. DeJongh, R. Kent, A. Kosky, S. Lewis, D. Hodnett, E. Neumann, F. Olken, D. Pathak, P. Tarczy-Hornoch, L. Toldo, & T. Topaloglou. An Evaluation of Ontology Exchange Languages for Bioinformatics. In *Proc. Conference on Intelligent Systems for Molecular Biology*, 2000. http://www.ai.sri.com/pubs/files/888.pdf.

[MELNIK *et al.*, 2003] Sergey Melnik, Erhard Rahm, & Philip A. Bernstein. Rondo: A Programming Platform for Generic Model Management. In *ACM Intl. Conference on Management of Data (SIGMOD)*, San Diego, CA, 2003. extended version: http://dol.uni-leipzig.de/pub/2003-3/en.

[NOY & MCGUINNESS, 2001] Natalya F. Noy & Deborah L. McGuinness. Ontology Development 101: A Guide to Creating Your First Ontology. Technical report, Knowledge Systems Laboratory, Stanford, 2001. http://protege.stanford.edu/publications/ontology_development/ontology101-noy-mcguinness.html.

[OIL, 2002] OilEd Project Page, 2002. http://oiled.man.ac.uk/index.shtml.

[OLKEN, 2003] Frank Olken. Biopathways Graph Data Manager (BGDM), 2003. http://pueblo.lbl.gov/~olken/graphdm/graphdm.htm.

[OWL, 2003] OWL Web Ontology Language Reference W3C Working Draft, 31 March 2003. http://www.w3.org/TR/owl-ref/.

[PATON *et al.*, 2000] Norman W. Paton, Shakeel A. Khan, Andrew Hayes, Fouzia Moussouni, Andy Brass, Karen Eilbeck, Carol A. Goble, Simon J. Hubbard, & Stephen G. Oliver. Conceptual Modelling of Genomic Information. *Bioinformatics*, 16(6):548–557, 2000. http://bioinformatics.oupjournals.org/cgi/reprint/16/6/548.pdf.

[PEIM *et al.*, 2002] Martin Peim, Enrico Franconi, Norman W. Paton, & Carole A. Goble. Query Processing with Description Logic Ontologies Over Object-Wrapped Databases. In *14th Intl. Conference on Scientific and Statistical Database Management (SSDBM)*, Edinburgh, Scotland, 2002. http://citeseer.nj.nec.com/peim01query.html.

[POTTINGER & BERNSTEIN, 2003] R.A. Pottinger & P. A. Bernstein. Merging Models Based on Given Correspondences. Technical Report UW-CSE-03-02-03, University of Washington, February 2003. ftp://ftp.cs.washington.edu/tr/2003/02/UW-CSE-03-02-03.pdf.

[PRO, 2003] Protégé 2000 Project Page, 2003. http://protege.stanford.edu/.

[PULLAN *et al.*, 2000] Martin R. Pullan, Mark F. Watson, Jessie B. Kennedy, Cédric Raguenaud, & Roger Hyam. The Prometheus Taxonomic Model: A Practical Approach to Representing Multiple Taxonomies. *Taxon*, 49(1):55–75, 2000. http://www.dcs.napier.ac.uk/~prometheus/prometheus_1/publications/ntct.ps.

[RDF, 2003] RDF Primer, W3C Working Draft, 23 January 2003. http://www.w3.org/TR/rdf-primer/.

[SHAPIRO, 2000] Stewart Shapiro. Classical Logic. Stanford Encyclopedia of Philosophy entry, 2000. http://plato.stanford.edu/entries/logic-classical/.

[SOWA, 2000a] John F. Sowa. *Knowledge Representation: Logical, Philosophical, and Computational Foundations*. Brooks Cole Publishing Co, 2000. see also http://users.bestweb.net/~sowa/krbook/ for the author's web site of the book, and http://www.cs.buffalo.edu/~shapiro/Papers/J01-2006.pdf for a review by Stuart C. Shapiro.

[SOWA, 2000b] John F. Sowa. Ontology. web resource, 2000. http://users.bestweb.net/~sowa/ontology/.

[SOWA, 2000c] John F. Sowa. Ontology, Metadata, and Semiotics, 2000. http://users.bestweb.net/~sowa/peirce/ontometa.htm.

[SOWA, 2001] John F. Sowa. Mathematical Background: Sets, Bags, and Sequences; Functions; Lambda Calculus; Graphs; Relations; Representing Relations by Graphs; Lattices; Propositional Logic; Predicate Logic; Axioms and Proofs; Formal Grammars; Game Graphs; Model Theory. web resource, 2001. http://users.bestweb.net/~sowa/misc/math.htm.

[SOWA, 2002] John F. Sowa. Signs, Processes, and Language Games – Foundations for Ontology. http://www.jfsowa.com/pubs/signproc.htm, 2002.

[STRUIK *et al.*, 2002] L.C. Struik, M.B. Quat, P.H Davenport, & A.V. Okulitch. A Preliminary Scheme for Multihierarchical Rock Classification for use with Thematic Computer-Based Query Systems. Technical Report 2002-D10, Geological Survey of Canada, Current Research, 2002. http://www.nrcan.gc.ca/gsc/bookstore/free/cr_2002/D10.pdf.

[STUCKENSCHMIDT, 2003a] Heiner Stuckenschmidt. *Ontology-Based Information Sharing in Weakly-Structure Environments*. PhD thesis, Vrije Universiteit Amsterdam, 2003. http://www.cs.vu.nl/~heiner/public/PhD.pdf.

[STUCKENSCHMIDT, 2003b] Heiner Stuckenschmidt. Query Processing on the Semantic Web. *KI – Special Issue on the Semantic Web*, 2003. to appear. http://www.cs.vu.nl/~heiner/public/KI-SW.pdf.

[TOS, 2003] ToscanaJ SourceForge Project, 2003. http://toscanaj.sourceforge.net/.

[TZITZIKAS *et al.*, 2001] Yannis Tzitzikas, Nicolas Spyratos, & Panos Constantopoulos. Mediators over Ontology-based Information Sources. In *Second Intl. Conf. on Web Information Systems Engineering (WISE)*, 2001. http://citeseer.nj.nec.com/tzitzikas01mediators.html.

[TZITZIKAS *et al.*, 2002] Yannis Tzitzikas, Nicolas Spyratos, & Panos Constantopoulos. Translation for Mediators over Ontology-based Information Sources. In *Second Hellenic Conf. on Artificial Intelligence (SETN)*, 2002. citeseer.nj.nec.com/tzitzikas02query.html.

[UML, 2003a] UMLS Knowledge Sources, 14th Edition – January Release 2003AA, Documentation, 2003. http://www.nlm.nih.gov/research/umls/UMLSDOC.HTML.

[UML, 2003b] Unified Medical Language System (UMLS). National Library of Medicine, 2003. http://www.nlm.nih.gov/research/umls/.

[USCHOLD & GRÜNINGER, 1996] Mike Uschold & Michael Grüninger. Ontologies: Principles, Methods, and Applications. *Knowledge Engineering Review*, 11(2):93–155, 1996. http://citeseer.nj.nec.com/uschold96ontologie.html.

[VISSER *et al.*, 2002] U. Visser, H. Stuckenschmidt, G. Schuster, & T. Vogele. Ontologies for Geographic Information Processing. *Computers and Geosciences*, 28:103–117, 2002. http://www.cs.vu.nl/~heiner/public/compgeosci.pdf.

[VOISARD, 1999] Agnès Voisard. Abduction and Deduction in Geologic Hypermaps. In Ralf Hartmut Güting, Dimitris Papadias, & Frederick H. Lochovsky, editors, *Advances in Spatial Databases, 6th International Symposium, SSD'99, Hong Kong, China, July 20-23, 1999, Proceedings*, volume 1651 of *Lecture Notes in Computer Science*, pages 311–329. Springer, 1999. http://citeseer.nj.nec.com/voisard99abduction.html.

[WACHE *et al.*, 2001] H. Wache, T. Vögele, U. Visser, H. Stuckenschmidt, G. Schuster, H. Neumann, & S. Hübner. Ontology-Based Integration of Information – A Survey of Existing Approaches. In *Proc. of the IJCAI-01 Workshop: Ontologies and Information Sharing*, 2001. http://www.cs.vu.nl/~heiner/public/ois-2001.pdf.

[WORBOYS & DUCKHAM, 2002] Michael Worboys & Matt Duckham. Integrating Spatio-Thematic Information. In Egenhofer & Mark [EGENHOFER & MARK, 2002].

[XTM, 2001] XML Topic Maps (XTM) 1.0 TopicMaps.Org Specification, 2001. http://www.topicmaps.org/xtm/.