## References


[ABA08] J. Atoum, D. Bader, and A. Awajan. Mining functional dependency from relational databases using equivalent classes and minimal cover. *Journal of Computer Science*, 4(6):421–426, 2008.

[Aik98] P.H. Aiken. Reverse engineering of data. *IBM Systems Journal*, 37(2):246–269, 1998.

[AK04] R. Al-Kamha. Grouping search-engine returned citations for person-name queries. Master's thesis, Brigham Young University, Department of Computer Science, June 2004.

[AK07] R. Al-Kamha. *Conceptual XML for Systems Analysis*. PhD dissertation, Brigham Young University, Department of Computer Science, June 2007.

[AKEL08] R. Al-Kamha, D.W. Embley, and S.W. Liddle. Foundational data modeling and schema transformations for XML data engineering. In *Proceedings of the 2nd International United Information Systems Conferences (UNISCON'08)*, pages 25–36, Klagenfurt, Austria, April 2008.

[AM04] M. Al-Muhammed. Dynamic matchmaking between messages and services in multi-agent systems. Master's thesis, Brigham Young University, Department of Computer Science, May 2004.

[AM07] M.J. Al-Muhammed. *Ontology Aware Software Service Agents: Meeting Ordinary User Needs on the Semantic Web*. PhD dissertation, Brigham Young University, Provo, Utah, August 2007.

[AME07] M. Al-Muhammed and D.W. Embley. Ontology-based constraint recognition for free-form service requests. In *Proceedings of the 23rd International Conference on Data Engineering (ICDE'07)*, pages 366–375, Istanbul, Turkey, April 2007.

[AriBC] Aristotle. *Metaphysics*. Oxford University Press, New York, about 350BC. (1993 translation).

[Ast04] I. Astrova. Reverse engineering of relational databases to ontologies. In *Proceedings of the First European Semantic Web Symposium*, pages 327–341, Heraklion, Crete, Greece, May 2004.

[BCHS09] P. Buitelaar, P. Cimiano, P. Haase, and M. Sintek. Towards linguistically grounded ontologies. In *Proceedings of the 6th European Semantic Web Conference (ESWC'09)*, pages 111–125, Heraklion, Greece, May/June 2009.

[BE03] J. Biskup and D.W. Embley. Extracting information from heterogeneous information sources using ontologically specified target views. *Information Systems*, 28(3):169–212, 2003.

[BL07] T. Berners-Lee. Future of the world wide web, March 2007. Testimony of Sir Timothy Berners-Lee Before the United States House of Representatives Committee on Energy and Commerce Subcommittee on Telecommunications and the Internet.

[Cha03] T. Chartrand. Ontology-based extraction of RDF data from the world wide web. Master's thesis, Brigham Young University, Provo, Utah, March 2003.

[Che04] X. Chen. Query rewriting for extracting data behind thml forms. Master's thesis, Brigham Young University, Department of Computer Science, March 2004.

[Cim06] P. Cimiano. *Ontology Learning and Population from Text: Algorithm, Evaluation and Applications*. Springer Verlag, New York, New York, 2006.

[DEG] Homepage for BYU Data Extraction Group. www.deg.byu.edu.

[DEL06] Y. Ding, D.W. Embley, and S.W. Liddle. Automatic creation and simplified querying of semantic web content: An approach based on information-extraction ontologies. In *Proceedings of the First Asian Semantic Web Conference (ASWC'06)*, pages 400–414, Beijing, China, September 2006.

[Din03] Y. Ding. Semiautomatic generation of relilient data-extraction ontologies. Master's thesis, Brigham Young University, Provo, Utah, June 2003.

[DRS09] N. Dalvi, C. Ré, and Dan Suciu. Probabilistic databases: Diamonds in the dirt. *Communications of the ACM*, 52(7):86–94, July 2009.

[ECJ+99] D.W. Embley, D.M. Campbell, Y.S. Jiang, S.W. Liddle, D.W. Lonsdale, Y.-K. Ng, and R.D. Smith. Conceptual-model-based data extraction from multiple-record web pages. *Data & Knowledge Engineering*, 31(3):227–251, November 1999.

[EHLN06] D.W. Embley, M. Hurst, D. Lopresti, and G. Nagy. Table-processing paradigms: A research survey. *International Journal of Document Analysis*, 8(2):66–86, 2006.

[EIV07] A.K. Elmagarmid, P.G. Ipeirotis, and V.S. Verykios. Duplicate record detection: A survey. *IEEE Transactions on Knowledge and Data Engineering*, 18(1):1–16, January 2007.

[EJX01] D.W. Embley, D. Jackman, and L. Xu. Multifaceted exploitation of metadata for attribute match discovery in information integration. In *Proceedings of the International Workshop on Information Integration on the Web (WIIW'01)*, pages 110–117, Rio de Janeiro, Brazil, April 2001.

[EJX02] D.W. Embley, D. Jackman, and L. Xu. Attribute match discovery in information integration: Exploiting multiple facets of metadata. *Journal of the Brazilian Computing Society*, 8(2):32–43, November 2002.

[EKW92] D.W. Embley, B.D. Kurtz, and S.N. Woodfield. *Object-oriented Systems Analysis: A Model-Driven Approach*. Prentice Hall, Englewood Cliffs, New Jersey, 1992.

[Emb80] D.W. Embley. Programming with data frames for everyday data items. In *Proceedings of the 1980 National Computer Conference*, pages 301–305, Anaheim, California, May 1980.

[ES07] J. Euzenat and P. Shvaiko. *Ontology Matching*. Springer-Verlag, Heidelberg, Germany, 2007.

[EX97] D.W. Embley and M. Xu. Relational database reverse engineering: A model-centric, transformational, interactive approach formalized in model theory. In *DEXA'97 Workshop Proceedings*, pages 372–377, Toulouse, France, September 1997.

[EXD04] D.W. Embley, L. Xu, and Y. Ding. Automatic direct and indirect schema mapping: Experiences and lessons learned. *SIGMOD Record*, 33(4):14–19, December 2004.

[EZ10] D.W. Embley and A. Zitzelberger. Theoretical foundations for enabling a web of knowledge. In *Proceedings of the Sixth International Symposium on Foundations of Information and Knowledge Systems (FoIKS10)*, Sophia, Bulgaria, February 2010. (to appear).

[Fel98] C. Fellbaum. *WordNet: An Electronic Lexical Database*. MIT Press, Cambridge, Massachussets, 1998.

[GMJ04] A. Gal, G.A. Modica, and H.M. Jamil. Ontobuilder: Fully automatic extraction and consolidation of ontologies from web sources. In *Proceedings of the 20th International Conference on Data Engineering*, page 853, Boston, Massachusetts, March/April 2004.

[HLF+08] L. Hunter, Z. Lu, J. Firby, W.A. Baumgartner Jr., H.L. Johnson, P.V. Ogren, and K.B. Cohen. OpenDMAP: An open source, ontology-driven, concept analysis engine, with applications to capturing knowledge regarding protein transport, protein interactions and cell-type-specific gene expression. *BMC Bioinformatics*, 9(8), 2008.

[Jac02] D. Jackman. Mapping target schemas to source schemas using WordNet hierarchies and structure context. Master's thesis, Brigham Young University, Provo, Utah, June 2002.

[Jha08] P. Jha. A layout independent representation of tables. Master's thesis, Department of Electrical, Computer, and Systems Engineering, Rensselaer Polytechnic Institute, Troy, New York, May 2008.

[JN08] P. Jha and G. Nagy. Wang notation tool: Layout independent representation of tables. In *Proceedings of the 19th International Conference on Pattern Recognition (ICPR08)*, Tampa, Florida, December 2008.

[LE09] S. Lynn and D.W. Embley. Semantically conceptualizing and annotating tables. In *Proceedings of the Third Asian Semantic Web Conference*, pages 345–359, Bangkok, Thailand, February 2009.

[LED+09] D.W. Lonsdale, D.W. Embley, Y. Ding, L. Xu, and M. Hepp. Reusing ontologies and language components for ontology generation. *Data & Knowledge Engineering*, 2009. (in press: doi:10.1016/j.datak.2009.08.003).

[Lia08] Z. Lian. A tool to support ontology creation based on incremental mini-ontology merging. Master's thesis, Department of Computer Science, Brigham Young University, Provo, Utah, March 2008.

[LMK03] K. Lerman, S.N. Minton, and C.A. Knoblock. Wrapper maintenance: A machine learning approach. *Journal of Artifical Intelligence Research*, 18:149–181, Feburary 2003.

[LTPE08] D.W. Lonsdale, C. Tustison, C. Parker, and D.W. Embley. Assessing clinical trial eligibility with logic formula queries. *Data & Knowledge Engineering*, 66(1):3–17, 2008.

[Lyn08] S. Lynn. Automating mini-ontology generation from canonical tables. Master's thesis, Department of Computer Science, Brigham Young University, Provo, Utah, 2008.

[ME06] W.Y. Mok and D.W. Embley. Generating compact redundancy-free XML documents from concptual-model hypergraphs. *IEEE Transactions on Knowledge and Data Engineering*, 18(8):1082–1096, August 2006.

[MHL03] X. Meng, D. Hu, and C. Li. Schema-guided wrapper maintenance for web-data extraction. In *Proceedings of the Fifth ACM International Workshop on Web Information and Data Management*, 2003.

[Min75] M. Minsky. A framework for representing knowledge. In P.H. Winston, editor, *The Psychology of Computer Vision*, pages 211–277. McGraw-Hill, 1975.

[NSD+01] N.F. Noy, M. Sintek, S. Decker, M. Crubezy, R.W. Fergerson, and M. Musen. Creating semantic web contents with Protègè-2000. *IEEE Intelligent Systems*, 16(2):60–71, March–April 2001.

[Pad09] R.K. Padmanabhan. Table abstraction tool. Master's thesis, Department of Electrical, Computer, and Systems Engineering, Rensselaer Polytechnic Institute, Troy, New York, May 2009.

[Par05] C.G. Parker. Generating medical logic modules for clinical trial eligibility. Master's thesis, Brigham Young University, Provo, Utah, November 2005.

[PJK+09] R. Padmanabhan, R.C. Jandhyala, M. Krishnamoorthy, G. Nagy, S. Seth, and W. Silversmith. Interactive conversion of large web tables. In *Proceedings of Eighth International Workshop on Graphics Recognition (GREC 2009)*, La Rochelle, France, July 2009.

[PlaBC] Plato. *Theaetetus*. BiblioBazaar, LLC, Charleston, South Carolina, about 360BC. (translated by Benjamin Jowett).

[PSC+07] A. Pivk, Y. Sure, P. Cimiano, M. Gams, V. Rajkovič, and R. Studer. Transforming arbitrary tables into logical form with TARTAR. *Data & Knowledge Engineering*, 60:567–595, 2007.

[Ros05] R. Rosati. On the decidability and complexity of integrating ontologies and rules. *Journal of Web Semantics*, 3(1):61–73, 2005.

[Sar08] S. Sarawagi. Information extraction. *Foundations and Trends in Databases*, 1(3):261–377, 2008.

[SWL09] W. Su, J. Wang, and F. Lochovsky. ODE: Ontology-assisted data extraction. *ACM Transactions on Database Systems*, 34(2):12.1–12.35, June 2009.

[Tao03] C. Tao. Schema matching and data extraction over HTML tables. Master's thesis, Brigham Young University, Department of Computer Science, September 2003.

[Tao08] C. Tao. *Ontology Generation, Information Harvesting and Semantic Annotation for Machine-Generated Web Pages*. PhD dissertation, Brigham Young University, Department of Computer Science, December 2008.

[TE07] C. Tao and D.W. Embley. Automatic hidden-web table interpretation by sibling page comparison. In *Proceedings of the 26th International Conference on Conceptual Modeling*, pages 556–581, Auckland, New Zealand, November 2007.

[TE09] C. Tao and D.W. Embley. Automatic hidden-web table interpretation, conceptualization, and semantic annotation. *Data & Knowledge Engineering*, 68(7):683–703, July 2009.

[TEL+05] Y.A. Tijerino, D.W. Embley, D.W. Lonsdale, Y. Ding, and G. Nagy. Toward ontology generation from tables. *World Wide Web: Internet and Web Information Systems*, 8(3):261–285, September 2005.

[TEL09] C. Tao, D.W. Embley, and S.W. Liddle. FOCIH: Form-based ontology creation and information harvesting. In *Proceedings of the 28th International Conference on Conceptual Modeling (ER2009)*, pages 346–359, Gramado, Brazil, November 2009.

[Vic06] M. Vickers. Ontology-based free-form query processing for the semantic web. Master's thesis, Brigham Young University, Provo, Utah, June 2006.

[W3C] W3C (World Wide Web Consortium) *Semantic Web Activity Page*. http://www.w3.org/2001/sw/.

[Wal04] T. Walker. Automating the extraction of domain-specific information from the web—a case study for the genealogical domain. Master's thesis, Brigham Young University, Department of Computer Science, August 2004.

[Wan96] X. Wang. *Tabular Abstraction, Editing, and Formatting*. PhD thesis, University of Waterloo, 1996.

[Wes05] A. Wessman. A framework for extraction plans and heuristics in an ontology-based data-extraction system. Master's thesis, Brigham Young University, Department of Computer Science, January 2005.

[WNB06] Y. Weidong, G. Ning, and S. Baile. Reverse engineering XML. In *Proceedings of the First International Multi-Symposiums on Computer and Computational Sciences (IMSCCS'06)*, volume 2, pages 447–454, Hangzhou, Zhejiang, China, June 2006.

[XE06] L. Xu and D.W. Embley. A composite approach to automating direct and indirect schema mappings. *Information Systems*, 31(8):697–732, December 2006.

[XE08] L. Xu and D.W. Embley. Categorization of web documents using extraction ontologies. *International Journal of Metadata, Semantics and Ontologies*, 3(1):3–20, 2008.

[Xu03] L. Xu. *Source Discovery and Schema Mapping for Data Integration*. PhD dissertation, Brigham Young University, August 2003.

[Yau01] S.H. Yau. Automating the extraction of data behind web forms. Master's thesis, Brigham Young University, Provo, Utah, December 2001.

[Zho05] Y. Zhou. Generating data-extraction ontologies by example. Master's thesis, Brigham Young University, Department of Computer Science, December 2005.