References

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

[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.

[BLHL01]   T. Berners-Lee, J. Hendler, and O. Lassila. The semantic web. *Scientific American*, 36(25):34–43, May 2001.

[BSST40]   R.L. Brooks, C.A.B. Smith, A.H. Stone, and W.T. Tutte. The dissection of rectangles into squares. *Duke Mathematics Journal*, 7:312–340, 1940.

[Buc06]   B.G. Buchanan. What do we know about knowledge? *AI Magazine*, 27(4):35–46, Winter 2006.

[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.

[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.

[ELL$^+$08] D.W. Embley, S.W. Liddle, D. Lonsdale, G. Nagy, Y. Tijerino, R. Clawson, J. Crabtree, Y. Ding, P. Jha, Z. Lian, S. Lynn, R.K. Padmanabhan, J. Peters, C. Tao, R. Watts, C. Woodbury, and A. Zitzelberger. A conceptual-model-based computational alembic for a web of knowledge. In *Proceedings of the 27th International Conference on Conceptual Modeling (ER08)*, pages 532–533, Barcelona, Spain, October 2008.

[ELN06]   D.W. Embley, D. Lopresti, and G. Nagy. Notes on contemporary table recognition. In *Proceedings of the Seventh International Association for Pattern Recognition Workshop on Document Analysis Systems*, pages 164–175, Nelson, New Zealand, February 2006.

[Emb98]   D.W. Embley. *Object Database Development: Concepts and Principles*. Addison-Wesley, Reading, Massachusetts, 1998.

[ETL05]   D.W. Embley, C. Tao, and S.W. Liddle. Automating the extraction of data from HTML tables with unknown structure. *Data & Knowledge Engineering*, 54(1):3–28, July 2005.

[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.

[GBH+07]   W. Gatterbauer, P. Bohunsky, M. Herzog, B. Krüpl, and B. Pollak. Towards domain-independent information extraction from web tables. In *Proceedings of the Sixteenth International World Wide Web Conference (WWW2007)*, pages 71–80, Banff, Alberta, Canada, May 2007.

[GO]   The gene ontology. http://www.geneontology.org/.

[Gru93]   T.R. Gruber. A translation approach to portable ontology specifications. *Knowledge Acquisition*, 5(2):199–220, 1993.

[Han01]   J.C. Handley. Table analysis for multi-line cell identification. In P.B. Kantor, D.P. Lopresti, and J. Zhou, editors, *Document Recognition and Retrieval VIII*, volume 4307 of *Proceedings of SPIE*, pages 34–43, 2001.

[HKL+01]   J. Hu, R. Kashi, D. Lopresti, G. Nagy, and G. Wilfong. Why table ground-truthing is hard. In *Proceedings of the Sixth International Conference on Document Analysis and Recognition (ICDAR'01)*, pages 129–133, Seattle, Washington, September 2001.

[HNP09]   A. Halevy, P. Norvig, and F. Pereira. The unreasonable effectiveness of data. *IEEE Intelligent Systems*, March/April 2009.

[Hur00]   M. Hurst. *The Interpretation of Tables in Texts*. PhD thesis, School of Cognitive Science, Informatics, The University of Edinburgh, United Kingdom, 2000.

[Ito93]   K. Itonori. A table structure recognition based on textblock arrangement and ruled line position. In *Proceedings of the Second International Conference on Document Analysis and Recognition (ICDAR'93)*, pages 765–768, Tsukuba Science City, Japan, October 1993.

[Jep09]   T.C. Jepsen. Just what is an ontology, anyway? *ITProfessional*, 11(5):22–27, September/October 2009.

[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.

[JNS+09] R.C. Jandhyala, G. Nagy, S. Seth, W. Silversmith, M. Krishnamoorthy, and R. Padmanabhan. From tessellations to table interpretation. In *Lecture Notes in Artificial Intelligence*, volume 5625, pages 422–437, Berlin, 2009. Springer-Verlag.

[KD98] T.G. Kieninger and A. Dengel. A paper-to-HTML table converting system. In *Proceedings of Document Analysis Systems (DAS'98)*, Nagano, Japan, 1998.

[KHG05] B. Krüpl, M. Herzog, and W. Gatterbauer. Using visual cues for extraciton of tabular data from arbitrary HTML documents. In *Proceedings of the 14th International World Wide Web Conference (WWW2005)*, Chiba, Japan, May 2005.

[KK01] S. Klink and T.G. Kieninger. Rule-based document structure understanding with a fuzzy combination of layout and textual features. *International Journal of Document Analysis and Recognition*, 4(1):18–26, 2001.

[KM88] D.A. Klarner and S.S. Magliveras. Tilings of a block with blocks. *European Journal of Combinatorics*, 9(4):317–330, July 1988.

[KNSV93] M. Krishnamoorthy, G. Nagy, S. Seth, and M. Viswanathan. Syntactic segmentation and labeling of digitized pages from technical journals. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(7):737–747, 1993.

[KO90] E.S. Kuh and T. Ohtsuki. Recent advances in VLSI layout. *Proceedings of the IEEE*, 78(2):237–263, 1990.

[Kyr90] G. Kyriazis. Table analysis. Technical report, Rensselaer Polytechnic Institute, Troy, New York, 1990. RPI DocLab Internal Report.

[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.

[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.

[LN99] D. Lopresti and G. Nagy. Automated table processing: An (opinionated) survey. In *Proceedings of the Third IAPR Workshop on Graphics Recognition*, pages 109–134, Jaipur, India, September 1999.

[LN00] D. Lopresti and G. Nagy. A tabular survey of table processing. In A.K. Chhabra and D. Dori, editors, *Graphics Recognition—Recent Advances*, Lecture Notes in Computer Science, LNCS 1941, pages 93–120. Springer Verlag, 2000.

[LV92] A. Laurentini and P. Viada. Identifying and understanding tabular material in compound documents. In *Proceedings of the Eleventh International Conference on Pattern Recognition (ICPR'92)*, pages 405–409, 1992.

[NL02] G. Nagy and D. Lopresti. Issues in ground-truthing graphic documents. In *Lecture Notes in Computer Science*, volume 2390, pages 46–66. Springer Verlag, Heidelberg, GermanyA, 2002. (selected papers from the Fourth International Workshop on Graphics Recognition).

[NS84]   G. Nagy and S. Seth. Hierarchical image representation with application to optically scanned documents. In *Proceedings of the International Conference on Pattern Recognition (ICPR'84)*, pages 347–349, Montreal, Canada, 1984.

[NV08]   G. Nagy and S. Veeramachaneni. Adaptive and interactive approaches to document analysis. In S. Marinai and H. Fujisawa, editors, *Machine Learning in Document Analysis and Recognition*, volume 90 of *Studies in Computational Intelligence*, pages 221–257. Springer Verlag, 2008.

[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.

[PC97]   P. Pyreddy and W.B. Croft. TINTIN: A system for retrieval in text tables. In *Proceedings of the 2nd ACM International Conference on Digital Libraries*, pages 193–200, Philadelphia, Pennsylvania, July 1997.

[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.

[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.

[Sam06]   H. Samet. *Foundations of Multidimensional and Metric Data Structures*. Morgan Kaufmann, 2006.

[SJT06]   A.C.e Silva, A.M. Jorge, and L. Torgo. Design of an end-to-end method to extract information from tables. *International Journal of Document Analysis and Recognition*, 8(2):144–171, 2006.

[STAa]   Canada statistics. http://www.libraries.rutgers.edu/cms/indexes/descriptions/estat, http://www40.statcan.gc.ca/l01/ind01/l3_1821_1825-eng.htm?hili_educ52.

[STAb]   http://www.indiastat.com/administrativesetup/1/stats.aspx.

[STAc]   http://www.bfs.admin.ch/bfs/portal/en/index/infothek/lexikon.html.

[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.

[UML]   Unified medical language system. http://www.nlm.nih.gov/research/umls/.

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

[XE03]   L. Xu and D.W. Embley. Using domain ontologies to discover direct and indirect matches for schema elements. In *Proceedings of the Workshop on Semantic Integration (WSI'03)*, pages 105–110, Sanibel Island, Florida, October 2003.

[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.

[ZBC04]   R. Zanibbi, D. Blostein, and J.R. Cordy. A survey of table recognition: Models, observations, transformations, and inferences. *International Journal of Document Analysis and Recognition*, 7(1):1–16, 2004.