

- 2010.
- [2] A.H.F. Laender, B.A. Ribeiro-Neto, A.S. Da Silva, and J.S. Teixeira. A brief survey of web data extraction tools. *ACM SIGMOD Record*, 31(2):84–93, June 2002.
 - [3] Robert Baumgartner, Georg Gottlob, and Marcus Herzog. Scalable web data extraction for online market intelligence. *Proceedings of the VLDB Endowment*, 2(2):1512–1523, August 2009. ISSN 2150-8097.
 - [4] L. Liu, C. Pu, and W. Han. XWRAP: An XML-enabled wrapper construction system for web information sources. In *Proceedings of the 16th International Conference on Data Engineering (ICDE)*, pages 611–621. IEEE, 2000.
 - [5] Ion Muslea, Steve Minton, and Craig Knoblock. Stalker: Learning extraction rules for semistructured, web-based information sources. In *Proceedings of AAAI-98 Workshop on AI and Information Integration*, pages 74–81, 1998.
 - [6] Chun-Nan Hsu and Ming-Tzung Dung. Generating finite-state transducers for semi-structured data extraction from the web. *Information Systems*, 23(8):521–538, 1998. ISSN 0306-4379. DOI 10.1016/S0306-4379(98)00027-1.
 - [7] A.H.F. Laender, B. Ribeiro-Neto, and A.S. Da Silva. DEByE: data extraction by example. *Data & Knowledge Engineering*, 40(2):121–154, 2002.
 - [8] Tobias Anton. XPath-wrapper induction by generalizing tree traversal patterns. In *Lernen, Wissensentdeckung und Adaptivität (LWA) 2005, GI Workshops, Saarbrücken*, pages 126–133, 2005.
 - [9] V. Crescenzi, G. Mecca, P. Merialdo, et al. Roadrunner: Towards automatic data extraction from large web sites. In *Proceedings of the International Conference on Very Large Data Bases (VLDB)*, pages 109–118, 2001.
 - [10] Nilesh Dalvi, Philip Bohannon, and Fei Sha. Robust web extraction: an approach based on a probabilistic tree-edit model. In *Proceedings of the 2009 ACM SIGMOD International Conference on Management of Data*, pages 335–348, 2009. ISBN 978-1-60558-551-2. DOI 10.1145/1559845.1559882.
 - [11] Thomas Kabisch, Eduard C. Dragut, Clement Yu, and Ulf Leser. Deep web integration with VisQI. *Proceedings of the VLDB Endowment*, 3(1-2):1613–1616, September 2010. ISSN 2150-8097.
 - [12] R.B. Trieschnigg, K.T.T.E. Tjin-Kam-Jet, and D. Hiemstra. Ranking XPath paths for extracting search result records. Technical Report TR-CTIT-12-08, Centre for Telematics and Information Technology, University of Twente, March 2012. ISSN 1381-3625.
 - [13] S.J. Grannis, J.M. Overhage, and C. McDonald. Real world performance of approximate string comparators for use in patient matching. *Medinfo*, 11(Pt 1):43–7, 2004.
 - [14] Book awards: 1001 books you must read before you die. <http://www.librarything.com/bookaward/1001+Books+You+Must+Read+Before+You+Die>, June 2012.
 - [15] Htmllcleaner. <http://htmllcleaner.sourceforge.net>, June 2012.
 - [16] Apache. Lucene core. <https://lucene.apache.org/core/>, May 2012.
 - [17] M. van Keulen. Managing uncertainty: The road towards better data interoperability. *IT - Information Technology*, 54(3):138–146, May 2012. ISSN 1611-2776. DOI 10.1524/itit.2012.0674.
 - [18] M. van Keulen and A. de Keijzer. Qualitative effects of knowledge rules and user feedback in probabilistic data integration. *The VLDB Journal*, 18(5):1191–1217, October 2009. ISSN 1066-8888. DOI 10.1007/s00778-009-0156-z.