

## REFERENCES

- [1] Ricardo Baeza-Yates and Berthier Ribeirmobileo-Neto, "Modern Information Retrieval", ACM Press/ Addison-Wesley, 1999.
- [2] Arvind Arasu, Junghoo Cho, Hector Garcia-Molina, Andreas Paepcke, and Sriram Raghavan, "Searching the Web", ACM Transactions on Internet Technology (TOIT), 1(1):2–43, August 2001.
- [3] Terrence A. Brooks, "Web Search: How the Web has changed information retrieval", Information Research, April 2003.
- [4] Michael K. Bergman, "The deep web: Surfacing hidden value", Journal of Electronic Publishing, 7(1), 2001.
- [5] Sergei Brin and Lawrence Page, "The anatomy of a large-scale hypertextual Web search engine", Computer Networks and ISDN Systems, 30(1–7):107–117, April 1998.
- [6] Ricardo Baeza-Yates, "Challenges in the interaction of information retrieval and natural language processing" In Proceedings of 5th international conference on Computational Linguistics and Intelligent Text Processing (CICLing), volume 2945 of Lecture Notes in Computer Science, pages 445–456. Springer, February 2004.
- [7] Ricardo Baeza-Yates and Carlos Castillo, "Crawling the infinite Web: five levels are enough", In Proceedings of the third Workshop on Web Graphs (WAW), volume 3243 of Lecture Notes in Computer Science, pages 156–167, Rome, Italy, October 2004. Springer.
- [8] Thanaa M. Ghanem and Walid G. Aref, "Databases deepen the Web", Computer, 37(1):116–117, 2004.
- [9] R. Kumar, P. Raghavan, S. Rajagopalan, D. Sivakumar, A. Tomkins, and E. Upfal, "Stochastic models for the web graph", In Proceedings of the 41st Annual Symposium on Foundations of Computer Science (FOCS), pages 57–65. IEEE CS Press, 2000.
- [10] Steve Lawrence and C. Lee Giles, "Accessibility of information on the web. Intelligence", 11(1):32–39, 2000.

- [11] Lipyew Lim, Min Wang, Sriram Padmanabhan, Jeffrey Scott Vitter, and Ramesh Agarwal, "Characterizing Web document change" In Proceedings of the Second International Conference on Advances in Web-Age Information Management, volume 2118 of Lecture Notes in Computer Science, pages 133–144, London, UK, July 2001. Springer-Verlag.
- [12] B. E. Brewington and G. Cybenko. "How Dynamic is the Web?" In Proceedings of the International World-Wide Web Conference, Amsterdam, The Netherlands, 2000.
- [13] J. Cho and H. Garcia-Molina. "The Evolution of the Web and Implications for an Incremental Crawler." In Proceedings of the Twenty-Sixth VLDB Conference, pp. 200–209, Cairo, Egypt, 2000.
- [14] J. Cho and H. Garcia-Molina. "Estimating Frequency of Change." Technical report, DB Group, Stanford University, Nov 2001.
- [15] J. Cho and A. Ntoulas. "Effective Change Detection Using Sampling." In Proceedings of the International Conference on Very Large Databases (VLDB), 2002.
- [16] Brightplanet's searchable databases directory. <http://www.completeplanet.com>.
- [17] Junghoo Cho, Hector Garcia-Molina, "Parallel Crawlers", *WWW2002*, May 7–11, 2002, Honolulu, Hawaii, USA.
- [18] S. Chakrabarti, K. Punera, and M. Subramanyam, "Accelerated focused crawling through online relevance feedback", In Proc. of WWW, pages 148–159, 2002.
- [19] S. Chakrabarti, M. van den Berg, and B. Dom, "Focused Crawling: A New Approach to Topic-Specific Web Resource Discovery" *Computer Networks*, 31(11-16):1623–1640, 1999.
- [20] K. C.-C. Chang, B. He, C. Li, M. Patel, and Z. Zhang, "Structured Databases on the Web: Observations and Implications", *SIGMOD Record*, 33(3):61–70, 2004.
- [21] K. C.-C. Chang, B. He, and Z. Zhang, "Toward Large-Scale Integration: Building a MetaQuerier over Databases on the Web", In Proc. of CIDR, pages 44–55, 2005.
- [22] J. Cope, N. Craswell, and D. Hawking, "Automated Discovery of Search Interfaces on the Web", In Proc. of ADC, pages 181–189, 2003.

- [23] M. Diligenti, F. Coetzee, S. Lawrence, C. L. Giles, and M. Gori, "Focused Crawling Using Context Graphs" In Proc. of VLDB, pages 527–534, 2000.
- [24] L. Gravano, P. G. Ipeirotis, and M. Sahami, "QProber: A system for automatic classification of hidden-Web databases" , ACM TOIS, 21(1):1–41, 2003.
- [25] B. He and K. C.-C. Chang, "Statistical Schema Matching across Web Query Interfaces", In Proc. of SIGMOD, pages 217–228, 2003.
- [26] H. He, W. Meng, C. T. Yu, and Z. Wu, "Automatic integration of Web search interfaces with WISE-Integrator", VLDB Journal, 13(3):256–273, 2004.
- [27] Profusion's search engine directory. <http://www.profusion.com/nav>.
- [28] S. Raghavan and H. Garcia-Molina, "Crawling the Hidden Web", In Proc. of VLDB, pages 129–138, 2001.
- [29] J. Rennie and A. McCallum, "Using Reinforcement Learning to Spider the Web Efficiently", In Proc. of ICML, pages 335–343, 1999.
- [30] Search engines directory. <http://www.searchengineguide.com/searchengines.html>.
- [31] S. Sizov, M. Biwer, J. Graupmann, S. Siersdorfer, M. Theobald, G. Weikum, and P. Zimmer, "The BINGO! System for Information Portal Generation and Expert Web Search", In Proc. of CIDR, 2003.
- [32] W. Wu, C. Yu, A. Doan, and W. Meng, "An Interactive Clustering-based Approach to Integrating Source Query interfaces on the Deep Web", In Proc. of SIGMOD, pages 95–106, 2004.
- [33] Luciano Barbosa, Juliana Freire, "Searching for Hidden-Web Databases", In Proc. of Web Database, 2005.
- [34] A. K. Sharma, J. P. Gupta, D. P. Agarwal, "An Alternative Scheme For Generating Document Fingerprints Of Static Documents", Journal of the CSI Vol. 35 No.1 – January-March 2005.
- [35] A. K. Sharma, J. P. Gupta, D. P. Agarwal, "A novel approach towards efficient management of volatile information", Journal of the CSI Vol. 33 No.1 – September 2003.
- [36] A. Bergholz and B. Chidlovskii, "Crawling for Domain-Specific Hidden Web Resources", Proceedings of the Fourth IEEE International Conference on Web Information Systems Engineering (WISE'03) 2003.

- [37] A. de Carvalho Fontes and F. Soares Silva, "SmartCrawl: A New Strategy for the Exploration of the Hidden Web", ACM Transaction on WIDM'04, Washington, DC, USA, November 2004.
- [38] A. Ntoulas, P. Zerfos, and J. Cho, "Downloading Textual Hidden Web Content Through Keyword Queries", ACM transaction on JCDL'05 Denver, Colorado, USA, June.2005.
- [39] D. Shestakov, Sourav S. Bhowmick, and Ee-Peng Lim, "DEQUE: Querying the Deep Web", Data & Knowledge Engineering Volume 52, Issue 3(March 2005).
- [40] J. Cho and H. Garcia-Molina, "Synchronizing a Database to improve Freshness", In Proc. of the ACM SIGMOD Conf. on Management of Data, 2000.
- [41] J. Cho, H. Garcia-Molina, and L. Page, "Efficient crawling through URL ordering", In Proc. of the 7th International WWW Conf., 1998.
- [42] J. Lage, A. Silva, P. Golgher, and A. Laender, "Automatic generation of agents for collecting hidden web pages for data extraction", Data & Knowledge Engineering Volume 49 , Issue 2 (May 2004)
- [43] J. Wang and F. H. Lochovsky, "Data Extraction and Label Assignment for Web Databases" WWW 2003, May 20-24, 2003, Budapest, Hungary. ACM 1-58113-680-3/03/0005
- [44] K.-I. Lin and H. Chen, "Automatic information discovery from the invisible web", In Proceedings of the International Conference on Information Technology: Coding and Computing (ITCC'02), pages 332–337, 2002.
- [45] L. Barbosa and J. Freire, "Siphoning hidden-web data through keyword-based interfaces", In SBBD, 2004.
- [46] L. Gravano, P. G. Ipeirotis, and M. Sahami, "QProber: A System for Automatic Classification of Hidden-Web Databases", ACM Transactions on Information Systems (TOIS), Vol. 21, No. 1, 2003.
- [47] S. Liddle, D. Embley, D. Scott, and S. H. Yau, "Extracting data behind web Forms", In Proceedings of the Workshop on Conceptual Modeling Approaches for e-Business, pages 38–49, 2002.
- [48] S. Lawrence, C. Lee Giles, "Searching the World Wide Web", Science, 280(5360):98, 1998.

- [49] S. Lawrence and C. Lee Giles, "Accessibility of information on the Web", *Nature*, 400:107–109, 1999.
- [50] A.K.Sharma, J. P. Gupta, "Design of a PARallel Crawler based on Augmented HYpertext Documents (PARCAHYD)", Ph.D. Thesis, IIT & M, Gwalior, Aug. 2003.
- [51] A. Bergholz and B. Chidlovskii, "Crawling for Domain-Specific Hidden Web Resources", *Proc. of the Fourth International Conference on Web Information Systems Engineering (WISE'03) 2003 IEEE*.
- [52] A. Ntoulas, P. Zerfos, and J. Cho, "Downloading Textual Hidden Web Content Through Keyword Queries", *ACM transaction on JCDL'05 Denver, Colorado, USA, June.2005*.
- [53] J. Wang and F. H. Lochovsky", "Data Extraction and Label Assignment for Web Databases", *WWW 2003, May 20-24, 2003, Budapest, Hungary. ACM 1-58113-680-3/03/0005*
- [54] K.-I. Lin and H. Chen, "Automatic information discovery from the invisible web" In *Proceedings of the International Conference on Information Technology: Coding and Computing (ITCC'02)*, pages 332–337, 2002.
- [55] L. Barbosa and J. Freire, "Siphoning hidden-web data through keyword-based interfaces", In *SBBB*, 2004.
- [56] Naresh Chauhan, A.K. Sharma, "Design of an Agent based Context Driven Focused Crawler", *Accepted in International Journal of Information Technology, Delhi, Jan. 2008*
- [57] A. K. Sharma, Komal Kumar Bhatia: "Automated Discovery of Task Oriented Search Interfaces through Augmented Hypertext Documents" *Proc. of First International Conference on Web Engineering & Application (ICWA2006)*.
- [58] Joachim Hammer, Jan Fiedler; "Using Mobile Crawlers to Search the Web Efficiently", In *International Journal of Computer and Information Science*, pages 36-58, 2000.
- [59] Hong-Hai Do, Erhan Rahm, "COMA-A system for flexible combination of schema matching approaches" In *Proc. 28th VLDB Conference 2002*.

- [60] Jayant Madhavan, Philip A. Bernstein, Erhard Rahm, "Generic Schema Matching with Cupid", Proc. of VLDB 2001.
- [61] Luciano Barbosa, Juliana Freire, "An Adaptive Crawler for Locating Hidden Web Entry Points", WWW 2007, May 8–12, 2007, Banff, Alberta, Canada.
- [62] L. Barbosa and J. Freire, "Searching for Hidden-Web Databases", In Proceedings of WebDB, pages 1–6, 2005
- [63] H He, W. Meng, C. Yu, and Z. Wu, "WISE-integrator: An automatic integrator of Web search interfaces for e-commerce", In Proc. of VLDB, 2003
- [64] Dinesh Sharma, A.K. Sharma, Komal Kumar Bhatia, "Web crawlers: a review", Proc. of NCTC-2007
- [65] Dinesh Sharma, A.K. Sharma, Komal Kumar Bhatia, "Search engines: a comparative review", Proc. of NGCIS-2007
- [66] M. Dell'Erba, O. Fodor, F. Ricci, H. Werthner, "Harmonise: A Solution for Data Interoperability". IFIP I3E 2002.
- [67] Dinesh sharma, A.K. Sharma, Komal Kumar Bhatia, "Crawling the hidden web resources using agent", Proc. of ETCC-07, NIT-Hamirpur, India.
- [69] Hong-Hai Do, Erhan Rahm, "COMA-A system for flexible combination of schema matching approaches", In Proc. 28th VLDB Conference.
- [70] AnHai Doan, P. Domingos, A. Halevy, "Reconciling Schemas of Disparate Data Sources: A Machine-Learning Approach", In SIGMOD Record, 2001
- [71] A.K. Sharma, Komal Kumar Bhatia, "A Framework for Task-Specific Interface Mapper (TSIMC)", Proc. of ICWA 2007.
- [72] Erhard Rahm, Philip A. Bernstein, "A survey of approaches to automatic schema matching", VLDB Journal 10, 2001.
- [73] M. Dell'Erba, O. Fodor, F. Ricci, H. Werthner, "Harmonise: A Solution for Data Interoperability", IFIP I3E 2002.
- [74] Doan, J. Madhavan, P. Domingos, and A. Halevy, "Learning to Map between Ontologies on the Semantic Web" In Proc. of WWW, 2002.
- [75] S. Bergamaschi, S. Castano, and M. Vincini, "Semantic Integration of Semistructured and Structured Data Sources", SIGMOD Record, 1999.

- [76] R. Dhamankar, Y. Lee, A. Doan, A. Halevy, and P. Domingos, “iMAP: Discovering Complex Mappings between Database Schemas”, In SIGMOD, 2004.
- [77] Doan, J. Madhavan, P. Domingos, and A. Halevy, “Learning to Map between Ontologies on the Semantic Web”, In WWW, 2002.
- [78] E. Dragut and R. Lawrence, “Composing mappings between schemas using a reference ontology”, In Proc. of ODBASE, pages 228–246, 2004.
- [79] Do H.-H., Melnik S., and Rahm E., “Comparison of Schema Matching Evaluations”, Proc. GI Workshop "Web and Databases", Erfurt, Oct. 2002.
- [80] He, K. Chang, and J. Han, “Discovering complex matchings across web query interfaces: A correlation mining approach” In Proc. of SIGKDD, 2004.
- [81] N. Noy and M. Musen, “PROMPT: Algorithm and Tool for Automated Ontology Merging and Alignment”, In Proc. of AAAI/IAAI, pages 450–455, 2000.
- [82] S. Melnik, E. Rahm, and P. Bernstein, „Rondo: A Programming Platform for Generic Model Management”, In Proc. of SIGMOD, pages 193–204, 2003.
- [83] He, K. Chang, and J. Han, “Discovering complex matching across web query interfaces: A correlation mining approach”, In Proc. of SIGKDD, 2004.
- [84] R. Pottinger and P. Bernstein, “Merging Models Based on Given Correspondences”, In VLDB, 2003.
- [85] S. Spaccapietra and C. Parent, “View integration: A step forward in solving structural conflicts” TKDE, 6(2), 1994.
- [86] J. Ullman, “Information Integration Using Logical Views”, In Proc. of ICDT, 1997.
- [87] W. Wu, C. Yu, A. Doan, and W. Meng, “An interactive clustering-based approach to integrating source query interfaces on the deep web”, In SIGMOD, 2004.
- [88] A.K.Sharma, J. P. Gupta, “An Architecture for Electronic Commerce on the Internet”, Journal of Continuing Engineering Education, Vol. 2, pp 10-15, Roorkee, July 2002.
- [89] Paolo Boldi, Bruno Codenotti, Massimo Santini, and Sebastiano Vigna, “Ubicrawler: A scalablefully distributed web crawler”, Proc. of Australian World Wide Web Conference (AusWeb), 2002.

- [90] Shaozi Ye, Ji-Rong Wen, "A Systematic study on parameter correlations in large scale duplicate documents detections", 10<sup>th</sup> Pacific-Asia Conference on KDD, April 2006.
- [91] Hong-Hai Do, Erhan Rahm, "COMA-A system for flexible combination of schema matching approaches." In Proc. of 28th VLDB Conference, 2004.
- [92] Ipeirotis, P., Gravano, L. & Mehran, S., "Probe, count, and classify: categorizing hidden web databases", ACM SIGMOD 30(2), 67 – 78.
- [93] Mike Burner, "Crawling towards Eternity: Building an archive of the World Wide Web", Web Techniques Magazine, 2[5], May 1997.
- [95] Sergey Melink, Hector Garcia-Molina, Erhard Rahm, "Similarity Flooding: A Versatile Graph Matching Algorithm and its Application to Schema Matching", In Proc. 18th International Conf. On Data Engineering, San Jose CA, 2002.
- [96] H. He, W. Meng, C. Yu, and Z. Wu, "WISE-integrator: An automatic integrator of Web search interfaces for e-commerce", In VLDB, 2003.
- [97] B. He, K. Chang, and J. Han, "Discovering complex matching across web query interfaces: A correlation mining approach", In SIGKDD, 2004.
- [98] R. Pottinger and P. Bernstein, „Merging Models Based on Given Correspondences”, In VLDB, 2003.
- [99] J. Ullman, "Information Integration Using Logical Views", In Proc. of ICDT, 1997.
- [100] W. Wu, C. Yu, A. Doan, and W. Meng, "An interactive clustering-based approach to integrating source query interfaces on the deep web", In SIGMOD, 2004. (ADC2003).
- [101] Ipeirotis, P., Gravano, L. & Mehran, S., "Probe, count, and classify: categorizing hidden web databases", ACM SIGMOD 30(2), 67 – 78, 2000.
- [102] B. He, Z. Zhang, K. Chang, "Knocking the Door to the Deep Web: Integration of Web Query Interfaces", SIGMOD Conference, Demo, 2004.
- [103] Z. Zhang, B. He, and K. Chang, "Understanding Web Query Interfaces: Best-Effort Parsing with Hidden Syntax", SIGMOD Conference, 2004.



- [104] Xiang Peisu, Tian Ke, Huang Qinzhen, "A Framework of Deep Web Crawler", Proceedings of the 27th Chinese Control Conference July 16-18, 2008, Kunming, Yunnan, China
- [105] Tiezheng Nie, Derong Shen, Ge Yu, Yue Kou, "Subject-Oriented Classification based on Scale Probing in the Deep Web", Ninth IEEE International Conference on Web-Age Information Management 2008
- [106] Ahmad Ibrahim, Syed Ahsan Fahmi, Sajid Ibrahim Hashmi, Ho-Jin Choi, "Addressing Effective Hidden Web Search Using Iterative Deepening Search and Graph Theory", IEEE 8th International Conference on Computer and Information Technology Workshops 2008.
- [107] Bao-hua Qiang, Jian-qing Xi, "Effective Schema Extraction of Query Interfaces on the Deep Web", Fifth IEEE International Conference on Fuzzy Systems and Knowledge Discovery, 2008.
- [108] Ying Wang, Wanli Zuo, Tao Peng, Fengling He, "Integration of Query Interfaces for Deep Web Databases", Fourth IEEE International Conference on Natural Computation 2008.
- [109] Xin Zhong, Yuchen Fu, Quan Liu, Xinghong Lin, Zhiming Cui, "A Holistic Approach on Deep Web Schema Matching", IEEE International Conference on Convergence Information Technology, 2007.
- [110] Peiguang Lin, Ruzhi Xu, Zhimin Hong, Yan Zhang, "Finding the WDB's Query Interface in Deep Web Automatically", IEEE International Conference on Internet Computing in Science and Engineering, 2008.
- [111] Komal Kumar Bhatia, A.K.Sharma, "A Framework for an Extensible Domain-specific Hidden Web Crawler (DSHWC)", communicated to IEEE TKDE Journal Dec 2008.
- [112] Komal Kumar Bhatia, A.K.Sharma, "A Framework for Domain-Specific Interface Mapper (DSIM)", International Journal of Computer Science and Network Security (IJCSNS 2008).
- [113] Komal Kumar Bhatia, A.K.Sharma, "Merging Query Interfaces in Domain-specific Hidden Web Databases", accepted in International Journal of Computer Science, 2008.

- [114] Komal Kumar Bhatia, A.K. Sharma, "AKSHR: A Novel Framework of a Domain-specific Hidden Web Crawler", accepted in IEEE International Conference of Advanced Computing 2009.
- [115] Komal Kumar Bhatia, A.K.Sharma, Neelam Duhan, "Page Ranking Algorithms: A Survey", accepted in IEEE International Conference of Advanced Computing 2009.
- [116] Komal Kumar Bhatia, A.K.Sharma, "A Task-specific Hidden Web Crawler", Proc. of CIC 2008, 17<sup>th</sup> International Conference on Computing, Dec.3 to 5, 2008 Mexico City, Mexico.
- [117] Komal Kumar Bhatia, A.K.Sharma, "Automated Discovery of Task Oriented Search Interfaces through Augmented Hypertext Documents", Proc. of First International Conference on Web Engineering & Application (ICWA2006).
- [118] Komal Kumar Bhatia, A.K.Sharma, "Extraction of the Hidden Web Pages with Automated Discovery Process" Proc. of ICACCT-2007, Panipat.
- [119] Komal Kumar Bhatia, A.K.Sharma, "Learning from Web for Database updation for Extraction of the Hidden Web Resources", Proc. of ICACCT-2007, Panipat.
- [120] Komal Kumar Bhatia, A.K.Sharma, "A Method to Generate Automatic Query for Information Retrieval from Structured Data Base Model in Hidden Web", Proc. of IISN 2007.
- [121] Komal Kumar Bhatia, A.K.Sharma "Downloading and Mapping of Web Forms for Information Retrieval from Hidden Web", Proc. of International Conference APIIT in 2007.
- [122] Dinesh Sharma, Komal Kumar Bhatia, A.K.Sharma, "Extraction of the Hidden Web Pages Through Single Attribute Forms with Automated Discovery Process", Proc. of International Conference of Information Technology (ICIT-2007).
- [123] A.K. Sharma, Komal Kumar Bhatia, "Crawling the hidden web resources", Proc. of NCIT-2007, Delhi.
- [124] Komal Kumar Bhatia, A.K.Sharma, "Merging Search Interfaces using Semantic Mappings in Domain-Specific Hidden Web Databases " in Proc. of National conference in 2007.

- [125] Komal Kumar Bhatia, A.K.Sharma, Anuradha, "Optimized Merging of Search Interfaces in Domain-Specific Hidden Web Databases", Proc. of International conference at APIIT, 2008.