Referências Bibliográficas

[Cai et. al., 2003] CAI, D.; YU, S.; WEN, J.-R. ; MA, W.-Y.. **Vips: a vision-based page segmentation algorithm**, 2003. 5.1

[Chakrabarti et. al., 2007] CHAKRABARTI, D.; KUMAR, R. ; PUNERA, K.. **Page-level template detection via isotonic smoothing**. In: WWW '07: PROCEEDINGS OF THE 16TH INTERNATIONAL CONFERENCE ON WORLD WIDE WEB, p. 61–70, New York, NY, USA, 2007. ACM. 1

[Chakrabarti et. al., 2008] CHAKRABARTI, D.; KUMAR, R. ; PUNERA, K.. **A graph-theoretic approach to webpage segmentation**. In: WWW '08: PROCEEDING OF THE 17TH INTERNATIONAL CONFERENCE ON WORLD WIDE WEB, p. 377–386, New York, NY, USA, 2008. ACM. 5.1

[Chisholm et. al., 2000] CHISHOLM, W.; VANDERHEIDEN, G. ; JACOBS, I.. **Html techniques for web content accessibility guidelines 1.0**. W3c note, W3C, Nov. 2000. http://www.w3.org/TR/2000/NOTE-WCAG10-HTML-TECHS-20001106. 1

[Chuang et. al., 2004] CHUANG, S.-L.; HSU, J. Y.-J.. **Tree-structured template generation for web pages**. In: WI '04: PROCEEDINGS OF THE 2004 IEEE/WIC/ACM INTERNATIONAL CONFERENCE ON WEB INTELLIGENCE, p. 327–333, Washington, DC, USA, 2004. IEEE Computer Society. 1

[Connolly, 2000] CONNOLLY, D.. **A little history of the world wide web**, 2000. 2.1

[Gatterbauer e Bohunsky, 2006] GATTERBAUER, W.; BOHUNSKY, P.. **Table extraction using spatial reasoning on the css2 visual box model**, 2006. 1.2, 4.1, 4.4

[Chrome] GOOGLE. **Chrome core principles**. 2.2

[Gusfield, 1997] GUSFIELD, D.. **Algorithms on Strings, Trees, and Sequences**. Cambridge University Press, 1997. 3.1.1

[Hu et. al., 2005] HU, Y.; XIN, G.; SONG, R.; HU, G.; SHI, S.; CAO, Y. ; LI, H.. **Title extraction from bodies of html documents and its application to web page retrieval**. In: SIGIR '05: PROCEEDINGS OF THE 28TH ANNUAL INTERNATIONAL ACM SIGIR CONFERENCE ON RESEARCH AND DEVELOPMENT IN INFORMATION RETRIEVAL, p. 250–257, New York, NY, USA, 2005. ACM. 1

[Kestern, 2009] KESTEREN, A.. **Html5 differences from html4**. W3c working draft, W3C, aug 2009. http://www.w3.org/TR/2009/WD-html5-diff-20090825/. 1

[Krüpl et. al., 2006] KRÜPL, B.; HERZOG, M.. **Visually guided bottom-up table detection and segmentation in web documents**. In: WWW '06: PROCEEDINGS OF THE 15TH INTERNATIONAL CONFERENCE ON WORLD WIDE WEB, p. 933–934, New York, NY, USA, 2006. ACM. 1, 1.1, 5

[Laber et. al., 2009] LABER, E. S.; DE SOUZA, C. P.; JABOUR, I. V.; DE AMORIM, E. C. F.; CARDOSO, E. T.; RENTERÍA, R. P.; TINOCO, L. C. ; VALENTIM, C. D.. **A fast and simple method for extracting relevant content from news webpages**. In: CIKM '09: PROCEEDING OF THE 18TH ACM CONFERENCE ON INFORMATION AND KNOWLEDGE MANAGEMENT, p. 1685–1688, New York, NY, USA, 2009. ACM. 1

[Liu et. al., 2003] LIU, B.; GROSSMAN, R. ; ZHAI, Y.. **Mining data records in web pages**. In: KDD '03: PROCEEDINGS OF THE NINTH ACM SIGKDD INTERNATIONAL CONFERENCE ON KNOWLEDGE DISCOVERY AND DATA MINING, p. 601–606, New York, NY, USA, 2003. ACM. (document), 1, 1.1, 3, 3.1, 3.1, 3.2, 3.1.1, 3.1.2, 5, 5.1

[Longman 1998] LONGMAN, A.. **Introduction to the world wide web**, 1998. 2.1

[Pemberton, 2000] PEMBERTON, S.. **Xhtml 1.0: The extensible hypertext markup language - a reformulation of html4 in xml1.0**. first edition of a recommendation, W3C, jan 2000. http://www.w3.org/TR/2000/REC-xhtml1-20000126. 2.1

[Pinto et. al., 2003] PINTO, D.; MCCALLUM, A.; WEI, X. ; CROFT, W. B.. **Table extraction using conditional random fields**. In: SIGIR '03: PROCEEDINGS OF THE 26TH ANNUAL INTERNATIONAL ACM SIGIR CONFERENCE ON RESEARCH AND DEVELOPMENT IN INFORMAION RETRIEVAL, p. 235–242, New York, NY, USA, 2003. ACM. 1.1, 4, 4.4

[Quinlan, 1993] QUINLAN, J. R.. **C4.5: Programs for Machine Learning**. Morgan Kaufmann, San Mateo, CA, 1993. 4.4

[Reis et. al., 2004] REIS, D. C.; GOLGHER, P. B.; SILVA, A. S. ; LAENDER, A. F.. **Automatic web news extraction using tree edit distance**. In: WWW '04: PROCEEDINGS OF THE 13TH INTERNATIONAL CONFERENCE ON WORLD WIDE WEB, p. 502–511, New York, NY, USA, 2004. ACM. 1

[Tengli et. al., 2004] TENGLI, A.; YANG, Y. ; MA, N. L.. **Learning table extraction from examples**. In: COLING '04: PROCEEDINGS OF THE 20TH INTERNATIONAL CONFERENCE ON COMPUTATIONAL LINGUISTICS, p. 987, Morristown, NJ, USA, 2004. Association for Computational Linguistics. (document), 1, 1.1, 4.1, 4, 4, 5

[Viera et. al., 2006] VIEIRA, K.; DA SILVA, A. S.; PINTO, N.; DE MOURA, E. S.; CAVALCANTI, J. M. B. ; FREIRE, J.. **A fast and robust method for web page template detection and removal**. In: CIKM '06: PROCEEDINGS OF THE 15TH ACM INTERNATIONAL CONFERENCE ON INFORMATION AND KNOWLEDGE MANAGEMENT, p. 258–267, New York, NY, USA, 2006. ACM. 1

[Wang e Hu, 2002] WANG, Y.; HU, J.. **A machine learning based approach for table detection on the web**. In: WWW '02: PROCEEDINGS OF THE 11TH INTERNATIONAL CONFERENCE ON WORLD WIDE WEB, p. 242–250, New York, NY, USA, 2002. ACM. (document), 1.2, 4, 4, 4.1, 4.2, 4.3, 4.4, 4.4, 4.4, 6

[WebKit] WEBKIT. **The webkit open source project**. 2.2

[Xue et. al., 2007] XUE, Y.; HU, Y.; XIN, G.; SONG, R.; SHI, S.; CAO, Y.; LIN, C.-Y. ; LI, H.. **Web page title extraction and its application**. Information Processing and Management, 43[5]:1332–1347, 2007. 1

[Yang, 1991] YANG, W.. **Identifying syntactic differences between two programs**. Software - Practice and Experience, 21:739–755, 1991. (document), 1.2, 3.1.1, 3.1.1, 3.4, 3.5

[Zhai et. al., 2005] ZHAI, Y.; LIU, B.. **Web data extraction based on partial tree alignment**. In: WWW '05: PROCEEDINGS OF THE 14TH INTERNATIONAL CONFERENCE ON WORLD WIDE WEB, p. 76–85, New York, NY, USA, 2005. ACM. 1, 1, 1.1, 1.2, 3.1.1, 3.1.1, 5, 5.1, 5.4