

Referências Bibliográficas

- [Ahu93] AHUJA, R.; MAGNANTI, T.; ORLIN, J. ; WEIHE, K.. **Network flows: theory, algorithms, and applications**. Prentice Hall, 1993. 1.1, 2.3.2, 4.1.1
- [Alm07] ALMEIDA, R.; MOZAFARI, B. ; CHO, J.. **On the evolution of wikipedia**. In: PROCEEDINGS OF THE INTERNATIONAL CONFERENCE ON WEBLOGS AND SOCIAL MEDIA, 2007. 1.3, 3.1, 6
- [Ara01] ARASU, A.; CHO, J.; GARCIA-MOLINA, H.; PAEPCKE, A. ; RAGHAVAN, S.. **Searching the web**. ACM Transactions on Internet Technology (TOIT), 1(1):43, 2001. 2
- [Bae99] BAEZA-YATES, R.; RIBEIRO-NETO, B. ; OTHERS. **Modern information retrieval**. Addison-Wesley Reading, 1999. 1, 2
- [Bae02] BAEZA-YATES, R.; CASTILLO, C.. **Balancing volume, quality and freshness in web crawling**. Soft Computing Systems-Design, Management and Applications, p. 565–572, 2002. 2.2
- [Bor02] BORTOLOSSI, H. J.. **Cálculo diferencial a vária variáveis: uma introdução à teoria da otimização**. Editora PUC-Rio, 2002. 2.3.1, 4.4
- [Bre02] BRETTHAUER, K.; SHETTY, B.. **The nonlinear knapsack problem – algorithms and applications**. European Journal of Operational Research, 138(3):459–472, 2002. 4.4
- [Bro97] BRODER, A.; GLASSMAN, S.; MANASSE, M. ; ZWEIG, G.. **Syntactic clustering of the web**. Computer Networks and ISDN Systems, 29(8-13):1157–1166, 1997. 1
- [Bro06] BRODER, A.; FONTURA, M.; JOSIFOVSKI, V.; KUMAR, R.; MOTWANI, R.; NABAR, S.; PANIGRAHY, R.; TOMKINS, A. ; XU, Y.. **Estimating corpus size via queries**. In: PROCEEDINGS OF THE 15TH ACM INTERNATIONAL CONFERENCE ON INFORMATION AND KNOWLEDGE MANAGEMENT, p. 603. ACM, 2006. 1

- [Cas04b] CASTILLO, C.; MARIN, M.; RODRÍGUEZ, A. ; BAEZA-YATES, R.. **Scheduling algorithms for web crawling**. In: PROCEEDINGS OF THE LATIN AMERICAN WEB CONFERENCE (WebMedia/LA-WEB), p. 10–17, Riberão Preto, Brazil, October 2004. IEEE CS Press. 1, 3.2
- [Cas04a] CASTILLO, C.. **Effective web crawler**. PhD thesis, Chile University, 2004. 2.2
- [Cha03] CHAKRABARTI, S.. **Mining the web: discovering knowledge from hypertext data**. Morgan Kaufmann, 2003. 2
- [Cho00] CHO, J.; GARCIA-MOLINA, H.. **The evolution of the web and implications for an incremental crawler**. In: PROCEEDINGS OF THE 26TH INTERNATIONAL CONFERENCE ON VERY LARGE DATA BASES, p. 200–209, 2000. 2
- [Cho01] CHO, J.. **Crawling the web: discovery and maintenance of large-scale web data**. PhD thesis, Stanford University, 2001. 1
- [Cho03a] CHO, J.; GARCIA-MOLINA, H.. **Effective page refresh policies for web crawlers**. ACM Transactions on Database Systems, 28(4):390–426, 2003. 1, 1, 1, 1, 1, 1.3, 2.1.1, 2.1.1, 2.1.2, 2.2, 2.3.1, 2.1, 2.2, 2.3.1, 2.3.2, 3, 3.3, 4, 4.1, 4.3, 4.4, 2a, 2b, 4.4, 5, 5.2, 6.2.1, 7
- [Cho03b] CHO, J.; GARCIA-MOLINA, H.. **Estimating frequency of change**. ACM Transactions on Internet Technology (TOIT), 3(3):256–290, 2003. 3.4
- [Cof98] COFFMAN, E. G.; LIU, Z. ; WEBER, R. R.. **Optimal robot scheduling for web search engines**. Journal of scheduling, 1(1):15–29, 1998. 1, 1, 1.3, 2.3.4, 4, 4.1.1, 5
- [Cor93] CORLESS, R. M.; GONNET, G. H.; HARE, D. E. G. ; JEFFREY, D. J.. **On the Lambert’s W function**. Technical report, CS-93-03, Department of Computer Science, University of Waterloo, 1993. 2.3.1
- [Cor98] CORMEN, T.; LEISERSON, C. ; RIVEST, R.. **Introduction to algorithms**. McGraw-Hill Book Company, 1998. 1.1
- [DOS10] **Wikipedia article: denial of service attack**. http://en.wikipedia.org/wiki/Denial-of-service_attack, 2010. 1, 2.2
- [Eck08] ECKSTEIN, J.; GAL, A. ; REINER, S.. **Monitoring an information source under a politeness constraint**. INFORMS Journal on Computing, 20(1):3, 2008. 1, 1.1, 2.1.4, 2.3.3, 3

- [Edw01] EDWARDS, J.; MCCURLEY, K. S. ; TOMLIN, J. A.. **An adaptive model for optimizing performance of an incremental web crawler**. In: PROCEEDINGS OF THE WORLD WIDE WEB CONFERENCE, p. 106–113, 2001. 2.3.5
- [Fet04] FETTERLY, D.; MANASSE, M.; NAJORK, M. ; WIENER, J.. **A large-scale study of the evolution of web pages**. *Software: Practice and Experience*, 34(2):213–237, 2004. 2
- [Hey99] HEYDON, A.; NAJORK, M.. **Mercator: a scalable, extensible web crawler**. In: PROCEEDINGS OF THE WORLD WIDE WEB CONFERENCE, p. 219–229, 1999. 2.2, 7.1
- [Kos96] KOSTER, M.. **A standard for robot exclusion**. <http://www.robotstxt.org/wc/robots.html>, 1996. 2.2
- [Kos93] KOSTER, M.. **Guidelines for robot writers**. <http://www.robotstxt.org/wc/guidelines.html>, 1993. 2.2
- [Kos95] KOSTER, M.. **Robots in the web: threat or treat?** *ConneXions*, 9(4), 1995. 2.2
- [Law91] LAW, A. M.; KELTON, W. D.. **Simulation modeling and analysis**. McGraw-Hill, 1991. 6.1.3
- [Lee02] LEE, J.; WHANG, K.; LEE, B. ; CHANG, J.. **An update-risk based approach to ttl estimation in web caching**. In: PROCEEDINGS OF THE THIRD INTERNATIONAL CONFERENCE ON WEB INFORMATION SYSTEMS ENGINEERING (WISE), p. 21–29, 2002. 1
- [Mon03] MONTGOMERY, D. C.; RUNGER, G. C.. **Applied statistics and probability for engineers**. John Wiley & Sons, Inc., 2003. 2.1.1, 5.2, 5.4
- [Mot95] MOTWANI, R.; RAGHAVAN, P.. **Randomized algorithms**. Cambridge University Press, New York, NY, USA, 1995. 5.4
- [Nto04] NTOULAS, A.; CHO, J. ; OLSTON, C.. **What's new on the web?: the evolution of the web from a search engine perspective**. In: PROCEEDINGS OF THE 13TH INTERNATIONAL CONFERENCE ON WORLD WIDE WEB, p. 1–12. ACM, 2004. 2
- [Ols08] OLSTON, C.; PANDEY, S.. **Recrawl scheduling based on information longevity**. In: PROCEEDING OF THE 17TH INTERNATIONAL CONFERENCE ON WORLD WIDE WEB, p. 437–446. ACM, 2008. 2.1.5, 7.1

- [Pag98] PAGE, L.; BRIN, S.; MOTWANI, R. ; WINOGRAD, T.. **The pagerank citation ranking: bringing order to the web**. Technical report, Stanford Digital Library Technologies Project, 1998. 1, 7.1
- [Ros07] ROSS, S.. **Introduction to probability models**. Academic Press, 2007. 1, 2.1.4
- [Rud87] RUDIN, W.. **Real and complex analysis**. McGraw-Hill, 1987. 4.1.1
- [Shi99] SHIVAKUMAR, N.; GARCIA-MOLINA, H.. **Finding near-replicas of documents on the web**. *The World Wide Web and Databases*, p. 204–212, 1999. 1
- [Sou07] SOUZA, C.; LABER, E.; VALENTIM, C. ; CARDOSO, E.. **A polite policy for revisiting web pages**. In: *PROCEEDINGS OF THE LATIN AMERICAN WEB CONFERENCE (LA-WEB)*, p. 128–135, Santiago, Chile, 2007. IEEE Computer Society. 5.3
- [Wal77] WALKER, A. J.. **An efficient method for generating discrete random variables with general distributions**. *ACM Transactions on Mathematical Software*, 3(3):253–256, 1977. 5.2
- [Web07] **The Stanford WebBase Project**,
<http://www-diglib.stanford.edu/~testbed/doc2/WebBase/>,
2007. 3.1
- [Wge07] **GNU Wget**,
<http://www.gnu.org/software/wget/>, 2007. 6.1.1
- [Wik07] **Wikipedia - The Free Encyclopedia**,
<http://wikipedia.org>, 2007. 1.3, 3, 3.1, 6, 6.1.1
- [Wir07] **The WIRE Crawler**,
<http://www.cwr.cl/projects/WIRE/>, 2007. 2
- [Wol02] WOLF, J. L.; SQUILLANTE, M. S.; YU, P. S.; SETHURAMAN, J. ; OZSEN, L.. **Optimal crawling strategies for web search engines**. In: *PROCEEDING OF THE INTERNATIONAL CONFERENCE ON WORLD WIDE WEB*, p. 136–147, New York, NY, USA, 2002. ACM Press. 1, 1.1, 2.1.3, 2.1.5, 2.3.2, 3, 4, 4.1.1, 4.1.1, 6, 6.1.3, 6.2.1, 6.3, 7.1