

Bibliography

- [AB92a] A. Amir and G. Benson. Efficient two-dimensional compressed matching. In *Proceedings of the 2th Data Compression Conference*, pages 279–288. IEEE Computer Society Press, 1992.
- [AB92b] A. Amir and G. Benson. Two-dimensional periodicity and its applications. In *Proceedings of the 3rd ACM-SIAM Annual Symposium on Discrete Algorithms*, pages 440–452. ACM Press, 1992.
- [ABF96] A. Amir, G. Benson, and M. Farach. Let sleeping files lie: Pattern matching in Z-compressed files. *Journal of Computer and Systems Sciences*, 52(2):299–307, 1996.
- [ABL00] A. Amir, A. Butman, and M. Lewenstein. Real scaled matching. In *Proceedings of the 11th ACM-SIAM Annual Symposium on Discrete Algorithms*, pages 815–816. ACM Press, 2000.
- [ABLX00] A. Apostolico, M. E. Bock, S. Lonardi, and X. Xu. Efficient detection of unusual words. *Journal of Computational Biology*, 7(1/2):71–94, 2000.
- [Abr87] K. Abrahamson. Generalized string matching. *SIAM Journal on Computing*, 16(6):1039–1051, 1987.
- [AC75] A. V. Aho and M. J. Corasick. Efficient string matching: an aid to bibliographic search. *Communications of the ACM*, 18(6):333–340, 1975.
- [ACR01] C. Allauzen, M. Crochemore, and M. Raffinot. Efficient experimental string matching by weak factor recognition. In *Proceedings of the 12th Annual Symposium on Combinatorial Pattern Matching*, number 2089 in Lecture Notes in Computer Science, pages 51–72. Springer-Verlag, 2001.
- [AF91] A. Amir and M. Farach. Efficient 2-dimensional approximate matching of non-rectangular figures. In *Proceedings of the 2nd ACM-SIAM Annual Symposium on Discrete Algorithms*, pages 212–223, 1991.
- [AG85] A. Apostolico and Z. Galil, editors. *Combinatorial Algorithms on Words*, volume 12. Springer-Verlag, 1985.
- [AG97] A. Apostolico and Z. Galil, editors. *Pattern Matching Algorithms*. Oxford University Press, 1997.
- [AGM⁺90] S. F. Altschul, W. Gish, W. Miller, E. W. Myers, and D. J. Lipman. A basic local alignment search tool. *Journal of Molecular Biology*, 215:403–410, 1990.
- [AHU83] A. V. Aho, J. E. Hopcroft, and J. D. Ullman. *Data Structures and Algorithms*. Addison-Wesley, 1983.
- [AL91] A. Amir and G. M. Landau. Fast parallel and serial multidimensional approximate array matching. *Theoretical Computer Science*, 81(1):97–115, 1991.
- [ALL97] A. Amir, M. Lewenstein, and N. Lewenstein. Pattern matching in hyper-text. In *Proceedings of the 5th Workshop on Algorithms and Data Structures*, number 1272 in Lecture Notes in Computer Science, pages 160–173. Springer-Verlag, 1997.

- [Aoe94] J.-I. Aoe, editor. *String pattern matching strategies*. IEEE Computer Society Press, 1994.
- [AR99] C. Allauzen and M. Raffinot. Factor oracle of a set of words. Technical report 99-11, Institut Gaspard-Monge, Université de Marne-la-Vallée, 1999.
- [AR00] C. Allauzen and M. Raffinot. Simple optimal string matching. *Journal of Algorithms*, 36:102–116, 2000.
- [ASU86] A. V. Aho, R. Sethi, and J. D. Ullman. *Compilers – Principles, Techniques and Tools*. Addison-Wesley, 1986.
- [Baa88] S. Baase. *Computer Algorithms – Introduction to Design and Analysis*. Addison-Wesley, 1988.
- [Bak78] T. P. Baker. A technique for extending rapid exact-match string matching to arrays of more than one dimension. *SIAM Journal on Computing*, 7(4):533–541, 1978.
- [BBE⁺87] A. Blumer, J. Blumer, A. Ehrenfeucht, D. Haussler, and R. McConnel. Complete inverted files for efficient text retrieval and analysis. *Journal of the ACM*, 34(3):578–595, 1987.
- [BBYDS96] V. Bruyère, R. A. Baeza-Yates, O. Delgrange, and R. Scheihing. About the size of Boyer-Moore automata. In N. Ziviani, R. Baeza-Yates, and K. Guimarães, editors, *Proceedings of the 3rd South American Workshop on String Processing*, pages 31–46, Recife, Brazil. Carleton University Press, 1996.
- [BCW90] T. Bell, J. Cleary, and I. Witten. *Text Compression*. Prentice Hall, 1990.
- [Ben98] G. Benson. An algorithm for finding tandem repeats of unspecified pattern size. In *Proceedings of the 2nd Annual International Conference on Computational Molecular Biology*, pages 20–29. ACM Press, 1998.
- [Ben99] G. Benson. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Research*, 27:573–580, 1999.
- [BH01] A. Bergeron and S. Hamel. Vector algorithms for approximate string matching. *International Journal of Foundations of Computer Science*, 2001. To appear.
- [Bir77] R. S. Bird. Two-dimensional pattern matching. *Information Processing Letters*, 6(5):168–170, 1977.
- [BK93] A. Brüggemann-Klein. Regular expressions into finite automata. *Theoretical Computer Science*, 120(2):197–213, 1993.
- [BLPS99] G. S. Brodal, R. B. Lyngsø, C. N. S. Pedersen, and J. Stoye. Finding maximal pairs with bounded gap. In *Proceedings of the 10th Annual Symposium on Combinatorial Pattern Matching*, number 1645 in Lecture Notes in Computer Science, pages 134–149. Springer-Verlag, 1999.
- [BM77] R. S. Boyer and J. S. Moore. A fast string searching algorithm. *Communications of the ACM*, 20(10):762–772, 1977.
- [Bre95] D. Breslauer. Dictionary matching on unbounded alphabets: uniform length dictionaries. *Journal of Algorithms*, 18(2):278–295, 1995.
- [BS86] G. Berry and R. Sethi. From regular expression to deterministic automata. *Theoretical Computer Science*, 48(1):117–126, 1986.
- [BY91] R. Baeza-Yates. Some new results on approximate string matching. In *Workshop on Data Structures*, Dagstuhl, Germany, 1991. Abstract.
- [BYCG94] R.A. Baeza-Yates, C. Choffrut, and G.H. Gonnet. On Boyer-Moore automata. *Algorithmica*, 12(4/5):268–292, 1994.

- [BYG89a] R. A. Baeza-Yates and G. H. Gonnet. Boyer-Moore automata. Report, University of Waterloo, 1989.
- [BYG89b] R. A. Baeza-Yates and G. H. Gonnet. A new approach to text searching. In N. J. Belkin and C. J. van Rijsbergen, editors, *Proceedings of the 12th International Conference on Research and Development in Information Retrieval*, pages 168–175. ACM Press, 1989.
- [BYG96] R. A. Baeza-Yates and G. H. Gonnet. Fast text searching for regular expressions or automaton searching on tries. *Journal of the ACM*, 43(6):915–936, 1996.
- [BYGR90] R. A. Baeza-Yates, G. H. Gonnet, and M. Régnier. Analysis of Boyer-Moore type string searching algorithms. In *Proceedings of the 1st ACM-SIAM Annual Symposium on Discrete Algorithms*, pages 328–343. ACM Press, 1990.
- [BYN97] R. A. Baeza-Yates and G. Navarro. Multiple approximate string matching. In *Proceedings of the 5th Workshop on Algorithms and Data Structures*, number 1272 in Lecture Notes in Computer Science, pages 174–184. Springer-Verlag, 1997. Extended version to appear in *Random Structures and Algorithms* (Wiley).
- [BYN99] R. A. Baeza-Yates and G. Navarro. Faster approximate string matching. *Algorithmica*, 23(2):127–158, 1999.
- [BYN00a] R. Baeza-Yates and G. Navarro. Block-addressing indices for approximate text retrieval. *Journal of the American Society for Information Science*, 51(1):69–82, 2000.
- [BYN00b] R. Baeza-Yates and G. Navarro. New models and algorithms for multidimensional approximate pattern matching. *Journal of Discrete Algorithms*, 1(1):21–49, 2000.
- [BYR92] R. A. Baeza-Yates and M. Régnier. Average running time of the Boyer-Moore-Horspool algorithm. *Theoretical Computer Science*, 92(1):19–31, 1992.
- [BYR93] R. A. Baeza-Yates and M. Régnier. Fast two-dimensional pattern matching. *Information Processing Letters*, 45(1):51–57, 1993.
- [BYRN99] R. A. Baeza-Yates and B. Ribeiro-Neto. *Modern Information Retrieval*. Addison-Wesley, 1999.
- [CCG⁺93] M. Crochemore, A. Czumaj, L. Gąsieniec, S. Jarominek, T. Lecroq, W. Plandowski, and W. Rytter. Fast practical multi-pattern matching. Rapport 93-3, Institut Gaspard Monge, Université de Marne-la-Vallée, 1993.
- [CCG⁺94] M. Crochemore, A. Czumaj, L. Gąsieniec, S. Jarominek, T. Lecroq, W. Plandowski, and W. Rytter. Speeding up two string matching algorithms. *Algorithmica*, 12(4/5):247–267, 1994.
- [CCG⁺99] M. Crochemore, A. Czumaj, L. Gąsieniec, T. Lecroq, W. Plandowski, and W. Rytter. Fast practical multi-pattern matching. *Information Processing Letters*, 71(3/4):107–113, 1999.
- [CFCH⁺01] R. Cole, M. Farach-Colton, R. Hariharan, T. Przytycka, and M. Thorup. An $O(n \log n)$ algorithm for the maximum agreement subtree problem for binary trees. *SIAM Journal on Computing*, 30(5):1385–1404, 2001.
- [CGR92] M. Crochemore, L. Gąsieniec, and W. Rytter. Turbo-BM. Rapport LITP 92.61, Universités Paris 6 et 7, France, 1992.
- [CH97] M. Crochemore and C. Hancart. Automata for matching patterns. In G. Rozenberg and A. Salomaa, editors, *Handbook of Formal Languages*, volume

- 2: Linear Modeling: Background and Application, chapter 9, pages 399–462. Springer-Verlag, 1997.
- [CH98] R. Cole and R. Hariharan. Approximate string matching: A simpler faster algorithm. In *Proceedings of the 9th ACM-SIAM Annual Symposium on Discrete Algorithms*, pages 463–472. ACM Press, 1998.
- [CHI99] R. Cole, R. Hariharan, and P. Indyk. Tree pattern matching and subset matching in deterministic $O(n \log^3 n)$ -time. In *Proceedings of the 10th ACM-SIAM Annual Symposium on Discrete Algorithms*, pages 245–254. ACM Press, 1999.
- [Cho90] C. Choffrut. An optimal algorithm for building the Boyer-Moore automaton. *Bulletin of the European Association of Theoretical Computer Science*, 40:217–225, 1990.
- [CL92] W. I. Chang and J. Lampe. Theoretical and empirical comparisons of approximate string matching algorithms. In *Proceedings of the 3rd Annual Symposium on Combinatorial Pattern Matching*, number 664 in Lecture Notes in Computer Science, pages 175–184. Springer-Verlag, 1992.
- [CL94] W. I. Chang and E. L. Lawler. Sublinear approximate string matching and biological applications. *Algorithmica*, 12(4/5):327–344, 1994.
- [CLR90] T. H. Cormen, C. E. Leiserson, and R. L. Rivest. *Introduction to Algorithms*. MIT Press, 1990.
- [CM94] W.I. Chang and T. Marr. Approximate string matching with local similarity. In *Proceedings of the 5th Annual Symposium on Combinatorial Pattern Matching*, number 807 in Lecture Notes in Computer Science, pages 259–273. Springer-Verlag, 1994.
- [Col94] R. Cole. Tight bounds on the complexity of the Boyer-Moore string matching algorithm. *SIAM Journal on Computing*, 23(5):1075–1091, 1994.
- [CP91] M. Crochemore and D. Perrin. Two-way string-matching. *Journal of the ACM*, 38(3):651–675, 1991.
- [CP92] C.-H. Chang and R. Paige. From regular expression to DFA's using NFA's. In *Proceedings of the 3rd Annual Symposium on Combinatorial Pattern Matching*, number 664 in Lecture Notes in Computer Science, pages 90–110. Springer-Verlag, 1992.
- [CR94] M. Crochemore and W. Rytter. *Text Algorithms*. Oxford University Press, 1994.
- [CR95] M. Crochemore and W. Rytter. Squares, cubes and time-space efficient string-searching. *Algorithmica*, 13(5):405–425, 1995.
- [Cro92] M. Crochemore. String-matching on ordered alphabets. *Theoretical Computer Science*, 92(1):33–47, 1992.
- [CV97a] M. Crochemore and R. V  rin. Direct construction of compact directed acyclic word graphs. In *Proceedings of the 8th Annual Symposium on Combinatorial Pattern Matching*, number 1264 in Lecture Notes in Computer Science, pages 116–129. Springer-Verlag, 1997.
- [CV97b] M. Crochemore and R. V  rin. On compact directed acyclic word graphs. In *Structures in Logic and Computer Science*, number 1261 in Lecture Notes in Computer Science, pages 192–211. Springer-Verlag, 1997.
- [CW79] B. Commentz-Walter. A string matching algorithm fast on the average. In

- Proceedings of the 6th International Colloquium on Automata, Languages and Programming*, number 71 in Lecture Notes in Computer Science, pages 118–132. Springer-Verlag, 1979.
- [DGM94] M. Dubiner, Z. Galil, and E. Magen. Faster tree pattern matching. *Journal of the ACM*, 41(2):205–213, 1994.
- [FLSS92] V. A. Fischetti, G. M. Landau, J. P. Schmidt, and P. H. Sellers. Identifying periodic occurrences of a template with applications to protein struture. In *Proceedings of the 3rd Annual Symposium on Combinatorial Pattern Matching*, number 664 in Lecture Notes in Computer Science, pages 111–120. Springer-Verlag, 1992.
- [FM00] P. Ferragina and G. Manzini. Opportunistic data structures with applications. In *Proceedings of the 41st IEEE Annual Symposium on Foundations of Computer Science*, pages 390–398. IEEE Computer Society Press, 2000.
- [FM01] P. Ferragina and G. Manzini. An experimental study of an opportunistic index. In *Proceedings of the 12th ACM-SIAM Annual Symposium on Discrete Algorithms*, pages 269–278. ACM Press, 2001.
- [FNU01] K. Fredriksson, G. Navarro, and E. Ukkonen. *Faster than FFT: Rotation Invariant Combinatorial Template Matching*, volume II. Transworld Research Network, 2001. To appear.
- [FP74] M. J. Fischer and M. Paterson. String matching and other products. In R. M. Karp, editor, *Proceedings SIAM-AMS Complexity of Computation*, pages 113–125. AMS, 1974.
- [FT98] M. Farach and M. Thorup. String matching in Lempel-Ziv compressed strings. *Algorithmica*, 20(4):388–404, 1998.
- [FU98] K. Fredriksson and E. Ukkonen. Rotation invariant filters for multidimensional string matching and orientation search. In *Proceedings of the 9th Annual Symposium on Combinatorial Pattern Matching*, number 1448 in Lecture Notes in Computer Science, pages 118–125. Springer-Verlag, 1998.
- [Gag94] P. Gage. A new algorithm for data compression. *The C Users Journal*, 12(2), 1994.
- [Gal79] Z. Galil. On improving the worst case running time of the Boyer-Moore string searching algorithm. *Communications of the ACM*, 22(9):505–508, 1979.
- [GBY90] G. H. Gonnet and R. A. Baeza-Yates. An analysis of the Karp-Rabin string matching algorithm. *Information Processing Letters*, 34(5):271–274, 1990.
- [GBYS92] G. Gonnet, R. Baeza-Yates, and T. Snider. *Information Retrieval: Data Structures and Algorithms*, chapter 3: New indices for text: Pat trees and Pat arrays, pages 66–82. Prentice-Hall, 1992.
- [GG97] R. Giancarlo and R. Grossi. Multi-dimensional pattern matching with dimensional wildcards: Data structures and optimal on-line search algorithms. *Journal of Algorithms*, 24(2):223–265, 1997.
- [Glu61] V-M. Gluskov. The abstract theory of automata. *Russian Mathematical Surveys*, 16:1–53, 1961.
- [GP90] Z. Galil and K. Park. An improved algorithm for approximate string matching. *SIAM Journal on Computing*, 19(6):989–999, 1990.
- [GS81] Z. Galil and J. Seiferas. Linear-time string matching using only a fixed number of local storage locations. *Theoretical Computer Science*, 13(3):331–336,

- 1981.
- [Gus97] D. Gusfield. *Algorithms on Strings, Trees and Sequences: Computer Science and Computational Biology*. Cambridge University Press, 1997.
- [GV00] R. Grossi and J.S. Vitter. Compressed suffix arrays and suffix trees with applications to text indexing and string matching. In *Proceedings of the 32th ACM Symposium on the Theory of Computing*, pages 397–406. ACM Press, 2000.
- [Han93] C. Hancart. On Simon’s string searching algorithm. *Information Processing Letters*, 47(2):95–99, 1993.
- [HBFB99] K. Hofmann, P. Bucher, L. Falquet, and A. Bairoch. The PROSITE database, its status in 1999. *Nucleic Acids Research*, 27:215–219, 1999.
- [HM98] C. Hagenah and A. Muscholl. Computing epsilon-free NFA from regular expressions in $O(n \log^2(n))$ time. In *Proceedings of the 23th Symposium on Mathematical Foundations of Computer Science*, number 1450 in Lecture Notes in Computer Science. Springer-Verlag, 1998.
- [HN01] H. Hyvärö and G. Navarro. Faster bit-parallel approximate string matching. Technical Report TR/DCC-2002-1, University of Chile, Santiago, Chile, 2002.
- [Hor80] R. N. Horspool. Practical fast searching in strings. *Software Practice and Experience*, 10(6):501–506, 1980.
- [HSW97] J. Hromkovic, S. Seibert, and T. Wilke. Translating regular expressions into small epsilon-free nondeterministic finite automata. In *Proceedings of the 14th Annual Symposium on Theoretical Aspects of Computer Science*, number 1200 in Lecture Notes in Computer Science, pages 55–66. Springer-Verlag, 1997.
- [HU79] J. E. Hopcroft and J. D. Ullman. *Introduction to Automata, Languages and Computations*. Addison-Wesley, 1979.
- [Huf51] D. A. Huffman. A method for the construction of minimum redundancy codes. *Proceedings of the Institute of Electrical and Radio Engineers*, 40:1098–1101, 1951.
- [Hyy01] H. Hyvärö. Explaining and extending the bit-parallel algorithm of Myers. Technical Report A-2001-10, University of Tampere, Finland, 2001.
- [IHS⁺01] S. Inenaga, H. Hoshino, A. Shinohara, M. Takeda, S. Arikawa, G. Mauri, and G. Pavesi. On-line construction of compact directed acyclic word graphs. In *Proceedings of the 12th Annual Symposium on Combinatorial Pattern Matching*, number 2089 in Lecture Notes in Computer Science, pages 169–180. Springer-Verlag, 2001.
- [JTU96] P. Jokinen, J. Tarhio, and E. Ukkonen. A comparison of approximate string matching algorithms. *Software Practice and Experience*, 26(12):1439–1458, 1996.
- [Kär99] J. Kärkkäinen. *Repetition-Based Text Indexing*. PhD thesis, Department of Computer Science, University of Helsinki, December 1999.
- [Kil92] P. Kilpeläinen. *Tree Matching Problems with Applications to Structured Text Databases*. PhD thesis, University of Helsinki, Finland, 1992.
- [KK99] R. Kolpakov and G. Kucherov. Finding maximal repetitions in a word in linear time. In *Proceedings of the 40th IEEE Annual Symposium on Foundations of Computer Science*. IEEE Computer Society Press, 1999.
- [KLPC99] D. K. Kim, J. S. Lee, K. Park, and Y. Cho. Efficient algorithms for approximate string matching with swaps. *Journal of Complexity*, 15:128–147, 1999.

- [KM92] P. Kilpeläinen and H. Mannila. Grammatical tree matching. In *Proceedings of the 3rd Annual Symposium on Combinatorial Pattern Matching*, number 644 in Lecture Notes in Computer Science, pages 162–174. Springer-Verlag, 1992.
- [KM93] S. K. Kannan and E. W. Myers. An algorithm for locating non-overlapping regions of maximum alignment score. In *Proceedings of the 4th Annual Symposium on Combinatorial Pattern Matching*, number 684 in Lecture Notes in Computer Science, pages 74–86. Springer-Verlag, 1993.
- [KM95] J. R. Knight and E. W. Myers. Approximate regular expression pattern matching with concave gap penalties. *Algorithmica*, 14(1):85–121, 1995.
- [KMP77] D. E. Knuth, J. H. Morris, Jr, and V. R. Pratt. Fast pattern matching in strings. *SIAM Journal on Computing*, 6(1):323–350, 1977.
- [Knu73] D. E. Knuth. *The Art of Computer Programming: Sorting and Searching*, volume 3. Addison-Wesley, 1973.
- [KNU00] J. Kärkkäinen, G. Navarro, and E. Ukkonen. Approximate string matching over Ziv-Lempel compressed text. In *Proceedings of the 11th Annual Symposium on Combinatorial Pattern Matching*, number 1848 in Lecture Notes in Computer Science, pages 195–209. Springer-Verlag, 2000.
- [Kos89] S. R. Kosaraju. Efficient tree pattern matching. In *Proceedings of the 30th IEEE Annual Symposium on Foundations of Computer Science*, pages 178–183. IEEE Computer Society Press, 1989.
- [KR87] R. M. Karp and M. O. Rabin. Efficient randomized pattern-matching algorithms. *IBM Journal of Research and Development*, 31(2):249–260, 1987.
- [KR95] G. Kucherov and M. Rusinowitch. Matching a set of strings with variable length don't cares. In *Proceedings of the 6th Annual Symposium on Combinatorial Pattern Matching*, number 937 in Lecture Notes in Computer Science, pages 230–247. Springer-Verlag, 1995.
- [Kri87] K. Krithivasan. Efficient two-dimensional parallel and serial approximate pattern matching. Technical Report CAR-TR-259, University of Maryland, 1987.
- [KS87] K. Krithivasan and R. Sitalakshmi. Efficient two-dimensional pattern matching in the presence of errors. *Information Science*, 43:169–184, 1987.
- [KS95] J. Kececiloglu and D. Sankoff. Exact and approximation algorithms for sorting by reversals, with application to genome rearrangement. *Algorithmica*, 13(1/2):180–210, 1995.
- [KS98] J. Kärkkäinen and E. Sutinen. Lempel-Ziv index for q -grams. *Algorithmica*, 21(1):137–154, 1998.
- [KTS⁺98] T. Kida, M. Takeda, A. Shinohara, M. Miyazaki, and S. Arikawa. Multiple pattern matching in LZW compressed text. In *Proceedings of the 8th Data Compression Conference*, pages 103–112. IEEE Computer Society Press, 1998.
- [KTSA99] T. Kida, M. Takeda, A. Shinohara, and S. Arikawa. Shift-And approach to pattern matching in LZW compressed text. In *Proceedings of the 10th Annual Symposium on Combinatorial Pattern Matching*, number 1645 in Lecture Notes in Computer Science, pages 1–13. Springer-Verlag, 1999.
- [KU94] J. Kärkkäinen and E. Ukkonen. Two and higher dimensional pattern matching in optimal expected time. In *Proceedings of the 5th ACM-SIAM Annual Symposium on Discrete Algorithms*, pages 715–723. ACM Press, 1994.
- [KU96] J. Kärkkäinen and E. Ukkonen. Lempel-Ziv parsing and sublinear-size index

- structures for string matching. In N. Ziviani, R. Baeza-Yates, and K. Guimarães, editors, *Proceedings of the 3rd South American Workshop on String Processing*, pages 141–155, Recife, Brazil. Carleton University Press, 1996.
- [Lev65] V. I. Levenshtein. Binary codes capable of correcting spurious insertions and deletions of ones. *Problems of Information Transmission*, 1:8–17, 1965.
- [LT97] D. P. Lopresti and A. Tomkins. Block edit models for approximate string matching. *Theoretical Computer Science*, 181(1):159–179, 1997.
- [LV89] G. M. Landau and U. Vishkin. Fast parallel and serial approximate string matching. *Journal of Algorithms*, 10(2):157–169, 1989.
- [LW75] R. Lowrance and R. A. Wagner. An extension of the string-to-string correction problem. *Journal of the ACM*, 22(2):177–183, 1975.
- [Man89] U. Manber. *Introduction to Algorithms*. Addison-Wesley, 1989.
- [Man97] U. Manber. A text compression scheme that allows fast searching directly in the compressed file. *ACM Transactions on Information Systems*, 15(2):124–136, 1997.
- [MBY91] U. Manber and R. A. Baeza-Yates. An algorithm for string matching with a sequence of don't cares. *Information Processing Letters*, 37(3):133–136, 1991.
- [McC76] E. M. McCreight. A space-economical suffix tree construction algorithm. *Journal of Algorithms*, 23(2):262–272, 1976.
- [Meh84] K. Mehlhorn. *Data Structures and Algorithms 1: Sorting and Searching*. Springer-Verlag, 1984.
- [Mel96] B. Melichar. String matching with k differences by finite automata. In *Proceedings of the 13th International Conference on Pattern Recognition*, volume II, pages 256–260, Vienna, Austria. IEEE Computer Society Press, 1996.
- [MKT⁺00] T. Matsumoto, T. Kida, M. Takeda, A. Shinohara, and S. Arikawa. Bit-parallel approach to approximate string matching in compressed texts. In *Proceedings of the 8th String Processing and Information Retrieval*, pages 221–228. IEEE Computer Society Press, 2000.
- [MM89] E. W. Myers and W. Miller. Approximate matching of regular expressions. *Bulletin of Mathematical Biology*, 51:7–37, 1989.
- [MM91] G. Mehlhorn and E. W. Myers. A system for pattern matching applications on biosequences. *Computer Applications in Biosciences*, 9(3):299–314, 1991.
- [MM93] U. Manber and E. W. Myers. Suffix arrays: a new method for on-line string searches. *SIAM Journal on Computing*, 22(5):935–948, 1993.
- [MM96] R. Muth and U. Manber. Approximate multiple string search. In *Proceedings of the 7th Annual Symposium on Combinatorial Pattern Matching*, number 1075 in Lecture Notes in Computer Science, pages 75–86. Springer-Verlag, 1996.
- [MNZBY00] E. Moura, G. Navarro, N. Ziviani, and R. Baeza-Yates. Fast and flexible word searching on compressed text. *ACM Transactions on Information Systems*, 18(2):113–139, 2000.
- [MP70] J. H. Morris, Jr and V. R. Pratt. A linear pattern-matching algorithm. Report 40, University of California, Berkeley, 1970.
- [MP94] S. Muthukrishnan and K. Palem. Non-standard stringology: Algorithms and complexity. In *Proceedings of the 26th ACM Symposium on the Theory of Computing*, pages 770–779. ACM Press, 1994.
- [MRS96] H. M. Mahmoud, M. Régnier, and R. T. Smythe. Analysis of Boyer-Moore-

- Horspool string-matching heuristic. Report 9634, The George Washington University, 1996.
- [MW94] U. Manber and S. Wu. GLIMPSE: A tool to search through entire file systems. In *Proceedings of the USENIX Winter 1994 Technical Conference*, pages 23–32, San Francisco, CA. USENIX Association, 1994.
- [Mye92] E. W. Myers. A four russians algorithm for regular expression pattern matching. *Journal of the ACM*, 39(2):430–448, 1992.
- [Mye94] E. W. Myers. A sublinear algorithm for approximate keyword searching. *Algorithmica*, 12(4/5):345–374, 1994.
- [Mye95] E. W. Myers. Approximately matching context-free languages. *Information Processing Letters*, 54(2):85–92, 1995.
- [Mye96] E. W. Myers. Approximate matching of network expression with spacers. *Journal of Computational Biology*, 3(1):33–51, 1996.
- [Mye99] E. W. Myers. A fast bit-vector algorithm for approximate string matching based on dynamic programming. *Journal of the ACM*, 46(3):395–415, 1999.
- [Nav98] G. Navarro. *Approximate Text Searching*. PhD thesis, Department of Computer Science, University of Chile, 1998.
- [Nav00] G. Navarro. Improved approximate pattern matching on hypertext. *Theoretical Computer Science*, 237:455–463, 2000.
- [Nav01a] G. Navarro. A guided tour to approximate string matching. *ACM Computing Surveys*, 33(1):31–88, 2001.
- [Nav01b] G. Navarro. Nr-grep: a fast and flexible pattern matching tool. *Software Practice and Experience (SPE)*, 2001. To appear.
- [Nav01c] G. Navarro. Regular expression searching over Ziv-Lempel compressed text. In *Proceedings of the 12th Annual Symposium on Combinatorial Pattern Matching*, number 2089 in Lecture Notes in Computer Science, pages 1–17. Springer-Verlag, 2001.
- [NBY99] G. Navarro and R. Baeza-Yates. Very fast and simple approximate string matching. *Information Processing Letters*, 72:65–70, 1999.
- [NBY00] G. Navarro and R. Baeza-Yates. A hybrid indexing method for approximate string matching. *Journal of Discrete Algorithms*, 1(1):205–239, 2000. Special issue on Matching Patterns.
- [NKT⁺01] G. Navarro, T. Kida, M. Takeda, A. Shinohara, and S. Arikawa. Faster approximate string matching over compressed text. In *Proceedings of the 11th Data Compression Conference*, pages 459–468. IEEE Computer Society Press, 2001.
- [NMN⁺00] G. Navarro, E. Moura, M. Neubert, N. Ziviani, and R. Baeza-Yates. Adding compression to block addressing inverted indexes. *Information Retrieval*, 3(1):49–77, 2000.
- [NR99a] G. Navarro and M. Raffinot. Fast regular expression search. In *Proceedings of the 3rd Workshop on Algorithm Engineering*, number 1668 in Lecture Notes in Computer Science, pages 199–213. Springer-Verlag, 1999.
- [NR99b] G. Navarro and M. Raffinot. A general practical approach to pattern matching over Ziv-Lempel compressed text. In *Proceedings of the 10th Annual Symposium on Combinatorial Pattern Matching*, number 1645 in Lecture Notes in Computer Science, pages 14–36. Springer-Verlag, 1999.

- [NR00] G. Navarro and M. Raffinot. Fast and flexible string matching by combining bit-parallelism and suffix automata. *ACM Journal of Experimental Algorithmics (JEA)*, 5(4), 2000. <http://www.jea.acm.org>.
- [NR01a] G. Navarro and M. Raffinot. Compact DFA representation for fast regular expression search. In *Proceedings of the 5th Workshop on Algorithm Engineering*, number 2141 in Lecture Notes in Computer Science, pages 1–12, 2001.
- [NR01b] G. Navarro and M. Raffinot. Fast and simple character classes and bounded gaps pattern matching, with application to protein searching. In *Proceedings of the 5th Annual International Conference on Computational Molecular Biology*, pages 231–240. ACM Press, 2001.
- [NSF01] P. Nicodème, B. Salvy, and P. Flajolet. Motif statistics. *Theoretical Computer Science*, 2001. To appear.
- [NT00] G. Navarro and J. Tarhio. Boyer-Moore string matching over Ziv-Lempel compressed text. In *Proceedings of the 11th Annual Symposium on Combinatorial Pattern Matching*, number 1848 in Lecture Notes in Computer Science, pages 166–180. Springer-Verlag, 2000.
- [Par96] K. Park. Analysis of two-dimensional approximate pattern matching algorithms. In *Proceedings of the 7th Annual Symposium on Combinatorial Pattern Matching*, number 1075 in Lecture Notes in Computer Science, pages 335–347. Springer-Verlag, 1996.
- [Pea91] W. R. Pearson. Searching protein sequence libraries: comparison of the sensitivity and selectivity of the Smith-Waterman and FASTA algorithms. *Genomics*, 11:635–650, 1991.
- [Pev00] P. A. Pevzner. *Computational Molecular Biology: An Algorithmic Approach*. MIT Press, 2000.
- [Pin85] R. Y. Pinter. Efficient string matching with don't care pattern. In A. Apostolico and Z. Galil, editors, *Combinatorial Algorithms on Words*, pages 11–29. Springer-Verlag, 1985.
- [PL88] W. R. Pearson and D. J. Lipman. Improved tools for biological sequence comparison. *Proceedings of the National Academy of Sciences of the U.S.A.*, 85:2444–2448, 1988.
- [Raf97] M. Raffinot. On the multi backward dawg matching algorithm (MultiBDM). In R. Baeza-Yates, editor, *Proceedings of the 4th South American Workshop on String Processing*, pages 149–165, Valparaíso, Chile. Carleton University Press, 1997.
- [Rég89] M. Régnier. Knuth-Morris-Pratt algorithm: an analysis. In *Proceedings of the 14th Symposium on Mathematical Foundations of Computer Science*, number 379 in Lecture Notes in Computer Science, pages 431–444. Springer-Verlag, 1989.
- [RH91] S. Ranka and T. Heywood. Two-dimensional pattern matching with k mismatches. *Pattern Recognition*, 24(1):31–40, 1991.
- [RS97] M. Régnier and W. Szpankowski. On the approximate pattern occurrence in a text. In *Proceedings Compression and Complexity of SEQUENCES'97*. IEEE Press, 1997.
- [Ryt80] W. Rytter. A correct preprocessing algorithm for Boyer-Moore string searching. *SIAM Journal on Computing*, 9(3):509–512, 1980.
- [Sad99] K. Sadakane. A modified Burrows-Wheeler transformation for case-

- insensitive search with application to suffix array compression. In *Proceedings of the 9th Data Compression Conference*, page 548, 1999. Poster. <http://www-imai.is.s.u-tokyo.ac.jp/~sada/papers/Sada99b.ps.gz>.
- [Sad00] K. Sadakane. Compressed text databases with efficient query algorithms based on the compressed suffix array. In *Proceedings of the 11st International Symposium on Algorithms and Computation*, number 1969 in Lecture Notes in Computer Science, pages 410–421. Springer-Verlag, 2000.
- [Sch98] J. P. Schmidt. All highest scoring paths in weighted grid graphs and their application to finding all approximate repeats in strings. *SIAM Journal on Computing*, 27(4):972–992, 1998.
- [Sed88] R. Sedgewick. *Algorithms*. Addison-Wesley, 1988.
- [Sel80] P. H. Sellers. The theory and computation of evolutionary distances: Pattern recognition. *Journal of Algorithms*, 1(4):359–373, 1980.
- [Sim93] I. Simon. String matching algorithms and automata. In R. Baeza-Yates and N. Ziviani, editors, *Proceedings of the 1st South American Workshop on String Processing*, pages 151–157, Brazil. Universidade Federal de Minas Gerais, 1993.
- [SK83] D. Sankoff and J. B. Kruskal. *Time Warps, String Edits, and Macromolecules: the Theory and Practice of Sequence Comparison*. Addison-Wesley, 1983.
- [SM98] M.-F. Sagot and E. W. Myers. Identifying satellites and periodic repetitions in biological sequences. *Journal of Computational Biology*, 5:539–554, 1998.
- [SMT⁺00] Y. Shibata, T. Matsumoto, M. Takeda, A. Shinohara, and S. Arikawa. A Boyer-Moore type algorithm for compressed pattern matching. In *Proceedings of the 11th Annual Symposium on Combinatorial Pattern Matching*, number 1848 in Lecture Notes in Computer Science, pages 181–194. Springer-Verlag, 2000.
- [Sri86] M. A. Sridhar. Efficient algorithms for multiple pattern matching. Technical Report 661, University of Wisconsin-Madison, 1986.
- [ST95] E. Sutinen and J. Tarhio. On using q -gram locations in approximate string matching. In *Proceedings 3rd Annual European Symposium*, number 979 in Lecture Notes in Computer Science, pages 327–340. Springer-Verlag, 1995.
- [Sun90] D. M. Sunday. A very fast substring search algorithm. *Communications of the ACM*, 33(8):132–142, 1990.
- [SV96] S.C. Sahinalp and U. Vishkin. Efficient approximate and dynamic matching of patterns using a labeling paradigm. In *Proceedings of the 37th IEEE Annual Symposium on Foundations of Computer Science*. IEEE Computer Society Press, 1996.
- [SW81] T. F. Smith and M. S. Waterman. Identification of common molecular sequences. *Journal of Molecular Biology*, 147:195–197, 1981.
- [SZ97] D. Shasha and K. Zhang. Approximate tree pattern matching. In *Pattern Matching Algorithms*, pages 341–371. Oxford University Press, 1997.
- [Tho68] K. Thompson. Regular expression search algorithm. *Communications of the ACM*, 11:419–422, 1968.
- [TSM⁺01] M. Takeda, Y. Shibata, T. Matsumoto, T. Kida, A. Shinohara, S. Fukamachi, T. Shinohara, and S. Arikawa. Speeding up pattern matching by text compression: The dawn of a new era. *Journal of the Information Processing Society of Japan (IPSJ)*, 42(3), 2001. To appear.

- [TU88] J. Tarhio and E. Ukkonen. A greedy approximation algorithm for constructing shortest common superstrings. *Theoretical Computer Science*, 57(1):131–145, 1988.
- [TU93] J. Tarhio and E. Ukkonen. Approximate Boyer-Moore string matching. *SIAM Journal on Computing*, 22(2):243–260, 1993.
- [Ukk85] E. Ukkonen. Finding approximate patterns in strings. *Journal of Algorithms*, 6(1–3):132–137, 1985.
- [Ukk92] E. Ukkonen. Approximate string matching with q -grams and maximal matches. *Theoretical Computer Science*, 92(1):191–212, 1992.
- [Wat96] B. W. Watson. A new regular grammar pattern matching algorithm. In *Proceedings of the 4th Annual European Symposium*, number 1136 in Lecture Notes in Computer Science, pages 364–377. Springer-Verlag, 1996.
- [WM92a] S. Wu and U. Manber. Agrep – a fast approximate pattern-matching tool. In *Proceedings USENIX Winter 1992 Technical Conference*, pages 153–162. USENIX Association, 1992.
- [WM92b] S. Wu and U. Manber. Fast text searching allowing errors. *Communications of the ACM*, 35(10):83–91, 1992.
- [WM94] S. Wu and U. Manber. A fast algorithm for multi-pattern searching. Report TR-94-17, Department of Computer Science, University of Arizona, Tucson, AZ, 1994.
- [WMB99] I. Witten, A. Moffat, and T. Bell. *Managing Gigabytes*. Van Nostrand Reinhold, 2nd edition, 1999.
- [WMM95] S. Wu, U. Manber, and E. W. Myers. A subquadratic algorithm for approximate regular expression matching. *Journal of Algorithms*, 19(3):346–360, 1995.
- [WMM96] S. Wu, U. Manber, and E. W. Myers. A subquadratic algorithm for approximate limited expression matching. *Algorithmica*, 15(1):50–67, 1996.
- [Yao79] A. C. Yao. The complexity of pattern matching for a random string. *SIAM Journal on Computing*, 8(3):368–387, 1979.
- [ZL77] J. Ziv and A. Lempel. A universal algorithm for sequential data compression. *IEEE Transactions on Information Theory*, 23:337–343, 1977.
- [ZL78] J. Ziv and A. Lempel. Compression of individual sequences via variable length coding. *IEEE Transactions on Information Theory*, 24:530–536, 1978.
- [ZS97] K. Zhang and D. Shasha. Tree pattern matching. In A. Apostolico and Z. Galil, editors, *Pattern Matching Algorithms*, chapter 11, pages 341–371. Oxford University Press, 1997.
- [ZSW94] K. Zhang, D. Shasha, and J. T. L. Wang. Approximate tree matching in the presence of variable length don't cares. *Journal of Algorithms*, 16(1):33–66, 1994.
- [ZT89] R. F. Zhu and T. Takaoka. A technique for two-dimensional pattern matching. *Communications of the ACM*, 32(9):1110–1120, 1989.
- [ZWM99] K. Zhang, L. Wang, and B. Ma. Computing similarity between RNA structures. In *Proceedings of the 10th Annual Symposium on Combinatorial Pattern Matching*, number 1645 in Lecture Notes in Computer Science, pages 281–293. Springer-Verlag, 1999.