

# Meilensteine der (Bio-)Informatik

Sommersemester 2012

## 1 Administratives

Di, 14-16 Uhr in V2-221

<http://gi.cebitec.uni-bielefeld.de/> → teaching → Meilensteine der (Bio-)Informatik

Roland Wittler, roland at cebitec.uni-bielefeld.de, Sprechzeit: Do, 14-15, Büro: U10-145

### Ausarbeitung:

- Latex-Vorlage verwenden (siehe Webseite)
- etwa 8–10 Seiten
- Abgabe zum Reviewing: 12.6. (ausgedruckt zum Seminar mitbringen)
- Finale Abgabe: 10.7.

### Vortrag:

- mindestens eine Woche vorher mit Veranstalter vorbesprechen
- 10 bis 15 Minuten

## 2 Fragestellungen für Vortrag und Ausarbeitung

### • Autoren:

- Werdegang
- Biologe oder Informatiker?
- zum Zeitpunkt des Papers: Alter, Karrierestand, Institut, ...
- Verhältnis der Koautoren, Erstautor
- Anekdoten?
- Zeitstrahl?

### • Paper:

- Journal/Konferenz
- Bio- oder Informatik-Paper?
- Entstehung, Anekdoten?
- Aufbau, Vorgehen, ...
- Stil, Besonderheiten
- Algorithmus wiedererkennbar?

### • Inhalt:

- Theorie, Algorithmus
- Anwendung/Ergebnisse im Paper
- historische Einordnung:  
Vorarbeiten, Weiterentwicklungen, Grundstein für ..., Nutzen/Anwendung heute, Zusammenhang zu anderen Veröffentlichungen, ...

### 3 Themen

---

Search, Sorting	
Quicksort	Hoare [1961]
Boyer-Moore(-Horspool)	Boyer and Moore [1977], Horspool [1980]
Knuth-Morris-Pratt	Knuth et al. [1977]
Aho-Corasick	Aho and Corasick [1975]
Suffix Trees: Mc Creight	McCreight [1976]
Suffix Trees: Ukkonen	Ukkonen [1995]
Suffix Arrays: Manber-Myers	Manber and Myers [1990]
Comparison	
Needleman-Wunsch	Needleman and Wunsch [1970]
Smith-Waterman	Smith and Waterman [1981a,b]
FASTA	Pearson and Lipman [1988]
BLAST	Altschul et al. [1990]
Carrillo-Lipman	Carrillo and Lipman [1988]
CLUSTAL-W	Thompson et al. [1994]
Mining, Compression	
Markov-Ketten	Markov [1906], Basharin et al. [2004]
GLIMMER	Salzberg et al. [1998], Delcher et al. [1999]
Nussinov	Nussinov and Jacobson [1980]
Lempel-Ziv	Ziv and Lempel [1977]
Burrows-Wheeler	Burrows and Wheeler [1994]
Phylogeny	
Fitch-Hartigan	Fitch [1971], Hartigan [1973]
Sankoff	Sankoff [1975], Sankoff and Rousseau [1975], Sankoff and Cedergren [1983]
Minimum Evolution	Desper and Gascuel [2004]
Neighbor-joining	Saitou and Nei [1987]
Maximum-Likelihood	Felsenstein [1981]
Graphs	
Hamiltonkreis	Hamilton [1853]
Eulerpfad	Euler [1741, 1956]

---

### Literatur

- A. V. Aho and M. J. Corasick. Efficient string matching: An aid to bibliographic search. *Commun. ACM*, 18(6):333–340, 1975.
- S. F. Altschul, W. Gish, W. Miller, E. W. Myers, and D. J. Lipman. Basic local alignment search tool. *J. Mol. Biol.*, 215(3):403–410, 1990.
- G. Basharin, A. Langville, and V. Naumov. The life and work of aa markov. *Linear Algebra and its Applications*, 386:3–26, 2004.
- R. S. Boyer and J. S. Moore. A fast string searching algorithm. *Commun. ACM*, 20(10):762–772, 1977.
- M. Burrows and D. J. Wheeler. A block sorting lossless data compression algorithm. Technical Report TR 124, Digital Equipment Corporation, Palo Alto, CA, 1994.
- H. Carrillo and D. Lipman. The multiple sequence alignment problem in biology. *SIAM J. Appl. Math.*, 48(5):1073–1082, 1988.
- A. L. Delcher, D. Harmon, S. Kasif, O. White, and S. L. Salzberg. Improved microbial gene identification with Glimmer. *Nucleic Acids Res.*, 27(23):4636–4641, 1999.

- R. Desper and O. Gascuel. Theoretical foundation of the balanced minimum evolution method of phylogenetic inference and its relationship to weighted least-squares tree fitting. *Mol. Biol. Evol.*, 21(3):587–598, 2004.
- L. Euler. Solutio problematis ad geometriam situs pertinentis. *Commentarii academiae scientiarum Petropolitanae*, 8:128–140, 1741.
- L. Euler. The seven bridges of Königsberg. In J. R. Newman, editor, *The World of Mathematics*, volume 1, pages 573–580. Simon and Schuster, New York, 1956.
- J. Felsenstein. Evolutionary trees from DNA sequences: A maximum likelihood approach. *J. Mol. Evol.*, 17:368–376, 1981.
- W. M. Fitch. Toward defining the course of evolution: Minimum change for a specific tree topology. *Syst. Zool.*, 20(4):406–416, 1971.
- W. R. Hamilton. Account of the icosian game. *Proc. Roy. Irish. Acad.*, 6:415–416, 1853.
- J. A. Hartigan. Minimum mutation fits to a given tree. *Biometrics*, 29(1):53–65, 1973.
- C. A. R. Hoare. Algorithm 64: Quicksort. *Commun. ACM*, 4(7):321, 1961. ISSN 0001-0782. doi: <http://doi.acm.org/10.1145/366622.366644>.
- R. N. Horspool. Practical fast searching in strings. *Softw. Pract. Exper.*, 10(6):501–506, 1980.
- D. E. Knuth, J. H. Morris Jr., and V. R. Pratt. Fast pattern matching in strings. *SIAM J. Computing*, 6:323–350, 1977.
- U. Manber and E. W. Myers. Suffix arrays: A new method for on-line string searches. In *Proc. First Annu. ACM-SIAM Symp. Discrete Algorithms, SODA 1990*, pages 319–327, 1990.
- A. Markov. Rasprostranenie zakona bol'shikh chisel na velichiny, zavisyaschie drug ot druga. *Izvestiya Fiziko-matematicheskogo obschestva pri Kazanskom universitete*, 2(15):135–156, 1906.
- E. M. McCreight. A space-economical suffix tree construction algorithm. *J. ACM*, 23(2):262–272, 1976.
- S. B. Needleman and C. D. Wunsch. A general method applicable to the search for similarities in the amino acid sequence of two proteins. *J. Mol. Biol.*, 48:443–453, 1970.
- R. Nussinov and A. B. Jacobson. Fast algorithm for predicting the secondary structure of single-stranded rna. *Proc. Natl. Acad. Sci. USA*, 77(11):6309–6313, 1980.
- W. R. Pearson and D. J. Lipman. Improved tools for biological sequence comparison. *Proc. Natl. Acad. Sci. USA*, 85(8):2444–2448, 1988.
- N. Saitou and M. Nei. The neighbor-joining method: A new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.*, 4(4):406–425, 1987.
- S. L. Salzberg, A. L. Delcher, S. Kasif, and O. White. Microbial gene identification using interpolated Markov models. *Nucleic Acids Res.*, 26(2):544–548, 1998.
- D. Sankoff. Minimal mutation trees of sequences. *SIAM J. Appl. Math.*, 28(1):35–42, 1975.
- D. Sankoff and R. J. Cedergren. Simultaneous comparison of three or more sequences related by a tree. In D. Sankoff and J. B. Kruskal, editors, *Time Warps, String Edits, and Macromolecules: The Theory and Practice of Sequence Comparison*, chapter 9, pages 253–263. Addison-Wesley, Reading, MA, 1983.
- D. Sankoff and P. Rousseau. Locating the vertices of a steiner tree in an arbitrary metric space. *Math. Program.*, 9:240–246, 1975.
- T. F. Smith and M. S. Waterman. Identification of common molecular subsequences. *J. Mol. Biol.*, 147(1):195–197, 1981a.
- T. F. Smith and M. S. Waterman. Comparison of biosequences. *Adv. Appl. Math.*, 2:482–489, 1981b.
- J. D. Thompson, D. G. Higgins, and T. J. Gibson. CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.*, 22(22):4673–4680, 1994.
- E. Ukkonen. On-line construction of suffix trees. *Algorithmica*, 14(3):249–260, 1995.
- J. Ziv and A. Lempel. A universal algorithm for sequential data compression. *IEEE Trans. Inf. Theory*, 23(3):337–343, 1977.