



Aalto University  
School of Science



Department of  
Computer Science

# String Algorithms Highlights

*Professor Jorma Tarhio*

# Research Group: String Algorithms

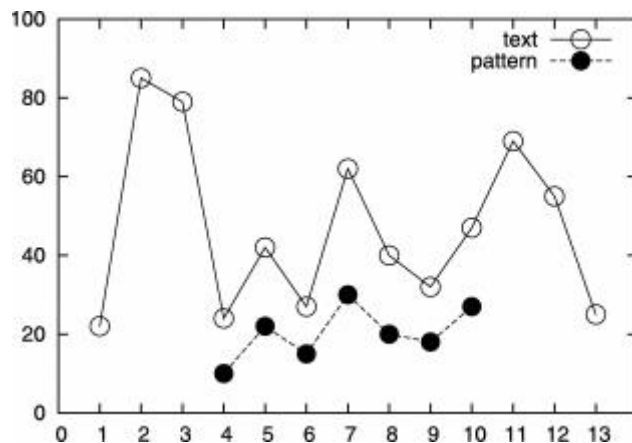
---

- ❑ We develop efficient algorithms for information retrieval and text compression. Most of our algorithms deal with online searching or indexing. Our perspective is algorithm engineering. Bioinformatics is one of the application areas.
- ❑ The following slides review some of our recent achievements.

Contact: [jorma.tarhio@aalto.fi](mailto:jorma.tarhio@aalto.fi)

# Order-Preserving Matching

T. Chhabra, J. Tarhio: A filtration method for order-preserving matching. *Information Processing Letters* 116, 2 (2016), 71–74.



- ❑ Strings of numbers are considered in order-preserving matching.
- ❑ The pattern  $P$  matches a text substring  $S$  if  $P$  and  $S$  have the same relative order.
- ❑ We present a practical filter to locate match candidates quickly. The filter can apply any algorithm  $X$  for exact string matching. If  $X$  is sublinear, the total method is sublinear on average.
- ❑ Application area: analysis of time series

# Order-Preserving Matching with SIMD

---

T. Chhabra, M. O. Kulekci, J. Tarhio: Alternative algorithms for order-preserving matching. In: Proc. Stringology 2015.

- ❑ Two filtration algorithms for order-preserving matching are implemented using two SIMD instruction sets, SSE (Streaming SIMD Extensions) and AVX (Advanced Vector Extensions).
- ❑ In most cases, the new algorithms are faster than previous algorithms.
- ❑ Later we have developed a still faster algorithm applying SSE technology (submitted).
- ❑ We also implemented an FM-index based solution for order-preserving matching.

# Approximate Order-Preserving Matching

---

T. Chhabra, E. Giaquinta, J. Tarhio: Filtration algorithms for approximate order-preserving matching. In: Proc. SPIRE 2015.

- ❑ In approximate order-preserving matching, the pattern  $P$  matches a text substring  $S$  if  $P$  and  $S$  have the same relative order after excluding up to  $k$  positions in both strings.
- ❑ We present two filtration methods for the problem. One of them is the first sublinear solution in the average case.
- ❑ Approximate order-preserving matching is more natural for applications than exact order-preserving matching.

# Approximate Circular Matching

---

T. Hirvola, J. Tarhio: Approximate online matching of circular strings. In: Proc. SEA 2014.

- ❑ In circular matching, the pattern  $P$  matches a text substring  $S$  if  $S$  is a rotation of  $P$ , e.g. NORMI is an occurrence of MINOR.
- ❑ In approximate circular matching, we allow errors in  $S$  besides a rotation, e.g. NORMA is an occurrence of MINOR with one mismatch.
- ❑ We present new bitparallel algorithms for the problem. The algorithms are sublinear on average.

# Approximate Longest Common Substring

---

T. Flouri, E. Giaquinta, K. Kobert, E. Ukkonen: Longest common substrings with  $k$  mismatches. *Information Processing Letters* 115, 6–8 (2015), 643–647.

- ❑ The Longest Common Substring problem consists in finding the longest common substring(s) of two strings. In the approximate variant, we allow up to  $k$  mismatches between two substrings.
- ❑ We present (i) a practical algorithm that runs in quadratic time in the length of the longer string, independent of  $k$ , and uses constant space and (ii) a theoretical algorithm that runs in quasilinear time and uses linear space in the length of the smaller string.
- ❑ Application area: computational biology – alignment free sequence comparison methods.