

Course code	CTU03
Course title	Text Searching Algorithms
Institution	Czech Technical University in Prague
Course address	Thákurova 7
City	Prague
Minimum year of study	3rd year
Minimum level of English	Good
Minimum level of French	None
Key words	Stringology, finite automata, exact and approximate string matching, forward and backward string matching, factor automata, subsequence automata
Language	English
Professor responsible	Ondřej Guth
Telephone	+420 224 359 867
Email	ondrej.guth@fit.cvut.cz
Participating professors	Bořivoj Melichar
Number of places	Minimum: 10, Maximum: 20, Reserved for local students: 0
Objectives	<p>Text is the simplest and most natural representation of information in a range of areas. Text is a linear sequence of symbols from some alphabet. The text is manipulated in many application areas: processing of text in natural and formal languages, study of sequences in molecular biology, music analysis, etc.</p> <p>The design of algorithms that process texts goes back at least thirty years. In particular, the 1990s produced many new results. This progress is due in part to genome research, where text algorithms are often used. The basic problem of text processing concerns string matching. It is used to access information and this operation is used very frequently. We have recognized while working in this area that finite automata are very useful tools for understanding and solving many text processing problems. We have found in some cases that well known algorithms are in fact simulators of non-deterministic finite automata serving as models of these algorithms. For this reason the material used in this course is based mainly on results from the theory of finite automata. Because the string is a central notion in this area, Stringology has become the nickname of this subfield of algorithmic research.</p>

Programme to be followed	<p>Five 3-hour lectures:</p> <ol style="list-style-type: none"> <li>1. Overview of Stringology, string matching problems, string matching and finite automata.</li> <li>2. Forward string matching, dynamic programming and bit parallelism.</li> <li>3. Factor automata, subsequence automata, repetition in text.</li> <li>4. Forward string matching, fail function.</li> <li>5. Backward string matching, models of backward string matching, Boyer-Moore algorithm.</li> </ol> <p>Four 2-hour seminars:</p> <ol style="list-style-type: none"> <li>1. Mastering finite automata: determinisation, union, intersection, <math>\varepsilon</math>-transitions removal, elimination of more than one initial states.</li> <li>2. Construction of string matching automata, their determinisation and simulation.</li> <li>3. Application of factor automata.</li> <li>4. Backward string matching.</li> </ol>
Prerequisites	Sets, relations, oriented graphs, finite automata, regular expressions.
Course exam	Written exam with the duration of 1 hour, evaluation of the results.
Publish on the web Session	<a href="http://www.stringology.org/athens/TextSearchingAlgorithms/">http://www.stringology.org/athens/TextSearchingAlgorithms/</a> March 2018