String Algorithms
String Algorithms

Lecturers

Thomas Mailund
E-mail: mailund@birc.au.dk
Office: 1110.326

Christian Storm
E-mail: cstorm@birc.au.dk
Office: 1110.325

Course homepage

http://users-birc.au.dk/cstorm/courses/StrAlg_f13
Lectures

Monday 10.15-12.00, 1110.214

Thursday 11.15-12.00, 1110.214

Full schedule available at the homepage

Topics for week $x$ should be considered tentative until week $x-1$

Slides will usually be put up after the lecture
Available in the GAD bookstore, additional papers will be available via the course www-page
A formal definition

A string is a collection of elements that obeys the following rules

1. every element has a label that is unique
2. every element with some label \( x \) (except at most one, called the \textit{leftmost}) has a unique determinable \textit{predecessor} labelled \( p(x) \)
3. every element with some label \( x \) (except at most one, called the \textit{rightmost}) has a unique determinable \textit{successor} labelled \( s(x) \)
4. whenever an element with label \( x \) is not leftmost, then \( x = s(p(x)) \)
5. whenever an element with label \( x \) is not rightmost, then \( x = p(s(x)) \)
6. for any two distinct elements with labels \( x \) and \( y \), there exist a positive integer \( k \) such that either \( x = s^k(y) \) or \( x = p^k(y) \)
But ...

... a string is still what it has always been ....

... a finite sequence of elements from some alphabet $A$, where the label of an element is its position/index ...

**Definition:** An element of $A^+$ is called a *linear string* on alphabet $A$. An element of $A^*$ is called a *finite string* on $A$. 
**Notation and terminology**

\(x: \text{array}\ [1..n]\) of \(A\) is a \textit{string} on alphabet \(A\) of \textit{length} \(n = |x|\).

The \textit{letter} at position \(i\) is \(x[i]\), i.e. \(x = x[1]x[2]...x[n]\)

\(x[i..j] = x[i] \ x[i+1] \ ... \ x[j]\) is a \textit{substring} of \(x\) of length \(j-i+1\), and a \textit{proper substring} if \(j-i+1 < n\). It occurs in \(x\) at position \(i\)

The substring \(x[1..i]\) is a \textit{prefix} of \(x\), and a \textit{proper prefix} if \(i < n\), and \(x[i..n]\) is a \textit{suffix} of \(x\), and a \textit{proper suffix} if \(i > 1\)
Mandatory Projects

There will be three mandatory projects

Constructing suffix trees (Apr 22 - May 2)
Finding tandem repeats (Apr 29 - May 16)
Exact pattern matching (May 13 - Jun 3)

Work in groups of 2-3 students

... implementation, experiments, discussion ...
**Mandatory Projects**

There will be three mandatory projects

- **Constructing suffix trees** (Apr 22 - May 2)
- **Finding tandem repeats** (Apr 29 - May 16)
- **Exact pattern matching** (May 13 - Jun 3)

Work in groups of 2-3 students

... implementation, experiments, discussion ...
Exam

Individual oral exam (20 minutes)

~6 exam questions related to the mandatory projects

Presentation of the mandatory project and related theory

... more info later ...