

Softwarepraktikum

“ Muster- und Bilderkennung”

am Lehrstuhl für Informatik VI

Suffix-Arrays

19.04.2007

Arne Mauser und David Vilar
Lehrstuhl für Informatik VI
{mauser,vilar}@i6.informatik.rwth-aachen.de

to be or not to be, that is the question.

to be or not to be, that is the question.
1 2 3 4 5 6 7 8 9 10 11 12

1. Indizieren

A B C D A B E F G H I J
1 2 3 4 5 6 7 8 9 10 11 12

Korpus

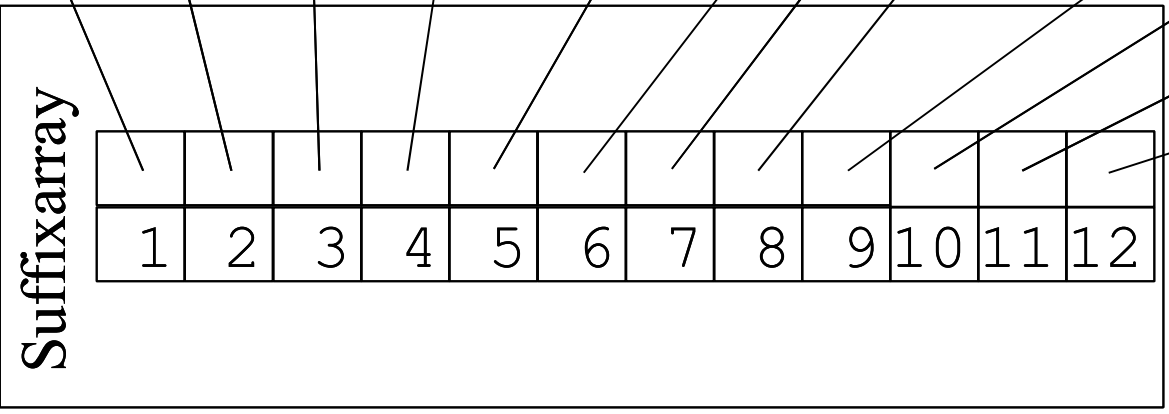
1. Indizieren
2. Lexicon

Datenstrukturen

Lexikon

A	→	to	F	→	that
B	→	be	G	→	is
C	→	or	H	→	the
D	→	not	I	→	question
E	→	,	J	→	.

A 1 B 2 C 3 D 4 A 5 B 6 E 7 F 8 G 9 H 10 I 11 J 12

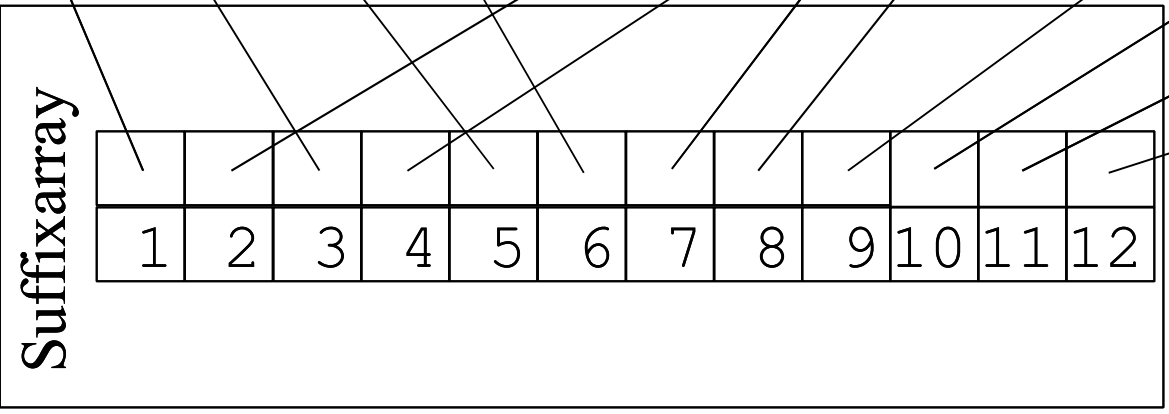


- 1. Indizieren
- 2. Lexicon
- 3. Suffix Array

Datenstrukturen

Lexikon	A	→	to	F	→	that
	B	→	be	G	→	is
	C	→	or	H	→	the
	D	→	not	I	→	question
	E	→	,	J	→	.

A 1 B 2 C 3 D 4 A 5 B 6 E 7 F 8 G 9 H 10 I 11 J 12



1. Indizieren
2. Lexicon
3. Suffix Array
4. Sortieren

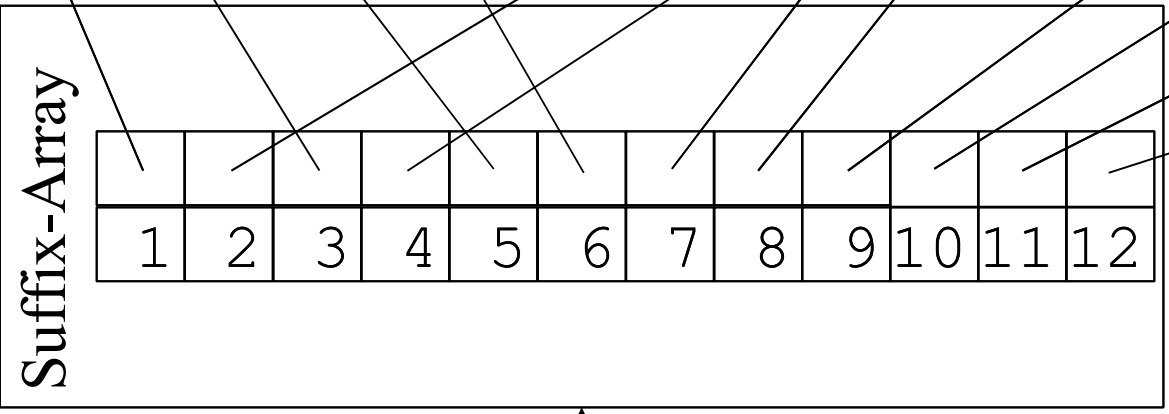
Sortierreihenfolge nach Suffix

Datenstrukturen

Lexikon	A	→	to	F	→	that
	B	→	be	G	→	is
	C	→	or	H	→	the
	D	→	not	I	→	question
	E	→	,	J	→	.

Korpus

A 1 B 2 C 3 D 4 A 5 B 6 E 7 F 8 G 9 H 10 I 11 J 12



Suche: be ,
Ergebnis: Position im Korpus

Datenstrukturen

Lexikon

A	→	to	F	→	that
B	→	be	G	→	is
C	→	or	H	→	the
D	→	not	I	→	question
E	→	,	J	→	.

Test des Suffix-Arrays

Testfile, Testanfragen

Auf der Seite des Praktikums

Bei Fragen:

mauser@i6.informatik.rwth-aachen.de

vilar@i6.informatik.rwth-aachen.de