

Softwarepraktikum

“ Muster- und Bilderkennung”

am Lehrstuhl für Informatik VI

Suffix-Arrays

20.04.2006

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

Korpus

Datenstrukturen

Lexikon

A → to

B → be

C → or

D → not

E → ,

F → that

G → is

H → the

I → question

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

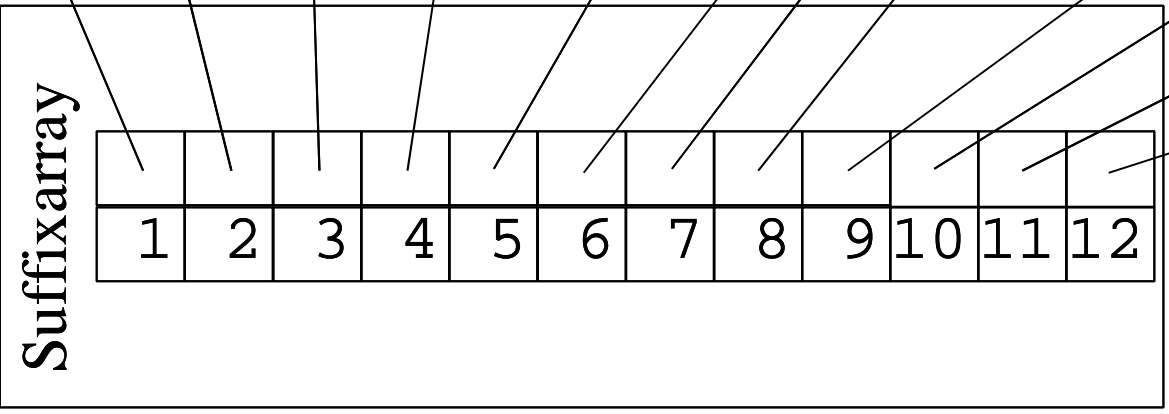
Korpus

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

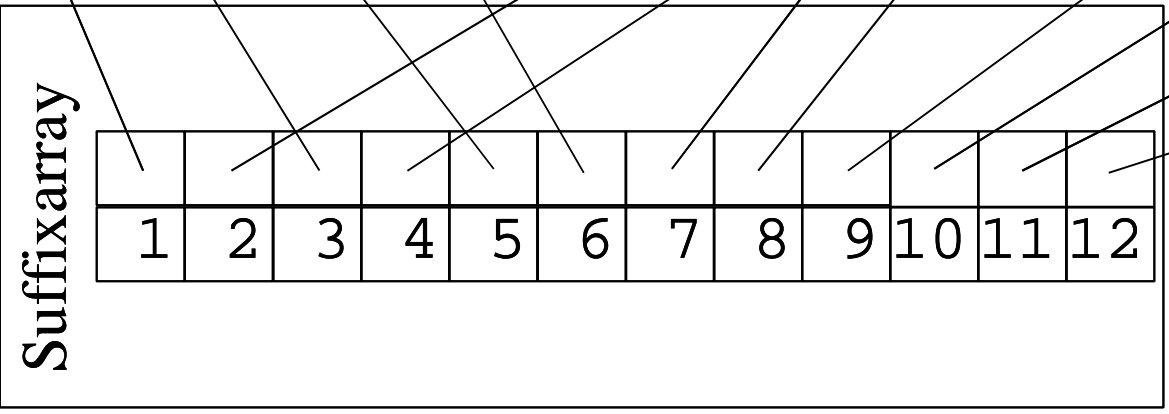


Datenstrukturen

Lexikon

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

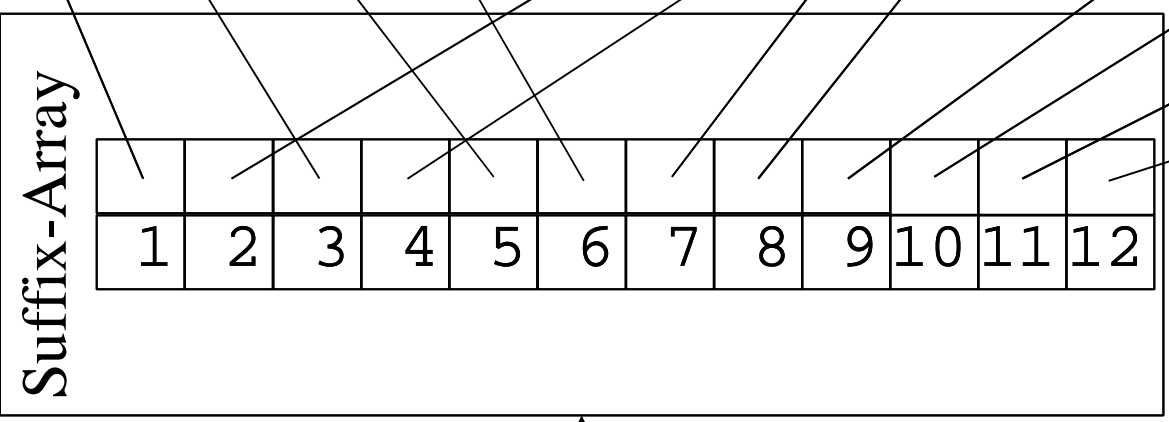
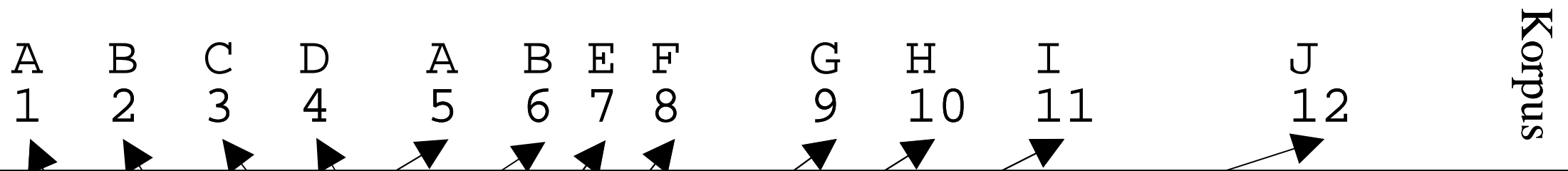
A 1 B 2 C 3 D 4 A 5 B 6 E 7 F 8 G 9 H 10 I 11 J 12



Datenstrukturen

Lexikon

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



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:

<http://www-i6.informatik.rwth-aachen.de/web/Teaching/LabCourses/SS05/Softwarepraktikum/Aufgabe2/80days.txt>

Bei Fragen:

mauser@i6.informatik.rwth-aachen.de
vilar@i6.informatik.rwth-aachen.de