

KMP-Algorithmus

Bioinformatik Bachelor

Propädeutikum WS 2019/2020

14. Januar 2020

Christian Hoffmann

Generell

- Algorithmus um einen Substring (Pattern P) innerhalb eines Strings (Text T) zu finden
- Vergleicht das Pattern mit dem Text von links nach recht
- Grundidee:
 - Wann immer wir einen mismatch (nach ein paar matches) haben, wissen wir schon ein paar Characters in dem Text des nächsten Fensters
 - Vorteil: Vermeidung von Vergleichen, die sowieso matchen
- Man benötigt eine Failure-Function!

Failure-Function

Vorgehensweise:

- Von links nach rechts durch das Pattern gehen
- Ist ein Präfix auch ein Suffix? → Wenn ja, Länge ist der Value

Beispielpattern: **AT**CATCGAT

P:	A	T	C	A	T	C	G	A	T
i:	0	1	2	3	4	5	6	7	8
Val:	0	0							

Failure-Function

Vorgehensweise:

- Von links nach rechts durch das Pattern gehen
- Ist ein Präfix auch ein Suffix? → Wenn ja, Länge ist der Value

Beispielpattern: **ATCATCGAT**

P:	A	T	C	A	T	C	G	A	T
i:	0	1	2	3	4	5	6	7	8
Val:	0	0	0						

Failure-Function

Vorgehensweise:

- Von links nach rechts durch das Pattern gehen
- Ist ein Präfix auch ein Suffix? → Wenn ja, Länge ist der Value

Beispielpattern: ATCATCGAT

P:	A	T	C	A	T	C	G	A	T
i:	0	1	2	3	4	5	6	7	8
Val:	0	0	0	1					

Failure-Function

Vorgehensweise:

- Von links nach rechts durch das Pattern gehen
- Ist ein Präfix auch ein Suffix? → Wenn ja, Länge ist der Value

Beispielpattern: ATCATCGAT

P:	A	T	C	A	T	C	G	A	T
i:	0	1	2	3	4	5	6	7	8
Val:	0	0	0	1	2				

Failure-Function

Vorgehensweise:

- Von links nach rechts durch das Pattern gehen
- Ist ein Präfix auch ein Suffix? → Wenn ja, Länge ist der Value

Beispielpattern: **ATCATCGAT**

P:	A	T	C	A	T	C	G	A	T
i:	0	1	2	3	4	5	6	7	8
Val:	0	0	0	1	2	3			

Failure-Function

Vorgehensweise:

- Von links nach rechts durch das Pattern gehen
- Ist ein Präfix auch ein Suffix? → Wenn ja, Länge ist der Value

Beispielpattern: ATCATCGAT

P:	A	T	C	A	T	C	G	A	T
i:	0	1	2	3	4	5	6	7	8
Val:	0	0	0	1	2	3	0		

Failure-Function

Vorgehensweise:

- Von links nach rechts durch das Pattern gehen
- Ist ein Präfix auch ein Suffix? → Wenn ja, Länge ist der Value

Beispielpattern: ATCATCGAT

P:	A	T	C	A	T	C	G	A	T
i:	0	1	2	3	4	5	6	7	8
Val:	0	0	0	1	2	3	0	1	

Failure-Function

Vorgehensweise:

- Von links nach rechts durch das Pattern gehen
- Ist ein Präfix auch ein Suffix? → Wenn ja, Länge ist der Value

Beispielpattern: **ATCATCGAT**

P:	A	T	C	A	T	C	G	A	T
i:	0	1	2	3	4	5	6	7	8
Val:	0	0	0	1	2	3	0	1	2

KMP

- Den eigentlichen Ablauf findet ihr auf den Folien von „Einführung in die Bioinformatik 1“ → wird hier nicht weiter darauf eingegangen
- KMP (vor Allem die Failure-Values berechnen) ist klausurrelevant für die Bioinformatik-Klausur!
- Weitere Ressourcen:
 - <https://www.geeksforgeeks.org/kmp-algorithm-for-pattern-searching/>
 - <http://whocouldthat.be/visualizing-string-matching/>
 - <http://jovilab.sinaapp.com/visualization/algorithms/strings/kmp>