

KMP-Algorithmus

Also kurz zur Zwei-Finger-Methode: Wir nehmen den rechten Finger auf die zu berechnende Stelle, den linken Finger positionieren wir eine Stelle daneben. Jetzt vergleichen wir, ob ein Mismatch vorliegt.

- Wenn ja, dann ist das schön und wir können mit der Parallelverschiebung beginnen, d.h. wir schieben beiden Finger synchron um eine Stelle nach links und testen auf Match.
 - Erreichen wir dann irgendwann den linken Rand, so war der Durchlauf erfolgreich und die Startposition des linken Fingers wird der Eintrag an der Startposition des rechten Fingers.
 - Erreichen wir den linken Rand nicht, d.h. es liegt irgendwo nochmal ein Mismatch vor, dann funktioniert dieser Durchlauf nicht, und wir gehen mit den Fingern wieder zurück zur Startposition.
- Da nun entweder der Mismatch-Test nicht geklappt hat oder die Parallelverschiebung nicht bis zum Schluss durchführbar war, müssen wir jetzt den linken Finger einmal nach links verschieben und dort wieder von Neuem beginnen, d.h. es startet wieder mit dem Mismatch-Test.

So jetzt schauen wir uns einmal die Aufgabe 3 von Blatt 9 (7.1.10) an. Gegeben ist die Tabelle

Position	0	1	2	3	4	5
Pattern	c	b				a
Tabelle	-1	0	-1	1	0	2

Wenden wir die Zyklensmethode an, dann lesen wir aus der Position 5 eine 2 aus, was uns sagt, dass die zwei Positionen davor mit dem Anfang des Patterns übereinstimmen müssen. Somit muss $\text{Pat}[3] = \text{Pat}[0] = c$ und $\text{Pat}[4] = \text{Pat}[1] = b$, also

Position	0	1	2	3	4	5
Pattern	c	b		c	b	a
Tabelle	-1	0	-1	1	0	2

Nun steht an Position 3 in der Tabelle eine 1, d.h. es muss einen 1-Zyklus vorher geben. Somit muss also die Position 2 mit dem Anfang übereinstimmen, ergo $\text{Pat}[2] = \text{Pat}[0] = c$

Position	0	1	2	3	4	5
Pattern	c	b	c	c	b	a
Tabelle	-1	0	-1	1	0	2

Schauen wir uns einmal die Zwei-Finger-Methode im Gegensatz dazu an. An Position 5 steht eine 2 in der Tabelle, d.h. also wir müssen an der Stelle, wo der linke Finger auf der 2 stand, einen erfolgreichen Mismatch-Test durchgeführt haben, d.h. $a = \text{Pat}[5] \neq \text{Pat}[2]$ und es kommen nur noch die Buchstaben b und c für $\text{Pat}[2]$ in Frage. An Position 2 steht eine -1, d.h. wir waren links davon nicht erfolgreich, weil entweder kein erfolgreicher Mismatch-Test vorlag oder aber die Parallelverschiebung nicht funktioniert hat. Angenommen wir hätten mit Position 1 keinen erfolgreichen Mismatch-Test, aber einen erfolgreichen mit $\text{Pat}[0]$, dann müsste $\text{Pat}[2] = b$ gelten. Dann hätten wir aber einen erfolgreichen Mismatch-Test und dementsprechend (weil wir schon am linken Rand sind) schon eine gültige Position und es sollte $\text{Tab}[2] = 0$ gelten, was falsch ist (ist ja schon mit -1 gegeben). Also muss $\text{Pat}[2] = c$ gelten. Warum passt das? Mit Position 1 ist der Mismatch-Test erfolgreich, beginnen wir mit der Parallelverschiebung, dann vergleichen wir $\text{Pat}[1]$ mit $\text{Pat}[0]$, was falsch ist. Damit kann $\text{Tab}[2] \neq 1$, mit Position 0 schlägt schon der Mismatch-Test fehl, also muss auch $\text{Tab}[2] \neq 0$ gelten. Damit bleibt also nur noch -1, also passt das.

Position	0	1	2	3	4	5
Pattern	c	b	c			a
Tabelle	-1	0	-1	1	0	2

Machen wir weiter mit Position 3. Wir müssen einmal weiter hinten schauen und lesen in der Tabelle an Position 5 eine 2 aus. Dementsprechend muss der Mismatch zwischen Position 5 und Position 2 vorliegen, was aber eher unwichtig ist. Viel wichtiger ist, dass die beiden Positionen links der 5 und links der 2 übereinstimmen müssen, damit während der Parallelverschiebung Matches auftreten. Somit muss also $\text{Pat}[4] = \text{Pat}[1] = b$ und $\text{Pat}[3] = \text{Pat}[0] = c$ gelten, sodass wir unsere beiden übrigen Plätze aufgefüllt haben.

Position	0	1	2	3	4	5
Pattern	c	b	c	c	b	a
Tabelle	-1	0	-1	1	0	2

Alle Angaben wie immer ohne Garantie.