

Selbsttest 10

1) Stringsuche nach Boyer-Moore

Gegeben sei der Text:

far out in the uncharted backwaters

of the unfashionable end of the western

spiral arm of the galaxy lies a small

unregarded yellow sun.

Suchen sie in diesem Text das erste Vorkommen der Zeichenfolge "regarded".

Lösung:

Zuerst das Aufstellen der Sprungtabelle entsprechend der Buchstabenpositionen im gesuchten Wort. Die Sprungtabelle wird dabei von vorne nach hinten durchgelaufen (bis zur vorletzten Position).

```
r e g a r d e d *
7 6 5 4 3 2 1 8 8
```

```
abcdefghijklmnopqrstuvxyz_.
48821858888888888838888888888
```

Bei einer Übereinstimmung wird von hinten nach vorne verglichen. Wenn ein Buchstabe unterschiedlich ist, wird entsprechend des Buchstabens im Text (Sprungtabelle) nach vorne gesprungen (die nächste Übereinstimmung kann frühestens an dieser Position beginnen).

far out in the uncharted backwaters

regarded regarded regarded

 regarded regarded regarde

of the unfashionable end of the western

d regarded regarded regarded

regarded regarded regarded

 regarde

spiral arm of the galaxy lies a small

d regarded regarded regard

regarded regarded regarded

unregarded yellow sun.

ed

regarded

2) Reguläre Ausdrücke

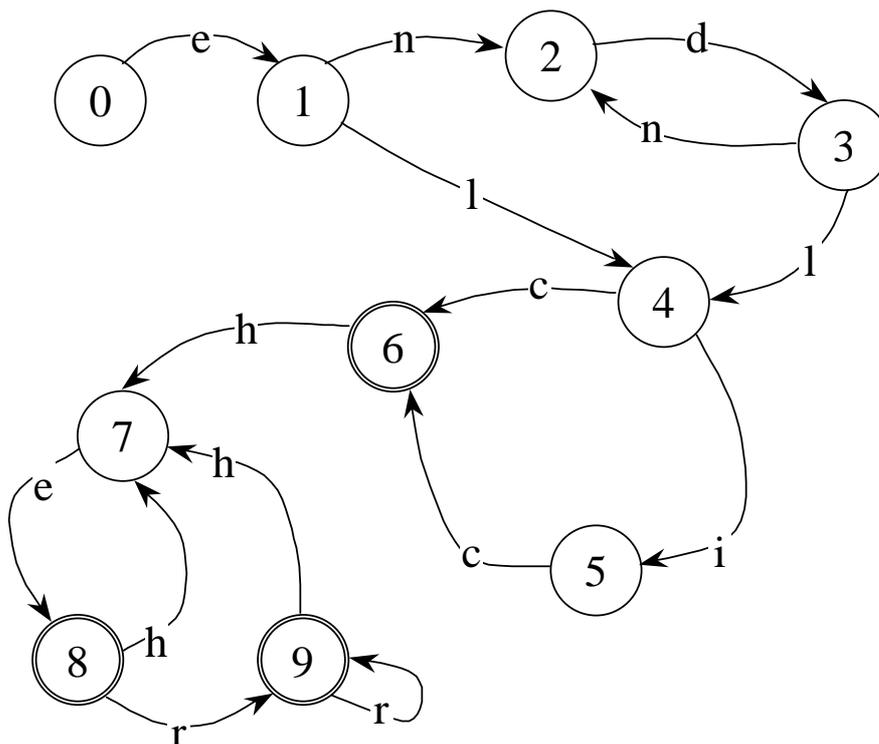
Konstruieren Sie einen endlichen Automaten, der den regulären Ausdruck $"e\{nd\}l\{i\}c\{he\{r\}"$ erkennt.

Durchsuchen Sie mit diesem endlichen Automaten den Text: "Heli's endloses Verstaendnis fuer endliche Automaten"

Lösung:

- 1) Man beginnt vor dem Start des regulären Ausdrucks bei 0 zu nummerieren (Startzustand)
- 2) Hinter jedem gelesenen Zeichen wird jeweils ein neuer Zustand eingetragen (fortlaufend)
- 3) Für jeden Zustand werden alle möglichen von dieser Position aus lesbaren Zeichen als Zustandsübergänge zum nachfolgenden Zustand eingezeichnet (Alternativen und Wiederholungen beachten!)
- 4) Alle Zustände, die ein Ende des regulären Ausdrucks kennzeichnen sind akzeptierende Zustände
- 5) Alle anderen Zustandsübergänge verweisen auf den Zustand 0 (und müssen nicht mit eingezeichnet werden, sondern werden implizit angenommen)

$0e_1\{n_2d_3\}l_4[i_5]c_6\{h_7e_8\{r_9\}\}$



Die Suche:

Üblicherweise wird versucht solange eine Übereinstimmung zu finden, bis Zustand 0 (durch implizite - d.h. nicht eingezeichnete - Zustandsübergänge => siehe Punkt 5) erreicht wird. Die Zustandsübergänge vom Zustand 0 bis zum letzten akzeptieren Zustand (sofern einer erreicht wurde) entsprechen dann der (längsten) Übereinstimmung. Wurde kein akzeptierender Zustand besucht, handelt es sich auch nicht um eine Übereinstimmung.

Zu beachten ist, dass die Suche mit dem endlichen Automaten alleine nur "akzeptiert" bzw. "nicht akzeptiert" als Ergebnis liefert. Der Rest (Position/Länge des Matches, etc.) muss außerhalb des endlichen Automaten ermittelt werden.