Chapter 12 - Textsuche

Naive Suche

```
Eingabe:
        t = t_1...t_n
        s = s_1...s_m  // Suchtext
Ausgabe:
        kleinstes i mit t_{i+1}...t_{i+m} = s

i <- 0
Solange i ≤ n - m
        // vgl. s mit t_{i+1}...t_{i+m}
        j <- m
        Solange j > 0 und s_j = t_{i+j}
                j <- j - 1
        Falls j = 0
                gib i aus; stop
        sonst i <- i + 1
```

Vorkommens-Heuristik

Vorkommensfunktion:

$$v : A \to 0, ..., m$$
$$a \mapsto v(a),$$
$$v(a) := min\{\ k | a \notin s_{k+1}...s_m\ und\ (a = s_k \vee k = 0)\}$$

Bem.

Kann vorberechnet werden in $O(|A| + m)$.

Änderungen in „Naive Suche“

Sei $t_{i+j} \neq s_j$ und $t_{i+j+1}...t_{i+m} = s_{j+1}...s_m$.
Dann ist $v := v(t_{i+j}) \neq j$ und

- falls $v < j$, kann i um $j - v$ erhöht werden
- falls $v > j$, kann i um $m - v + 1$ erhöht werden

Bem.

Für kleines m und großes Alphabet A nun Laufzeit $O(\frac{n}{m})$

Suffixfunktion

Def.

Ein Präfix eines Wortes w, das zugleich ein Suffix von w ist, heißt *Präsuffix*. $geps(w)$ ist das größte echte Präsuffix von w.
$w_j := s_{j+i}...s_m$.

$$\gamma(j) := |geps(w_j)|$$

Suffixfunkion σ wird dargestellt:

$$\sigma(j) = min(\{k \in \{1, ..., j\} |\ \gamma(j - k) = m - j\} \cup \{m - \gamma(0)\})$$

Algorithmus für σ (Laufzeit $O(m)$)

```
Eingabe:
        γ(0 ... m)
Ausgabe:
        σ(1 ... m)

For j = 1, ..., m
        σ(j) ← m - gamma(0)
For i = 0, ..., m-1
        k ← m - γ(i) - i
        j ← m - γ(i)
        // Es gilt jetzt γ(j-k) = m - j
        falls σ(j) > k
                dann σ(j) ← k
```

Algorithmus für γ (Laufzeit $O(m)$)

```
Eingabe:
        s_1...s_m
Ausgabe:
        γ(0 ... m-1)

γ(m-1) ← 0
For i = m-1, ..., 1
        j ← m - γ(i)
        Solange s_i ≠ s_j und m ≠ j
                j ← m - γ(j)
        falls s_i = s_j
        dann γ(i-1) ← m-j+1
        sonst γ(i-1) ← 0
```

Boyer-Moore-Algorithmus

```
// wie in "Naiver Suche", letzte Zeile
// durch folgendes ersetzt
v ← v(t_{i+j})
falls v < j
        i ← i + max{j-v, σ(j)}
sonst
        i ← i + max{m-v+1, σ(j)}
```