Formale Sprachen

und

Strukturelle Induktion

Korrekte arithmetische Terme

- Betrachten wir den Wert des folgenden Terms:

$$4 + 2 \cdot 3 = 18$$

$$4 + (2 \cdot 3) = 10$$

$$(4 + 2) \cdot 3 = 18$$

- Und dann:

$$(14295 + 55328) \cdot 4672$$

- Wie steht es mit diesem?

$$14295 \cdot +55328\ 4672()$$

Alphabete, Zeichenketten, formale Sprachen

- **Definition**. Ein *Alphabet* ist eine endliche Menge von Symbolen.
- Die Symbole nennt man auch „Zeichen“ und Alphabete daher auch „Zeichenvorräte“.
- **Definition (Zeichenketten).** Sei A ein Alphabet. Wir bezeichnen mit $A^k = \{a_1 \dots a_k | a_i \in A, 1 \leq i \leq k\}$ die Menge aller Folgen der Länge k von Zeichen aus A. Wir definieren weiterhin: $A^+ = \bigcup_{k \in \mathbb{N}} A^k = A^1 \cup A^2 \cup A^3 \cup \cdots$
- **Definition**. Sei A ein Alphabet. Eine *(formale) Sprache (über A)* ist eine Teilmenge $L \subseteq A^+$. Die Elemente von L nennt man *Worte*.

Beispiel

- Sei $A = \{\text{Alice, Bob, ignoriert, liebt}\}$ ein Alphabet.
- Dann ist:
  - $A^1 = A$
  - $A^2 = \{\text{Alice Alice, Alice Bob, Alice ignoriert, Alice liebt, Bob Alice, Bob Bob, Bob ignoriert, Bob liebt, ignoriert Alice, ignoriert Bob, ignoriert ignoriert, ignoriert liebt, liebt Alice, liebt Bob, liebt ignoriert, liebt liebt}\}$
- Die Menge $L = \{\text{Bob liebt Alice, Alice ignoriert Bob}\}$ ist eine formale Sprache über A, denn $L \subseteq A^3 \subseteq A^+$.
- Quiz: Was sind die *Worte* von L? "*Bob liebt Alice*" und "*Alice ignoriert Bob*"

Syntax und Semantik

- Aus jedem Alphabet A kann man unendlich viele Zeichenfolgen bilden.
  - Beispiel: $A = \{a\} \Rightarrow A^+ = \{a, aa, aaa, aaaa, \ldots\}$
- In konkreten Anwendungen sind nicht alle Zeichenfolgen interessant. Die Definition einer Sprache $L \subseteq A^+$ erlaubt uns, uns auf gewisse Worte zu konzentrieren.
- *Syntax*: gibt an, welche Zeichenketten zulässig sind und welche nicht, allerdings ohne eine Aussage über deren Bedeutung zu machen.
  - „Farblose grüne Ideen schlafen wutentbrannt" ✓
  - „Ging hallo mit und unter des Frieden gestern Fußball" ✖
- *Semantik*: gibt an, welche Bedeutung eine syntaktisch korrekte Zeichenkette besitzt
- Eine formale Sprache definiert also zunächst lediglich eine Syntax ohne Semantik.

Ein etwas komplexeres Beispiel

- **Definition.** Sei $\hat{F}$ eine endliche Menge von Bezeichnern und sei $V = \{X_1, \dots, X_n\}$ für ein $n \in \mathbb{N}$. Wir betrachten das Alphabet $A = \{\mathbf{0}, \mathbf{1}, \sim, \wedge, \vee, (,), ,\} \cup V \cup \hat{F}$ und definieren die Sprache B der *vollständig geklammerten Boole'schen Ausdrücke* wie folgt rekursiv:
  - $\{\mathbf{0}, \mathbf{1}\} \subseteq B$
  - $V \subseteq B$
  - Ist $b \in B$, so ist auch $(\sim b) \in \mathrm{B}$
  - Sind $b_1, b_2 \in B$ so sind auch $(b_1 \vee b_2) \in B$ und $(b_1 \wedge b_2) \in B$
  - Sind $b_1, \dots, b_k \in B$ und ist $f \in \hat{F}$, so ist auch $f(b_1, \dots, b_k) \in B$.
  - Ansonsten enthält B keine weiteren Worte.

Beispiele

- Die vier Zeichenketten $\mathbf{0}$, $\mathbf{1}$, X_{42} und X_{101} sind alle vollständig geklammerte Boole'sche Ausdrücke.
- Daher sind auch $(\sim\mathbf{1})$ und $(\sim X_{42})$ v.g.B.A.
- Daher sind auch $(\mathbf{0} \wedge (\sim\mathbf{1}))$ und $((\sim X_{42}) \vee X_{101})$ v.g.B.A.
- Daher ist auch $\left((\mathbf{0} \wedge (\sim\mathbf{1})) \vee ((\sim X_{42}) \vee X_{101})\right)$ ein v.g.B.A.

Verwendung Boole’scher Ausdrücke

- Das war erst einmal alles zum Thema “Formale Sprachen”.
- Wir werden uns noch etwas weiter mit B.A. beschäftigen, da diese durchaus nützlich sind:
- Aussagenlogik
  - Ausdrücke ohne Funktionssymbole und meist ohne **0**, **1**, da redundant
  - Ausdrücke werden dort üblicherweise “Formeln” genannt
- Schaltkreistheorie
  - Einfaches Bindeglied zwischen Schaltkreisen und Schaltfunktionen

Strukturelle Induktion

- Will man Aussagen über rekursiv definierte Strukturen (wie etwa vollständig geklammerte Boole'sche Ausdrücke) beweisen, bietet sich folgendes Verfahren an:
  - Beweis der Aussage für die Anfangsfälle (oft nur ein einziger)
  - Beweis der Aussage für alle der rekursiv definierten Fälle unter Rückgriff auf die bereits bewiesenen Fälle

Beispiel

- Behauptung: Besteht ein v.g.B.A. b aus n_b Zeichen $a_1, \dots, a_{n_b} \in A$, so kommen in b höchstens $\frac{n_b+1}{2}$ Variablen $X_i \in V$ vor.
- Beweis durch strukturelle Induktion über b.
  Sei v_b die Anzahl der in b enthaltenen Variablen.

  Induktionsanfang:
  - Fall: $b = \mathbf{0}$ oder $b = \mathbf{1}$. Dann besteht b aus genau einem Zeichen, enthält aber keine Variablen: $v_b = 0 < \frac{1+1}{2} = \frac{n_b+1}{2}$ ■
  - Fall: $b \in V$. Dann besteht b ebenfalls aus einem Zeichen und enthält genau eine Variable: $v_b = 1 = \frac{1+1}{2} = \frac{n_b+1}{2}$ ■

- $\left((\sim\mathbf{0}) \vee X_5\right)$
- 8 Zeichen
- 1 Variable
- $1 \le \frac{8+1}{2} = 4.5$

Beispiel (Fortsetzung)

Induktionsschritt

1. Hat b die Form $(\sim b')$, wobei b' eine v.g.B.A. ist, dann ist $n_b = n_{b'} + 3$.

   Es gilt: $v_b = v_{b'} \leq \frac{n_{b'}+1}{2} \leq \frac{n_{b'}+4}{2} = \frac{n_b+1}{2}$ ■

2. Hat b die Form $(b_1 \vee b_2)$ oder $(b_1 \wedge b_2)$, wobei b_1, b_2 v.g.B.A. sind, dann ist $n_b = n_{b_1} + n_{b_2} + 3$.

   Es gilt: $v_b = v_{b_1} + v_{b_2} \leq \frac{n_{b_1}+1}{2} + \frac{n_{b_2}+1}{2} \leq \frac{n_{b_1}+n_{b_2}+4}{2} = \frac{n_b+1}{2}$ ■

3. Hat b die Form $f(b_1, \dots, b_k)$, wobei $b_1, \dots, b_k$ v.g.B.A. sind, dann ist $n_b = n_{b_1} + \cdots + n_{b_k} + k + 2$ und es gilt:

$$
\begin{aligned}
v_b &= v_{b_1} + \cdots + v_{b_k} \\
&\leq \frac{n_{b_1}+1}{2} + \cdots + \frac{n_{b_k}+1}{2} \\
&= \frac{n_{b_1}+\cdots+n_{b_k}+k}{2}
\end{aligned}
$$

$$
\begin{aligned}
&\leq \frac{n_{b_1}+\cdots n_{b_k}+k+3}{2} \\
&= \frac{n_b+1}{2}
\end{aligned}
$$ ■

qed