

Formale Grundlagen der Informatik 1

Kapitel 7

Kontextfreie Sprachen

Frank Heitmann
heitmann@informatik.uni-hamburg.de

25. April 2016

Formalismen für REG

Satz

Zu jeder regulären Sprache L gibt es

- einen DFA A mit $L(A) = L$
- einen NFA B mit $L(B) = L$
- einen NFA mit λ -Kanten C mit $L(C) = L$
- einen regulären Ausdruck D , der L beschreibt ($M_D = L$)

DFAs, NFAs, NFAs mit λ -Kanten und reguläre Ausdrücke sind also äquivalent.

Abschlusseigenschaften

Satz

Die regulären Sprachen sind abgeschlossen gegenüber

- Vereinigung \cup
- Konkatenation \cdot
- "hoch +"
- "hoch *"
- Komplementbildung
- Durchschnitt \cap
- Reverse
- ...

Das Pumping Lemma

Lemma

Sei $L \in REG$ eine reguläre Sprache. Dann gibt es ein $n \geq 1$, so dass jedes Wort $z \in L$ mit $|z| \geq n$ in die Form $z = uvw$ zerlegt werden kann, wobei

- 1 $|uv| \leq n$
- 2 $|v| \geq 1$
- 3 $uv^i w \in L$ für jedes $i \in \mathbb{N}$ (inkl. der 0)

(Äquivalent zur letzten Aussage ist die Aussage $\{u\}\{v\}^*\{w\} \subseteq L$.)

Das Pumping Lemma - Der Ablauf

Der Ablauf beim Pumping Lemma:

- 1 Annehmen L wäre regulär
- 2 Die Zahl aus dem Pumping Lemma benennen (z.B. k).
- 3 **Ein Wort** z finden mit $|z| \geq k$ und $z \in L$.
 - Dieses Wort muss gut gewählt sein, damit der nachfolgende Widerspruch gelingt! Hier muss man also u.U. experimentieren!
- 4 **Für alle Zerlegungen** von z in uvw zeigen, dass sie im Widerspruch zur ersten, zweiten oder dritten Bedingung sind.
 - Üblicherweise betrachtet man alle Zerlegungen, die die erste und zweite Bedingung erfüllen und zeigt, dass man dann einen Widerspruch zur dritten Bedingung erhält.
- 5 Also: Alle Zerlegungen $z = uvw$ mit $|uv| \leq k$ und $|v| \geq 1$ betrachten.
- 6 Zeigen, dass man **ein** i findet mit $uv^i w \notin L$. (Dies kann auch $i = 0$ sein!)

Das Pumping Lemma - Beispiel 1

Behauptung

$L := \{a^n b^n \mid n \in \mathbb{N}\}$ ist nicht regulär.

Beweis.

Angenommen L wäre regulär. Dann gilt das Pumping Lemma. Sei k die Zahl aus dem Pumping Lemma. Wir betrachten das Wort $z = a^k b^k$. Es gilt $z \in L$ und $|z| \geq k$. Damit muss es eine Zerlegung $z = uvw$ geben, die die drei Eigenschaften erfüllt. Wir betrachten nun nur jene Zerlegungen, die die erste und zweite Bedingung erfüllen und zeigen, dass diese im Widerspruch zur dritten stehen. Wir haben dann gezeigt, dass jede Zerlegung von z in uvw im Widerspruch zu einer der drei Bedingungen steht und sind dann fertig. ...

Das Pumping Lemma - Beispiel 1

Behauptung

$L := \{a^n b^n \mid n \in \mathbb{N}\}$ ist nicht regulär.

Beweis.

Sei also $z = a^k b^k = uvw$ eine Zerlegung mit *i*) $|uv| \leq k$ und *ii*) $|v| \geq 1$. Es muss dann $v \in \{a\}^+$ gelten, da uv ja die ersten Buchstaben von z ausmacht, aber davon nur maximal k einnehmen kann und da ferner v mindestens die Länge 1 hat. Es gibt sogar ein j mit $1 \leq j \leq k$ derart, dass $v = a^j$ gilt. Wir betrachten nun das Wort uv^2w , das nach der dritten Bedingung in L sein müsste. Es ist aber $uv^2w = a^{k+j}b^k$ und somit $uv^2w \notin L$. Die ursprüngliche Annahme muss also falsch sein und daher ist L nicht regulär. \square

Fragen...

Wovon muss das gewählte Wort z abhängen?

- 1 Von der Anzahl der Zustände $|Z|$ des DFA.
- 2 Von der Zahl aus dem Pumping Lemma k , die noch genauer spezifiziert werden muss.
- 3 Von der Zahl aus dem Pumping Lemma k , die nicht genauer spezifiziert werden muss.
- 4 Das Wort muss lediglich in der Sprache sein.

Fragen...

Wieviele Zerlegungen von z müssen betrachtet werden?

- 1 Eine muss zum Widerspruch geführt werden!
- 2 Alle Zerlegungen von z müssen zum Widerspruch gebracht werden!
- 3 Zerlegungen mit bestimmten Eigenschaften müssen zum Widerspruch gebracht werden!
- 4 Man muss eine Zerlegung mit bestimmten Eigenschaften finden und diese zum Widerspruch führen!

Fragen...

Kann mit dem Pumping Lemma gezeigt werden, dass eine Sprache regulär ist?

- ① Ja!
- ② Nein!

Fragen...

Ist $L := \{ww^{rev} \mid w \in \{0,1\}^*\}$ regulär?

- 1 Ja!
- 2 Nein!

Zur Nachbereitung

Zur Nachbereitung

Richtige Antworten sind:

- ① 3
- ② 2 und 3 ist beides ok
- ③ 2
- ④ 2 (Beweis folgt)

Das Pumping Lemma - Beispiel 2

Behauptung

$L := \{ww^{rev} \mid w \in \{0,1\}^*\}$ ist nicht regulär.

Beweis.

Angenommen L wäre regulär. Dann gilt das Pumping Lemma. Sei r die Zahl aus dem Pumping Lemma und $k \geq r$ der kleinste gerade Nachfolger. Wir betrachten $z = 0^{k/2}1^k0^{k/2}$. Es ist $z \in L$ und $|z| \geq k$. Sei $z = uvw$ eine Zerlegung von z mit $|uv| \leq k$ und $|v| \geq 1$. v kann nun in den ersten 0en liegen, in den 0en und den 1en oder in den 1en. Wenn v in den 0en liegt ($v = 0^j$), dann ist $uv^2w = 0^{k/2+j}1^k0^{k/2}$ nicht mehr in L und wir haben einen Widerspruch. Wenn v aus 0en und 1en besteht erreichen wir ebenfalls mit uv^2w einen Widerspruch. Wenn v aus 1en besteht ($v = 1^j$), dann ist $uv^2w = 0^{k/2}1^{k+j}0^{k/2}$... kein Widerspruch, wenn j gerade ist! Auch mit uv^0w oder einem anderen $uv^i w$ erreichen wir dann keinen Widerspruch. ... Die Wahl von z war vielleicht schlecht ... ein neuer Versuch ... □

Das Pumping Lemma - Beispiel 2

Behauptung

$L := \{ww^{rev} \mid w \in \{0,1\}^*\}$ ist nicht regulär.

Beweis.

Angenommen L wäre regulär. Dann gilt das Pumping Lemma. Sei k die Zahl aus dem Pumping Lemma. Wir betrachten $z = 0^k 1^k 1^k 0^k$. Es ist $z \in L$ und $|z| \geq k$. Sei $z = uvw$ eine Zerlegung von z mit $|uv| \leq k$ und $|v| \geq 1$. Hieraus folgt $v = 0^j$ für ein j mit $1 \leq j \leq k$. Nun ist aber $uv^0w = 0^{k-j} 1^k 1^k 0^k$ nicht mehr in L im Widerspruch zum Pumping Lemma. Die Annahme ist daher falsch und L somit nicht regulär. \square

Das Pumping Lemma - Nochmal der Ablauf

Der Ablauf beim Pumping Lemma:

- 1 Annehmen L wäre regulär
- 2 Die Zahl aus dem Pumping Lemma benennen (z.B. k).
- 3 **Ein Wort** z finden mit $|z| \geq k$ und $z \in L$.
 - Dieses Wort muss gut gewählt sein, damit der nachfolgende Widerspruch gelingt! Hier muss man also u.U. experimentieren!
- 4 **Für alle Zerlegungen** von z in uvw zeigen, dass sie im Widerspruch zur ersten, zweiten oder dritten Bedingung sind.
 - Üblicherweise betrachtet man alle Zerlegungen, die die erste und zweite Bedingung erfüllen und zeigt, dass man dann einen Widerspruch zur dritten Bedingung erhält.
- 5 Also: Alle Zerlegungen $z = uvw$ mit $|uv| \leq k$ und $|v| \geq 1$ betrachten.
- 6 Zeigen, dass man **ein** i findet mit $uv^i w \notin L$. (Dies kann auch $i = 0$ sein!)

Das Pumping Lemma - Anmerkung zum Ablauf

Anmerkung

Bei der Zerlegung (vorletzter Schritt eben) gelingt oft eine Parametrisierung. Bspw. konnte in den Beispielen oben $v = 0^j$ bzw. $v = a^j$ mit einem j mit $1 \leq j \leq k$ gesetzt werden. Das j wandert quasi den Bereich von 1 bis k ab. Für jedes j ist dies eine mögliche Zerlegung, die die ersten beiden Bedingungen erfüllt – und man führt im Grunde genommen *alle* diese Zerlegungen zum Widerspruch!

Zusammenfassung - Überblick

- DFA, NFA, (NFA mit λ -Kanten)
 - Zustände, Startzustand, Endzustände
 - Überföhrungsfunktion
 - erweiterbare Überföhrungsfunktion
 - vollständig, initial zusammenhängend
 - Konfiguration, Konfigurationsübergang
 - Rechnung, Erfolgsrechnung
 - akzeptierte Sprache
- reguläre Ausdröcke
- Potenzautomatenkonstruktion
- Konstruktionstechniken und $L(A) = M$
- Abschlusseigenschaften von *REG*
- Grenzen von *REG*
 - Pumping Lemma
 - (Kellerautomaten)

Motivation

Bisher hatten wir

- **Automaten**, die Wörter **akzeptieren**

Wir werden nun ein neues Modell kennenlernen:

- **Grammatiken**, die Wörter **generieren**

Informales Beispiel

Ihr kennt vielleicht schon Beispiele bei der Definition von Programmiersprachen! Hier z.B. um mögliche Identifier abzuleiten, die Ziffern und Buchstaben enthalten dürfen, aber mit einem Buchstaben beginnen müssen:

$$\langle \text{identifier} \rangle ::= \langle \text{letter} \rangle \mid \langle \text{identifier} \rangle \langle \text{letter} \rangle \mid \langle \text{identifier} \rangle \langle \text{digit} \rangle$$
$$\langle \text{letter} \rangle ::= A \mid B \mid C$$
$$\langle \text{digit} \rangle ::= 0 \mid 1 \mid 2 \mid 3 \mid 4$$

oder kompakter:

$$I \longrightarrow L \mid IL \mid ID$$
$$L \longrightarrow a \mid b \mid c$$
$$D \longrightarrow 0 \mid 1 \mid 2 \mid 3 \mid 4$$

Informales Beispiel

$$I \longrightarrow L \mid IL \mid ID$$

$$L \longrightarrow a \mid b \mid c$$

$$D \longrightarrow 0 \mid 1 \mid 2 \mid 3 \mid 4$$

Eine mögliche Ableitung eines Identifiers (Ziel: $abc1$):

$$I \Rightarrow ID \Rightarrow ILD \Rightarrow ILLD \Rightarrow LLLD$$

$$\Rightarrow aLLD \Rightarrow abLD \Rightarrow abL1 \Rightarrow abc1$$

Hin zu Grammatiken

Dies ist genau die Idee bei Grammatiken!

- Die Großbuchstaben, die ersetzt werden können, sind die **Nonterminale**.
- Die Kleinbuchstaben, aus denen sich das abgeleitete Wort zusammensetzt, sind die **Terminale**.
- Die $I \rightarrow IL$ usw. sind die **Regeln** oder **Produktionen**.

Wir definieren nun zunächst die Grammatik *sehr allgemein* und schränken diese später ein...

Grammatiken. Formale Definition

Definition (Grammatik)

Eine **Grammatik** ist ein Quadrupel $G = (V_N, V_T, P, S)$ mit

- 1 Dem endlichen Alphabet von **Nonterminalen** V_N .
- 2 Dem endlichen Alphabet von **Terminalen** V_T mit $V_T \cap V_N = \emptyset$. Das *Gesamtalphabet* wird mit $V := V_T \cup V_N$ bezeichnet.
- 3 Der endlichen Menge von **Produktionen** (oder *Regeln*) $P \subseteq (V^* \setminus V_T^*) \times V^*$.
- 4 Dem **Startsymbol** $S \in V_N$.

Grammatiken. Anmerkungen

Bemerkung

- Auf der linken Seite einer Regel steht stets mindestens ein Nonterminal!

$$P \subseteq (V^* \setminus V_T^*) \times V^*$$

Notationen

- Eine Regel $(u, v) \in P$ wird meist als $u \rightarrow v$ notiert.
- Mehreren Regeln $u \rightarrow v$ und $u \rightarrow w$ mit gleicher linker Seite werden als $u \rightarrow v \mid w$ abgekürzt.
- Bei $u \rightarrow v \mid w, v \rightarrow w \mid s \mid t$ folgen nach dem Komma Regeln mit neuer linker Seite. Gemeint ist dann also, dass es die Regeln $(u, v), (u, w)$ sowie $(v, w), (v, s)$ und (v, t) in P gibt.

Ableitung

Definition (Ableitung)

Die **einschrittige Ableitung** eines Wortes v aus einem Wort u mittels einer Produktion einer Grammatik G wird notiert als

$u \xrightarrow{G} v$. Dabei ist die Relation $\xrightarrow{G} \subseteq V^* \times V^*$ für alle

$u, v \in V^*$ definiert durch: $u \xrightarrow{G} v$ gdw.

$$\exists u_1, u_2 \in V^* \exists (w_l, w_r) \in P : u = u_1 w_l u_2 \text{ und } v = u_1 w_r u_2$$

Ist der Kontext klar, wird das tief gestellte G weggelassen. Ferner bedienen wir uns wieder der reflexiven, transitiven Hülle \xrightarrow{G}^* für **mehrschrittige Ableitungen**.

Grammatiken. Beispiel

Vorgehen

Eine Ableitung funktioniert also so:

- Das aktuelle Wort w betrachten
- Treten in w linke Seiten von Regeln auf?
- Falls ja, wähle eine dieser Regeln und
- ersetze die in w auftretende linke Seite der Regel
- durch die rechte Seite dieser Regel

Beispiel

Mit $S \rightarrow ASB \mid AB$, $AB \rightarrow c$, $AcB \rightarrow c$ gilt

$$S \Rightarrow ASB \Rightarrow AABB \Rightarrow AcB \Rightarrow c$$

Generierte Sprache

Definition (Generierte Sprache)

Sei $G = (V_N, V_T, P, S)$ eine Grammatik. Die von G **generierte** oder **erzeugte** Sprache ist

$$L(G) := \{w \in V_T^* \mid S \xrightarrow[G]{*} w\}$$

Definition (Äquivalenz von Grammatiken)

Zwei Grammatiken G_1 und G_2 werden genau dann als **äquivalent** bezeichnet, wenn $L(G_1) = L(G_2)$ gilt.

Anmerkung

Eine Zeichenkette $u \in V^*$ mit $S \xrightarrow[G]{*} u$ nennt man auch *Satzform*.

Fragen...

Wenn man zwei Regeln mit gleichen linken Seite hat, wie wird ausgewählt welche Regel benutzt wird?

- 1 Das darf es nicht geben!
- 2 Darf man frei wählen!
- 3 Die, die am weitesten links steht!
- 4 Es werden alle benutzt!

Fragen...

Kann man zwei Regeln mit gleicher rechten Seite haben?

- 1 Ja!
- 2 Nein!

Fragen...

Kann es mehrere Möglichkeiten geben, um das gleiche Wort abzuleiten?

- 1 Ja!
- 2 Nein!

Zur Nachbereitung

Zur Nachbereitung

Richtige Antworten sind:

- ① 2
- ② 1
- ③ 1

Kontextfreie Grammatiken

Die bisher eingeführten Grammatiken sind sehr mächtig.

Wir wollen sie jetzt zunächst einschränken und bei den eingeschränkten Grammatiken gucken, was man damit alles machen kann...

Kontextfreie Grammatiken. Formale Definition

Definition (Kontextfreie Grammatik (CFG))

Eine **kontextfreie Grammatik** (CFG) ist ein Quadrupel $G = (V_N, V_T, P, S)$ mit

- 1 Dem endlichen Alphabet von **Nonterminalen** V_N .
- 2 Dem endlichen Alphabet von **Terminalen** V_T mit $V_T \cap V_N = \emptyset$. Das *Gesamtalphabet* wird mit $V := V_T \cup V_N$ bezeichnet.
- 3 Der endlichen Menge von **Produktionen** (oder *Regeln*) $P \subseteq V_N \times V^*$.
- 4 Dem **Startsymbol** $S \in V_N$.

Kontextfreie Grammatiken werden auch als **Typ-2**-Grammatiken bezeichnet.

Grammatiken vs. kontextfreie Grammatiken

Anmerkung

Der Unterschied zwischen (allgemeinen) Grammatiken und kontextfreien Grammatiken liegt lediglich in der Menge der Produktionen P :

- bei allgemeinen Grammatiken:

$$P \subseteq (V^* \setminus V_T^*) \times V^*$$

- bei kontextfreien Grammatiken:

$$P \subseteq V_N \times V^*$$

Bei einer kontextfreien Grammatik ist bei einer Regel also stets genau ein Nonterminal auf der linken Seite! Daher auch der Name kontextfrei. Das Nonterminal wird unabhängig von dem Kontext, in dem es steht, ersetzt.

CFG - das λ Definition (λ -Produktionen, λ -frei)

- Eine kontextfreie Produktion (A, λ) wird als λ -**Produktion** bezeichnet.
- Besitzt eine CFG keine λ -Produktionen, so heißt sie λ -**frei**.

Sprachfamilie CF

Definition (Sprachfamilie CF)

Ableitung und akzeptierte Sprache ist wie bei Grammatiken definiert. Die **Familie der kontextfreien Sprachen** ist dann jene Familie von Sprachen, für die es eine kontextfreie Grammatik gibt, die sie generiert. Abgekürzt wird diese Sprachfamilie mit CF.

Wichtige Bemerkung

Neben REG haben wir nun mit CF unsere zweite Sprachfamilie kennengelernt!

Fragen...

Welche der folgenden Grammatiken ist *nicht* kontextfrei?

- ① $S \rightarrow AB, A \rightarrow aA \mid \lambda, B \rightarrow bB \mid \lambda$
- ② $S \rightarrow Ab \mid aB, A \rightarrow aA \mid a, B \rightarrow bB \mid b$
- ③ $S \rightarrow AB, AB \rightarrow aABb \mid ab$
- ④ zwei davon sind nicht kontextfrei
- ⑤ alle drei sind nicht kontextfrei
- ⑥ keine Ahnung...

Fragen...

Welche Ableitung ist mit

$$S \rightarrow AB \mid cC, A \rightarrow aAb \mid \lambda, B \rightarrow bB \mid C, C \rightarrow cC \mid \lambda$$

möglich?

- ① $S \Rightarrow AB \Rightarrow AC \Rightarrow AcC \Rightarrow cC \Rightarrow c$
- ② $S \Rightarrow AB \Rightarrow aAbB \Rightarrow aAbC \Rightarrow aAb \Rightarrow ab$
- ③ $S \Rightarrow AB \Rightarrow aAbB \Rightarrow aAbC \Rightarrow aAbc \Rightarrow abc$
- ④ zwei davon sind möglich
- ⑤ alle drei sind möglich
- ⑥ keine Ahnung...

Fragen...

Was gilt für die Menge der Produktionen einer kontextfreien Grammatik?

- ① $P \subseteq (V^* \setminus V_T^*) \times V^*$
- ② $P \subseteq V_N \times V_T^*$
- ③ $P \subseteq V_N \times V^*$
- ④ $P \subseteq (V^* \setminus V_N^*) \times V_T^*$
- ⑤ $P \subseteq V_N \times V$
- ⑥ keine Ahnung...

Zur Nachbereitung

Zur Nachbereitung

Richtige Antworten sind:

- 1 3
- 2 4 (nämlich 1 und 2; die 3 ist aber nur falsch, weil $aAbC \Rightarrow aAbc$ nicht in einem Schritt (sondern nur in zweien) möglich ist)
- 3 3

Zusammenfassung

Definitionen bisher:

- Grammatik
 - Nonterminal, Terminal, Produktion, Startsymbol
 - Ableitung
 - (Generierte) Sprache, Äquivalenz
- Kontextfreie Grammatik / Typ-2-Grammatik
 - λ -Produktionen, λ -frei
- Sprachfamilie CF

Morgen:

- Grammatiken zu Sprachen konstruieren
- Eigenschaften kontextfreier Grammatiken
- Grenzen kontextfreier Grammatiken