Kapitel 1

Einleitung

1.1 Formale Sprachen

Die Lehre von den formalen Sprachen ist entstanden aus der (algorithmischen) Theorie von endlich erzeugten Algebren, Halbgruppen, Gruppen. Wir befinden uns also sehr nahe der Mathematik. Das ist gut, wir verwenden ihre exakte und klare Schreib- und Denkweise.

Die Lehre von den formalen Sprachen ist damit ein sehr altes (100 Jahre) Teilgebiet der Theoretischen Informatik. Auch das ist gut, die Anfangsschwierigkeiten sind überwunden, es herrscht Einvernehmen über die wesentlichen Definitionen, es gibt klassische Sätze, sowohl einfache (die zeigen, daß die Begriffe gut gewählt sind) als auch tiefliegende (die zeigen, daß wir interessante Strukturen untersuchen).

1.2 Maschinen, auch endliche

Informatik ist Lehre von den Algorithmen. Im Allgemeinen laufen diese auf Turing-Maschinen. Damit sind Programm-Eigenschaften nicht entscheidbar. (Bsp: ob eine Eingabe von einer Maschine akzeptiert wird.)

Das ist schade. Es wäre z. B. schön, für ein Programm P zu zeigen, daß es die korrekte Ausgabe liefert oder nicht abstürzt usw. (Bsp: die Eingabe akzeptiert.)

Wir vereinfachen P zu einem Programm P' mit beweisbaren Eigenschaften. Beispielsweise: schränken die Ressourcen ein. Sehr starke Einschränkung: nur endlichen Speicher: endlicher Automat.

Dafür ist Akzeptanz der Eingabe entscheidbar. Das hätten wir auch anders haben können (kontextfrei CF, kontextsensitiv CS).

Der Vorteil in REG ist: akzeptieren geht schnell (Linearzeit) und reguläre Sprachen sind boolesch abgeschlossen. (Übung: CF nicht — CS ja. Aber wie lange dauert Akzeptieren?)

Endliche Automaten sind also eine handliche Maschinenklasse.

1.3 Bäume

Die klassische Theorie formaler Sprachen behandelt Mengen von Wörtern. Es ist aber oft so, daß die Daten, auf denen Programme operieren, eine hierarchische Struktur haben: sie sind Bäume. Deswegen interessieren wir uns für Automaten auf Bäumen.

Wir können viele der Resultate (oder wenigstens Methoden) von Wortsprachen übertragen, aber benötigen manchmal Modifikationen, und auch völlig neue Ideen.

1.4 Endlich gleich einfach?

Ein endlicher Automat (reguläre Sprache) erscheint als etwas sehr Einfaches. Das täuscht. Neben den erwähnten klassischen Resultaten gibt es auch viele schwere und offene Probleme.

1.5 Erweiterte Sternhöhe

Ein (erweiterter) regulärer Ausdruck beschreibt eine reguläre Wortsprache.

1. Zeichen $\emptyset$ für leere Wortmenge
2. Zeichen ϵ für leeres Wort
3. für jeden Buchstaben $a \in \Sigma$: er selbst
4. Aneinanderhängen von Wörtern
5. Vereinigen von Wortmengen

6. Kleene-Stern einer Wortmenge

7. (nur in erweiterten Ausdrücken erlaubt:) Komplement einer Wortmenge

Beispiel: $\Sigma = \{a, b\}$. Menge aller nichtleeren Wörter der Form $abab \dots abab$ ist $(ab)^*$.

Die Sternhöhe eines (erweiterten) regulären Ausdrucks ist die maximale Anzahl geschachtelter Sterne in ihm. (Das Beispiel hat Sternhöhe 1.)

Die gleiche Menge wird beschrieben durch $a\overline{\emptyset}b \setminus \overline{\emptyset}(aa \cup bb)\overline{\emptyset}$. (Dieser erweiterte Ausdruck hat Sternhöhe 0.)

Die erweiterte Sternhöhe einer regulären Sprache ist das Minimum der Sternhöhen aller erweiterten regulären Ausdrücke, die diese Sprache beschreiben. (Die Beispielsprache hat erweiterte Sternhöhe 0.)

Eine Sprache mit erweiterter Sternhöhe 1 ist $(aa)^*$.

Es ist bis heute unbekannt, ob es Sprachen mit erweiterter Sternhöhe > 1 gibt. Übung: erweiterte Sternhöhe von $((aa)^*b)^*$.

1.6 Baumgleichungen

Eine Baumgleichung ist ein Paar (l, r) von Termen aus $T(\Sigma \cup \Delta)$, wobei Σ eine Signatur von Konstanten und Δ eine Signatur von Variablen.

Eine Lösung der Gleichung $l = r$ ist ein Homomorphismus $h : T(\Sigma \cup \Delta) \to T(\Sigma)$, der auf $T(\Sigma)$ konstant ist, mit $h(l) = h(r)$.

Es ist bis heute unbekannt, ob die Lösbarkeit einer Baum-Gleichung entscheidbar ist.

Für den Wort-Fall (die Signaturen enthalten nur Symbole der Stelligkeit 1) gibt es ein Entscheidungsverfahren. Das hat aber eine Laufzeit von

$$\exp(\exp(\exp(\exp(\exp(n))))),$$

wenn n die Eingabegröße ist.