

2. Übungsblatt - Abgabe: 15.11.2011

Aufgabe 2.1 Morphologie

Natürlich-sprachliche Systeme verwenden entweder Wortformenlexika, oder sie verwenden eine Flexionsmorphologie (Lemmatisierer), und brauchen dann nur die Wortstämme explizit als Einträge aufzuführen (Stammlexika). Die Ersparnis durch eine Flexionsmorphologie ist von Sprache zu Sprache stark verschieden: Bei der flexionsarmen englischen Sprache ist das Verhältnis Wortformen: Stämme und damit der Einsparungsfaktor < 2 . Bei Sprachen wie dem Türkischen und Finnischen liegt er eher in der Gegend von 50 oder 100. Beim Deutschen liegt er irgendwo zwischen den Extremen.

Versuchen Sie, die Relation Wortformen:Stämme für das Deutsche abzuschätzen, und beschreiben Sie Ihr Vorgehen.

Hinweis: Wichtig ist, dass Sie nachvollziehbar darstellen, wie Sie zu Ihrem Resultat gekommen sind. Es ist nicht erforderlich, dass Sie bei einer genauen Ziffer ankommen. Eine Schätzung der Größenordnung reicht. Wenn Sie Schwierigkeiten oder Probleme für die Schätzung sehen, stellen Sie sie bitte ebenfalls dar. Falls Sie in Ihrer Abschätzung Angaben aus Büchern oder aus dem Internet nutzen, geben Sie bitte die vollständige Quelle an.

Aufgabe 2.2 Endliche Automaten und Morphologie

- (a) Das Adjektiv-Endungs-Diagramm auf der Vorlesungsfolie akzeptiert das Wort *st*. Die Superlativendung *kleinst* kommt aber für sich allein gar nicht vor. Korrigieren Sie den NEA mit möglichst wenig Aufwand so, dass *st* für sich genommen nicht mehr akzeptiert wird (alle übrigen Endungen dagegen wie bisher).
- (b) Entwerfen Sie einen möglichst einfachen NEA, der die einfachen flektierten Formen des Verbs *reden* spezifiziert (Präsens und Präteritum, ohne Partizipialendungen!).

Aufgabe 2.3 Endliche Automaten und Syntax

- (a) In der Vorlesung haben wir einen NEA betrachtet, der einige zulässige Wortartketten für Nominalausdrücke spezifiziert. Der NEA spezifiziert Nominalausdrücke mit Artikel, Gattungssubstantiv, und einem potentiellen pränominalen adjektivischen Attribut, möglicherweise mit Gradpartikeln (GPRT). Erweitern bzw. modifizieren Sie den Automaten so, dass es auch Nominalausdrücken mit post-nominalen Präpositionalausdrücken umfasst. Also z.B.:

ART NN (*das Auto*)

ART ADJA ADJA NN (*das neue schnelle Auto*)

ART ADJA NN APPR ART NN (*das grüne Auto auf dem Parkplatz*)

ART NN APPR ART NN APPR ART ADJA NN (*das Auto auf dem Parkplatz bei dem neuen Institutsgebäude*)

- (b) Versuchen Sie, Nominalausdrücke mit Personalpronomen (PPER) und Eigennamen (NE) im Diagramm mit zu berücksichtigen und geben Sie jeweils ein Beispiel an, das von Ihrem Automaten akzeptiert wird.
- (c) Geben Sie zwei unterschiedliche Beispiele für Nominalausdrücke, die durch das Zustandsdiagramm nicht abgedeckt sind. – Erweitern Sie das Diagramm so, dass es auch diese Fälle akzeptiert.

Aufgabe 2.4 Endliche Automaten und Zahlausdrücke

- (a) Entwerfen Sie einen Automaten, der alle durch 2 teilbaren natürlichen Zahlen (als Zeichenkette in Dezimalschreibweise) akzeptiert - in zwei Varianten: Nullen als echte Präfix sind zulässig/ nicht zulässig. (im ersten Fall sind also z.B. 0, 00, 14, 0014, 888, 0888 $\in L(A)$; im zweiten Fall sind 00, 0014, 0888 $\notin L(A)$).
- (b) Entwerfen Sie einen entsprechenden NEA für durch 4 teilbare Zahlen (Hinweis: Die letzten beiden Ziffern müssen eine durch 4 teilbare Zahl ergeben).
- (c) Entwerfen Sie einen entsprechenden NEA für durch 3 teilbare Zahlen (Hinweis: Teilbarkeit durch 3 ist aus der Quersumme ablesbar!).

Der Einfachheit halber können Sie sich bei (b) und (c) auf die Variante beschränken, in der Nullen als Präfixe zugelassen sind.

Aufgabe 2.5 Formale Schreibweise von Automaten

Gegeben sei der Automat $A = \langle K, \Sigma, \Delta, s, F \rangle$ mit

$$K = \{1, 2, 3, 4\}$$

$$\Sigma = \{a, h, !\}$$

$$s = 1$$

$$F = \{1, 4\}$$

$$\Delta = \{\langle 1, h, 2 \rangle, \langle 2, a, 3 \rangle, \langle 3, h, 2 \rangle, \langle 3, !, 4 \rangle\}$$

- (a) Beschreiben Sie informell die Sprache, die der Automat akzeptiert. Wie lang ist das kürzeste, wie lang das längste akzeptierte Wort?
- (b) Geben Sie drei verschiedene Möglichkeiten an, A so zu verändern, dass auch Wörter erkannt werden, die kein Ausrufezeichen am Ende haben. Zulässige Veränderungen dabei sind: Das Einfügen einer neuen Transition in Δ , sowie das Hinzufügen eines Zustandes zu F .

- (c) Ändern sie A so, dass Wörter mit beliebig vielen a's (z.B. haahaaa!) erkannt werden. Ändern Sie dabei nur Δ und lassen Sie den Automaten ansonsten unverändert.

In b) und c) sollen Sie keine Zustandsdiagramme zeichnen, sondern die neuen Automaten formal angeben.

Abgabe in Gruppen von bis zu drei Studierenden bis **15.11.2011** 10 Uhr entweder als Email im pdf-Format an **i2cl@coli.uni-sb.de** oder auf Papier im Briefkasten an der Tür von Raum 1.04 in C7.2.