

# Schlussprüfung Finite-State-Methoden in der Sprachtechnologie FS 13

Aufgabenstellung: Simon Clematide

Prüfung vom 3. Juni 2013  
Institut für Computerlinguistik  
Universität Zürich

Vorname \_\_\_\_\_ Matrikelnummer \_\_\_\_\_

Nachname \_\_\_\_\_

## Für Studierende der folgenden Studiengänge:

- ☐ BA - Studiengang Computerlinguistik (Phil. Fakultät)
- ☐ BA - Studiengang Computerlinguistik und Sprachtechnologie (Phil. Fakultät)
- ☐ BA-Studierende (Wirtschaftswissenschaftliche Fakultät)
- ☐ Studierende des Nebenfachs Informatik mit Studienbeginn ab WS 04/05
- ☐ Andere:

## Nur für Lizentiatsstudierende der Computerlinguistik als ein Fach aus der Phil. Fakultät:

Strasse: \_\_\_\_\_ Hauptfach: \_\_\_\_\_

PLZ/Ort: \_\_\_\_\_ E-Mail: \_\_\_\_\_

Aufgabe Nr.:	1	2	3	4	5	6	7	8	Summe
Punktzahl:	5	4	5	5	8	11	10	12	60
Davon erreicht:									

Note SU: \_\_\_\_\_ Note SP: \_\_\_\_\_

Endnote: \_\_\_\_\_ Bestanden: ☐ Ja ☐ Nein

Bitte auf jedes separate, d.h. nicht angeheftete Blatt mit Lösungen den Nachnamen schreiben.

**Viel Erfolg!**

## Wichtige Hinweise

Punkte-Maximum: 60 (pro Minute 1 Punkt)

Hinweis: Bitte schreiben Sie in einem überlegten und knappen, aber verbalen Stil (keine Stichwortsammlungen). Bei inhaltlichen Auswahlendungen, wo einfach mal alles spontan hingeschrieben wird und Falsches wie Korrektes munter vermischt sind, behalte ich mir Abzüge vor. Das Zeitbudget ist so berechnet, dass man vor dem Schreiben kurz überlegen darf.

- 5 1. **Lemmatisierung (5 Punkte)** Wieso macht man überhaupt **Lemmatisierung** mit morphologischer Analyse in der NLP? Geben Sie 1 gutes und motivierendes Beispiel für die **Vorteile**. Für welche **Typen von Sprachen** ist sie sinnvoll (Stichworte)? Welche **Alternativen zur Lemmatisierung** gibt es (Stichworte)?

.....

.....

.....

.....

.....

.....

.....

.....

- 4 2. **Morphologische Prozesse (4 Punkte)** Nennen Sie 2 **unterschiedliche morphologische Prozesse** und geben Sie je 1 **gutes Beispiel** dafür.

.....

.....

.....

.....

.....

.....

- 5 3. **lexc (5 Punkte)** Was sind die **wesentlichen** Eigenschaften des LEXC-Formalismus? Worin liegen seine sprachtechnologischen **Vorteile**? Inwiefern unterstützt der lexc-Formalismus auch den **Umgang mit Suppletiv-Formen**?

.....

.....

.....

.....

.....

.....

.....

.....

**5** 4. (A) Frage zu gewichteten Endliche Automaten oder (B) Alternativfrage (5 Punkte)

(A) Was sind die wesentlichen **Unterschiede** zwischen normalen EA und gewichteten EA? Geben Sie eine mögliche **Anwendung** von gewichteten Automaten in der NLP.

(B) Alternativfrage: Was versteht man unter morphologischer **Über- bzw. Unteranalyse**? Geben Sie realistische Beispiele! (Beantworten Sie entweder A oder B!)

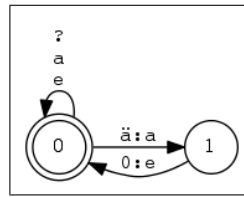
This image shows a full page of white paper with horizontal dotted lines, typical of primary school writing paper. The lines are evenly spaced and run across the width of the page. There are no margins, text, or other markings on the paper.

**8 5. Zwei-Ebenen-Regeln vs. Ersetzen und Komponieren (8 Punkte)** Geben Sie die wesentlichen Eigenheiten, Unterschiede und Gemeinsamkeiten von beiden Ansätzen zur Bildung von Morphologiesystemen an. Geben Sie auch eine schematische Zeichnung.

[illegible]

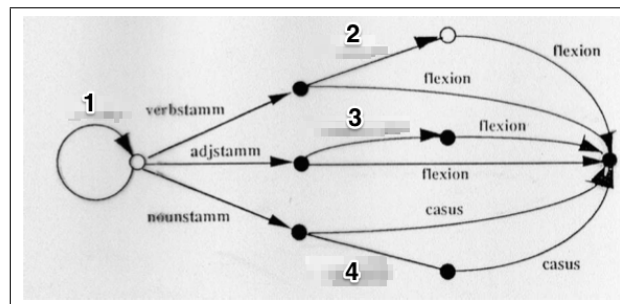
## 6. Reguläre Ausdrücke und Übergangsdiagramme (11 Punkte)

- 3 (a) Schreiben Sie einen **regulären Ausdruck**, der in XFST den rechts abgebildeten Automaten ergibt:



- 8 (b) Zeichnen Sie das **Zustandsdiagramm** eines möglichst kleinen Automaten, der die Sprache eines in XFST mittels `read regex ?* - {ei};` definierten regulären Ausdrucks akzeptiert. Geben Sie auch das **Sigma** Ihres Automaten an. Hinweis: Bitte benutzen Sie "?" als Kantenbeschriftung für das UNKNOWN-Symbol wie im Automaten von Teilaufgabe (a).

## 7. Modell für Deutsch (10 Punkte)



- 1 (a) Wie heisst der Fachbegriff, den obige Abbildung exemplifiziert? \_\_\_\_\_
- 8 (b) Ergänzen Sie die 4 **fehlenden nummerierten Beschriftungen** mit der für die deutsche Sprache passende morphologische Kategorie. Geben Sie je 1 **Beispiel**, welche Morphe durch die 4 nummerierten angesprochen werden und **markieren Sie die Morphe** im Wort.

.....

.....

.....

.....

.....

.....

- 1 (c) Geben Sie 1 **Beispiel**, wo dieses Modell nicht passt.

.....

### 8. Ein Problem (12 Punkte)

Das "Translation Bureau" der kanadischen Regierung hat folgende Regel publiziert auf ihrer Homepage:

*Double the final consonant before y or before a suffix beginning with a vowel in a word of one syllable that ends in a single consonant preceded by a single vowel:*

*bed* → *bedded*, *dip* → *dipped*, *fat* → *fatty*, *fit* → *fitted*, *flit* → *flitting*, *gum* → *gummy*, *log* → *logged*, *mad* → *madden*,  
*rot* → *rotted*

## Exceptions

*Do not double the final consonant in a word of one syllable if the vowel sound is long:*

*boat* → *boating*, *light* → *lighten*, *stoop* → *stooped*, *read* → *reading*

3

(a) Wie lautet der Fachbegriff für solche Phänomene? Worum geht es dabei?

.....

.....

.....

9

(b) Ihr Praktikant Peter sollte ursprünglich einen Transduktor bauen, der auf der Unterseite die orthographisch korrekt derivierten Wortformen und auf der Oberseite ihre morphologische Struktur enthält, welche durch obige Suffigierungen entstehen können (z.B. {bed+ed} : {bedded}). Er hat aus einem Aussprachewörterbuch alle einsilbigen Wörter extrahiert und jeweils vor dem Wort markiert, ob der Vokal kurz ("=") oder lang "===" ist. Leider endete Peters Praktikum, bevor er fertig implementiert hatte. Programmieren Sie seine Aufgabe fertig und verwenden Sie wie im Beispiel das Zeichen + als Trennsymbol. Hinweis: Übergenerierung (z.B. "ready") ist erlaubt, da ja keine Part-Of-Speech-Information vorhanden ist.

```
define OneSyllable [ "="[{bed}|{dip}|{fat}|{fit}|{gum}|{log}|{mad}]
                    |["=" [{boat}|{light}|{stoop}|{read}]]];
```

This image shows a full page of white paper with horizontal dotted lines. The lines are evenly spaced and run across the width of the page, providing a guide for handwriting practice. There are no margins, text, or other markings on the page.