

Entwicklung einer Computermorphologie

In dieser Übung sollen Sie mit Hilfe der SFST-Werkzeuge (<http://www.cis.lmu.de/~schmid/tools/SFST>) eine einfache Morphologie für eine morphologisch interessante Sprache (also nicht Englisch, Chinesisch) implementieren. Die morphologische Analyse eines Wortes soll die Grundform und die morphologischen Merkmale liefern.

Schritte

- Finden Sie heraus, welche **Flexionsklassen** in der Sprache existieren. Hier sind Wörterbücher und Webseiten wie Wiktionary hilfreich.
- Wählen Sie eine Teilmenge der Flexionsklassen aus.
- Implementieren Sie die Flexionsklassen und die erforderlichen **morphologischen Regeln**. (Mindestens eine Regel sollte dabei sein.)

Empfohlene Schritte:

- Erstellen Sie eine Datei mit **Stammlexikon**-Einträgen für jede Flexionsklasse. Sie können hier auch Stämme (glaub) auf Grundformen (glauben) abbilden durch einen Eintrag der Form: `glaube:<n:<>`
Beachten Sie dabei, dass in einer Lexikodatei nur die Symbole “:”, “<”, “>” und “\” als Operatoren interpretiert werden. Sie können also nicht schreiben `glaub{en}:{}`
 - Erstellen Sie für jede Flexionsklasse einen **Flexions-Transducer**, der morphologische Merkmale auf Flexionsendungen abbildet, bspw. `{<pres><1><sg>}:{e} | {<pres><2><sg>}:{st}|...`
 - Lesen Sie jedes Stammlexikon in einen Transducer ein und **konkatenieren** Sie diesen Transducer mit dem entsprechenden Flexions-Transducer.
 - **Vereinigen** Sie die Transducer für die verschiedenen Flexionsklassen mit dem Disjunktions-Operator.
 - Schreiben Sie die notwendigen **morphophonologischen Regeln**, um die korrekten Wortformen zu erhalten (bspw. für `happy+er` → `happier`), und wenden Sie sie an.
- Kommentieren Sie Ihren Code. Schreiben Sie zu jeder Ersetzungsregel ein Beispiel für die gewünschte Ersetzung als Kommentar dazu.
 - Dokumentieren Sie, welche Phänomene Sie erfasst haben, und an welcher Quelle (Buch, Webseite) Sie sich orientiert haben.

Schicken Sie Ihr SFST-Programm und Ihre Dokumentation an schmid@cis.lmu.de.

Tipp

Manchmal sind Triggersymbole nützlich: Deutsche Nomen wie “Haus” oder “Buch” bilden den Plural mit Umlautung. Hier kann man zu der Endung -er ein Triggersymbol hinzufügen: {<pl>}:{er} und dann später (nach der Konkatination von Stämmen und Endungen) eine morphophonologische Regel anwenden, die den Stamm nur dann umlautet, wenn das Triggersymbol nachfolgt. Gleichzeitig oder anschließend wird das Triggersymbol gelöscht. Das folgende Beispiel benutzt temporär das Symbol , um den Wortanfang zu markieren und löscht es am Ende wieder.

```
ALPHABET = [A-Za-zäöüÄÖÜß] <UL>:<> <B>:<>
$cons$ = [bcd fghjklmnpqrstvwxyz]
$Umlaut$ = ([AOUaou]:[ÄÖÜäöü]|{aa}:ä|{oo}:ö|{au}:{äu}) ^-> \
            (($cons$|<B>) __ $cons$* (e[r1])? <UL>)
$Morph$ = <>:<B> $Morph$ || $Umlaut$
```