



## Digitale Typografie (SS 2016)

Abgabe bis zum 8. Juli 2016, 10:00 Uhr

### Lernziele:

- Entwicklung eines Trennungsalgorithmus mit Gewichten

### Aufgabe 11: Trennen mit Gewichten

Grundsätzlich kann mit Gewichten getrennt werden, indem die entsprechenden Sollbruchstellen mit unterschiedlichen Strafwerten versehen werden. Dabei stellt sich die Frage, auf welcher Grundlage die Strafwerte zu bestimmen sind.

Da grundsätzlich Trennstellen zu bevorzugen sind, die weiter innen in einem Wort liegen, könnte dies ein recht einfaches Kriterium sein.

Alternativ bietet sich die Auswertung von Trennungslisten an, die auf vollständigen Wörtern beruhen, wie beispielsweise die von Werner Lemberg gepflegte Liste für die deutsche Sprache. Hier ist ein Beispiel für einen typischen durch Semikolons getrennten Eintrag:

```
Altersteilzeitregelung;Al-ters==teil=zeit===re-ge-lung
```

Ein Bindestrich deutet eine hohe Bindung an, „=" weist auf eine Bruchstelle zwischen Wortteilen mit niedriger Bindung an, die bei Trennungen eher zu bevorzugen sind. Wenn es mehrere solcher Bruchstellen gibt, werden diese ggf. noch hierarchisiert, d.h. je länger das Trennsymbol ist, umso eher ist dort zu trennen. Die vollständige Dokumentation zu der Trennliste findet sich in der Datei *dokumente/README.wortliste*. Die Wortliste bietet auch unterschiedliche Trennungen an, je nach alter, neuer oder schweizerischer Rechtschreibung. Hier ist ein Beispiel dazu:

```
abhängstigen;-2-;ab<äng-sti-gen;ab<ängs-ti-gen
```

Die „-2-“ im zweiten Feld bedeutet hier, dass das Feld unbenutzt ist. In so einem Fall findet sich im dritten Feld die Trennung nach alter Rechtschreibung und im vierten Feld die Trennung nach der neuen. Allerdings gibt es auch Regeln, bei denen im dritten Feld „-3-“ steht, dann ist dies ebenfalls leer.

Entwickeln Sie einen Parser für die Lemberg'sche Wortliste, die eine entsprechende Tabelle füllt, die dann die Schnittstelle des *WeightedHyphenators* unterstützt. Analog zu *SequenceHyphenator* sollte es dann eine Klasse *SequenceWeightedHyphenator* geben, die gewichtete Trennungen unterstützt und entsprechende mit Strafwerten versehende Sollbruchstellen erzeugt.

Experimentieren Sie mit der Bestimmung der Strafwerte, bis Sie mit den Ergebnissen zufrieden sind. Dies ist nicht einfach, da die Bevorzugung einzelner Trennstellen ggf. gegen Zeilen mit zu großen oder zu kleinen Leerbereichen zwischen den Wörtern entgegengerechnet werden muss. Selbst bei längeren Paragraphen gibt es nicht ohne weiteres Unterschiede, ggf. muss mit unterschiedlichen Breiten experimentiert werden, bis Fälle gefunden werden, die tatsächlich unterschiedlich getrennt werden. Da ein *WeightedHyphenator* eine Erweiterung von *Hyphenator* ist, können Sie immer auch vergleichenderweise den *SequenceHyphenator* verwenden.

**Hinweis:** Die Wortliste können Sie mit dem Befehl

```
git clone git://repo.or.cz/wortliste.git
```

herunterladen. Dann liegen alle zugehörigen Dateien im Verzeichnis *wortliste*.

Wenn Sie mit Ihrer Lösung fertig sind, sollten Sie alle notwendigen Dateien mitsamt den Quellen in ein ausführbares Java-Archiv *hyphen.jar* packen und zusammen mit einer beispielhaften Ausgabe Ihres Programms mit dem Kommando

```
submit typo 11 hyphen.jar paragraph.ps
```

auf der Thales einreichen.

Dies ist das letzte Übungsblatt im Rahmen dieser Vorlesung.

**Viel Erfolg!**