

Übung zur Vorlesung "Computerlinguistik II / Sprachtechnologie"

Sommersemester 2010, Prof. Dr. Udo Hahn, Erik Fäßler

Übungsblatt 8 vom 23.06.2010

Abgabe bis 28.06.2010, 23.59 Uhr; per Email (Standard Dateiformat: ps, pdf, oder doc) an erik.faessler@uni-jena.de

Aufgabe 1 : Annotationstools: MMAX2

In der Vorlesung wurden unterschiedliche Werkzeuge für die manuelle Annotation von Texten vorgestellt (Alembic Workbench, Wordfreak und MMAX2). MMAX2 ist inzwischen ein OpenSource Projekt¹ und wird noch immer weiter entwickelt.

In dieser Übungsserie sollen Sie sich mit MMAX2 vertraut machen und einen kurzen Beispieltext annotieren. Laden Sie dazu die neueste Version von MMAX2 von der MMAX2-Homepage herunter. Die Installation geschieht durch einfaches Entpacken des heruntergeladenen Archivs.

Die Dokumentation des Programms finden sie im Unterordner "Docs" der entpackten Dateien. Lesen Sie zunächst die dort befindliche Datei "readmefirst.txt". Anschließend sollten Sie sich mit dem Quickstart ("mmax2quickstart.pdf") beschäftigen, um mit dem Programm vertraut zu werden.

Hinweis: Es scheint nicht möglich zu sein, das Programm ohne ein konkretes Projekt zu starten. Es muss stets ein bestehendes Projekt als Kommandozeilenargument übergeben werden. Der Aufruf könnte etwa so aussehen:

```
mmax2 Samples\HTC\002_htc_abn.mmax (Windows)
./mmax2.sh Samples/HTC/002_htc_abn.mmax (Linux, Unix, MacOS)
```

Legen Sie ein eigenes Projekt mit Hilfe des Project Wizards an. Starten Sie MMAX2, indem Sie eines der Beispielprojekte öffnen. Klicken Sie im Hauptfenster dann auf Tools, Project Wizard. Verwenden sie als Eingabe die Beispieldatei, die Sie zusammen mit dem Übungsblatt von der JULIE Homepage herunter laden können. Als Encoding wird UTF-8 verwendet. Eine Beschreibung des Project Wizard finden sie im QuickStart unter Abschnitt 6.2.

Hinweis: Bitte verwenden Sie für alle Projektdateien (ganz unten im Project Wizard) den Button "Use for all", so dass alle Projektdateien im gleichen Verzeichnis abgelegt werden. Legen Sie zuvor für das Projekt ein eigenes, leeres Verzeichnis an.

Definieren Sie Annotationslevels, um folgende Annotationen anlegen zu können:

- Satzannotationen; sorgen Sie zudem dafür, das MMAX hinter jedem Satz einen Zeilenumbruch im Annotationsfenster einfügt.
- MUC-Named-Entities²,
- die Relation, miteinander verwandt zu sein,
- die Relation, das eine Person bei einem Arbeitgeber oder Ausbildungsinstitut beschäftigt ist.

Die Details der Annotationslevels können (noch) nicht über der Project Wizard oder eine andere grafische Oberfläche definiert werden. Ziehen Sie die Dokumentationsdatei "annoschemes.pdf" zu Rate, um die Aufgabe zu erfüllen.

¹<http://mmax2.sourceforge.net/>

²Sie finden eine Beschreibung der originalen MUC-NEs unter http://www-nlpir.nist.gov/related_projects/muc/proceedings/ne_task.html, Abschnitte 2.2 bis 2.4.

Annotieren Sie den Beispielttext gemäß der definierten Annotationslevels.

Beschreiben Sie kurz:

- In welcher Form speichert MMAX2 den Eingabetext und wie wird er in MMAX2 genannt?
- In welcher Form speichert MMAX2 die eingegebenen Annotationen?

Packen Sie anschließend das fertige Projekt zusammen mit Ihren Antworten in ein Archiv (z.B. ZIP) und lassen Sie es Ihrem Übungsleiter zukommen.

Viel Spaß beim Annotieren! :-)