

1. Übungsblatt - Abgabe: 4.11.2014

Aufgabe 1.1 Maschinelle Übersetzung

- (a) Auf den Webseiten der Computerlinguistik findet sich unter <http://www.coli.uni-saarland.de/msc/page.php?id=department> ein Absatz mit der Überschrift “Computational Linguistics Department in Saarbrücken”. Kopieren Sie diesen Absatz (bis zu “... multimodal interfaces”) in das Textfenster des google translators (unter <http://translate.google.com/>), und lassen Sie ihn ins Deutsche übersetzen. Wie würden Sie Qualität und Stil der Übersetzung generell kommentieren? (1-2 Sätze)
- (b) Korrigieren Sie den übersetzten Text: Markieren Sie fehlerhafte oder stilistisch schlechte Übersetzungen, geben Sie an, welche Art von Fehler gemacht wurde (ähnlich, wie Sie das aus der Schule noch kennen: „Ausdruck“, „Satzbau“ etc.), und geben Sie bessere Übersetzungsalternativen an (für die beanstandeten Stellen oder für ganze Sätze).
- (c) Probieren Sie das System auch mit ein paar (mindestens 2) englischen Texten von anderen Textsorten aus: News, Gedicht, persönliche Email. Wie ist das Ergebnis hier im Vergleich? Geben Sie Originaltext und Übersetzung an begründen Sie Ihre Einschätzung mit Beispielen.

Aufgabe 1.2 Dialogsysteme

- (a) Testen Sie den „sprechenden Fahrstuhl“ in Gebäude C 7.4! Sie können Ihr Fahrtziel auf unterschiedliche Weise sowie in Deutsch und Englisch spezifizieren. Probieren Sie eine Reihe von Möglichkeiten aus! Versuchen Sie auch, dem Fahrstuhl (plausible) Aufgaben zu geben, die er nicht erfüllen kann! Berichten Sie und geben Sie insbesondere ein Beispiel für einen geglückten Dialog an und ein Beispiel, in dem der Aufzug Ihre Aufgabe nicht erfüllen konnte.
- (b) Versuchen Sie, möglichst kurz zu beschreiben, wie der Fahrstuhl Ihrer Meinung nach funktioniert: Was versteht er? Was weiß er? Was kann er?
- (c) Versuchen Sie, das Muster, das der Dialogführung des Fahrstuhls zu Grunde liegt, möglichst übersichtlich aufzuschreiben (nicht unbedingt vollständig): Wie reagiert der Fahrstuhl in bestimmten Situationen auf bestimmte Benutzeräußerungen? Wie sieht dementsprechend der „Dialogplan“ des Fahrstuhls aus? Sie können ein Diagramm angeben, in dem Sie mögliche Äußerungen von Fahrstuhl und Benutzer in Abhängigkeit voneinander angeben.

Aufgabe 1.3 Informationssuche

- (a) Benutzen Sie Google (www.google.com), um die Antwort auf die Frage „Wer war US-Präsident als Barack Obama geboren wurde?“ bzw. “Who was American president when Barack Obama was born?” zu erhalten. Beschreiben Sie, wie Sie vorgegangen sind, und die Schwierigkeiten, die man hat, wenn man die Frage durch eine einzige Google-Anfrage beantworten möchte.
- (b) Finden Sie eine weitere Frage, deren Antwort aus möglichst anderen Gründen nur schwer oder gar nicht mit Google zu finden ist (also möglichst nicht nur, weil man mehrere Suchanfragen miteinander kombinieren muss), deren Antwort aber grundsätzlich im Internet vorkommt. Erläutern Sie, warum diese Fragen schwierig ist.

Aufgabe 1.4. Ambiguität

Lesen Sie aus Jurafsky/Martin (Neu-Ausgabe von 2009) Kapitel 1, S. 35-43, Abschnitte 1.1, 1.2, 1.4 und 1.5. Die in 1.3 verwendeten Begriffe sind nicht ohne weiteres verständlich, sie werden im Lauf der Vorlesung eingeführt; 1.3 kann übersprungen werden. Der Text steht in Kopie in der Institutsbibliothek (Vorlesungsordner).

- (a) Versuchen Sie, für die 5 Lesarten des Satzes “I made her duck” aus 1.2 möglichst eindeutige und gut verständliche deutsche Formulierungen zu finden.
- (b) Versuchen Sie, einen deutschen Satz mit maximal 4 Wörtern zu finden, der möglichst viele, aber mindestens 2 Lesarten besitzt. Paraphrasieren Sie die einzelnen Lesarten und erklären Sie, wodurch die Ambiguität zustande kommt.

Aufgabe 1.5. Unix-Tools

Loggen Sie sich in Ihren Coli-Account ein. Speichern Sie die auf der Vorlesungs-Homepage unter diesem Übungsblatt verlinkte Datei *Beispieltext.txt* in Ihr Homeverzeichnis und bearbeiten Sie dann folgende Aufgabe. Geben Sie immer auch an, welchen Befehl mit welchen Argumenten Sie benutzt haben (also z.B. `cp Datei1 Datei2`).

Benutzen Sie den Befehl `wc` um angezeigt zu bekommen, wie viele Wörter die Datei enthält. Schauen Sie sich die Datei (z.B. mit `less`, `cat` oder im `emacs` an) und zählen Sie die Wörter. Stimmen die beiden Werte überein? Wenn nicht, äußern Sie eine Vermutung wieso.

Hinweis: Sie können sich mit dem Befehl `man` zusätzliche Informationen zu einem Kommando anzeigen lassen, z.B. mit `man wc`.

Abgabe in Gruppen von bis zu drei Studierenden am **4.11.2014** vor der Vorlesung.