

Einführung in die Computerlinguistik

Übung 2

Jan Hendrik Dithmar - Matrikelnummer 2031259

Aufgabe 2.1

(b)

Das Dialogverhalten finde ich „kundenfreundlicher“. Das System versteht nach meiner Erfahrung ganze Sätze wesentlich besser als das Bundesbahnsystem. Allerdings geht das System - anders als das System der Deutschen Bahn - davon aus, dass seine zweite Annahme richtig ist. So sieht z.B. ein Dialog mit dem System aus:

Sys: Von wo nach wo möchten Sie fahren?

Ich: Ich möchte von Würzburg an der Saar nach Saarbrücken fahren.

Sys: Von Andernach, richtig?

Ich: Nein.

Sys: Wann möchten Sie **von Würzburg Hauptbahnhof** nach Saarbrücken Hauptbahnhof fahren?

Dies kann zum einen natürlich von Vorteil sein, da es nicht wie das System der Bundesbahn zuerst alle 200 Möglichkeiten vorspricht und jedes Mal fragt, ob der genannte Bahnhof der richtige ist. Ich vermute, dass dieses Verhalten auch damit zusammenhängt, dass das Philips-System nur einen begrenzten Datensatz zur Verfügung hat, worauf es am Anfang auch hinweist („... Sie können Zugverbindungen zwischen 1000 Bahnhöfe in Deutschland erfragen ...“). Versteht es nach mehrmaligem Nachfragen immer noch nicht, wird man auf das System der Deutschen Bahn verwiesen („Bitte rufen Sie die normale Fahrplanauskunft an.“).

Des weiteren kann es bei der Sprachausgabe nicht mit Jahreszahlen umgehen. Gibt man als Datum z.B. den 02.01.2007 an, so sagt das System, dass der aktuelle Fahrplan am 2. Januar <beep> nicht mehr gültig ist, sondern nur bis zum 9. Dezember <beep>. Es ist auch nicht zwingend erforderlich, dass man das Jahr nennt. Alle Daten, die in der Zukunft liegen werden automatisch mit der nächstmöglichen Jahreszahl vorgegeben. Nenne ich nun z.B. den 2. Januar, so geht das System vom Jahr 2007 aus, da 2008 wohl zu weit in der Zukunft liegt und 2006 als Option nicht mehr zur Verfügung stehen kann.

(c)

Wir haben in der Übung erfahren, dass das System der Deutschen Bahn nicht als endlicher Automat realisiert ist, sondern über eine Liste mit Fragen, welche eine gewisse Priorität haben. Zum einen könnte ich mir vorstellen, dass das Philips-System ebenfalls so arbeitet, jedoch könnte es auch als endlicher Automat modelliert sein.

Die Antworten sind anders gelöst als beim System der Bundesbahn. Sie wirken teilweise wie schlecht geschnittene Audio-Samples, die „ohne Rücksicht auf Verluste“ aneinander gehängt werden. Ist z.B. eine Abfahrtszeit um 23 Uhr, so beginnt das System, wie wenn es 23.27 Uhr sagen wollte, stoppt aber umgehend nach der 23. Ein weiterer Unterschied wurde von mir bereits in Aufgabenteil (b) etwas überspitzt formuliert dargestellt: das System geht nach einer Korrektur des Abfahrtsbahnhofes davon aus, dass die zweite Möglichkeit, die es glaubt verstanden zu haben, die richtige ist und nimmt diese für den weiteren Verlauf an. Das System der Deutschen Bahn hingegen stellt eine Liste mit mehreren Bahnhöfen zur Verfügung und arbeitet diese Bahnhof für Bahnhof ab.

Aufgabe 2.2

(a)

Dieser Satz enthält mehrere Hürden für ein TTS-System. Der orthografische Punkt muss beim ersten und zweiten Auftreten als Nummerierung, beim dritten als Teil einer Abkürzung, beim vierten als Tausender-Trennzeichen in der Geldsumme und beim fünften Auftreten als Satzende erkannt werden. Zum anderen muss das System erkennen, dass es sich bei „Shooting-Star“ um einen englischen Begriff, bei „Tour de France“ um einen französischen Begriff handelt und diese idealerweise auch so auszusprechen sind. Eine weitere Hürde ist die Zeitangabe. Diese muss als Zeit im Format **Stunden:Minuten:Sekunden** gesehen werden und nicht etwa als 48 Uhr 34 Minuten 12 Sekunden. Auch das Euro-Zeichen sollte (ähnlich wie z.B. ein Prozent-Zeichen) richtig erkannt und als Euro ausgesprochen werden. Eine letzte Hürde, die ich entdeckt habe, ist die Abkürzung „USA“, welche Buchstabe für Buchstabe ausgesprochen werden sollte. Als kleinen Stolperstein könnte man auch den Namen „Frans van der Waal“ sehen, da dies ein ausländischer Name ist und dieser auch so ausgesprochen werden sollte. Da allerdings das Niederländische in der Aussprache dem Deutschen teilweise sehr ähnlich ist, stellt dies meiner Meinung nach keine große Hürde für ein (deutsches) TTS-System dar.

(b)

- Transporterfahrer:

Tran [Nom-Stamm] + *sport* [Nom-Stamm] + *erfahr* [Vb-Stamm] + *er* [Pron]
Tran [Nom-Stamm] + *sport* [Nom-Stamm] + *er* [Pron] + *fahr* [Vb-Stamm] + *er* [Pron]
Transport [Nom-Stamm] + *erfahr* [Vb-Stamm] + *er* [Pron]
Transport [Nom-Stamm] + *er* [Pron] + *fahr* [Vb-Stamm] + *er* [Pron]
Transporter [Nom-Stamm] + *fahr* [Vb-Stamm] + *er* [Pron]
Transporter [Nom-Stamm] + *fahrer* [Nom-Stamm]

- Alleinerziehender:

All [Nom-Stamm] + *ein* [Art] + *er* [Pron] + *ziehen* [Vb-Stamm] + *der* [Art]
All [Nom-Stamm] + *ein* [Art] + *erziehen* [Vb-Stamm] + *der* [Art]
All [Nom-Stamm] + *ein* [Art] + *erziehender* [Nom-Stamm]
Allein [Adj-Stamm] + *er* [Pron] + *ziehen* [Vb-Stamm] + *der* [Art]
Allein [Adj-Stamm] + *erziehen* [Vb-Stamm] + *der* [Art]
Allein [Adj-Stamm] + *erziehender* [Nom-Stamm]

(c)

1. Als das neue Gebäude fertig war, kamen die Menschen, ihre Begeisterung auszudrücken.
 - Welches Verb gehört zu welchem Teilsatz?
 - Welche Teilsätze gehören zusammen?
 - Ist der letzte Teilsatz ein vollständiger Satz?
 - Wo kann man eine Pause setzen, damit es nicht unnatürlich klingt? Eine Pause nach dem ersten Komma? Oder nach dem zweiten? Oder nach beiden?
2. Dürfte ich um etwas mehr Ruhe bitten?
 - Als Frage formuliert. Streng genommen muss die Stimme am Ende angehoben werden.
 - Eigentlich eine Aussage: Bitte etwas mehr Ruhe! Die Stimme sollte am Ende nicht wie bei einer Frage gehoben werden.

(d)

CTS bietet dadurch, dass die sprachgenerierende Komponente des CTS-Systems „weiß“, was gesagt werden soll und wie es gesagt werden soll, eine wesentlich höhere Natürlichkeit der Sprachausgabe, was ein TTS-System auf Grund der Bereitstellung größtmöglicher Flexibilität nicht bieten kann. Es wird eine linguistische Repräsentation des Satzes generiert, welche eine Verbesserung der Qualität der synthetischen Sprache bietet, da z.B. die Intonation dadurch verbessert werden kann. CTS-Systeme können jedoch keinen orthografischen Text in angemessener Form wiedergeben, da dieser eine stark verarmte Repräsentation der Sprache darstellt und somit die nötigen Informationen (z.B. Intonation) fehlen, um den Text angenehmer als ein TTS-System darzustellen. Das bedeutet, dass eine aus einem Text generierte linguistische Struktur immer Defizite gegenüber der linguistischen Struktur aus einem CTS-System hat.

Aufgabe 2.3

(a)

Mary: Getestet wurde mit den Stimmen de7 und de6.

- France wurde als „Fränke“ ausgesprochen.
- Die Gesamtzeit wurde nicht als Zeit sondern als einzelne Zahlen mit kurzen Pausen dazwischen vorgelesen.
- Das Preisgeld wurde als „Elf Punkt Null Null Null Euro“ ausgesprochen.

AT&T: Getestet wurden beide deutschen Stimmen (männlich als auch weiblich).

- Die Gesamtzeit wurde als Uhrzeit dargestellt: 48 Uhr 34 Minuten und 12 Sekunden.
- Unmittelbar vor „11.000 €“ wurde eine etwas unnatürliche Pause eingelegt.

(b)

Leider konnte ich bei Mary nicht viel finden, was besser geklappt hätte oder was ausschließlich dem System in Bezug auf Können vorbehalten ist. Lediglich konnte Mary „ICE“ als Abkürzung erkennen und hat die Buchstaben einzeln ausgesprochen. Dieses Vorgehen hat das System von AT&T nicht verfolgt. Es wurde „ICE“ als ein Wort gelesen (irgendetwas in Richtung „Mieze“ nur ohne „M“). Probiert habe ich auch eMails. Ausgehend vom Klang der Systeme liegt das System von AT&T meiner Meinung nach vorne. Es klingt verständlich und relativ natürlich, obwohl das System vor Zahlen gerne eine etwas unnatürlich klingende Pause einlegt. Eine eMail, die ich als Eingabe benutzt habe, hatte keine Umlaute, sondern deren Schreibweise mit zwei Buchstaben (also ae anstatt ä). Mary konnte damit leider auch nichts anfangen, das System von AT&T schon. Eine ganz klare Schwäche von Mary (und gleichzeitig eine Stärke des anderen Systems) ist die Melodik im Sprechen. Das System von AT&T versucht die Stimme zu heben und zu senken, während Mary immer ganz monoton spricht.

Zum Schluss noch ein Beispiel. Ich habe u.a. folgende (fiktive) Aussage benutzt: „Ihre Verbindung lautet wie folgt. Abfahrt um 23:00 ab Saarbrücken Hauptbahnhof mit Regionalexpress nach Kaiserslautern Hauptbahnhof. Ankunft um 23:56. Dort weiter mit ICE, Abfahrt 00:12 in Richtung Mannheim. Ankunft Mannheim Hauptbahnhof um 01:13. Möchten Sie, dass ich die Verbindung für Sie wiederhole?“

Einzige Stolperfalle für das System von AT&T war der ICE. Mary hingegen hatte dies richtig gemacht, dafür aber folgende Fehler: Die Uhrzeiten wurden nicht richtig erkannt. So hat das System z.B. folgende Aussage getroffen: „... Abfahrt

um 23–00 ab Saarbrücken Hauptbahnhof mit Regionalexpress nach Kaiserslautern Hauptbahnhof. Ankunft um 23. Sechsfünfzigste dort weiter mit I–C–E, Abfahrt 0–0–12 ...“ (die – stellen Pausen dar).