

# Strukturerkennung von Bedingungen in gesprochener Sprache

Dokumentenart: Exposé für eine Bachelorarbeit  
Autor: Vanessa Steurer  
Matrikel-Nr.: 1703644  
Studiengang: Informatik Bachelor  
Betreuer: Sebastian Weigelt  
Datum: 8. Juli 2017

## 1 Motivation

Staubsauger-Roboter oder autonome Rasenmäher sind bereits in vielen Haushalten integriert. Insbesondere robotische Assistenzsysteme, welche ein größeres Spektrum an Aufgaben bewerkstelligen können, liegen im Fokus zahlreicher internationaler Forschungsprojekte. Laut dem Jahresbericht<sup>1</sup> der *World Robotics Conference* sollen bis spätestens 2018 über 8000 Serviceroboter auf dem Markt sein.

Dabei wird für eine benutzerfreundliche Interaktion zwischen Mensch und Maschine die Kommunikation über Spracheingabe realisiert. Um weiterhin auch in individuellen Situationen einsetzbar zu sein, soll die Programmierung von Robotern mit gesprochener Sprache ermöglicht werden.

Ein solcher Haushaltsroboter erhält seine Anweisungen in Form von mündlichen Befehlssträngen. Die hiesige Abschlussarbeit soll es ermöglichen, auch komplexere Befehlsfolgen in Form von bedingten Anweisungen behandeln zu können. Eine solche Anfrage könnte wie folgt aussehen:

*„If the dishes are dirty, then put them into the dishwasher.  
Else put them into the cupboard please.“*

Der Mensch erkennt hierbei sofort die sich in der Aussage befindliche logische Aufteilung der Handlungsstränge. Für eine Maschine ergeben sich jedoch aufgrund der Komplexität der natürlichen, gesprochenen Sprache viele Herausforderungen. Fehlende Satzzeichen, zahlreiche äquivalente Formulierungen oder Mehrdeutigkeiten erfordern Flexibilität.

---

<sup>1</sup><http://www.ifr.org/service-robots/statistics/>

## 2 Projekt PARSE

Das Projekt *PARSE* befasst sich damit, Zielsysteme, wie den am KIT entwickelten humanoiden Haushaltshelfer ARMAR-III [? ], um neue Anweisungen zu erweitern. Unbekannte Kommandos sollen dabei mit natürlichsprachlicher Programmierung, durch Extraktion von Arbeitsabläufen und Kontrollstrukturen, erlernt werden.

PARSE gliedert sich in mehrere Teilprojekte, deren Zusammenhang in Abbildung 1 illustriert ist. Zunächst müssen die englischen Sprachdateien von einem automatischen Spracherkenner (*Automatic Speech Recognition*) erfasst und in Text überführt werden. Dieser wird mit einer einfachen Sprachverarbeitung (*Shallow Natural Language Processing*) aufbereitet, welche die enthaltenen Wortarten sowie die Wortgruppen bestimmt. Im Rahmen eines weiteren Projekts werden die Eingabedaten auf eine logische Graphenrepräsentation abgebildet. Sie dient als gemeinsame Datenstruktur. Agenten für Sprachverständnis (*Natural Language Understanding*) sollen diesen Graphen anschließend durch einzelne, unabhängige Analyseschritte modifizieren. Auf diese Weise kann die Sprachaufnahme mit zusätzlichen Informationen angereichert werden. Im letzten Schritt muss aus den extrahierten Informationen für den Roboter ausführbaren Code generiert werden.

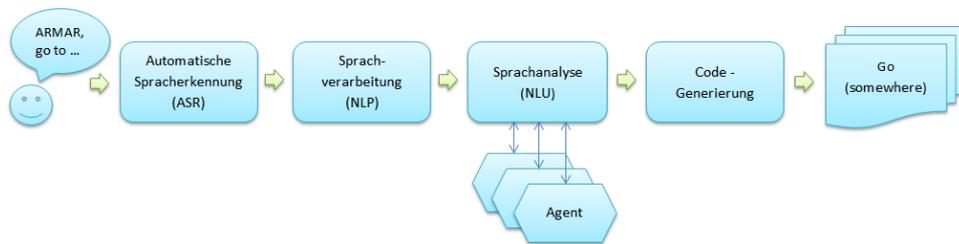


Abbildung 1: Schematische Darstellung von Projekt PARSE

## 3 Zielsetzung und Anforderungen

Das Ziel dieser Bachelorarbeit ist es, ein Werkzeug zu implementieren, welches bedingte Anweisungen erkennen kann. Die Spracheingabe des Nutzers wird dafür auf Kommandos untersucht, die auf bestimmte „Wenn-Dann“-Kontrollstrukturen abgebildet werden können. Anschließend müssen die gefundenen Bedingungen im Graphen abgespeichert werden.

Es wird ein Agent implementiert, der für die Sprachanalyse tätig ist. Da dieser Agent als Werkzeug in Projekt Parse (Abbildung 1) integriert wird, müssen dessen Entwicklungskriterien übernommen werden. Es ergeben sich folgende Anforderungen:

- Domänenunabhängigkeit: Das Werkzeug soll auch mit Eingabedaten

anderer Domänen (z.B. neue Einsatzumgebung, neuer Roboter) umgehen können.

- Modularität: Der Agent muss sich mit geringem Aufwand einzeln betrachten, debuggen und austauschen lassen. Das Ergebnis anderer Agenten darf nicht vorausgesetzt werden.
- Erweiterbarkeit: Es wird ermöglicht, zusätzliche Informationen (z.B. die Koreferenzauflösung) anderer Agenten einbinden zu können.

Als optionales Ziel könnte man sich nach der Erkennung von Bedingungen auch weitere, möglicherweise ableitbare, Kontrollstrukturen ansehen.

## 4 Vorgehensweise

Ansatz für die Strukturerkennung der Bedingungssätze bildet die Suche nach Schlüsselwörtern, wie „wenn“ (if), „dann“ (then) und „andernfalls“ (else). Dabei muss zunächst erörtert werden, welche Schlüsselwörter im Sprachgebrauch bei einer bedingten Anweisung verwendet werden. Diese Suche setzt eine Vorverarbeitung [?] der verbalen Anweisung voraus. Als Eingabe dient deshalb die textuelle Ausgabe des automatischen Spracherkenners, welche zusätzlich mit den Ergebnissen einer einfachen Sprachverarbeitung (*Shallow-NLP*) annotiert wurde. Dazu gehört die Erkennung der Wortarten (*Part-Of-Speech-Tagging*) sowie die Bestimmung von sinnabhängigen Worteinheiten (*Chunk Parsing*).

Für die Bedingungserkennung wird ein heuristischer Ansatz gewählt. Es wird das Wissen genutzt, dass ein Bedingungssatz stets durch die Schlüsselphrase „if“, oder dessen Synonyme, eingeleitet wird. Basierend auf einem Abgleich der Wortgruppen (*Chunks*) wird eine weitere Annahme getroffen: Der Bedingungssatz ist mindestens aus einem Nomen und einem darauf folgenden Verb aufgebaut, welche die Rolle von Subjekt und Prädikat einnehmen. Ein Beispiel soll dies verdeutlichen:

[SBAR <i>if</i> ] [NP <i>the door</i> ] [VP <i>is closed</i> ]	Chunk	Beschreibung
	SBAR	Konjunktion
	NP	Nominalphrase
	VB	Verbalphrase

Auf eine Abfrage folgt der Anweisungsblock, welcher unter bewahrheiteter Bedingung ausgeführt wird. Da die noch nicht implementierte Koreferenzauflösung die Eingrenzung von abhängigen Anweisungssträngen deutlich präzisieren würde, werden diese Informationen als gegeben angenommen. So kann man ansetzen, dass die bedingte Anweisung sich bis zum letzten Auftreten des Subjekts (hier: „the door“) erstreckt.

Einen anderen Ausgangspunkt für die Strukturerkennung von Bedingungssätzen bildet ein maschinelles Lernverfahren (*Maximum Entropy*), welches

grammatikalische Zusammenhänge aus einem korrekt annotierten Trainingskorpus gewinnt. Die Übereinstimmung der Ausgabe von beiden Ansätzen kann dabei genutzt werden, um eine erfolgte Strukturerkennung zu verifizieren.

Wird in der Eingabe eine bedingte Anweisung erkannt, muss diese im Graphen gespeichert werden. Eine geeignete Transformation des Graphen zur Informationsrepräsentation muss im Zuge dieser Bachelorarbeit erarbeitet werden.

In einer Befehlsfolge können vier unterschiedliche Befehlsmuster auftreten:

- Typ Bedingung, boolesche Abfrage
- Typ Anweisung, falls Bedingung bewahrheitet
- Typ Alternativ-Anweisung, falls Bedingung unzutreffend
- Typ Befehl, falls ein bedingungs-unabhängiges Kommando gestellt wird

Herausforderungen ergeben sich beispielsweise durch fehlerhaft erkannte Wörter oder Satzenden. Es sollte daher ein entsprechend robustes System entwickelt werden, das die Fehlerrate des Spracherkenners sowie eine falsche Punktuation angemessen kompensiert. Die wohl größte Hürde bildet jedoch die Vielfältigkeit der natürlichen Sprache, welche durch ihre zahlreichen, äquivalenten Formulierungen eines einzigen Satzes, die Mustererkennung erschwert.

## 4.1 Beispiel

Zur Verdeutlichung der Aufgabenstellung wird eine beispielhafte Anfrage an den Küchenroboter ARMAR-III betrachtet:

*„ARMAR please go to the table if the dishes are dirty then put them into the dishwasher else put them into the cupboard“*

Der zu entwickelnde Agent sucht in der Spracheingabe nach zutreffenden Befehlsmuster-Typen aus {**Bedingung**, **Anweisung**, **Alternativ-Anweisung**, **Befehl**}. Dabei fällt auf, dass die obige Befehlsfolge bereits mehrere Interpretationsmöglichkeiten zulässt, was insbesondere durch die fehlenden Satzzeichen verstärkt wird. Eine erfolgreiche Strukturerkennung könnte folgendes Ergebnis liefern:

*ARMAR, please go to the table **if the dishes are dirty then put them into the dishwasher else put them into the cupboard***

Abbildung 2 zeigt einen Ausschnitt einer möglichen Graphentransformation durch die Bedingungs-Erkennung:

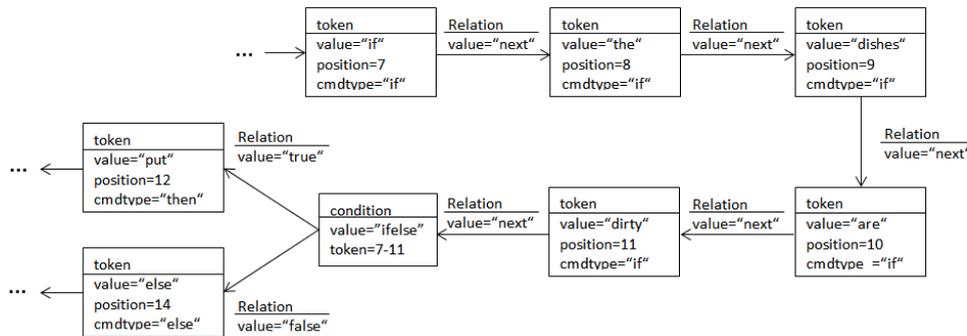


Abbildung 2: exemplarischer Ausschnitt der Graph-Datenstruktur

Eine anschließende Visualisierung als If-Else-Anweisung sähe wie folgt aus:

```

ARMAR, please go to the table
if the dishes are dirty then
  | put them into the dishwasher
else
  | put them into the cupboard
end
  
```

Da zu diesem Zeitpunkt zunächst eine Sprachverarbeitung und -analyse und noch keine Informations-Extraktion durchgeführt wurde, ist der Inhalt der verschiedenen Anweisungsblöcke als noch nicht interpretierte Zeichenketten zu verstehen.

## 5 Evaluation

In der Evaluation wird die Funktionalität des entworfenen Werkzeuges zur Erkennung von bedingten Anweisungen überprüft. Dabei kann der, von einer weiteren Bachelorarbeit entwickelte, Sprachkorpus [? ], welcher sowohl händische Transkriptionen als auch Ausgaben eines automatischen Spracherkenners enthält, als Evaluationsgrundlage verwendet werden. Zu jedem Eingabetext wird eine Musterlösung erstellt. Diese kann mit der von dem Werkzeug generierten Lösung verglichen werden.

Um eine angemessene Evaluation zu ermöglichen, muss mit hoher Wahrscheinlichkeit der Korpus um weitere eigens konzipierte Transkriptionen erweitert werden. Grund dafür ist, dass die vorhandenen Sprachdaten unzureichend viele und komplexe Bedingungssätze beinhalten.

Des Weiteren müssen der Einfluss von Fehlerraten auf die Ausgabe des Werkzeuges untersucht werden. Die Ergebnisse können dabei mit den Bewertungsmetriken Ausbeute (*Recall*) und Präzision (*Precision*) [? ] bewertet werden.