

# Synthese von Methodendeklarationen aus gesprochener Sprache

Dokumentenart: Exposé für eine Masterarbeit  
Autor: Vanessa Steurer  
Matrikel-Nr.: 1703644  
Studiengang: Informatik Master  
Betreuer: Sebastian Weigelt, Hans-Peter Zorn  
Datum: 22. Oktober 2018

## 1 Motivation

Intelligente Assistenten wie Amazon Echo, OK Google oder Siri werden immer häufiger in den Alltag integriert. Sie beantworten Fragen oder erledigen vermehrt auch kleine Aufgaben, wie die Erstellung eines Kalendereintrags im persönlichen Terminplaner. Ihre zunehmende Beliebtheit verdanken sie vor Allem der Einfachheit ihrer Steuerung durch natürlichsprachliche Anweisungen. Eine bestehende Herausforderung der Schnittstellenentwicklung ist dabei die Komplexität der natürlichen Sprache.

Die Problemstellung, das Verständnis beliebiger Ausdrucksweisen und komplexer Anweisungen zu ermöglichen, wird als Programmieren in natürlicher Sprache bezeichnet. Damit jene freien Formulierungen auch in ausführbare Aktionen intelligenter Systeme umgesetzt werden können, ist die Abbildung menschlicher Sprachkonstrukte auf die Funktionen des Zielsystems erforderlich. Dabei soll es auch möglich sein, nicht nur bereits einprogrammierte Kommandos abzurufen, sondern auch neue Funktionalitäten hinzuzufügen. Diese Funktionserweiterung soll wiederum in natürlicher Sprache und überdies ohne Kenntnis von Programmierkonzepten ermöglicht werden.

Mögliche Anwendungsgebiete würde dieses Konzept dabei nicht nur in der im Projekt PARSE betrachteten Haushaltsroboter-Domäne finden. Auch beliebte Sprachassistenten wie Alexa müssen bisher für das Hinzufügen von nutzereigenen Fähigkeiten über Quelltext konfiguriert werden.

## 2 Projekt PARSE

Das Projekt PARSE (Programming ARchitecture for Spoken Explanations) beschäftigt sich mit der Entwicklung einer Architektur zur Steuerung und Funktionserweiterung objektorientierter Zielsysteme mit Hilfe gesprochener, natürlicher Sprache (siehe Abbildung 1). Ein besonderer Fokus liegt dabei

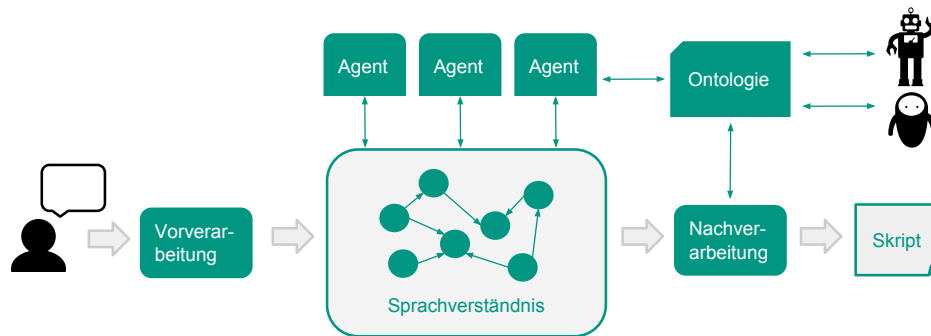


Abbildung 1: Schematische Darstellung der Architektur von PARSE

auf der modularen, agentenbasierten Architektur ProNat [WT15], welche die Integration beliebiger Eingabequellen und Zielsysteme ermöglichen soll. Als exemplarisches Zielsystem wird der am KIT entwickelte humanoide Haushaltsroboter Armar-III [ARAS06] betrachtet. Liegt die Eingabe in gesprochener Form vor, wird diese mittels eines automatischen Spracherkenners (ASR) zunächst in Text überführt. Die gewonnenen Texte werden anschließend mit einer seichten Sprachverarbeitung (SNLP) analysiert und in eine Graphenrepräsentation überführt. Dieser Graph dient in den anschließenden Analyseschritten durch Sprachverständnisagenten (NLU) als gemeinsamen Datenstruktur und wird fortgehend mit extrahierten Informationen angereichert. Dabei ist die PARSE-Architektur so konzipiert, dass die einzelnen Agenten unabhängig voneinander sind und somit eine parallele Verarbeitung der Eingabe möglich ist. Die Modellierung des Zielsystems wird mit einer Domänen-Ontologie umgesetzt und enthält die Attribute, Methoden und zugehörigen Parameter des Roboters. In der Nachverarbeitung wird diese Ontologie zusammen mit dem annotierten Graphen schließlich als Informationsquelle für die Generierung von ausführbarem Quellcode genutzt.

### 3 Zielsetzung

Das Ziel dieser Masterarbeit ist die Entwicklung eines Werkzeuges, welches in gesprochenen Eingaben Anweisungen für die Funktionserweiterung des Zielsystems detektiert und in Methodendeklarationen überführt.

Dafür soll ein Ansatz verfolgt werden, der die Intention des Nutzers, dem Zielsystem neue Funktionen beizubringen, anhand von syntaktischen und semantischen Mustern in der Spracheingabe prüft. Zunächst werden die beschriebenen Arbeitsabläufe auf definierte Aktionen und zugehörige Aktionssequenzen untersucht. Aus den identifizierten Elementen wird eine Methodendeklaration und ein Methodenrumpf aus den Teilinstruktionen aufgebaut. Eine Teilinstruktion besteht dabei wiederum aus Funktionsnamen und mög-

lichen Parametern, welche in der Spracheingabe erkannt werden müssen. Um diese als neue Funktionalitäten in das Zielsystem integrieren zu können, müssen die extrahierten Teilinstruktionen auf bestehende Methoden der Programmierschnittstelle des Roboters abgebildet werden. Zuletzt soll die Persistierung der zu lernenden Methoden für spätere erneute Verwendungen durch den Nutzer untersucht werden.

### 3.1 Verwandte Arbeiten

Bisherige Arbeiten richten ihren Lösungsansatz für die Erkennung von Methodendeklarationen häufig auf spezifische Domänen, wie beispielsweise der Excel-Dokumentenmodifikation oder der Smartphone-Programmierung.

Bei dem textuellen Dialogsystem JustLingo [WWVT15] werden zur Initiierung einer Funktionserweiterung des Excel-Zielsystems durch den Nutzer vordefinierte Floskeln genutzt. Während JustLingo Mustervorlagen modelliert, basiert das System NLyze [GM14] auf einer entworfenen formalen Sprache, um die textuellen Spezifikationen des Nutzers in Operationen des Excel-Zielsystems umzuwandeln. Weiterhin stützt sich das System, wie auch viele anderen Arbeiten [LGS13, LL05, KM06] auf charakteristische Schlüsselwörter für das Diktieren von neuen Funktionen.

## 4 Vorgehen

Zur Lösung der Problemstellung sollen in einer Vorstudie zunächst mögliche Formulierungen von Funktionserweiterungen, im Folgenden als Lehrsequenzen bezeichnet, im Umfeld der Haushaltsroboter-Domäne auf syntaktische und semantische Muster untersucht werden. Da diese Problematik in der Forschung bisher kaum untersucht wurde und auch der bestehende PARSE-Korpus [Gü15] bisher keine geeigneten Sprachaufnahmen enthält, sollen dafür neue Szenarios konzipiert und Aufnahmen getätigt werden. Eine weitere Möglichkeit bietet der Internet-Marktplatz Amazon Mechanical Turk<sup>1</sup>, durch welchen die Aquirierung der Spracheingaben als sogenannter „Human Intelligence Task“ auf dessen große Nutzerbasis ausgelagert werden könnte. Basierend auf diesen Spracheingaben soll ein Ansatz für das Erkennen von durch den Nutzer neu definierten Arbeitsabläufen entworfen werden. Dazu gehört eine linguistische Analyse möglicher Formulierungen der für eine Definition sowie der zugehörigen Beschreibung von Arbeitsabläufen im Hinblick auf Satzbau und Blockgröße. Aus den identifizierten Elementen der Lehrsequenz sollen daraufhin eine Methodendeklaration einschließlich Methodenrumpf aus Teilinstruktionen aufgebaut werden. Zusätzlich muss die Spracheingabe hinsichtlich möglicher Methoden und Parameter untersucht

---

<sup>1</sup>Quelle: <https://www.mturk.com/>, zuletzt besucht am 21.10.18

und mit den vorhandenen Zielsystemfunktionen abgeglichen werden. Ferner kann auf die Vollständigkeit nötiger Parameter in der Eingabe geprüft und fehlende Parameter mithilfe des PARSE-Dialogagenten [Sch17] erfragt werden. Die beschriebene Funktionalitätserweiterung könnte dabei als Ergänzung der bestehenden Zielsystem-Ontologie umgesetzt werden. Das beschriebene Vorgehen soll daraufhin als Agent für die Rahmenarchitektur PARSE [WT15] umgesetzt und evaluiert werden.

## 4.1 Beispiel

Zur Verdeutlichung der Aufgabenstellung wird eine beispielhafte Anfrage an den Küchenroboter ARMAR-III [ARAS06] betrachtet.

Der zu entwickelnde Agent soll die in der Spracheingabe befindliche Funktionserweiterung, genannt Lehrsequenz, erkennen. Dabei gilt es zu unterscheiden, welcher Teil der Nutzereingabe die *Methodendefinition bzw. Abstraktion [DEF]* und welcher Block die *Definitionsbeschreibung [DESC]* aus Teilinstruktionen und zugehörigen Parametern repräsentiert.

### Annotierte Beispielanfrage

```
„ [when I say you should empty the dishwasher]DEF [you have to open  
the dishwasher grab every piece of dishware in it and put it into the  
cupboard then close the dishwasher]DESC“
```

Die vom Nutzer erwartete Ergänzung der Zielsystemfunktionalitäten um den Befehl „*empty dishwasher*“ könnte in Pseudocode formuliert wie folgt aussehen:

### Pseudocode der beschriebenen Spracheingabe

```
# Methodendefinition  
empty_dishwasher:  
# Beschreibung von Teilinstruktionen  
open(dishwasher)  
foreach piece_of_dishware in dishwasher:  
    grab(piece_of_dishware, dishwasher)  
    put(piece_of_dishware, cupboard)  
close(dishwasher)
```

## 4.2 Herausforderungen

Für das zu entwickelte Werkzeug ergeben sich im Hinblick auf die Themenstellung folgende Herausforderungen:

- Teilinstruktionen und Parameter:
  - Methoden-Abbildung: Zuordnung der Teilinstruktionen der Spracheingabe auf bestehende Methoden des Zielsystems
  - Parameter-Abbildung: Zuordnung möglicher Entitäten der Methodendeklaration auf Parameter der beschriebenen Teilinstruktionen sowie deren Abbildung auf die Parameter bekannter Zielsystem-Methoden
  - Fehlende Parameter: Dabei auftretende fehlende Parameter bei einer Teilinstruktion könnten, beispielsweise durch eine Dialog-Interaktion, behandelt werden.
  - Unbekannte Eingaben: Dabei auftretende unbekannte, nicht auf die API-abbildbare, Teilinstruktionen könnten beispielsweise durch einen rekursiven Aufruf des Werkzeugs zur Erkennung von Methodendeklarationen, initiiert durch eine Dialog-Interaktion, erfragt werden.
- Synonyme: Das Hinzufügen von Synonymen zu neu programmierten Funktionalitäten soll behandelt werden.
- Kontrollstrukturen: Die Formulierung von Kontrollstrukturen (z.B. Schleifen, Bedingungen) in der Definitionsbeschreibung soll integriert werden können.

## 5 Evaluation

Zur Evaluation des entwickelten Werkzeuges sollen die erzeugten Ausgaben des Agenten für gegebene Sprachaufnahmen mit händisch erstellten Musterlösungen abgeglichen werden. In der Vorstudie wird dazu der bestehende PARSE-Korpus bereits um neue Spracheingaben für die behandelte Problemstellung erweitert. Um jedoch eine unbeeinflusste Evaluationsgrundlage zu erhalten, müssen für die Bewertung der Werkzeuggüte erneute Aufnahmen getätigt werden. Die Bewertung der Ergebnisse sollen anhand der Gütemaße Genauigkeit, Präzision, Ausbeute und das F1-Maß bestimmt werden.

## Literatur

- [ARAS06] ASFOUR, T. ; REGENSTEIN, K. ; AZAD, P. ; SCHRÖDER, J.: ARMAR-III: An integrated humanoid platform for sensory-motor control. In: *Humanoid Robots, 2006 6th IEEE-RAS International Conference on IEEE*, 2006, S. 169–175
- [GM14] GULWANI, Sumit ; MARRON, Mark: NLyze: Interactive Programming by Natural Language for Spreadsheet Data Analysis

- and Manipulation. In: *Proceedings of the 2014 ACM SIGMOD International Conference on Management of Data*, 2014 (SIGMOD '14). – ISBN 978-1-4503-2376-5, 803–814
- [Gü15] GÜNES, Zeynep: *Aufbau eines Sprachkorpus zur Programmierung autonomer Roboter mittels natürlicher Sprache*, Karlsruher Institut für Technologie (KIT) – IPD Tichy, Bachelor's Thesis, 2015. [https://code.ipd.kit.edu/weigelt/parse/wikis/Theses/guenes\\_ba](https://code.ipd.kit.edu/weigelt/parse/wikis/Theses/guenes_ba)
- [KM06] KNÖLL, Roman ; MEZINI, Mira: Pegasus: First Steps Toward a Naturalistic Programming Language. In: *Companion to the 21st ACM SIGPLAN Symposium on Object-oriented Programming Systems, Languages, and Applications*, 2006 (OOPSLA '06). – ISBN 1-59593-491-X, 542–559
- [LGS13] LE, Vu ; GULWANI, Sumit ; SU, Zhendong: SmartSynth: synthesizing smartphone automation scripts from natural language. In: *The 11th Annual International Conference on Mobile Systems, Applications, and Services, MobiSys'13, Taipei, Taiwan, June 25-28, 2013*, 2013, 193–206
- [LL05] LIU, Hugo ; LIEBERMAN, Henry: Metafor: Visualizing Stories As Code. In: *Proceedings of the 10th International Conference on Intelligent User Interfaces*, 2005 (IUI '05). – ISBN 1-58113-894-6, 305–307
- [Sch17] SCHLERETH, Mario: *Entwicklung eines Dialogagenten für dialogbasierte Programmierung*, Karlsruher Institut für Technologie (KIT) – IPD Tichy, Master's Thesis, 2017. [https://code.ipd.kit.edu/weigelt/parse/wikis/Theses/schlereth\\_ma](https://code.ipd.kit.edu/weigelt/parse/wikis/Theses/schlereth_ma)
- [WT15] WEIGELT, S. ; TICHY, W.F.: Poster: ProNat: An Agent-Based System Design for Programming in Spoken Natural Language. In: *Software Engineering (ICSE), 2015 IEEE/ACM 37th IEEE International Conference on* Bd. 2, 2015, S. 819–820. – [http://ps.ipd.kit.edu/163\\_422.php](http://ps.ipd.kit.edu/163_422.php)
- [WWVT15] WACHTEL, Alexander ; WEIGELT, Sebastian ; VOIGT, Philipp ; TICHY, Walter F.: Prototyp einer natürlichsprachlichen Schnittstelle für Tabellenkalkulation. In: *Multikonferenz Software Engineering & Management 2015* Bd. 239, Gesellschaft für Informatik (GI), Bonn, 2015. – ISBN 978-3-88579-633-6, S. 1617–5468