

Erkennung von Aktionen in gesprochener Sprache

Dokumentenart: Exposé für eine Bachelorarbeit
Autor: Yue Ou
Matrikel-Nr.: 1519440
Studiengang: Informatik Bachelor
Betreuer: Sebastian Weigelt
Datum: 27. Mai 2016

1 Motivation

Um die Kommunikation zwischen Menschen und Maschinen mittels natürlichen Sprachen zu ermöglichen, muss die Semantik der gesprochenen Sprache von den Maschinen erfasst werden. Ein wesentlicher Bestandteil der Semantik sind die in der Eingabe enthaltenen Aktionen. Die Maschinen müssen die Aktionen in den Sätzen erkennen und verstehen, sodass sie diese ausführen können.

Eine Aktion besteht aus einer Tätigkeit, die von einem Akteur ausgeübt wird. Dabei können zusätzlichen Informationen enthalten sein, z.B. das Objekt, der Zeitpunkt und der Ort. Damit die Maschinen die Aktionen verstehen, ist es sinnvoll, diese Informationen aus unterschiedlichen konstruierten Sätzen zu extrahieren und einheitlich darzustellen.

Es gibt vier Konstruktionen für Sätze in Englisch: deklarative Sätze, imperative Sätze, Ja/Nein-Fragen und W-Fragen[MJ00]. Diese Arbeit wird sich mit Aktionen beschäftigen und daher nur die imperative Sätze und deklarative Sätze betrachten.

Es handelt sich bei einem imperativen Satz meistens um einen Befehl. Ein Beispiel dafür ist:

„Get me an orange juice.“ [1]

Um diesen Satz semantisch zu erfassen ist es wichtig zu erkennen, dass es um die Tätigkeit „get“ geht und dass „me“ und „orange juice“ zusätzliche Informationen sind. Ein Akteur ist hier nicht explizit angegeben.

In einem deklarativen Satz ist meistens eine Beschreibung enthalten. Allerdings kann die Beschreibung auch eine Aktion enthalten. Folgenden Beispiele enthalten die gleiche Aktion:

„Leia kissed Luke.“ [2]
„Luke was kissed by Leia.“ [3]

In beiden Sätzen handelt es sich um die gleiche Aktion: „kiss“, den gleichen Akteur: „Leia“ und das gleiche Objekt: „Luke“. [2] ist eine aktive Schreibweise und [3] dahingegen passiv. Trotz der unterschiedlichen syntaktischen Strukturen soll für solche Sätzen eine einheitliche semantische Darstellung konstruiert werden, da die Semantik gleich ist.

2 Projekt PARSE

Das Projekt PARSE wurde entwickelt, um Programmieren mit gesprochenen natürlichen Sprachen in verschiedenen Bereichen zu verwirklichen. Ein Zielsystem ist ARMARIII[ARA⁺06]. Hier soll einem Haushaltsroboter neue Anweisungen mit gesprochener Sprache beigebracht werden.

Zuerst wird das akustische Signal mithilfe von einem automatischen Spracherkennung (engl. Automatic Speech Recognizer - ASR) in Text transformiert. Dann wird aus dem Text mittels seichtler Sprachverarbeitung (engl. Shallow Natural Language Processing - SNLP)[Koc15] ein Graph erzeugt, der als initiale Handlungsrepräsentation dient. Der Graph enthält Informationen über Wortarten und Phrasen. PARSE basiert auf mehrere Agenten, die die initiale Handlungsrepräsentation modifizieren und unterschiedlichen Funktionalitäten ins System integrieren. Die aufgearbeitete Repräsentation kann dann auf verschiedenen Zielsysteme abgebildet werden.

Der Ablauf wird in der Abbildung 1 illustriert.

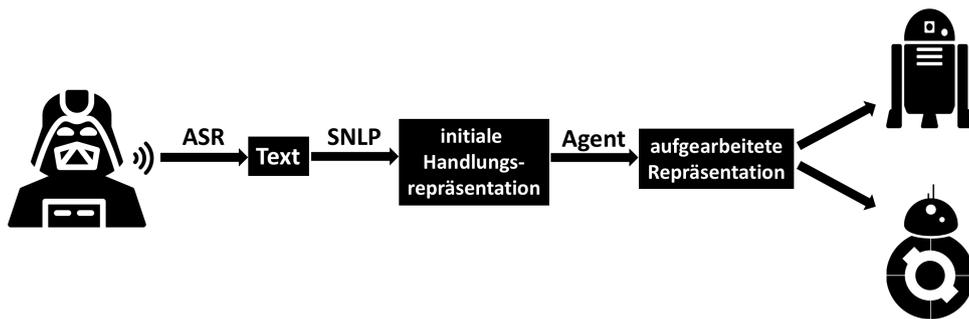


Abbildung 1: Ablauf von Parse

3 Zielsetzung

Das Ziel dieser Bachelorarbeit ist, einen Agent für das Projekt PARSE zu entwickeln. Der Agent soll die von SNLP erzeugte initiale Handlungsrepräsentation entgegennehmen und einzelne Aktionen daraus extrahieren. Eine Aktion enthält die folgenden Informationen: der Akteur, das Prädikat und die Parameter. Der Akteur führt die Aktion aus. Das Prädikat bestimmt die Tätigkeit und besteht immer aus einem Verb. Die Parameter sind die Objekte des Satzes und beantworten die Fragen „was“, „wer“, „wo“, „wann“, „wie“ und „warum“ (die W-Fragen [Kli76]). Der Agent modifiziert die initiale Handlungsrepräsentation um diese Informationen im Graph darzustellen.

3.1 Beispiel

Die modifizierte graphische Darstellung wird anhand des folgenden Beispiels illustriert. Zur Veranschaulichung werden die Prädikate rot, Akteure blau und Parameter grün markiert.

„Ben killed Han with a light saber and Chewie cried.“ Dieses Beispiel enthält zwei Aktionen mit mehreren Parametern und einen Akteur.

In der Abbildung 2 wird der modifizierte Graph angezeigt.

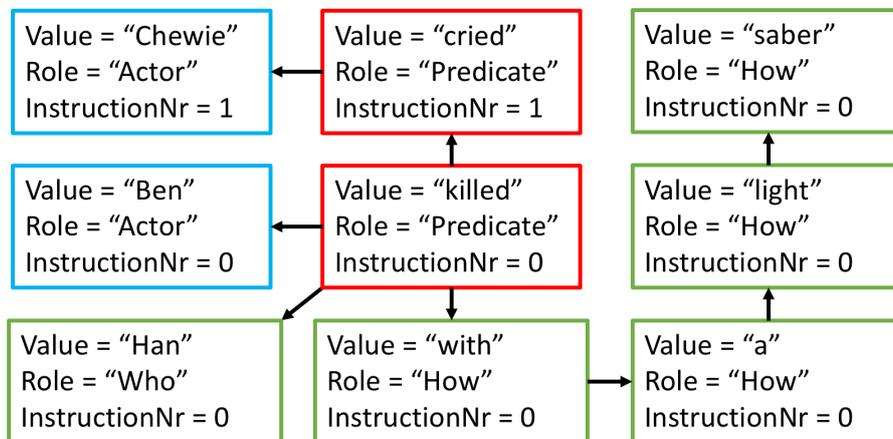


Abbildung 2: Der modifizierte Graph für „Ben killed Han with a light saber and Chewie cried“

Die erste Aktion „Ben killed Han with a light saber“ kann durch die Befehlsnummer „0“ erkannt werden. In dieser Aktion ist „Ben“ der Akteur und „killed“ das Prädikat. „Han“ ist ein Parameter der Art „who“. „With a light saber“ beantwortet die Frage „how“. Durch das „and“ wird eine zweite Aktion verknüpft, nämlich „Chewie cried“. Es wird von „killed“ auf „cried“ gezeigt. Das Wort „and“ wird aber selbst nicht in dem Graph dargestellt,

weil es nur zur Verknüpfung dient. Die zweite Aktion „Chewie cried“ hat die Befehlsnummer „1“.

4 Vorgehensweise

Zwei Aufgaben müssen erledigt werden um das Ziel zu erreichen: die Aktionen in der initialen Handlungsrepräsentation erkennen und die Informationen über die Aktionen in der Repräsentation darstellen.

Die Eingabe wird bereits von SNLP tokenisiert und jedem Wort werden bereits seine Wortart, Befehlsnummer und Phrase zugeordnet. Die Befehlsnummer kann direkt übernommen werden. Es müssen nur Rollen und Kanten zu dem eingegebenen Graphen hinzugefügt werden.

Um die erste Aufgabe zu erfüllen kommt Zuweisung von semantischen Rollen (engl. Semantic Role Labeling - SRL)[CM05] zum Einsatz. Eine semantische Rolle ist die Beziehung zwischen einer Phrase und einem Prädikat. [CM05] Die Phrasen dienen als Argumente des Prädikats. Typische semantische Argumente sind Agent, Patient, Instrument, Zeit, Ort, Anlass usw. Die Zuweisung von semantischen Rollen ist eine wichtige Aufgabe, um die W-Fragen zu beantworten.

Nachdem die Rolle jeder Phrase identifiziert wurde, soll das Attribut „Role“ jeden Knoten hinzugefügt werden. Dann werden Kanten hinzugefügt, um die Prädikate und die dazugehörigen Akteure und Parameter zu verbinden. Es wird zwischen den Prädikaten in mehreren Aktionen auch Kanten eingefügt.

5 Herausforderungen

Sätze, die das Wort „and“ enthalten, müssen besonders betrachtet werden. „And“ verbindet oft zwei separaten Aktionen wie in dem obigen Beispiel. Es gibt aber auch andere Fälle. In dem Beispiel „Leia kissed Luke and Han“ ist „Han“ keine zweite Aktion. Der Satz enthält im Grunde zwei Aktionen, nämlich „Leia kissed Luke“ und „Leia kissed Han“. SNLP trennt verschiedene Befehle nur anhand der Schlüsselwörter „and“, „if“ und „when“. Dieses Defizit soll in dieser Arbeit verbessert werden.

Die Aufgabenstellung beinhaltet zudem weitere Herausforderungen.

Wie in der Motivation schon erwähnt kann ein deklarativer Satz entweder eine Beschreibung sein, oder auch eine Aktion enthalten. „Han Solo flies the Millennium Falcon“ ist z.B. eine Zustandbeschreibung. „Han Solo is flying the Millennium Falcon“ ist dahingegen eine Aktion. Es kann anhand der Zeitform unterschieden werden. Der Agent sollte erkennen, ob ein Satz eine Aktion enthält. Sätze ohne Aktion sollen nicht bearbeitet werden.

Eine weitere Herausforderung ist Substantivierung. „R2 turned around.“ und „R2 made a turn.“ haben die gleiche Bedeutung, und sollten daher die

gleiche semantische Darstellung haben.

6 Evaluation

Um die Ergebnisse des Agents zu evaluieren wird der in der Bachelorarbeit von Zeynep entstandene Sprachkorpus[G15] verwendet. Es werden Musterlösungen für die Aktionen erstellt. Die Korrektheit wird bewertet, indem die Ergebnisse des Agenten mit den Musterlösungen verglichen werden. Dabei werden unterschiedliche Maße, wie F-Maß, Präzision und Ausbeute[MJ00] verwendet.

Literatur

- [ARA⁺06] ASFOUR, Tamim ; REGENSTEIN, Kristian ; AZAD, Pedram ; SCHRÖDER, Joachim ; BIERBAUM, Alexander ; VAHRENKAMP, Nikolaus ; DILLMANN, Rüdiger: 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
- [AZCA08] ATSERIAS, Jordi ; ZARAGOZA, Hugo ; CIARAMITA, Massimiliano ; ATTARDI, Giuseppe: Semantically Annotated Snapshot of the English Wikipedia. In: *LREC*, 2008
- [CDP00] COOPER, Stephen ; DANN, Wanda ; PAUSCH, Randy: Alice: a 3-D tool for introductory programming concepts. In: *Journal of Computing Sciences in Colleges* Bd. 15 Consortium for Computing Sciences in Colleges, 2000, S. 107–116
- [CM05] CARRERAS, Xavier ; MÀRQUEZ, Lluís: Introduction to the CoNLL-2005 shared task: Semantic role labeling. In: *Proceedings of the Ninth Conference on Computational Natural Language Learning* Association for Computational Linguistics, 2005, S. 152–164
- [Cul15] CULPEPER, Jonathan: *History of English*. Routledge, 2015
- [G15] GÜNES, Zeynep: *Aufbau eines Sprachkorpus zur Programmierung autonomer Roboter mittels natürlicher Sprache*, Karlsruher Institut für Technologie (KIT) – IPD Tichy, Bachelor’s Thesis, Mai 2015. https://code.ipd.kit.edu/weigelt/parse/wikis/Theses/guenes_ba
- [HMK⁺14] HOSEK, Jiri ; MASEK, Pavel ; KOVAC, Dominik ; RIES, Michal ; KRÖPFL, Franz: IP home gateway as universal multi-purpose

enabler for smart home services. In: *e & i Elektrotechnik und Informationstechnik* 131 (2014), Nr. 4-5, S. 123–128

- [Kli76] KLINKE, Wolfgang: Wie heißt die Antwort auf diese Frage? In: *Zum Status von Fragen und Antworten in einer Sprechakttheorie*. In: Weber, Heinrich/Weydt, Harald (eds.). *Sprachtheorie und Pragmatik. Akten des 10* (1976), S. 123–132
- [Koc15] KOCYBIK, Markus: *Projektion von gesprochener Sprache auf eine Handlungsrepräsentation*, Karlsruher Institut für Technologie (KIT) – IPD Tichy, Bachelor’s Thesis, 2015. https://code.ipd.kit.edu/weigelt/parse/wikis/Theses/kocybik_ba
- [MJ00] MARTIN, James H. ; JURAFSKY, Daniel: Speech and language processing. In: *International Edition* (2000)
- [QCE85] QUIRK, Randolph ; CRYSTAL, David ; EDUCATION, Pearson: *A comprehensive grammar of the English language*. Bd. 397. Cambridge Univ Press, 1985
- [Sus99] SUSSEX, Roland: DAVID CRYSTAL, English as a global language. Cambridge & New York: Cambridge University Press, 1997. Pp. x, 150. Hb. In: *Language in Society* 28 (1999), Nr. 01, S. 120–124