

Visuelle Unterstützung von Patternregeln-basierten Dialogsystemen

Immer mehr Dialogsysteme, auch Chatbots genannt, versuchen, auf natürlich-sprachliche Eingaben mit natürlich-sprachlichen Ausgaben zu reagieren. Sie basieren häufig auf Patternregeln (vgl. ELIZA, AIML). Die Patterns in solchen Regeln modellieren im Allgemeinen eine unendliche Sprache. Ein eingegebener Text wird auf die Zugehörigkeit zu den modellierten Sprachen geprüft. Im positiven Fall kann dies zur Ausführung der Regelschlussfolgerung und zur Erzeugung einer Reaktion auf die Eingabe führen. Dies ist abhängig von der implementierten Konfliktlösungsstrategie des Regelinterpreters. Im einfachsten Fall wird die erste Patternregel ausgewählt, zu der der eingegebene Text gehört. Allerdings gibt es auch andere Konfliktlösungsstrategien. Dann bestimmt der Regelinterpreter alle Patternregeln, zu denen ein eingegebener Text gehört. Die Konfliktlösungsstrategie nutzt verschiedene Heuristiken, um eine konkrete Regel aus der Kandidatenmenge auszuwählen.

Bei der Modellierung der konkreten Patternregeln für einen Anwendungsbereich muss man die vom Regelinterpreter genutzte Konfliktlösungsstrategie berücksichtigen. Dennoch verliert man bei vielen Regeln schnell die Übersicht, welche Regel zur Ausführung kommt. Als Konsequenz entstehen Reaktionen, die nicht beabsichtigt waren und die sehr häufig als unangemessen auf den Benutzer wirken.

Deshalb soll im Rahmen dieses Themas eine graphische Visualisierung der entstehenden Halbordnung zwischen den Patternsprachen erzeugt werden. Zunächst ist dazu im theoretischen Teil der Arbeit das Konzept der Inklusion (Teilmengenbeziehung) zu untersuchen.

Sind in Patterns lediglich reguläre Ausdrücke zugelassen, werden reguläre Sprachen modelliert. Für reguläre Sprachen ist bekannt, dass das Wort-, Leerheits-, Endlichkeits- und Äquivalenzproblem entscheidbar sind und diese Sprachen abgeschlossen sind gegenüber Vereinigung, Durchschnitt, Komplement und Verkettung. Damit kann auch entschieden werden, ob eine reguläre Sprache eine Teilmenge einer anderen regulären Sprache ist.

„Echte“ Patternsprachen, die in ihren Mustern Variablen (mehrfach) verwenden, liegen orthogonal zur Chomsky-Hierarchie. Sind Variablenersetzungen mit dem leeren Wort zugelassen, spricht man von E-Patternsprachen, sonst von NE-Patternsprachen. Sie sind in der Lage, einige reguläre Sprachen zu modellieren, aber auch kontextsensitive Sprachen. Für Patternsprachen mit Variablen und der Auflage, dass eine Variable mindestens an ein Zeichen gebunden werden muss, ist bereits das Wortproblem NP-vollständig. Dominik D. Freydenberger hat sich in seiner Dissertation mit der Inklusion von Patternsprachen auseinandergesetzt.

Basierend auf der Ergebnisanalyse soll zunächst eine Familie von Patternsprache mit maximaler Ausdrucksfähigkeit festgelegt werden, für die die Inklusion entscheidbar ist. Ggf. sind weitere Einschränkungen nötig, um die Laufzeit zur Berechnung zu begrenzen. Damit soll die Grundlage geschaffen

werden, um eine Anwendung zu konzipieren und zu entwickeln, die ein Textfile mit Patternregeln eines Dialogsystems so aufbereitet, dass die enthaltene Halbordnung mit graphviz visualisiert werden kann. Ein zweiter Schritt kann darin bestehen, die Priorisierung der Regeln in der Halbordnung graphisch darzustellen.

Die Umsetzung schließt ein Software-Design, die Programmierung und eine Funktionsdemonstration ein. Entsprechende Einstiegsliteratur kann bei Übernahme des Themas zur Verfügung gestellt werden.

Betreuende Hochschullehrerin: Prof. Dr. Oksana Arnold

April 2014