

**Name:** Shamsan, Mohammed

**Matrikel:** 2615892

**Ausgabe:** 10.01.2013

**Fakultät 3**  
**Lehrstuhl Kommunikationstechnik**

---

Prof. Dr.-Ing. habil. Matthias Wolff  
Lehrstuhlinhaber  
T +49 (0)355 69 21 28  
F +49 (0)355 69 21 50  
E matthias.wolff@tu-cottbus.de  
www.tu-cottbus.de/kommunikationstechnik

## **Untersuchungen zur Modellierung von semantischer Information am Beispiel des Sprachverarbeitungssystems SIRI**

*Investigations on modeling of semantic information based on the speech processing system SIRI*

Es ist allgemein anerkannt, dass ein tiefgreifender Fortschritt der Sprachtechnologie zu erwarten ist, wenn die Systeme „verstehen“ was sie erkennen bzw. synthetisieren. Solche Systeme benötigen neben den Funktionen zur Analyse und Synthese von Sprache auch eine semantische Ebene. Damit wird es insbesondere bei Sprachdialogsystemen möglich, die Konfiguration der Spracherkennung bzw. Sprachsynthese semantik- bzw. dialoggetrieben zu gestalten.

In der vorliegenden Studienarbeit soll der gegenwärtige Stand der Technik am Beispiel der SIRI-Software untersucht werden, welche in der Produktpalette von iPhone Anwendung findet. Von besonderem Interesse sind dabei die Datenorganisation und der Fluss der semantischen Information. Beide Schwerpunkte sollen am Beispiel von einschlägigen Online- und Offline- Spracherkennungsdiensten analysiert werden. Nachfolgend sind aussagekräftige Modelle zu entwickeln, mit denen die Stärken und Schwächen des Systems identifiziert und bewertet werden können. Mit den gewonnenen Einsichten können anschließend zukünftige Forschungsschwerpunkte für die semantische Modellierung herausgearbeitet werden.

Im Einzelnen sind folgende Teilaufgaben zu behandeln:

- Untersuchung zum Funktionsumfang
  - technische Kommunikationsstruktur, Bedienung, Ressourcen
  - Dialog mit Online-Diensten
  - Dialog mit Offline-Applikationen
- Entwicklung von dienstspezifischen ER-Diagrammen
  - Bewertung der Güte der Modellierung
  - direkte und indirekte Identifikation von Datenobjekten

Betreuer: Dr. Ing. Ronald Römer

# Studienarbeit

am

Lehrstuhl für Kommunikationstechnik

Prof. Dr. -Ing. habil. Matthias Wolff



Brandenburgische Technische Universität Cottbus  
Fakultät für Maschinenbau, Elektrotechnik und  
Wirtschaftsingenieurwesen  
Institut für Elektronik und Informationstechnik

## **Untersuchung zur Modellierung von semantischer Information am Beispiel des Sprachverarbeitungssystems Siri**

Eingerichtet von:

Betreut von:

Name: Mohammed Shamsan

Dr. Ing. Ronald Römer

Matrikel Nr.: 2615892

Studiengang : Elektrotechnik

Cottbus den 13.06.2013

# Inhalt

1. Einleitung .....	4
1.1 Motivation und Ziel .....	4
1.2 Aufbau der Arbeit .....	5
2. Theoretische Grundlagen.....	7
2.1 Sprachverarbeitungssystem .....	7
2.1.1 Sprachsynthese .....	8
2.1.2 Spracherkennung .....	9
2.1.3 Dialogsysteme .....	11
2.2 Siri.....	14
2.2.1 Vorgeschichte.....	14
2.2.2 Mensch-Maschine-Interaktion.....	15
2.2.3 Dialogverarbeitungssystem .....	16
2.2.4 System Semantic Siri .....	17
2.2.5 Gesamtsystem Siri.....	19
2.3 semantische Modellierung.....	21
2.3.1 Statische Modellierung .....	21
2.3.1.1 Entity-Relationship-Modell .....	22
2.3.1.2 Integriertes Struktur- und Funktionsmodell .....	26
2.3.1.3 Die Assoziations- und Verknüpfungstabelle .....	28
2.3.2 Dynamische Modellierung .....	29
2.3.3 Direkte und indirekte Identifikation .....	31
2.3.4 Beispiel Telefonbuch.....	33
3. Stand der Technik .....	38
3.1 Siri und seine Anwendungen.....	38
4. Untersuchung.....	39
4.1 Dialog mit Offline-Applikationen.....	39
4.1.1 Das Telefonbuch .....	40
4.1.2 Analyse .....	43
4.2 Dialog mit Online-Diensten .....	47
4.2.1 Sport (Yahoo Sportacular):.....	47
4.2.2 Analyse .....	51

5. Bewertung der Güte der Modellierung für Siri bei den getesteten Diensten .....	53
6. Zusammenfassung und Ausblick.....	56
Literaturverzeichnis.....	59
Abbildungsverzeichnis.....	61

# 1. Einleitung

Maschinen, die mit Menschen kommunizieren, waren seit langer Zeit ein fester Bestandteil der alten Science-Fiction-Filme. Schon seit über 30 Jahren versuchen Wissenschaftler solche Maschinen zu verwirklichen, Maschinen, die in der Lage sind sprachlich mit Menschen zu kommunizieren. Solche Systeme sind bis heute nicht wirklich realisiert worden, dennoch hat es gerade im Bereich des automatischen Verstehens von gesprochener Sprache in den vergangenen Jahren erhebliche Fortschritte gegeben, die in Form von Produkten auch dem Endverbraucher zugänglich gemacht werden. Hierzu gehören beispielsweise Mobiltelefone, die bei der Nennung eines Namens automatisch die entsprechende Nummer wählen und Diktiersysteme, die das Gesprochene mit durchaus überzeugenden Ergebnissen in geschriebenen Text umwandeln. Sogar natürlich sprachliche Dialoge mit Computern sind heute bereits möglich. So existieren produktive Systeme in den unterschiedlichsten Bereichen, die das Verhalten einer menschlichen Auskunftsperson erfolgreich simulieren (vgl. [24]).

Heutzutage ist die Mobiltelefonie eine der interessantesten Märkte für Sprachsysteme. Durch die Entwicklung von intelligenten Handys (Smartphone) wird es möglich solche Systeme im Handy zu integrieren. Diese Systeme sollen dem Verbraucher ermöglichen eine akustische Kommunikation mit dem Handy zu führen und es leichter zu bedienen.

## 1.1 Motivation und Ziel

Wegen der zahlreichen verschiedenen Spracherkennungssysteme in allen Bereichen, wie zum Beispiel in Computer, Smartphones, Tablets und auch im Smart-TV, sollen diese Systeme die Kommunikation zwischen Mensch und Maschine bzw. System verbessern. Smartphones mit einem Spracherkennungssystem sollen nicht nur Diktieren, Datenerfassen oder Sprachsteuern, solche Systeme sollen auch Sprachdialoge mit den Nutzern ausführen und ihre Absichten verstehen. Um das zu verwirklichen brauchen solche Systeme neben Spracherkennung und Sprachsynthese eine semantische Ebene. Jetzt stellt sich die Fragen, wie gut kann ein Sprachsystem seinen Nutzer verstehen und welche Strukturen seine semantischen Daten haben. Um diese Fragen beantworten zu können, wird in dieser Arbeit ein aktuelles Sprachsystem untersucht. Dabei ist es das Ziel anhand von Testversuchen seine semantische Ebene zu entschlüsseln. Als Beispiel für solche Systeme soll der gegenwertige Stand der Technik am Beispiel des Siri-System untersucht werden. Dieses System hatte Apple Ende 2010 gekauft (vgl. [9]) und für ihre iPhone-Nutzer zur Verfügung gestellt. In dieser Studienarbeit wird die Struktur der semantischen Informationen des Siri-Systems untersucht und anhand verschiedener Anwendungen (Dienste) werden seine Fähigkeiten erforscht. Interessant ist zu wissen, wie weit dieses System die Nutzerinformationen verarbeiten kann und ob es die Bedeutung der Nutzeraussagen verstehen kann.

Um das zu erkennen, musste man vorher die semantische Struktur von Siri analysieren, ein passendes Weltmodell entwickeln und versuchen zu verstehen wie Siri die Nutzeraktionen versteht und die Wunschobjekte identifiziert. Siri soll letztendlich drei Fragen beantworten können, damit es die Bedeutung der Nutzerwünsche erfüllen kann. Die Fragen lauten wie folgt(vgl. [1]):

- Was muss Siri tun? (Aktion-ID),
- Welche Parameter braucht es dafür? und
- Womit macht es das? (Objekt-ID)

Diese Fragen werden in dieser Arbeit diskutiert und analysiert.

## 1.2 Aufbau der Arbeit

Im Voraus der Untersuchungen wird zunächst das Kapitel der theoretischen Einführung der drei Ebenen des Sprachverarbeitungssystems behandelt. Die Ebenen gliedern sich folgendermaßen: Synthese, Erkennung und Dialogsysteme.

Da in dieser Arbeit nur die semantische Ebene eines Sprachsystems behandelt wird, werden die Synthese und Erkennung nicht in der Tiefe untersucht. In Dialogsystemen werden die Dialogarten erwähnt und jeweils ein Beispiel vorgestellt.

In dem zweiten Teil der theoretischen Grundlagen wird die Geschichte von Siri's Entwicklung erläutert. Hier wird Bezug auf die Absichten zur Entwicklung eines solchen System genommen und Vorteile der Mensch-Maschinen-Interaktion benannt.

Als nächstes wird das Sprachverarbeitungssystem Siri erläutert und in einem Schema dargestellt um visuell zu veranschaulichen wie bei Siri ein Dialogablauf stattfindet. Für das Siri-System hatten die Entwickler eine Vorstellung bezüglich der Semantik, diese wird in der semantischen Ebene diskutiert. Anschließend wird das Gesamtsystem von Siri in zwei Ebenen, der Nutzer-Ebene und der System-Ebene, dargestellt.

Im dritten Teil der theoretischen Grundlagen wird die semantische Modellierung beschrieben, die in zwei Teile unterteilt ist, die statische- und dynamische Modellierung.

In der statischen Modellierung wird mit Hilfe des Entity-Relation-Modells und anhand von ER-Diagrammen ein Weltmodell in der Systemebene aufgeführt. Als Beispiel für dieses Weltmodell wird ein Telefonbuch-Dienst eines Handys genommen. Zu dem Weltmodell gehört auch in der Systemebene ein Integrations- und Funktionsmodell für die Identifikation von Aktionen und Objekten, wie auch Berichtigungs- und Assoziationstabellen, welche die Struktur- und Funktionsmodelle verbindet.

Die dynamische Modellierung dagegen versucht während eines Dialogablaufes eine Strategie zu finden, mit der das System effektiv die passende Antwort geben könnte. Als nächstes wird auf die Problematik von Datenidentifikationen eingegangen und auf die möglichen Methoden zur Identifizierung eines Objekts. Hier werden eine direkte und indirekte Objektidentifikation im Idealfall beschrieben und in Modellen dargestellt. Abschließend wird das Telefonbuch als Beispiel genommen um die theoretischen Kenntnisse zu praktizieren.

In Kapitel 3 wird der Stand der Technik und Siri- Aufgaben erklärt, außerdem wird darauf eingegangen, was Siri den Nutzern verspricht. Wie man das System benutzt und welche Dienste mit diesem System arbeiten, werden ebenfalls aufgeführt.

Im 4. Kapitel werden zwei unterschiedliche Dienste mit Siri untersucht. Hier lag der Schwerpunkt in der Analyse, ob diese Dienste semantisch arbeiten und die jeweiligen Objekte direkt und indirekt identifizieren können. Als ein Offline-Dienst wird das Telefonbuch für das iPhone untersucht. Als Online-Dienst dient der Sportdienst von Yahoo (*Yahoo Sportacular*).

In Kapitel 5 wird Siri anhand der untersuchten Dienste bewertet um dann in dem Kapitel 6 eine abschließende Schlussbemerkung erarbeiten und Zukunftsvisionen benennen zu können.

## 2. Theoretische Grundlagen

### 2.1 Sprachverarbeitungssystem

Unter dem Begriff Sprachverarbeitung versteht man die Verarbeitung lautlicher als auch geschriebener Sprache. Die Idee für eine Sprachverarbeitung gab es schon vor etwa 130 Jahren, als der Mensch von den „Sprechenden“-Maschinen inspiriert war. Damals, im Jahr 1880 begann die Idee nicht nur als Mittel der Interaktion zwischen Mensch und Maschine, sondern auch als Möglichkeit zur Herstellung von Sprachen. In den frühen 1950er Jahren entwickelten drei Forscher der Bell Laboratories in den USA die erste automatische Spracherkennung. Hierbei entstand ein System, das in der Lage ist, isolierte Ziffern von 0 bis 9 aus einem einzigen Lautsprecher durch Best-Matching sprachabhängige Standard-Ziffer-Muster zu erkennen(vgl.[2]). Ein paar Jahre später kam der Durchbruch. 1971 förderte die US Defense Advanced Research Projects Agency (DARPA) eine Forschungs- Initiative auf dem Gebiet. Fünf Jahre später ist es der Pennsylvania Carnegie Mellon University "Harpy" gelungen die DARPA Anforderungen zu erfüllen. Sie entwickelten ein System das in der Lage ist, verbundene Rede mit einem Wortschatz von über 1000 Worten zu erkennen. Seine Leistung war noch sehr langsam, dennoch wurde der Grundstein für zukünftige Erfolge gelegt[2].

Der wichtigste Gegenstand der Sprachverarbeitung ist das Sprachsignal. Es entsteht, wenn eine Person etwas spricht und die produzierten Schallwellen über einen elektroakustischen Wandler (Mikrophon) in ein elektrisches Signal umgewandelt werden. Das Sprachsignal wird also durch das, was die Person sagt (Aussage), geprägt. Die Abbildung 1 von Beat Pfister und Tobias Kaufmann in ihrem Buch: *Sprachverarbeitung, Grundlagen und Methoden der Sprachsynthese und Spracherkennung* verdeutlicht, dass das Sprachsignal die Ausgabe und auch die Zusage eines Sprachprozesses ist, der selbst von verschiedenen Faktoren beeinflusst werden kann und damit das Sprachsignal steuert [3].

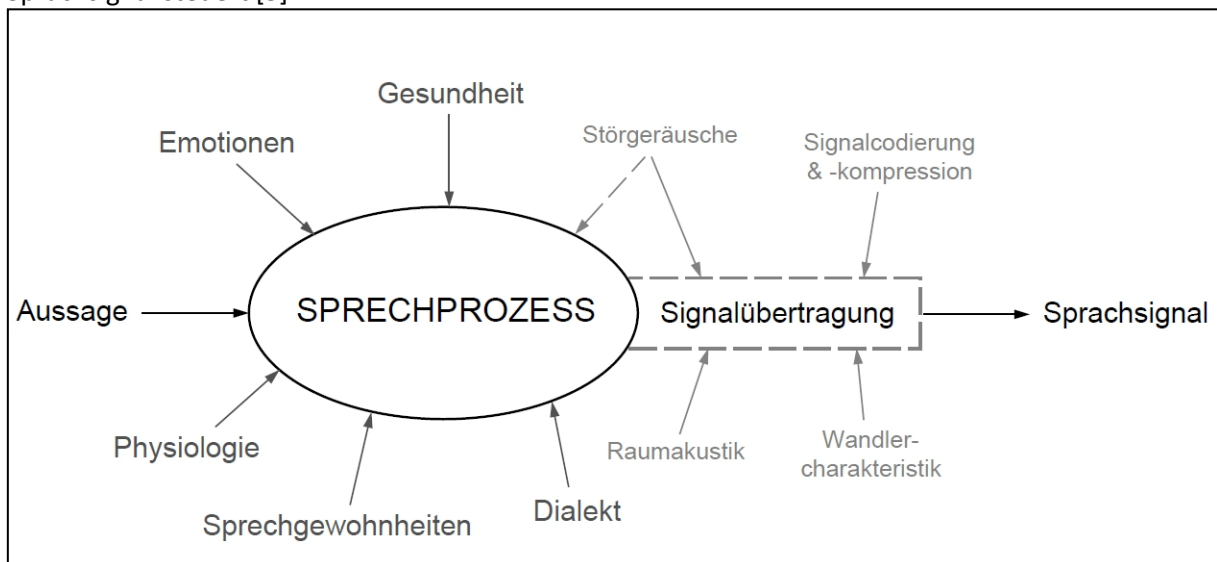


Abbildung 1: Auswirkungen auf Sprachsignal (Pfister & Kaufmann 2008, Sprachverarbeitung,S. 25)

Die maschinelle Verarbeitung natürlicher Sprache umfasst sehr unterschiedliche Bereiche, aber in dem folgenden Abschnitt der Arbeit befassen wir uns mit textlichen und akustischen Erscheinungsformen der Sprache. Es werden deshalb vorwiegend die Spracherkennung und die Sprachsynthese behandelt



### 2.1.1 Sprachsynthese

Bei der Sprachsynthese wird eine Aussage, die in einer symbolischen Notation vorliegt, zum Beispiel eine geschriebene Sprache, in ein Sprachsignal umgewandelt (vgl.[3] S.27). Die Sprachsynthese unterscheidet sich hier in zwei Arten, je nach Form der der eingegebenen Daten.

**TTS-Synthese:** Wenn die Eingabe in Form einer geschriebenen Sprache (Text) erfolgt, dann wird der Antworttext mit Hilfe einer „Text-To-Speech (TTS)-Engine“ in eine Sprachausgabe transformiert. Dabei werden entweder einzelne Phoneme zu ganzen Wörtern und Sätzen kombiniert oder fertige Sprachbausteine zusammengefügt.

**CTS-Synthese :** dieses System (eng. Concept-to-Speech) wird verwendet, wenn die Eingabedaten hierarchisch strukturiert sind, z.B. Syntaxbaum (formale Beschreibung des Satzaufbaus).

Von den Eingabedaten bis zum fertigen Sprachsignal werden zwei Stufen durchlaufen: Die Transkriptionsstufe und die Phonoakustische Stufe (Abbildung 2).

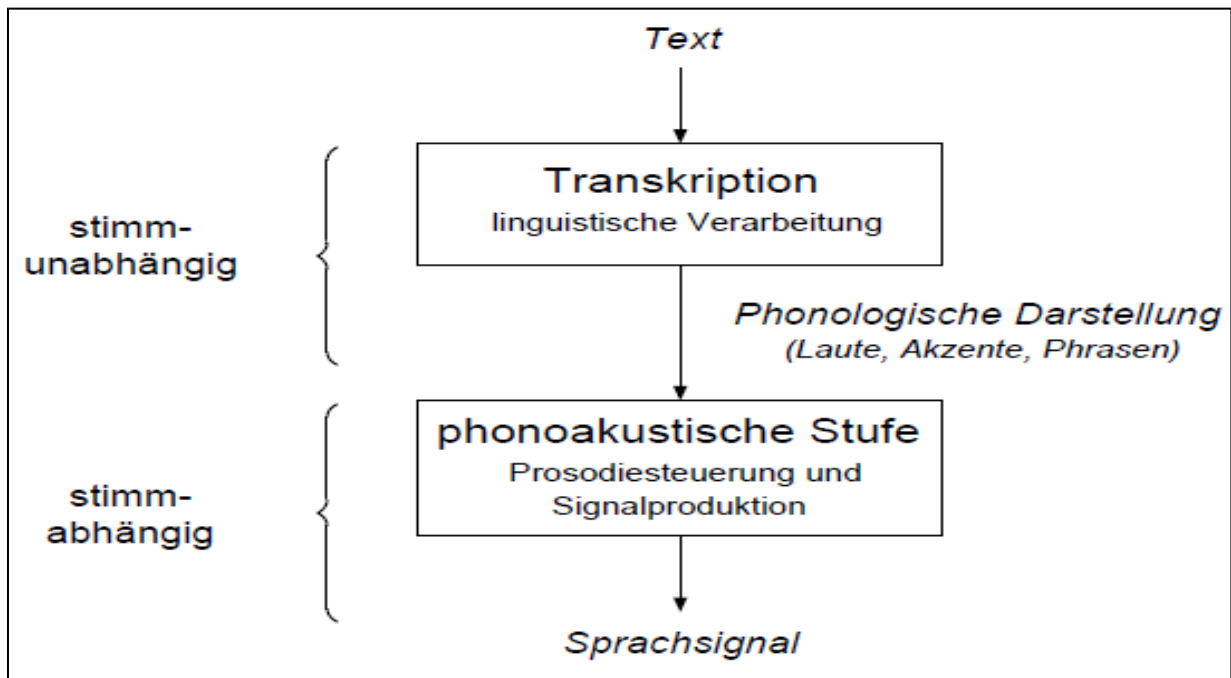


Abbildung 2: Die zwei Stufen des Sprachsynthesensystem (Pfister & Kaufmann 2008, Sprachverarbeitung S.197)

Die phonologische Darstellung des Textes, das Ergebnis der Transkription, ist genauso wie der Text selbst noch stimmabhängig, es werden nur die aufeinander folgenden Laute, Akzente, Phrasen definiert. Erst in der phonoakustischen Stufe wird durch Prosodiesteuerung (Dauer, Frequenz, Intensität) und die Signalproduktion das stimmabhängige Sprachsignal erzeugt (vgl. [3] S.196). Eine weitere Möglichkeit wäre, einen vorher aufgenommenen Text auszugeben. Ein noch ausführlicheres Diagramm haben Pfister und Kaufmann in ihrem Buch erläutert (Abbildung 3 und 4).

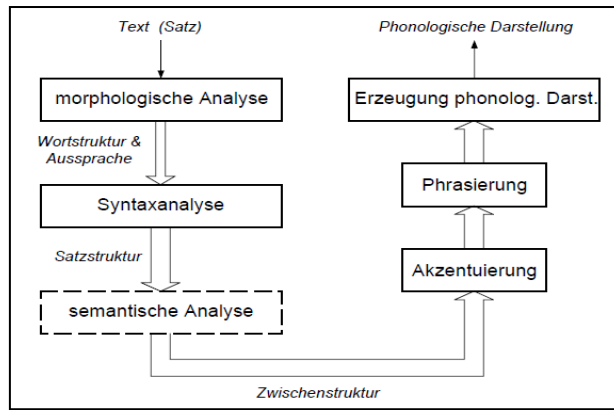


Abbildung 3: Transkriptionsstufe der Sprachsynthese (Pfister & Kaufmann 2008, *Sprachverarbeitung*, S 199)

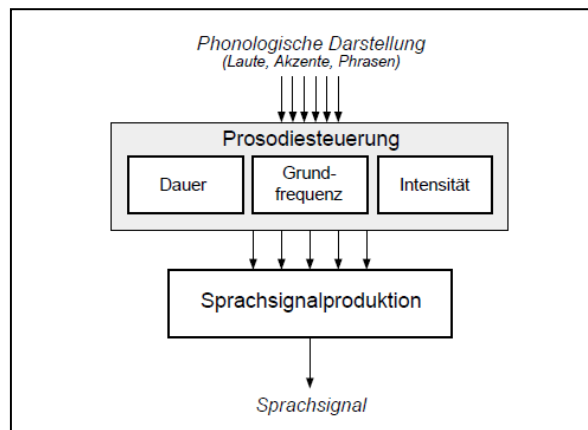


Abbildung 4: Phonoakustische Stufe der Sprachsynthese (Pfister & Kaufmann 2008, *Sprachverarbeitung*, S 200)

## 2.1.2 Spracherkennung

Bei der Spracherkennung werden die Sprachsignale (die gesprochene Sprache) von einer Spracherkennungssoftware (Speech Recognition Engine) in eine textuelle Form umgewandelt. Dabei wird entweder der Ansatz des Mustervergleichs (Sprachsignal bzw. Merkmalssequenz wird mit abgespeicherten Mustern verglichen) oder der Ansatz der statistischen Spracherkennung (Anwendung einer Statistik die angibt, mit welcher Wahrscheinlichkeit bestimmte Wörter vorkommen oder einander folgen können – Hidden Markov Modelle) verwendet (vgl. [4] und [5]). Ein allgemeiner Ablauf für den Spracherkennungsprozess könnte dieses Schema verdeutlichen (Abbildung 5).

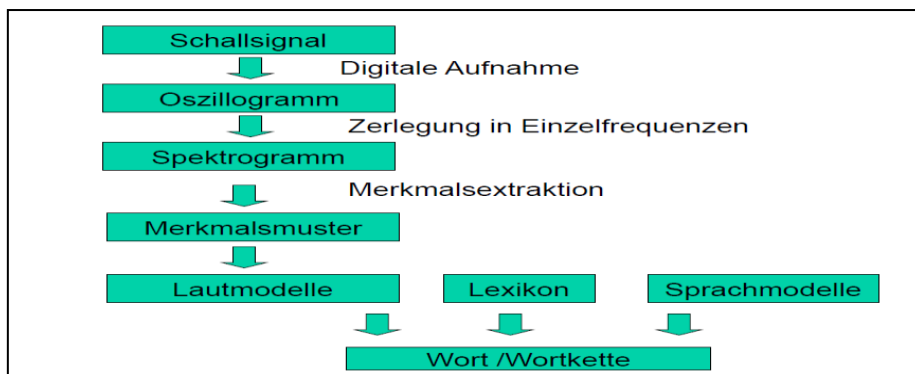


Abbildung 5: Spracherkennungsprozess (Manfred Pinkal (2011): *Verarbeitung gesprochene Sprache*, S.8)

Nachdem ein kontinuierliches Schallsignal gegeben wurde, werden diese Signale (Laute) in eine technische Form (z.B. *Oszillogramm*) digitalisiert (Abbildung 6) damit die Sprachanalyse durchzuführen. Diese Laute werden durch eine Kombination aus Schwingungen verschiedener Frequenzen, die aber wegen der Überlagerung in dem Oszillogramm schwer erkennbar sind, dargestellt. Aus diesem Grund muss das Signal noch geschickter repräsentieren. Durch eine Kombinationsanalyse (Fourier-Transformation) wird das Oszillogramm in Form eines Zeit-Frequenz-Diagrammes (*Spektrogramm*) (Abbildung 7) dargestellt. Da in Spektrogrammen, aufgrund der Varianz des Signals (gleiche Laute – gleiches Wort wird nicht immer gleich ausgesprochen) und der Kontinuität des Signals (die Laute eines Wortes lassen sich schwer gegeneinander abgrenzen), die Laute noch nicht identifizieren werden können, wird eine statische Modellierung durchgeführt. Dadurch kann man die wahrscheinlichsten Wortkette  $W = w_1 w_2 \dots w_n$ , die einem beobachteten akustischen Signal entspricht, ermitteln und die akustische Information für statistische Berechnungen differenzieren. Mit einer Merkmalsextraktion wird dann eine handhabbare Charakterisierung (*Merkmalmuster*) des akustischen Signals erzeugt. Das heißt, dass die Schallenergie in einzelnen Frequenzfenstern und Zeitfenstern abgestimmt wird und eine Folge von (Einzel-) Beobachtungen geschehen  $O = o_1 o_2 \dots o_m$ . Dabei ist jedes  $o$  ein Merkmalsvektor, der die Schallenergie für die unterschiedlichen Frequenzfenster in einem bestimmten Zeitfenster angibt (Abbildung 8).

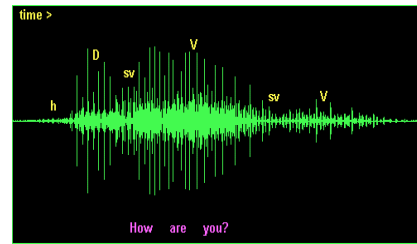


Abbildung 6: Oszillogrammdarstellung eines Schallsignals ([6], S.2)

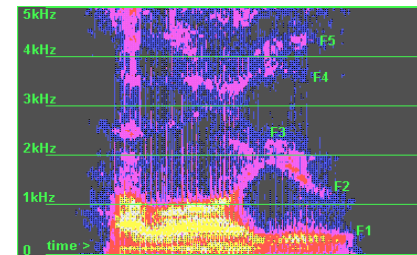


Abbildung 7: Spektrogrammdarstellung ([6], S.3)

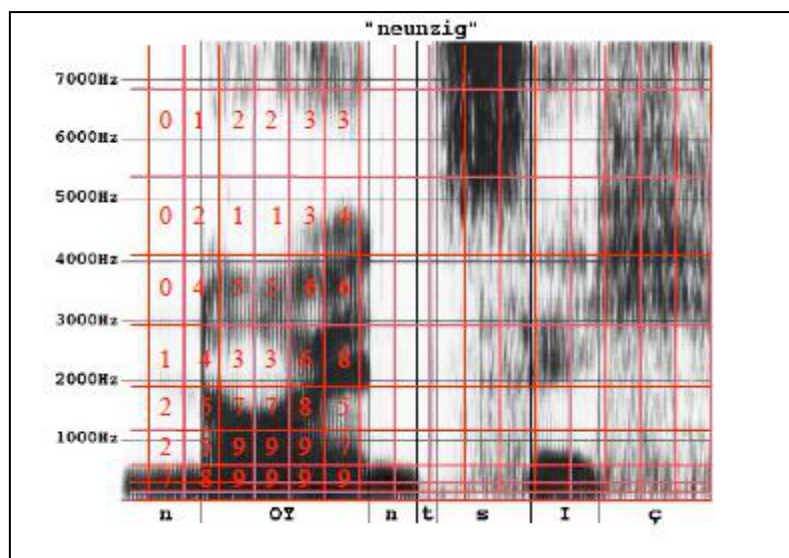


Abbildung 8: Darstellung von Merkmalsvektoren ([6], S.5)

Jetzt muss aus dem Merkmalmuster  $O$  eine möglichst gute Schätzung  $\hat{W}$  der geäußerten Wortfolge  $W$  ermittelt werden. Mit anderen Worten kann dies wie folgt formuliert werden: Suche zu einem gegebenen Merkmalsvektor  $O$  diejenige Wortfolge, welche von allen möglichen Folgen von Wörtern aus einem vorgegebenen beobachteten akustischen Signal  $W$  mit der kleinsten Wahrscheinlichkeit zu einem Fehlentscheid führt. Ein Fehlentscheid tritt dann auf, wenn mindestens eines der Wörter

falsch ist. Dies wird mit Hilfe der Entscheidungsregel **Maximum-a-posteriori-Regel (MAP-Regel)** erreicht:

$$\hat{W} = \operatorname{argmax} P(W|O) \quad (1)$$

Diese Formel stellt, um die Wahrscheinlichkeit eines Fehlentscheid zu minimieren, dar, dass bei dem gegebenen Merkmalsvektor  $O$  diejenigen Wörter/Wortketten  $\hat{W}$  auszuwählen sind, welche die höchste A-Posteriori-Wahrscheinlichkeit  $P(W|O)$  aller möglichen Folgen von Wörtern aufweist [5].

Unter der Verwendung des *Satzes von Bayes* (mehr dazu siehe [5] ab S. 327 ), kann die A-Posteriori-Wahrscheinlichkeit ausgedrückt werden als

$$P(W|O) = \frac{P(O|W).P(W)}{P(O)} \quad (2)$$

Durch das Einsetzen dieser Gleichung in (1) wird  $P(O)$  weggelassen, da diese Wahrscheinlichkeit für die Maximierung konstant ist und die Entscheidung nicht beeinflussen kann, und kann die erkannte Wort/Wortkette so erfasst werden:

$$\hat{W} = \operatorname{argmax} P(O|W).P(W) \quad (3)$$

Wobei :

**$P(O|W)$** :ist die bedingte Wahrscheinlichkeit, die aussagt, mit welcher Wahrscheinlichkeit der Merkmalsfaktor  $O$  zu beobachten ist, wenn die Wortfolge  $W$  gesprochen wird. Aufgrund vieler Beispiele von Sprachsignalen kann eine statistische Beschreibung erzeugt werden und angeben, welcher Merkmalsfaktor für welche Wortfolgen wie wahrscheinlich ist. So eine Beschreibung wird auch als „Akustisches Model“ bezeichnet. Die Bestimmung von  $P(O|W)$  geschieht in der Regel in zwei Schritten. Als erstes werden die Wörter als Ketten von Grundeinheiten, zum Beispiel Phonemen, dargestellt. Dabei werden zunächst die Lautmodelle in einer Trainingsphase in einer Datensammlung für gesprochene Sprache aufgenommen und diese Aufnahmen von Sprachlauten mit ihrer phonetischen Kategorie hat die Wahrscheinlichkeit mit bestimmten Lauten durch Merkmalsmuster realisiert zu werden. Für die statistische Zuordnung von Merkmalsmustern und Wörtern wird die HMM-Methode („Hidden Markov Models“) verwendet.

**$P(W)$** :ist die Wahrscheinlichkeit, dass eine bestimmte Wortfolge geäußert wird: „Sprachmodell“[5]. Eine allgemeine Methode um  $P(W)$  zu berechnen, stellt das Multiplikationsgesetz dar. Mit dessen Hilfe lässt sich die Verbundwahrscheinlichkeit in die bedingte Wahrscheinlichkeit der einzelnen Wortfolgen zerlegen und mit der N-Gram Methode  $P(W)$  berechnen (mehr dazu siehe [3] ab Seite 370).

### 2.1.3 Dialogsysteme

Ein Dialogsystem ist ein interaktives System, das dem Benutzer die Möglichkeit gibt, Zugang zu Systemfähigkeiten durch das Führen eines Dialogs zu erlangen, mit einer Maschine interagieren zu können und in gesprochener Sprache miteinander (Mensch und System) zu kommunizieren. Es sollte in der Lage sein einen Überblick über die Grundlagen auf einem Dialog basierten Schnittstellen zu geben. Im Idealfall soll ein Dialogsystem außerdem in der Lage sein, über eigene Fähigkeiten, Ziele und Wünsche „nachzudenken“ und dem Dialogpartner das Resultat dieser Überlegungen mitzuteilen.

Diese Eigenschaften sind derzeit noch nicht umsetzbar, jedoch ein erstrebenswertes Ziel für die Zukunft. Abbildung 9 zeigt den Ablauf der Sprachanalyse eines typischen Dialogsystems:

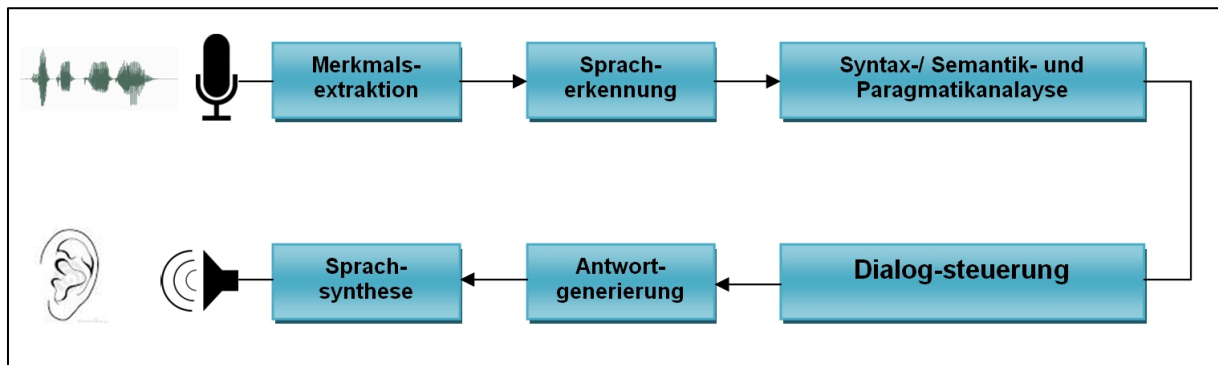


Abbildung 9: Dialogsystem der Sprachanalyse

Der erste Schritt nach der Spracheingabe ist die Extraktion von Merkmalen, wie beispielsweise Wörtern, Silben und Lauten. Auf Basis dieser Merkmale werden Wörter gebildet (Spracherkennung), die im Lexikon vorhanden sind und im nächsten Schritt, der syntaktischen Analyse, werden diese in eine grammatikalisch richtige und sinnvolle Reihenfolge gebracht. Die zusammengesetzte Wortreihe wird anschließend auf ihrer Semantik und Pragmatik überprüft. In den weiteren Schritten, der Dialogsteuerung, Antwortgenerierung und Sprachsynthese, erarbeitet das System ein passendes sinngemäßes Feedback auf die Ansage des Benutzers. (vgl. [6], S. 214). Dialogsteuerung spielt hier eine große Rolle in dem gesamten Dialogsystem, denn seine Aufgabe besteht in der Kontrolle des Dialogflusses und involviert folgendes:

- Bestimmung der Dialogart, des Dialogablaufes und des Dialogzweckes
- Verzeichnet und aktualisiert den bisherigen Verlauf des Dialoges
- Kommunikation mit dem jeweiligen Domänenserver um die Fragen der Benutzer weiter zu leiten und entsprechend die geeignete Antwort abzurufen
- Planung des nächsten Dialogschritts ( vgl. [7] S. 123)

Die Dialogsysteme zwischen Anwender und System können auf verschiedene Arten gestaltet werden und basieren im Allgemeinen auf dem Frage-/Antwort System. Im Hinblick auf die Dialogsteuerungsart, der Benutzerfreundlichkeit und die mögliche Freiheit der Benutzereingaben könnten die Dialogsysteme in folgende Arten unterteilt werden:

- **geführte Dialoge:** sind Dialoge bei denen das System den Ablauf des Dialogs vollständig kontrolliert. Schrittweise wird der Anwender durch eine Folge von Fragen geführt. Nachteil solcher Systeme ist, dass das Erkennen nur für einzelne Wörter bzw. Wortketten betrieben werden kann und dass der Benutzer keine Freiheit im Dialog hat. Beispiel dafür ist folgender Dialog

**System:** *Guten Tag. Hier ist das Sprachsystem deines Smartphones.*

*Was kann ich für Sie tun? Um jemanden anzurufen, sagen Sie Anrufen. Um Nachrichten zu verschicken, sagen Sie Nachrichten.*

**Benutzer:** *Anrufen.*

**System:** *Sagen Sie bitte die Telefonnummer oder nennen Sie aus Ihrer Kontaktadresse einen Namen.*

*Benutzer: Sebastian.*

*System: Welche Telefonnummer von Sebastian soll ich anrufen? Handy Privat oder Arbeit?*

*Benutzer: Handy.*

*System: Die Handynummer von Sebastian wird angerufen.*

- **gemischte initiative Dialoge:** Hier wird beim Dialog die Eingabe in ganzen Sätzen erlaubt. Damit hat der Benutzer mehr Freiheit im Dialog und kann selbst den Ablauf des Dialogs mitbestimmen. Beispiel dafür ist u.a. ein Info-System für die Universität oder FH, Auskunft-Systeme für Bahn oder Flugverbindungen. Als Beispiel Dafür ist folgende Dialog angeführt:

*System: Guten Tag. Was kann ich für Sie tun?*

*Benutzer: Hallo, ich möchte meinen Freund Sebastian anrufen.*

*System: Welchen Sebastian meinen Sie?*

*Benutzer: Sebastian Müller.*

*System: Welche Telefonnummer von Sebastian Müller soll ich anrufen?*

*Benutzer: Die Handynummer.*

*System: Die Handynummer von Sebastian Müller wird angerufen.*

- **Freie Dialogsysteme:** Zurzeit sind Dialoge, bei denen sich das Dialogziel nicht in einfacher Weise strukturieren lässt und damit den Nutzer ohne Einschränkungen frei sprechen kann, noch nicht 100% realisierbar. (Vgl. [8] S.155). Das obige Dialogbeispiel könnte zum Beispiel so ablaufen

*System: Guten Tag. Was kann ich für Sie tun?*

*Benutzer: Hallo. Kannst du die Handynummer von Sebastian, der in Schiller Straße wohnt anrufen?*

*System: Ok, die Handynummer von Sebastian Müller wird angerufen.*

Dialogsysteme sollen allgemein in der Lage sein, die Absichten der Benutzer zu verstehen, deshalb erfordert ein funktionsfähiges Dialogsystem z.B. eine Buchführung über die Dialoggeschichte, Wissen über die Struktur von Dialogen und ebenso Wissen über den Benutzer. Um die Funktionalität eines Dialogsystems zu erlangen, basieren die Dialoge wie am Anfang erwähnt auf einem Frage-/Antwort System und ermöglichen dem Benutzer im Idealfall die kooperative Lösung von Problemen. Kooperativität heißt einerseits das Erreichen eines gemeinsamen Zieles und andererseits, dass der Dialog von beiden Seiten verstanden werden könnte. Und dies erfordert die Bestimmung der Dialogsteuerung (siehe Abbildung 9).

Anhand des obigen Dialogbeispiels für gemischter initiative Dialoge könnte die Aufgaben der Dialogsteuerung verfolgt und verstanden werden. Nach der Abfrage des Systems, gibt der Benutzer Informationen ein und teilte dem System mit, welche Dienste er nutzen möchte und gleichzeitig

spezifiziert er bereits Detailinformation. Das System erkennt den Wunsch des Benutzers und stellt fest, dass die Angaben noch nicht ausreichend sind und fragt daraufhin gezielt nach den fehlenden Angaben (Nachname oder Adresse). Da die gewünschte Kontaktadresse mehrere Nummern hat, fragte das System noch einmal nach der gewünschte Telefonnummer. Diese Dialogstrategie, das Frage-/Antwort System, verfolgt in der Dialogtheorie das Ziel, die benötigten Informationen für eine Abfrage in der der Gesamtanwendung zugrunde liegenden Datenbank zu sammeln. Solange benötigte Felder nicht gefüllt sind, wird das System durch gezielte Fragen versuchen, diese Lücken zu schließen.

## 2.2 Siri

In diesem Abschnitt werden am Anfang die Vorgeschichte und die Entwicklungsprozesse von Siri erwähnt, wie auch die Gründe für die Entwicklung solcher Systeme, die Semantisch interagieren können. Danach wird die Mensch-System-Interaktion bei Siri erklärt und die Unterschiede zwischen akustischer und motorischer Kommunikation aufgezeigt. In den nächsten zwei Punkten wird Siri in zwei verschiedenen Perspektiven betrachtet. Als erstes das Siri-System und sein Dialogverarbeitungssystem, anschließend Siri als Semantisches System. Am Ende wird Siri als Gesamtsystem betrachtet und, als Vorbereitung für die Datenmodulation im darauffolgenden Abschnitt, ihre Kommunikation mit der äußeren Welt dargestellt.

### 2.2.1 Vorgeschichte

Der Grundstein von Siri wurde Anfang 2000 gelegt, als der Defense Advanced Research Projects Agency (DARPA) *SRI International* fragte, inwiefern die Möglichkeit zur Erstellung eines lernfähigen Personal Assistent, besteht. Dieser sollte für die militärische Nutzung angewendet werden. In einem Artikel vom 14.06.2010 erzählen William Mark und Norman Winarsky<sup>1</sup> die Geschichte von Siri's Entwicklung (siehe [9])<sup>2</sup>. Die Idee eines solchen lernfähigen System war, wie William Mark das ausgedrückt hat, der Schlüssel „Es ist eine enorme Belastung, das gesamte Wissen was wir tatsächlich brauchen in einem System zu bekommen und dabei es auf dem Laufenden zu halten, es sei denn es gibt ein Art des Lernens“ sagte er (Quelle:<http://www.xconomy.com/san-francisco/2010/06/14/the-story-of-siri-from-birth-at-sri-to-acquisition-by-apple-virtual-personal-assistants-go-mobile/2/>). Aus diesem Grund startete SRI ein Projekt für DARPA, um verschiedene Arten von Computer-Lernfähigkeiten zu erkunden. Dieses Projekt hieß CALO (Cognitive Assistant that Learns and Organizes). Etwa zur gleichen Zeit, aber getrennt von dem CALO-Projekt, starteten Winarsky und Mark eine interne SRI-Studie, die sie mit dem Codenamen Vanguard benannten. Dieses Projekt war ein Versuch die rätselhafte Wirtschaftlichkeit der Mobilfunk-Industrie zu verstehen und zu verbessern. „Dieses Projekt ging aufgrund der niedrigen Einnahmen aus den Sprachdiensten exponentiell runter“ sagt Mark. „Das Problem war, dass die Einnahmen aus Datendiensten tatsächlich nicht auftraten oder sie wurden nicht annähernd genug erhöht, um die Kosten zu kompensieren“ [9]. Schließlich kooperierten Vanguard und CALO, um das unglaubliche Bedürfnis, eine bessere Kommunikation der Dienstleistungen in der mobilen Welt zu erreichen, in die Hand zu nehmen. So wurde das CALO zu einem Konzept eines Office-Assistenten. Es wurde zur Unterstützung in der mobilen Welt und schuf die Basis Siri-Konzept. Die Algorithmen die dieses Projekt

---

<sup>1</sup>Norman Winarsky und William Mark waren die verantwortlichen für CALO-Projekt in SRI International

<sup>2</sup> Diese Informationen ergeben sich aus der Quelle [9]: <http://www.xconomy.com/san-francisco/2010/06/14/the-story-of-siri-from-birth-at-sri-to-acquisition-by-apple-virtual-personal-assistants-go-mobile>

funktionsfähig machen, sind das Produkt von jahrelanger Forschung in SRI International und anderen Instituten. Das Siri-Team hatte immer gewusst, dass Smartphones die richtige Plattform für einen virtuellen persönlichen Assistenten (VPA) darstellen, deswegen war das Debüt von iPhone 2007 und sein iTunes App Store eine ideale Umgebung für diese Software. Das iPhone 3GS, das im Juni 2009 kam, war die erste Version des iPhone, die sowohl die interne Rechenleistung als auch die drahtlose Bandbreite des Siri-Teams benötigt. Im April 2010 wurde *Siri inc.* von Apple aufgekauft“ (vgl. [9]). Apple hat das ganze Programm nochmal bereinigt, überarbeitet und in die iPhone 4S Software integriert [10].

## 2.2.2 Mensch-Maschine-Interaktion

Die Interaktion zwischen System und Nutzer, und der wechselseitige Datenaustausch zwischen den beiden Kommunikationsteilnehmern ist ein Teil der Mensch-Maschine-Interaktion. Ziel solcher Interaktionen ist, das System möglichst benutzerfreundlich zu gestalten und an den Nutzer anzupassen.

Siri ist eine Software für Smartphones und die Kombination dieser Software mit Smartphone-Geräten sollte deren Bedienung möglichst vereinfachen. Da Siri seit April 2010 ein Produkt der Firma Apple ist, wurde Siri seitdem auf der vierten Generation von iPhone (ab iPhone 4s) und danach von Nachfolgemodellen, integriert. Grundsätzlich ist beim Smartphone von Apple das Touchscreen die alleinige Schnittstelle zwischen dem Nutzer und dem Gerät. Obwohl das Touchscreen mittlerweile weiter entwickelt wurde und sehr schnell arbeitet ( kurze Reaktionszeit ), sind dennoch für gewöhnlich die Tasten deutlich kleiner als die meisten Finger, was häufig zu Fehlern bei der Eingabe führt. Siri soll diese Art Interaktion ersparen und ein schnelles und bequemes Arbeiten ermöglichen. Die Benutzerschnittstelle soll möglichst ergonomisch an der Benutzer angepasst werden und eine akustische Kommunikation (Sprachsteuerung) als Schnittstelle zwischen den Nutzern und ihrer Smartphone Geräte ist zumindest, ergonomisch gesehen, vorteilhafter als eine motorische Kommunikation (in diesen Fall Tippen auf den Touchscreen). Das ist aber nur der Fall, wenn die sprachlichen Interaktionen einfacher und ohne bestimmte Anweisungen von beiden Seiten verstanden werden und das ist was Apple mit Siri auf dem Markt verkauft. Damit Siri als eine benutzerfreundliche Schnittstelle betrachtet wird und ergonomisch an dem Nutzer angepasst ist, muss sie die sieben Usability-Grundsätze erfüllen(vgl. [11]):

- 1- **Angemessenheit:** *„ein Dialog ist aufgabenangemessen, wenn er den Benutzer unterstützt, seine Arbeitsaufgabe effektiv und effizient zu erledigen.“ (Balzert, 2004, S.71)*
- 2- **Selbstbeschreibungsfähigkeit:** *„Ein Dialog ist selbstbeschreibungsfähig, wenn jeder einzelne Dialogschritt durch Rückmeldung des Dialogsystems unmittelbar verständlich ist oder dem Nutzer auf Anfrage erklärt wird.“ (Balzert, 2004, S.71)*
- 3- **Erwartungskonformität:** *„Ein Dialog ist erwartungskonform, wenn er konsistent ist und den Merkmalen des Benutzers entspricht, zum Beispiel seinen Kenntnissen aus dem Arbeitsgebiet und seinen Erfahrungen sowie den allgemein anerkannten Konventionen.“ (Balzert, 2004, S. 72)*
- 4- **Fehlertoleranz:** *„Ein Dialog ist fehlertolerant, wenn das beabsichtigte Arbeitsergebnis trotz erkennbarer Fehleingaben entweder mit keinem oder mit minimalem Korrekturaufwand seitens des Benutzers erreicht werden kann.“ (Balzert, 2004, S.72)*
- 5- **Steuerbarkeit:** *„Ein Dialog ist steuerbar, wenn der Benutzer in der Lage ist, den Dialogablauf zu starten sowie seine Richtung und Geschwindigkeit zu beeinflussen, bis das Ziel erreicht ist.“ (Balzert, 2004, S.71)*



**6- Individualisierbarkeit:** „Ein Dialog ist individualisierbar, wenn das Dialogsystem Anpassungen an die Erfordernisse der Arbeitsaufgabe sowie an die individuellen Fähigkeiten und Vorlieben des Benutzers zulässt.“ (Balzert, 2004, S.72)

**7- Lernförderlichkeit:** „Ein Dialog ist lernförderlich, wenn er den Benutzer beim Erlernen des Dialogsystems unterstützt und anleitet.“ (Balzert, 2004, S.73) [12].

### 2.2.3 Dialogverarbeitungssystem

Das Dialogsystem bei Siri ist eine Schnittstelle, bei der der Nutzer in gesprochener Sprache mit dem Smartphone interagiert. Dieses Dialogsystem wurde von den Entwicklern von Siri als geführtes Dialogsystem eingestuft, das bedeutet dass die Dialogabläufe vom System geführt und kontrolliert werden.

Siri soll laut ihren Entwicklern ein Virtual personal Assistent „VPA“ sein, der auf Basis eines Sprachverarbeitungssystem basiert. VPAs können dank semantischer Technologien und Kontextsensibilität die Absicht der Nutzer verstehen und die Aufgaben durch Internetdienste ausführen. Die Delegation von Aufgaben an Siri erfolgt über einen geführten Dialog, Nutzer geben eine gesprochene Anweisungen ein, die gesprochenen Anweisungen werden mittels eines kontinuierlichen Spracherkennungssystem ausgewertet. Das Sprachverstehen wird durch Siri's Aufgaben- und Domäne-Modelle ermöglicht. Ein geführter Dialog soll Siri ermöglichen, die Absicht der Nutzer zu erkennen. Praktisch bedeutet das, dass die Eingabe der Nutzer einer Domäne bzw. Aufgabe zugeordnet wird. Bei Mehrdeutigkeit bietet Siri dem Nutzer unterschiedliche Interpretationen an, bis sie die Absicht der Nutzer versteht [13]. Der Dialogablauf wird nur dann ausgeführt, wenn SIRI über eine Internetverbindung verfügt. Da Siri laut der Entwickler([14]S.48) zur Spracherkennung und Sprachsynthese auf ein Domänen-Server und andere Dienste zugreifen muss, benötigen sie eine gute, stabile 3G- oder 4G-Funk- oder WLAN-Verbindung zum Internet. Ein Dialogablauf eines Siri Dialogsystems könnte eventuell in 4 Ebenen unterteilt.

- I. Nutzer Ebene und Sein Umgebung
- II. Akustische Schnittstelle (das Smartphone)
- III. Domänen Server (Apple Server)
- IV. Internetdienste.

Das folgende Schema stellt die Ebenen dar (Abbildung 10):

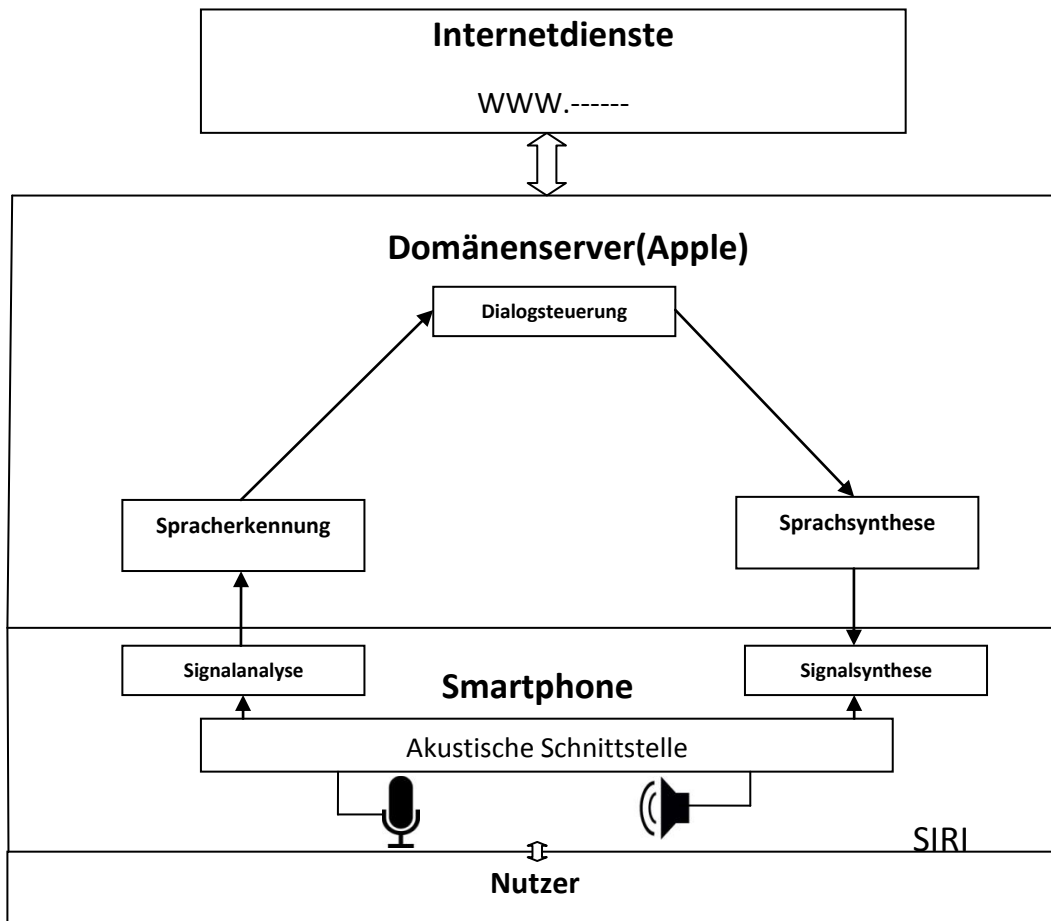


Abbildung 10: Dialogsystem Siri

## 2.2.4 System Semantic Siri

Wie in 2.2.1 erwähnt wurde, war Siri ein Ergebnis langjähriger Forschungsarbeit der DARPA (<http://www.darpa.mil/>) über Artificial Intelligence Center der SRI International (<http://www.ai.sri.com/>) durch CALO, eines von zwei Projekten, die für DARPA entwickelt wurde. Aber welche Semantik verbirgt sich hinter diesem System?

Im Grunde ist das CALO-Projekt ein Teil des *Person Perceptive Assistant that Learns* (PAL) Programmes der DARPA und hat das Ziel ein kognitives Software-System zu entwickeln, welches in der Lage ist:

- Dialoge und natürliche Sprachen zu verstehen
- Beweise und Wahrscheinlichkeiten, Argumentationen zu verstehen
- Ontologie und Wissensrepräsentation zu verstehen
- Aus Erfahrungen zu lernen
- Sagen was zu tun ist
- Erklären was sie tut und
- Planen, delegieren und überlegen.

Aus all diesen Anforderungen bestand CALO's Auftrag, dem Benutzer als persönlicher Assistent zu dienen und die Zusammenarbeit in allen Bereichen des Arbeitslebens zu erleichtern, z.B.: Organisieren von Informationen, die Vorbereitung Artefakte für Informationen; Person-Person-Wechselwirkungen Vermittlung; Organisation und Planung in der Zeit, die Überwachung und die Verwaltung von Aufgaben und den Erwerb und die Zuteilung von Ressourcen. Um all diese Anforderungen zu realisieren und ein semantisches System zu entwickeln, sollte diese Semantik eine Voraussetzungen erfüllen, damit es für CALO's Projekt, und später auch für Siri, geeignet ist. Ein paar dieser Voraussetzungen sind zum Beispiel:

- Das System soll in der Lage sein, einfach mit *Java* zu interagieren. Diese Forderung ergibt sich aus der Tatsache, dass viele der Maschinen Learning-Komponenten für CALO's Projekt in *Java* implementiert haben.
- Das Semantische System sollte ein Ontologie-Knowledge-Basis(Ontologie- KB) haben. Das heißt, dieses System besitzt die Fähigkeit, reichhaltige Semantik Strukturen zu modellieren, die jeden Aspekt eines Benutzers Arbeitsumgebung erfassen kann. Und um diese KB zu kodieren sollte diese semantische Repräsentation in der Lage sein mit einer so genannten Knowledge Machine (KM) zu interagieren<sup>3</sup>.
- Das System sollte auch in der Lage sein die persönlichen Wissensvermögenswerte zu unterstützen. Das bedeutet, die Fähigkeit des Nutzers ihre eigenen Informationsquellen so zu organisieren, dass sie zu ihren individuellen Bedürfnissen passt, unter Beibehaltung der Kompatibilität mit anderen semantische CALO's Einrichtungen.

Diese Forderungen und Voraussetzungen werden zusammen die Semantik für CALO und auch Siri gründen. In Grunde ist das System Semantik in 3 Ebenen gestuft, diese Ebenen sind:

- ONTOLOGIE EBENE
- KB-EBENE oder DATA EBENE
- USER INTERFACE EBENE mit

## 1. Ontologie-Ebene

Ontologie (oft auch als Domain Model bezeichnet) kann eine entscheidende Rolle bei der Ermöglichung der Verarbeitung und Weitergabe von Wissen (Knowledge) zwischen den Programmen im Web zu spielen. Ontologien sind in der Regel als "Darstellung einer gemeinsamen Konzeption einer bestimmten Domäne" definiert ([25], S. 2). Sie bieten ein gemeinsames Verständnis einer Domäne, die über Menschen und Anwendungssysteme kommuniziert. Sie sind für Artificial Intelligence (AI) entwickelt worden, um den Wissensaustausch zu erleichtern und wiederverwenden, und sind in Prinzip die Grundlage für die Erstellung intelligenter Anwendungen. Bei dem CALO Projekt wurde von einer so genannten *Active Ontologie* gesprochen und definiert das als formale Vertretung für Domäne Wissen (Knowledge) mit verschiedenen *Klassen*, *Attributen* und *Beziehungen* zwischen den Klassen. Aktive Ontologie ist ein formaler Prozess, bei dem unterschiedliche Verarbeitungselemente nach Ontologie Vorstellung angewendet werden. Die Active Ontologie besteht aus relationalen Netzen von Konzepten, diese Konzepte dienen für die Identifikation sowohl der Datenstrukturen in der Domäne (Z.B. eine Nachricht hat einen Sender und Empfänger, einen

---

<sup>3</sup> KM ist eine Wissensrepräsentation und ein Schlussfolgerungssystem mit klarer Logik der ersten Ordnungssemantik und dieses System ist der Gehirnmeister der CALO verwendet. (Quelle [16])

Inhaltstext und eine Sendeart) als auch die damit verbundenen Regelsätze, die der Aktionen innerhalb und zwischen den Konzepten dienen [25]. Diese genannte Ontologie ist auch eine Art Datenmodellierung die wir in Kapitel( 2.3.1) diskutiert werden soll.

## 2. KB-Ebene oder Data-Ebene

Eine Wissensbasis (KB) bietet ein einheitliches Datenmodell, Persistenz Speicher und Abfrage-Mechanismen über die Informationsressourcen und semantischen Beziehungen zwischen ihnen.

## 3. USER INTERFACE EBENE

Die Siri-Benutzeroberfläche ermöglicht Plug-in-Anwendungen, um ihre eigenen Schnittstellen innerhalb Siri einzubetten und mit globalen Internetdiensten zusammenzuarbeiten<sup>4</sup>.

### 2.2.5 Gesamtsystem Siri

In dem oberen Teil wurde Siri sowohl als Dialogverarbeitungssystem als auch als ein Semantik-System betrachtet. Nunmehr wird das Gesamtsystem von Siri beschrieben, das als gesamte Einheit in der Lage ist mit der Außenwelt zu agieren. Im Juni 2009 hatte Tom Gruber (CTO & cofounder,Siri) Siri *in Semantic Technologie Conference* als VPA, als Anwendung für das Smartphone von Apple, präsentiert [13]. Dieser Assistent soll, laut Gruber, mit Hilfe von Apple Smartphone folgendes können:

- „Does things for You“: Hilft uns Dinge zu tun durch zahlreiche Aufgaben im Smartphone und die Verwendung von vielen Online und Offline Diensten.
- „Gets what you say“ Versteht unsere Absichten über Konversation.
- „Gets to know You“ Lernt und verwendet unsere persönlichen Informationen.

Mit dieser Vorstellung von Siri, und der Implementierung mit dem iPhone Gerät konnte das Gesamtsystem und seine Umgebung dargestellt werden (Abbildung11).

In Grunde soll das Gesamtsystem eine sprachliche Kommunikation zwischen dem Siri-System und dem Nutzer, der dieses System bedient, beschreiben. In diesem Sinn wird das Gesamtsystem in zwei Ebenen betrachtet: der Nutzer-Ebene in seiner Umgebung und der System-Ebene. Da diese Arbeit hauptsächlich die Prozesse der semantischen Datenmodellierung behandelt, wird davon ausgegangen, dass bei der Kommunikation zwischen den beiden Ebenen keine Spracherkennungsprobleme existieren. Die Nutzer-Ebene beschreibt die Elemente und alle Einflüsse mit denen der Nutzer das System von außen steuert. Die System-Ebene beinhaltet alle Datenbanken des Systems, die Offline- und Online-Dienste, so wie ein Funktionsmodell, mit dem das Gerät die Wünsche des Nutzers erfüllen kann.

---

<sup>4</sup> Mehr dazu siehe auch Literatur [15] und [16]

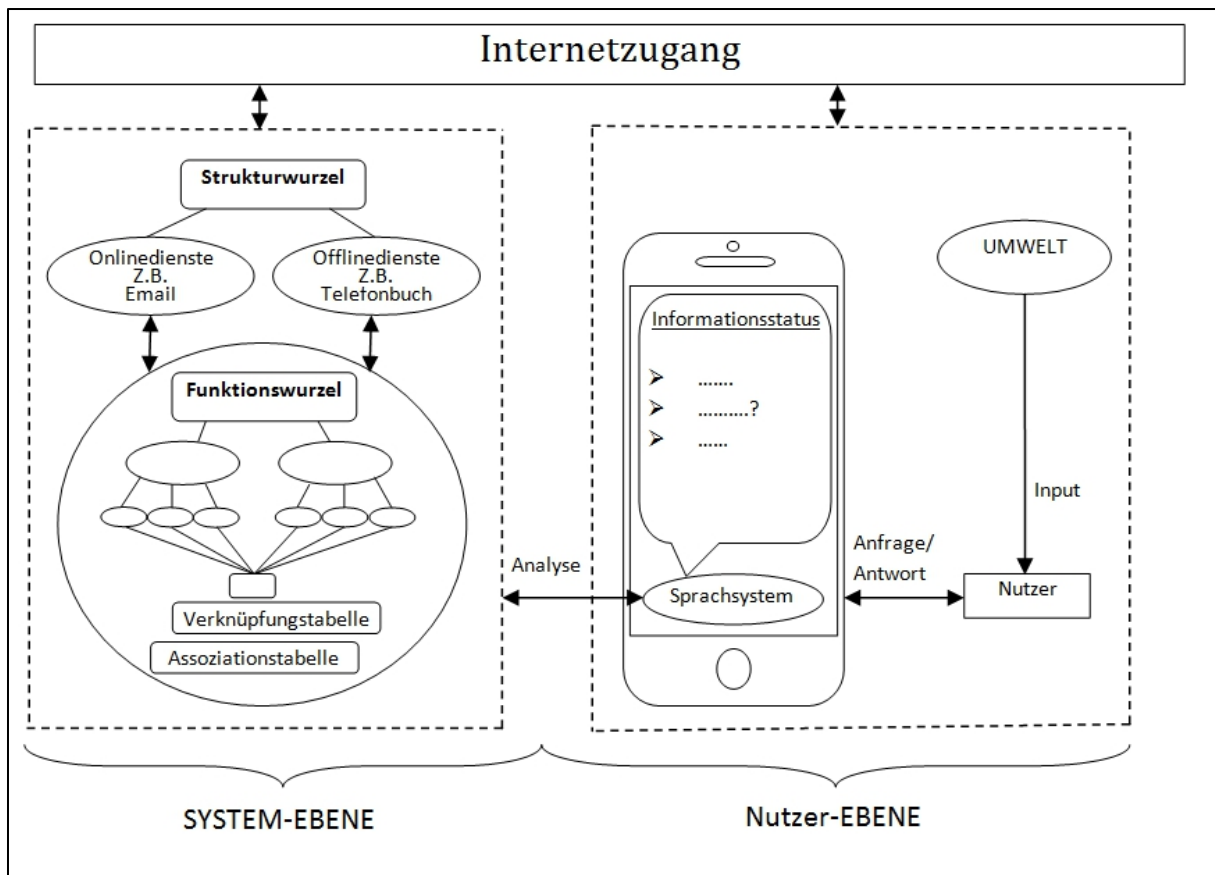


Abbildung 11: Siri als Gesamtsystem

- **Nutzer Ebene**

In der Nutzer-Ebene werden die Nutzer-Daten und -Informationen mit dem System ausgetauscht. Die Angaben des Nutzers werden mit Hilfe des Smartphone-Gerätes per Sprache eingegeben, diese Eingabe könnte eine Anregung oder eine Frage sein und wird von vielen Komponenten und der Weltumgebung beeinflusst. Zum Beispiel vom sozialen Umfeld oder von den Bedienungshilfen des Geräts. Die Information die vom System kommt, kann gleich eine Antwort sein oder auch eine Rückfrage und ist immer von der Nutzereingabe abhängig. Für einen erfolgreichen Datenaustausch zwischen dem Nutzer und dem System ist ein Sprachsystem für die Bearbeitung erforderlich, gleichzeitig muss der Nutzer über eine Internetverbindung verfügen. Bei einem Datenaustausch werden Nutzerinformationen gesammelt und diese Informationen werden mithilfe einer dynamischen Modellierung gespeichert und als Informationsstatus betrachtet (Kapital 2.3.2). Dieser Informationsstatus wird erst nach Erfüllung des Nutzerwunsches (Ende der Aufgabe) auf null gesetzt und bei der nächsten Aufgabe oder Nutzeraktion baut das System einen neuen Informationsstatus auf.

- **System Ebene**

Die System-Ebene wird in zwei Hauptkomponente unterteilt: Eine Funktions- und eine Strukturwurzel (Kapital 2.3.1.2). Die Funktionswurzel ist eine Art Arbeitsspeicher, welcher die gefundene Objekt- und Aktions-ID verwaltet. Diese werden in der Assoziationstabelle gespeichert und mit der Verknüpfungstabelle abgeglichen (siehe Abschnitt 2.3.1.3). Die Strukturwurzel beinhaltet Online- und Offline-Dienste auf denen die Objekte verwaltet werden.

Mittels eines uneingeschränkten Zugriffs des Arbeitsspeichers auf die beiden Dienste kann das System stets die Datenabgleiche vornehmen. Die Identifikation von Objekt- und Aktionsmodell wird statisch modelliert (siehe Kapitel 2.3.1) und die identifizierte Komponente, A-ID und O-ID, wird durch die Berechtigungstabelle entweder miteinander verknüpft oder in einer Assoziationstabelle gesammelt [1].

Dieses Gesamtsystem soll Grubers Vorstellungen erfüllen. Siri könnte dem Nutzer bei vielen Aufgaben helfen, z.B.: Jemanden Anrufen, Emails Aufrufen oder den Weg zur Arbeit Zeigen. Anhand eines Sprachsystems soll Siri mit dem Nutzer sprachlich kommunizieren und seine Absichten verstehen können. Durch die persönlichen gespeicherten Informationen hat Siri die Möglichkeit, diese zu lernen und zu verwenden.

Auf die Modellierung solcher Funktionen in diesem Gesamtsystem wird in dem nächsten Abschnitt eingegangen.

## 2.3 semantische Modellierung

Ein Modell ist laut der Definition des Dudens „ein vereinfachtes Abbild der Realität von einem Objekt-/Objektbereich mit Betonung seiner speziellen Eigenschaften“ (Schülerduden Informatik)<sup>5</sup>. In der Informatik bezeichnet man die Modellierung als Vorgang eines vereinfachten Abbildes eines realen Systems um es besser zu verstehen und deren Aufgaben und Lösungsvorschläge darstellen zu können, bevor mit der Herstellung einer Lösung in der Software bzw. Hardware begonnen wird ([17], S.16). Durch die Modellierung kann das System anhand der Entwicklungsphase überprüft werden, ob die Aufgabenstellung richtig verstanden wurde, welche Eigenschaften die vorgeschlagenen Lösungen haben und ob diese akzeptabel sind. Eine Modellierung, die sich mit dem Verständnis der Bedeutung eines realen Modelles beschäftigt und spezialisiert, ist eine *semantische Modellierung*. Diese semantische Datenmodellierung ist eine reine formale Beschreibung und Darstellung eines Modells in einem bestimmten Zusammenhang der realen Welt und soll in frühen, noch nicht IT-technisch bestimmten Projektphasen, alle fachlich relevanten Aspekte um das Thema 'Daten', wie z.B wie wird's gemacht, was brauche ich dafür und womit mache ich das, möglichst genau und vollständig sowie einheitlich strukturiert aufzunehmen und darzustellen. Für das Modellieren eines Systems sollte zwischen statischer und dynamischer Modellierung unterteilt werden.

### 2.3.1 Statische Modellierung

Eine statische Modellierung ist ein strukturelles Modell, das sich, unabhängig von der Zeit, mit den Eigenschaften von Objekten und den dazwischen entstehenden Beziehungen beschäftigt. In dieser Arbeit wird solche Modellierung benötigt, um eine Kommunikation zwischen einem System und einem Nutzer, der das System bedient, herzustellen. Um eine erfolgreiche Kommunikation zwischen den beiden Hauptkomponenten sicherstellen zu können, benötigt das System eine Aktionsidentifikation (A\_ID) und eine Objektidentifikation (O\_ID), und diese beiden IDs werden von dem System erkannt nur wenn eine Mensch-Maschine-Interaktion stattgefunden hat. Laut [1] konnte der Zusammenhang zwischen den beiden Komponenten wie folgt dargestellt werden

---

<sup>5</sup> "Schülerduden Informatik", Dudenverlag, Mannheim Leipzig Wien Zürich, 4. aktualisierte Auflage, 2003

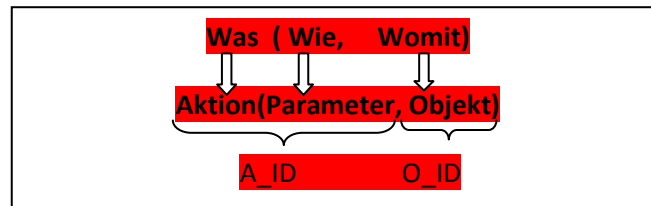


Abbildung 12: Struktur- und Funktionskomponente und ihre Zusammengehörigkeit([1], S.12)

Mit dem oben gezeigten Zusammenhang, setzt sich die Aktionsidentifikation A\_ID aus den Fragen „Was soll mit einem Objekt passieren?“, und „Wie soll die Aktion durchgeführt werden?“ zusammen. Für die Objektidentifikation O\_ID stellte sich die Frage, womit eine Aktion durchgeführt werden soll.

Bevor die Aktions- und die Objektidentifikation näher betrachtet werden können, muss zuerst die Basis-Datenbank für das System entworfen und dargestellt werden. Anhand einer Datenbankmodellierung werden die Objekte eines Systems und die Eigenschaften und Beziehungen zwischen ihnen bestimmt. Bestandteil dieser Arbeit ist eine bestimmte Art der strukturellen Modellierung: das ER-Modell (*Entity-Relationship-Modell*) und dessen graphisches ER-Diagramm.

### 2.3.1.1 Entity-Relationship-Modell

ER-Modell beschreibt einen Ausschnitt der realen Welt in Rahmen der semantischen Datenmodellierung [20]<sup>6</sup>. Es wurde im Jahr 1976 von Peter Chen vorgeschlagen. U.Kastens und H.K.Büningin definieren in ihrem Buch *Modellierung Grundlagen und formale Methoden* [17] das ER-Modell „ist ein formales Kalkül zur Modellierung von Themenbereichen mit ihren Objekten, deren Eigenschaften und Beziehungen zwischen ihnen“([17], S. 202). ER-Modell gilt bis heute als Standardmodell für die frühe Entwicklungsphase in Datenbankentwicklung und kann in vielen Bereichen eingesetzt werden. Es kann entweder grafisch (ER-Diagramme) oder als textuelle Notation dargestellt werden. Da die grafische Notation besonders anschaulich ist, wird in dieser Arbeit nur mit ER-Diagrammen gearbeitet. Das ER-Modell basiert auf drei Grundbegriffen:

- Entity: Ist ein Objekt oder Ding (Entity) der realen oder virtuellen Welt, für das Informationen zu speichern sind. Jedes Entity ist von einem bestimmten Entity-Typ.
- Attribut: Eine Eigenschaft von Objekten bzw. Beziehungen beschrieben durch einen Wert.
- Relation: Eine Beziehung zwischen Objekten.

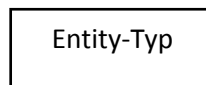
Für diese drei Begriffe sind Datentypen erforderlich, was zur Definition von Wertbereichen dient. Ein Datentyp ist gekennzeichnet durch einen Wert, den ein Objekt annehmen kann und durch den Zustand, in dem dieser Wert angenommen wird.

#### **Entity-Typ:**

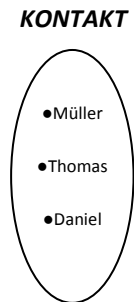
Die Entity-Menge deklariert eine Menge von Objekten, mit bestimmten Eigenschaften die als gleichartig angesehen werden und repräsentiert werden sollen. Sie wird deshalb durch die gleichen Attribute charakterisiert und kann an den gleichen Relationen beteiligen. Diese Menge wird meist nicht explizit festgestellt sondern als genügend groß. Entity-Typen müssen nicht disjunkt sein, d.h. im

<sup>6</sup> Quelle: <https://de.wikipedia.org/wiki/Entity-Relationship-Modell>

Allgemein kann ein Entity Element mehrerer Entity-Typen sein. Graphisch wird eine Entity-Menge durch ein Rechteck mit einem Namen darin angegeben



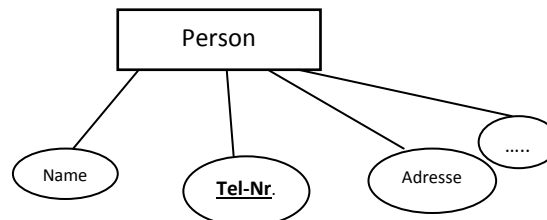
In der konkreten Ausprägung wird die Entity-Menge als Ellipse und die Entities als Punkte mit identifizierten Namen angegeben. Hat zum Beispiel eine Entity-Menge bei einem Telefon den Namen KONTAKT, wird die Ellipse wie folgt dargestellt: "(vgl. [17], S. 204):



Die Entity-Menge KONTAKT aus unserem Beispiel steht also für die Menge aller Personen die man haben möchte. Objekte im ER-Modell (auch *Entities* genannt) repräsentieren Gegenstände des modellierten Themenbereiches. Jedes Objekt hat eine eindeutige Identität, die es von allen anderen Objekten unterscheidet. Zwei Objekte sind also immer verschieden, auch wenn alle ihre Eigenschaften übereinstimmen.

### Attribute und Attributentypen

„Ein Attribut definiert die Eigenschaft für alle Instanzen eines Entity-Typs oder Beziehungstyps. Es kann deshalb als Funktion verstanden werden, die jeder Entity aus der konkreten Ausprägung einer Entity-Menge einen Wert zuordnet“ ([17], S.205). Graphisch werden Attribute durch Ellipsen notiert, die mit dem Rechteck ihrer Entity-Menge verbunden sind und den Attributnamen als Inschrift enthalten.



Der Attributtyp bezeichnet den Typ der Ausprägung der Eigenschaft und ist üblicherweise ein Standard-Datentyp, wie z.B. Integer für die Attribut *Tel-Nr* oder String für das Attribut *Name*. Es kann sein dass manchmal zwei oder mehrere Entities denselben Attributwert haben aber wird denn bestimmte Attributwerte geben der die Entity von den anderen identifizieren und das wird Schlüsselattribute genannt. U Kastens & H.K Büning haben das Schlüsselattribut wie folgt definiert:

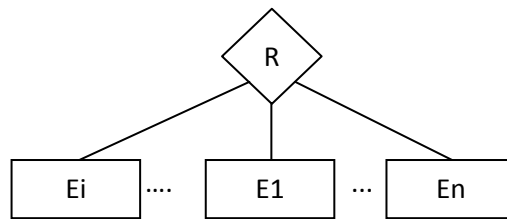
„Ein Attribut, dessen Wert jede Entity einer konkreten Ausprägung einer Entity-Menge eindeutig identifiziert, heißt Schlüsselattribut. Sein Name wird im Modell durch Unterstreichen hervorgehoben. Auch die Werte mehrerer Attribute können zusammen die eindeutige Unterscheidung leisten. Sie werden dann alle als Schlüsselattribute gekennzeichnet“ ([17] S.206).

In dem Beispiel oben ist das Attribut "Tel-Nr." das Schlüsselattribut für die Entity Kontaktperson.

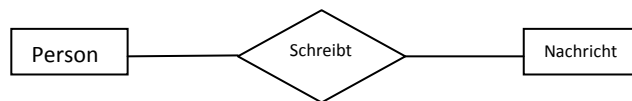
### Relation-Type

Ein Beziehungstyp deklariert eine Beziehung zwischen Entity-Typen. Es kann eine beliebige Anzahl  $n \geq 2$  von Entity-Typen an einem Beziehungstyp teilhaben. Sie werden graphisch repräsentiert durch eine Raute, die mit den  $n$  Entity-Typen verbunden ist.



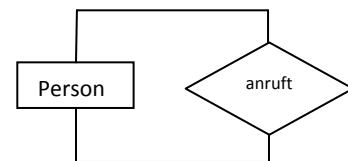


Beispiel dafür ist der zweistellige Relations-Typ, die zwischen den Entity-Typ Handy und Person gehört:



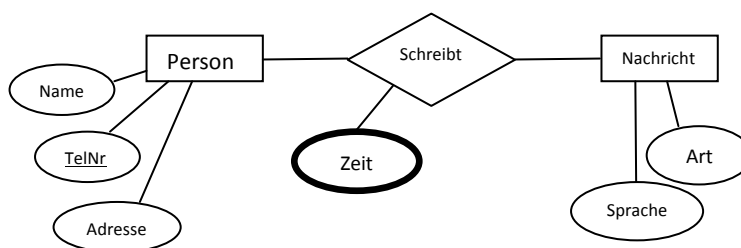
Häufig kann man aus den Namen der Relation und der beteiligten Entity-Mengen erkennen, welche Rollen die Entities jeweils in der Relation spielen: Zum Beispiel ist in der Relation gehört in klar, dass ein Person eine Nachricht schreibt und nicht umgekehrt.

Sind einzelne Entity-Typen mehrfach in einen Beziehungstyp eingebunden, legen Rollenbezeichner die Rollen der beteiligten Objekte in der Beziehung fest. Beispiel :



Wie das obige Beispiel zeigt, kann ein Entity-Typ mit sich selbst in Beziehung gesetzt werden.

In den Beziehungen können auch Attribute zugeordnet werden. Diese Attribute beschreiben die Eigenschaften, die Beziehungen und ihre Werte, sind nur mit der Beziehung gebunden und unabhängig von den jeweiligen verbundenen Entity-Typen. Ein Beispiel dafür ist folgendes:



Das Attribut Zeit in dem Beispiel ist eine Eigenschaft von Schreiben und erklärt wann das Schreiben geschah, diese Eigenschaft ist aber mit der Entities *Nachricht* oder *Person* unabhängig.

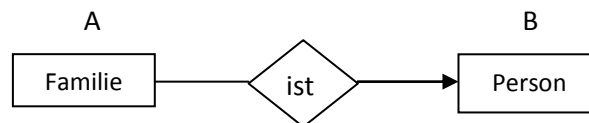
Ein wichtiges Charakteristikum von Relation-Typen ist ihre Funktionalität oder Kardinalität. Dadurch wird zum Ausdruck gebracht, mit wie vielen Entities ein gegebenes Entity in Beziehung stehen kann. Beim zweistelligen Relation-Typ ist hier die häufig verwendete Semantik:

- 1:1-Relation-Typ: Jedem Entity  $e_1$  in Entity-Typ  $E_1$  ist höchstens ein Entity  $e_2$  in Entity-Typ  $E_2$  zugeordnet.

- 1:n- Relation-Typ: Jedem Entity  $e_1$  in Entity-Typ  $E_1$  ist beliebig viele Entities  $e_2$  aus Entity-Typ  $E_2$  zugeordnet, und jedem Entity  $e_2$  in Entity-Typ  $E_2$  ist höchstens ein Entity  $e_1$  in Entity-Typ  $E_1$  zugeordnet.
- n:1- Relation-Typ: jedem Entity  $e_1$  in Entity-Typ  $E_1$  ist höchstens ein Entity  $e_2$  in Entity-Typ  $E_2$  zugeordnet, und jedem Entity  $e_2$  in Entity-Typ  $E_2$  sind beliebig viele Entities  $e_1$  aus Entity-Typ  $E_1$  zugeordnet.
- n:m-Relation-Typ: jedem Entity  $e_1$  in Entity-Typ  $E_1$  kann mit beliebig vielen Entities  $e_2$  in Entity-Typ  $E_2$  in Beziehung stehen und umgekehrt<sup>7</sup> [17].

Für die semantische Modellierung sind bestimmte Beziehungstypen von großer Bedeutung, welche folglich in diesen Modellen häufig vorkommen. Die sogenannte *ist*-Beziehung hat die Aufgabe, eine bestimmte Entity Menge zu spezialisieren bzw. generalisieren und wird laut([17], S.212) so definiert: „Zwei Entity-Mengen  $A$  und  $B$  stehen in der Beziehung  $A$  IST  $B$ . Es bedeutet, dass in konkreten Ausprägungen jedes Element aus  $A$  auch Element aus  $B$  ist.  $B$  modelliert die allgemeinere Entity-Menge,  $A$  eine speziellere dazu. Die Beziehung wird durch einen Pfeil von  $A$  nach  $B$  notiert mit einer Route, die mit *IST* beschriftet ist. Die Entities in  $A$  erben alle Attribute von  $B$  und können noch weitere Attribute haben, die spezielle Eigenschaften von  $A$  beschreiben“.

Ein Beispiel für diesen Beziehungs-Typ ist folgendes:



dieses Beispiel zeigt, dass die *Familie* eine speziellere Menge von der Entity-Menge *Person* ist, das heißt, Entities der Menge *Familie* erben alle Attribute der allgemeineren Menge *Person*.

Mit den *ist*-Beziehungen werden Teilmengenbeziehungen zwischen Entity-Mengen eingeführt. Auch wenn manche Entities in mehreren Entity-Mengen enthalten sein können, so gilt doch weiterhin, dass jede Entity eine eindeutige Identität hat.

Um eine bessere Vorstellung in solcher Modellierung zu ergreifen, wird ein Beispiel anhand eines Weltmodelles in ER-Diagramm für das Telefonbuch eines Smartphones dargestellt (Abbildung13).

<sup>7</sup> Beispiele dafür finden sie im Literatur [17] ab Seite 209

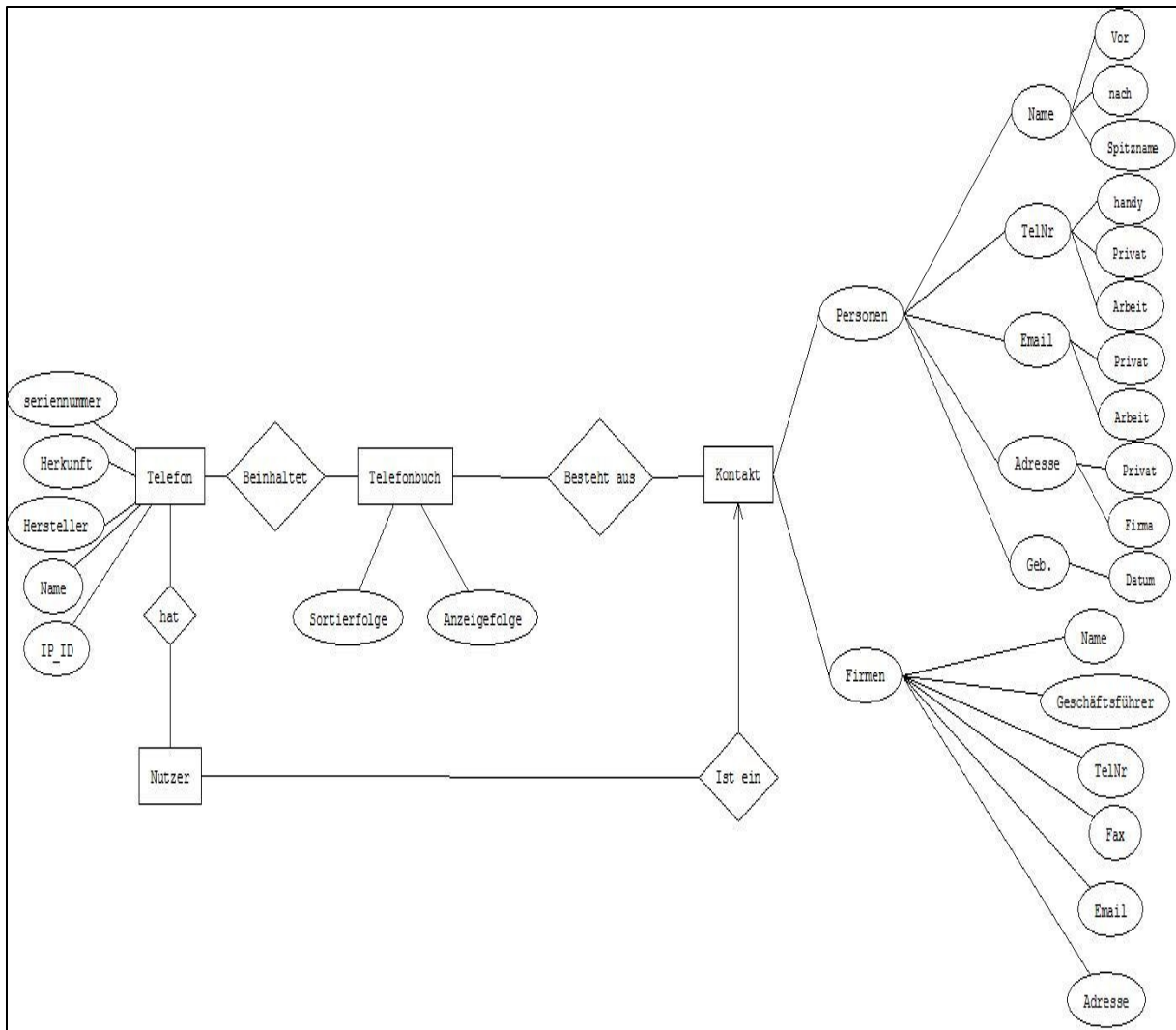


Abbildung 13: Weltmodell für ein Telefonbuch

### 2.3.1.2 Integriertes Struktur- und Funktionsmodell

Nachdem ein Weltmodell als Datenmodell bestimmt wurde, stellt sich nun die Frage, wie die Aktion und das Objekt identifiziert werden können. Um diese Frage zu beantworten, werden laut [1] eine integrierte Struktur und ein Funktionsmodell entwickelt, mit dem man das Weltmodell überprüfen kann.

- **Aktionsidentifikation(A\_ID)**

für die Identifikation einer Aktion setzt sich die Fragen aus der Aktion, was mit einem Objekt passiert, und einem Parameter, wie die Aktion durchgeführt wird, zusammen (siehe Abbildung 12). Diese Identifikation ist wichtig, damit das System erkennt, welche Parameter für welche Aktion benötigt werden. Um diese zwei Fragen, was? und wie? beantworten zu können, muss ein Aktionsmodell entwickelt werden. Dieses Aktionsmodell wird laut [1] dargestellt (Abbildung 14).

Anhand des Beispiels für das Telefonbuch-Weltmodell, wird ein Aktionsmodell entworfen. Das Hauptziel eines Aktionsmodells ist es zu wissen, welche Parameter für welche Aktion benötigt werden. Funktionen werden durch Aktionen (WAS) und Parametern (WIE) in ein Funktionsmodell gefasst. Sobald Aktionen mit den Parametern verknüpft sind, können Aktions-IDs gebildet werden. In unserem Telefonbuch-Beispiel wird zum Beispiel für den Aktionsfall Anrufen eine Nummer, je

nachdem welche Nummer man anrufen möchte, benötigt. Für den Aktionsfall Schreiben wird eine Nummer oder ein Name für die A\_ID benötigt. Für eine A\_ID ist es auch wichtig zu wissen, ob das benötigte Parameter intern, in dem Gerät angelegt, oder extern, also in einer externen Datenbank, ist.

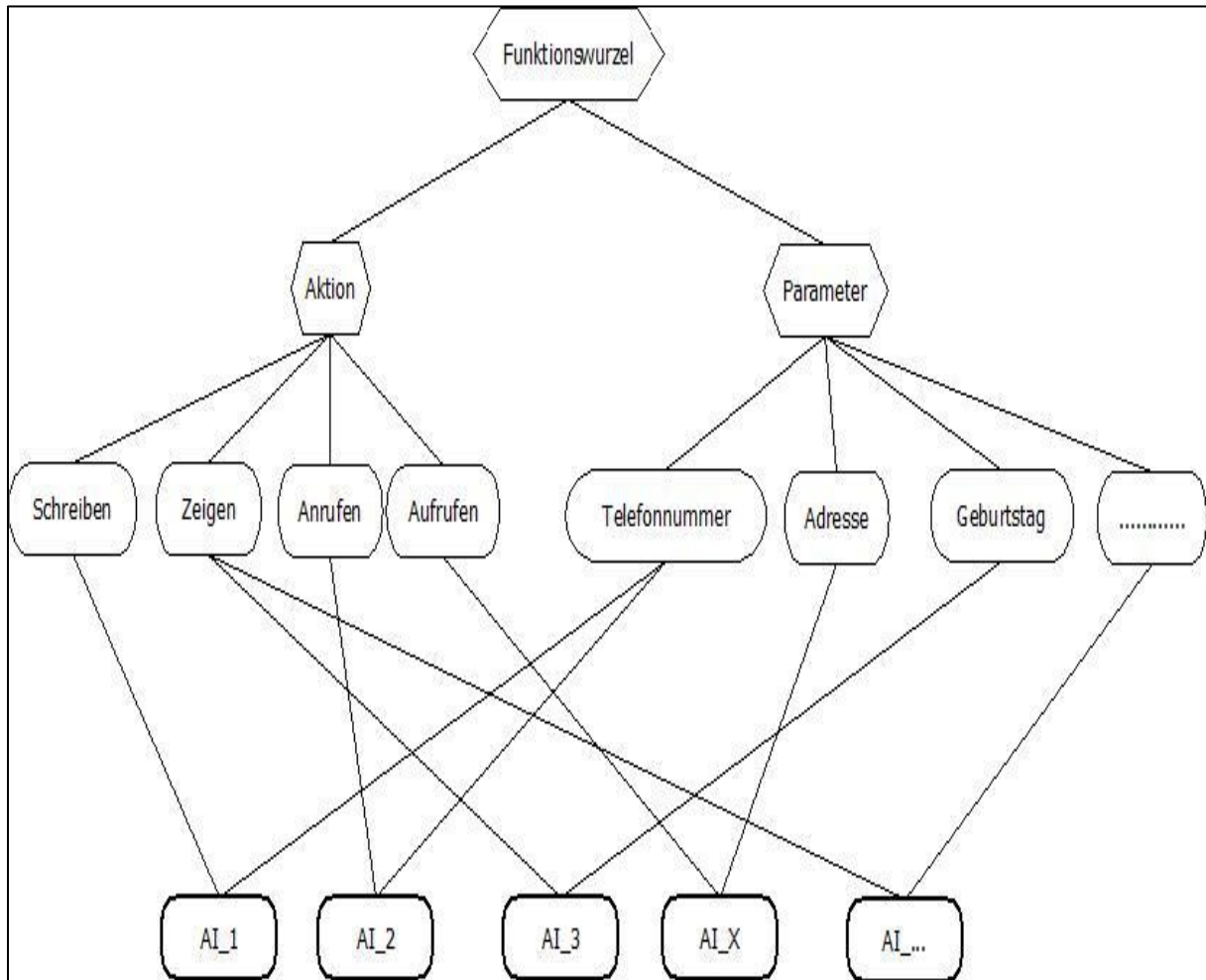


Abbildung 14: Aktionsmodell zur Identifikation von Aktionen-ID für das Telefonbuch-Beispiel

### Objektidentifikation (O\_ID)

Die Identifikation eines Objekts wird durch ein Strukturmodell gebildet (Abbildung 15). Dieses Modell wird auf Basis des ER-Modells geschaffen, das in der Datenbank gespeichert wird und ermöglicht die Antwort auf die Frage, womit eine Aktion durchgeführt werden soll. Zudem ordnet es allen in ER-Modellen befindlichen Objekten ihre Attributwerte zu.

In der Regel sind in Objektmodell alle Objekte in ER-Modell „Entities“, mit allen ihrem Attributwerten, mit einer Funktionswurzel verknüpft. In der Abbildung 15 sieht man das Objektmodell für das Telefonbuch-Beispiel. Durch eine zahlreiche Verbindung verschiedener Daten und mit Hilfe von Merkmal-Wert-Relation (siehe Kapitel 2.3.2) kann jedes beliebige Objekt eindeutig identifiziert werden.

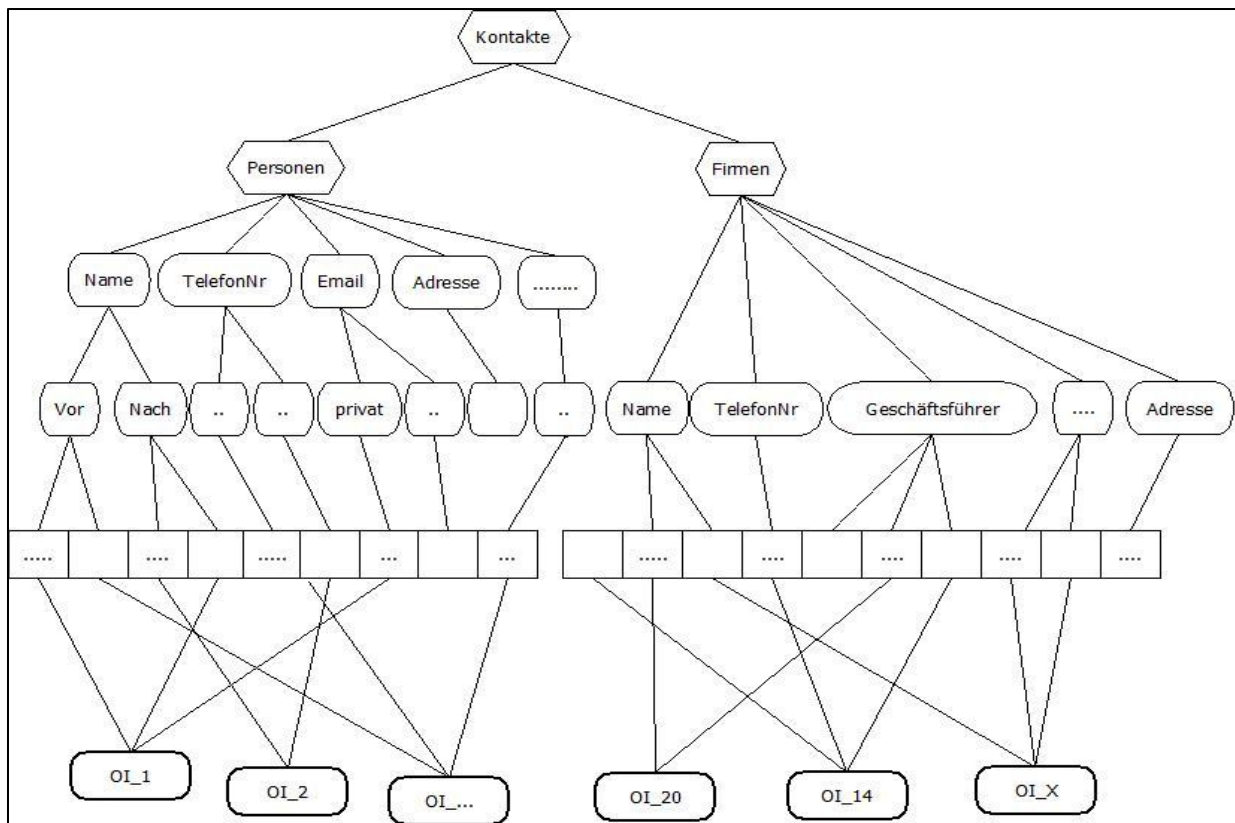


Abbildung 15: Objektmodell für Identifikation von Objekt-ID

Die Reihenfolge, was zuerst identifiziert würde, ist nicht relevant. Wichtig ist, dass die beiden Komponenten, Aktion-/ und Objekt-ID, gefunden werden.

### 2.3.1.3 Die Assoziations- und Verknüpfungstabelle

Nach der Bildung von Aktions- und Objekt-IDs werden diese in einer Assoziationstabelle eingetragen (Tabelle 1). Eine Assoziation entsteht erst, wenn zu jedem identifizierten Objekt eine Aktion, mit der das Objekt ausgeführt werden soll, zugeordnet ist (vgl. [22])

Assoziation \ ID	Aktion-ID	Objekt-ID
1	A_1	O_1
2	A_2	O_2
3	...	...

Tabelle 1: Assoziationstabelle

Jetzt sind alle Aktionen und Objekte in der Tabelle eingetragen, das heißt allerdings nicht, dass all diese Aktionen und Objekte miteinander verknüpfen dürfen. Für die Berichtigung der Assoziationstabelle ist eine Verknüpfungstabelle notwendig. Diese Verknüpfungstabelle verhindert, dass Aktionen unerlaubt auf Objekten durchgeführt werden. Zum Beispiel soll im Siri-System keine Email angerufen werden oder Nachricht an eine Adresse gesendet werden. Somit funktioniert die Verknüpfungstabelle an diese Stelle als ein Art Filter, der auf die Assoziationstabelle angewendet wird. Nach der Filterung der Assoziationstabelle entsteht eine Art aktuelle Berichtigungstabelle (Tabelle 2). Diese Tabelle zeigt welche Aktion mit welchem Objekt ausgeführt werden kann (vgl. [1]).

Aktion \ O_ID	Anrufen	Schreiben	Aufrufen	Zeigen	.....	.....	
O_1	X						
O_2			X				
.....							

Tabelle 2: Berichtigungstabelle

### 2.3.2 Dynamische Modellierung

Nach der Entwicklung und dem Entwurf einer statischen Modellierung und nachdem alle Objekte und Aktionen des Weltmodell identifiziert werden können, stellt sich jetzt die Frage wann welche Aktion auszuführen ist und wie ein Objekt ausgewählt wird. Während eines Dialogablaufs zwischen Nutzer und einem System ist eine dynamische Modellierung notwendig. Das Ziel dieser Modellierung ist es während der Interaktion eine Dialogstrategie zu finden, mit der man ein Szenario aufbauen kann und die gewünschte Aktionen und Objekte zu bestimmen. Bevor eine dynamische Modellierung entwickelt wird, muss zuerst das gewünschte Systemverhältnis bestimmt werden. Das Dialogsystem bei Siri (siehe Abbildung 10) soll neben der Analyse von Benutzereingaben (Erkennung) und der Synthese von Systemausgaben auch Entscheidungsprozesse (Dialogsteuerung) modellieren. Dabei hat das System nach der Eingabe des Benutzers die korrespondierende optimale Reaktion auszuwählen, so dass der Benutzer möglichst ohne Umwege sein aktuelles Ziel erreichen kann. Solche Entscheidungsprozesse beschreiben die Arbeitsweise eines kognitiven Systems.

Zur Modellierung derartiger Prozesse wurde zunächst der Markov Entscheidungsprozess (*Markov Decision Process, MDP*) behandelt. Im Grunde besteht ein MDP aus:

- Eine Zustandsmenge  $S$
- Eine Aktionenmenge  $A$
- Eine reellwertige Belohnungsfunktion  $R(s, a)$
- Eine Abbildung  $T$  jeder Aktionsauswirkungen in jedem Zustand

Ein wichtige Eigenschaft bei der MPD ist, dass die Zustandsübergang  $s'$  nur vom vorhergehenden Zustand  $s$  und nicht von dessen Vorgängern abhängig ist. Die Abbildung eines neuen Zustands kann durch  $T: S \times A \rightarrow Prop(S)$  dargestellt werden. Für jeden Zustand und jede Aktion wird eine Wahrscheinlichkeitsverteilung für den nächsten Zustand festgelegt, wobei  $Prop(S)=P(s'/s,a)$  ist die Wahrscheinlichkeit  $P$  von Zustand  $s$  ist und die Ausführung von Aktion  $a$  in Zustand  $s'$  zu gelangen. Die Repräsentation eine Lösung für MDP kann Zum Beispiel durch die Value-Iteration-Verfahren dargestellt werden[1]:

$$\pi: S \rightarrow a$$

$\pi$  ist die Strategie, welche den aktuellen Zustand auf eine optimale Aktion abbildet, sodass sich die Kosten über die Zeit minimieren. Aber wegen der fehlenden Beobachtungen werden Schwierigkeiten

im Umgang mit unsicherer Information entstehen und daher kann der Systemzustand nicht immer ermittelt werden.

Erst mit der Einführung der sogenannten POMDP-Modellierung (*Partially Observable Markov Decision Process*), bei dem die Kommandos des Benutzers als Beobachtungen aufgefasst wurden, konnte man dieser Problemstellung gerecht werden. In der Abbildung 16 wird ein Dialogsystem mit der Modellierung eines Entscheidungsprozesses auf POMDP-Basis dargestellt.

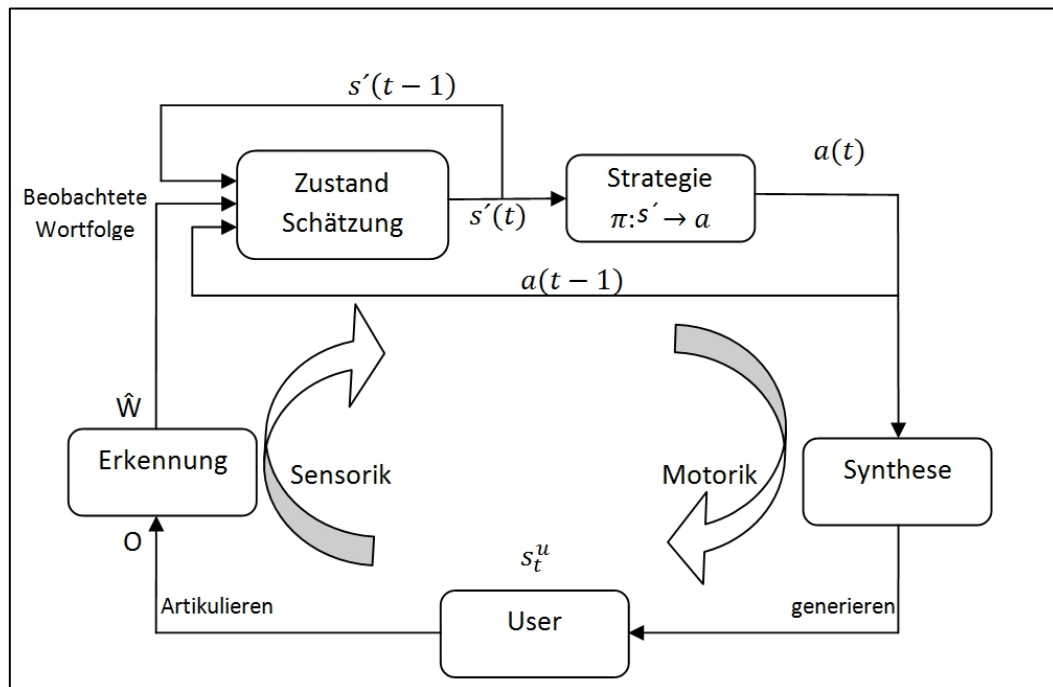


Abbildung 16: Modellierung der Entscheidungsprozess ([17], S.2)

Dieses System verhält sich im Betriebszustand wie folgt:

Nachdem die Sprachsignaleingabe vom System erkannt wird, entsteht als aktuelle Beobachtung eine Wortfolge vom System  $\hat{W}$  (siehe Spracherkennung). Das System wird nicht direkt auf die Nutzerabsicht schließen, es schätzt dessen Absicht durch Beobachtung. Diese Beobachtung wird an eine Zustandsschätzung weitergeleitet. Die Zustandsschätzung und die Verfügbarkeit einer optimalen Strategie, ausgehend von dem geschätzten Zustand, werden die Reaktion des Systems bestimmen. Die Zustandsschätzung ist von der aktuellen Beobachtung  $\hat{W}$  sowie der vorangegangenen Aktion des Systems  $a(t-1)$  und der vorangegangenen Zustandsschätzung  $s(t-1)$  abhängig.

$$s(t) \propto \hat{W}, s(t-1), a(t-1)$$

Jetzt kann das System Anhand des geschätzten Zustands und der aktuellen Beobachtung eine optimale Reaktion hervorrufen. Die optimale Strategie ergibt sich aus einer Funktion mit der, die aktuelle Zustandsschätzung  $s(t)$  auf eine Aktion des Systems  $a(t)$ , abgebildet wird ( $\pi: s \rightarrow a$ ) und in eine vorab durchgeführte Trainingsphase ermittelt wurde [18]. Jede entstehende Aktion ( $a(t)$ ) des Systems wird mit entsprechenden Kosten beaufschlagt, so dass eine optimale Strategie die Kosten der Mittel minimiert. Die gewählte Reaktion  $a(t)$  wird schließlich über Sprachsynthese als Wortfolge an den Nutzer übergeben. Diese Modellierung kann das Verhältnis des Benutzers in der Übergangswahrscheinlichkeitsverteilung mit Hilfe des *Satzes von Bayes* erfasst werden:

$$s(t) = P\{s(t) | \hat{W}, a(t-1), s(t-1)\}$$

### **Merkmal-Werte-Relation:**

Die Identifikation von Objekten und Aktionen in dieser Arbeit basieren auf Basis der so genannten Merkmal-Werte-Relation (MWR). Laut [19] wird MWR als eine azyklisch gerichtete graphische Darstellung definiert, bestehend aus einer Hierarchie von semantische Kategorien, genannt *Merkmale*, wie zum Beispiel Kontakt, Name, Adresse, usw. und für diese Merkmale werden wiederum Datenwerte zugeordnet, sogenannte *Werte*, wie zum Beispiel *Sebastian*, oder *Karl-Marx-Straße*. Aus diesen beiden Komponenten werden dann die ID's gebildet. Durch eine gezielte Eingabe des Nutzer werden nach und nach Informationen zwischengespeichert, der sogenannte *Informationsstatus*, bis die gesuchten Objekte und Aktionen identifiziert werden.

### **Informationsstatus:**

Bei einer dynamischen Modellierung werden die Daten so lange gesammelt bis die Identifikation von den gewünschten Aktionen und Objekten erfolgt. Diese Informationssammlung aktualisiert sich mit jeder neuen Eingabe des Nutzers und wird als Informationsstatus bezeichnet. Deswegen bildet sich der aktuelle Informationsstatus  $I'$  aus einem alten Zustand  $I$  und eine neue Nutzereingabe  $E$

$$I' = f(I, E)$$

Und die Systemausgabe  $A$  wird denn von den Aktuellen Zustand  $I'$  und der Nutzereingabe abhängig

$$A = g(I', E)$$

Sobald ein Nutzerwunsch erfüllt ist, setzt sich der Informationsstatus auf null und wird für einen neuen Wunsch oder Aufgabe bereit. Mit Unterbrechung einer geführten Aufgabe, setzt sich wieder der Informationsstatus auch auf null und es werden erst Information bei einer neue Eingabe aufgenommen[1].

### **2.3.3 Direkte und indirekte Identifikation**

Die Identifikation von Objekten und Aktionen geschieht in der Regel über die Zuordnung von Attributen. Um die Effektivität des Dialoges zwischen den beiden Ebenen Mensch und System zu garantieren, gibt es verschiedene Methoden. In das Weltmodell in Abbildung 13 kann man Beziehungen zwischen Objekten einfügen. Mit Hilfe dieser Beziehungen ist das gewünschte Objekt schneller und nicht mehr nur durch die ihm zugeordneten Attribute, identifizierbar (Abbildung 17).



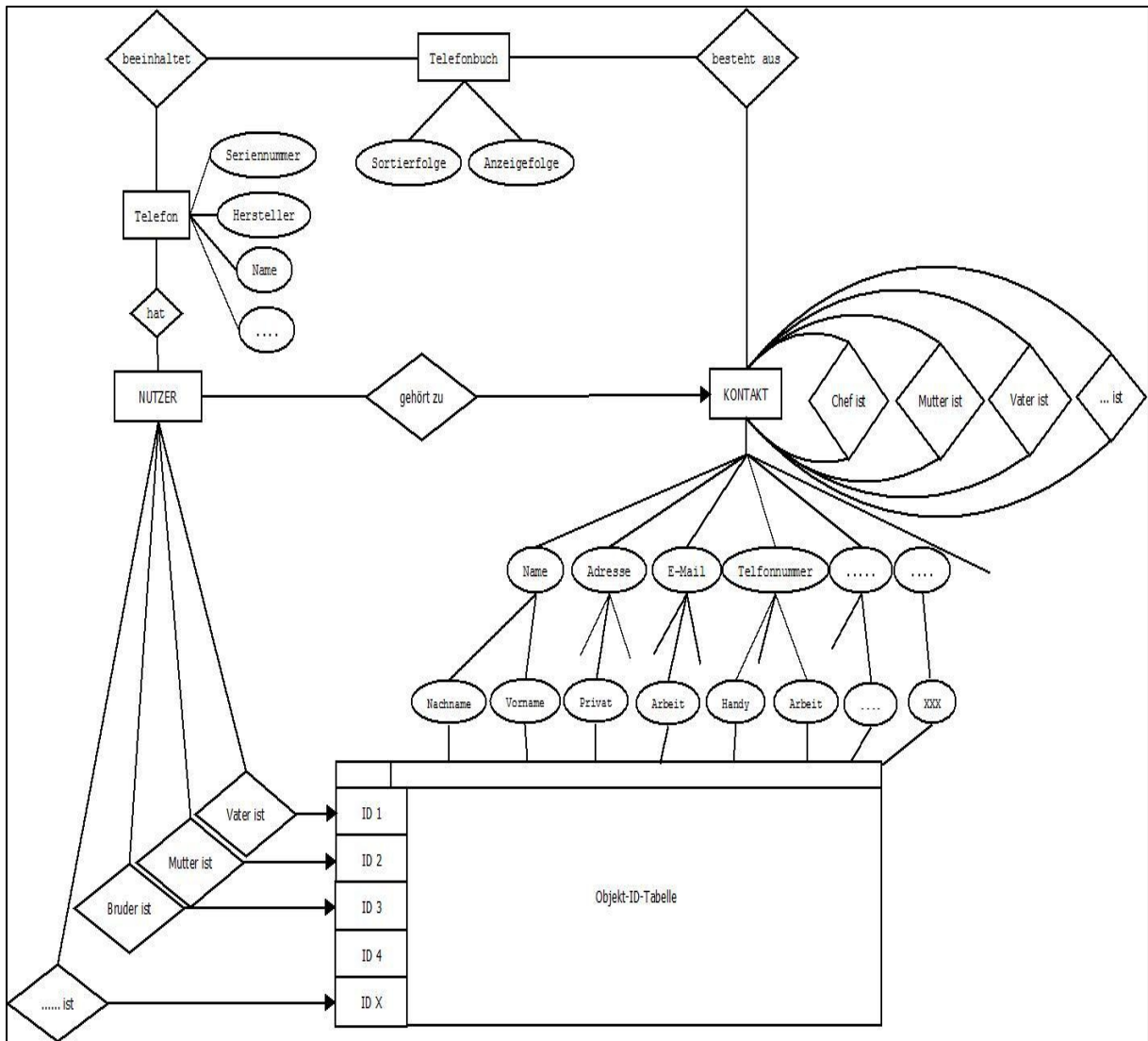


Abbildung 17: Erweiterter Weltmodell mit direkte und indirekte Objekt-ID

Durch diese zusätzlichen Informationen und die Beziehungen zwischen den Objekten, sind jetzt verschieden Methoden zur Identifikation von Objekt-ID:

- Direktionale Identifikation: über Attribute
- Indirekte Identifikation: über Beziehungen
- Kombination Identifikation: über Attribute und Beziehungen

### 2.3.4 Beispiel Telefonbuch

Dieses Beispiel soll während eines Dialogablaufs zwischen Nutzer und System zeigen, wie die ID's identifiziert werden und wie der Informationsstatus mit Hilfe eines MWR abgebildet wird:

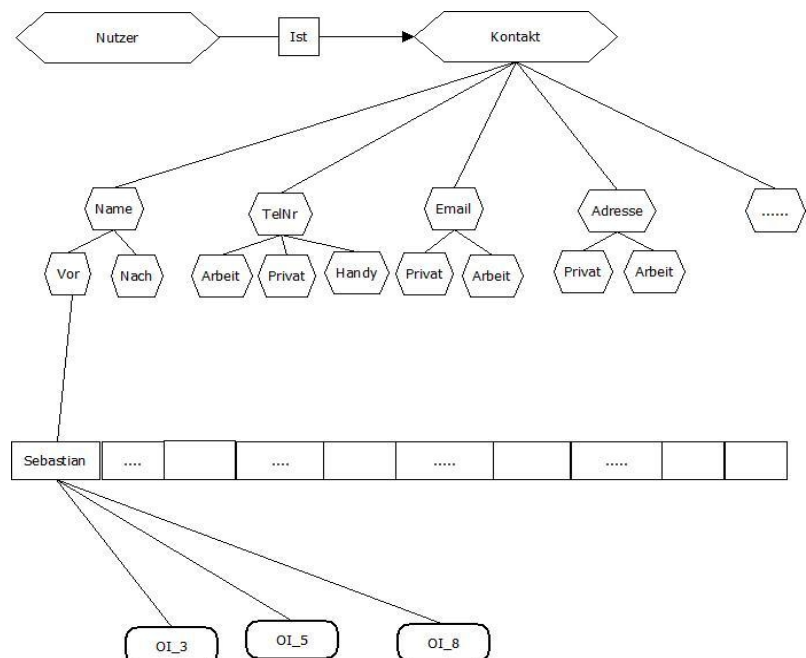
Der Nutzer möchte die Handynummer einer Kontaktperson namens **Sebastian Schmidt** anrufen. Dieses Objekt hat die ID (O\_83). Der Nutzer weiß nicht viel über der gewünschte Kontaktperson und hat nur begrenzte Informationen. Um die gewünschte Kontaktperson zu identifizieren werden hier drei Szenarien ausgedacht. Das erste Szenario besteht aus der direkten Eingabe der Objekt-Attribute, das Zweite aus der Beziehung der Objekte miteinander, und das Dritte aus der Kombination von beiden. Diese drei Szenarien sollen dann die Flexibilität des Systems zeigen und wie das System im Idealfall reagieren sollte. Am Anfang jeder Äußerung oder jedes Wunsches, welcher von dem Nutzer geäußert wird, ist der Informationsstatus noch auf null, weil sich das System noch keine Werte gemerkt hat. Im Laufe der Interaktion merkt sich das System mehrere Informationen und ändert sich mit jeder neuen Eingabe des Nutzers. Diese Informationen und Werte, die das System sammelt, können bei der Identifikation der Aktionen und Objekte helfen. Das heißt, dass in dem Informationsstatus Informationen sowohl für Aktions- als auch für Objektidentifikation sind und dazu gehören zwei Modelle: Ein Modell für die Aktionsidentifikation und das andere für die Objektidentifikation. In den drei ausgedachten Szenarien ist der Aktionswunsch Anrufen identisch, deshalb werden nur die verschiedenen Objektidentifikationen beobachtet.

#### Szenario 1 (direkte Identifikation):

Der Nutzer kennt nur den Vornamen der gewünschte Kontaktperson und sein Wohnadresse. Um den Wunsch des Nutzers an das System weiterzuleiten, ist ein Dialog dieser Art denkbar:

**Nutzer:** ich möchte Sebastian anrufen

**System:** es gibt verschiedene Personen mit dem Vornamen Sebastian ich brauche mehr Informationen.



**Nutzer:** ich glaube er wohnt in Karl-Marx-Straße

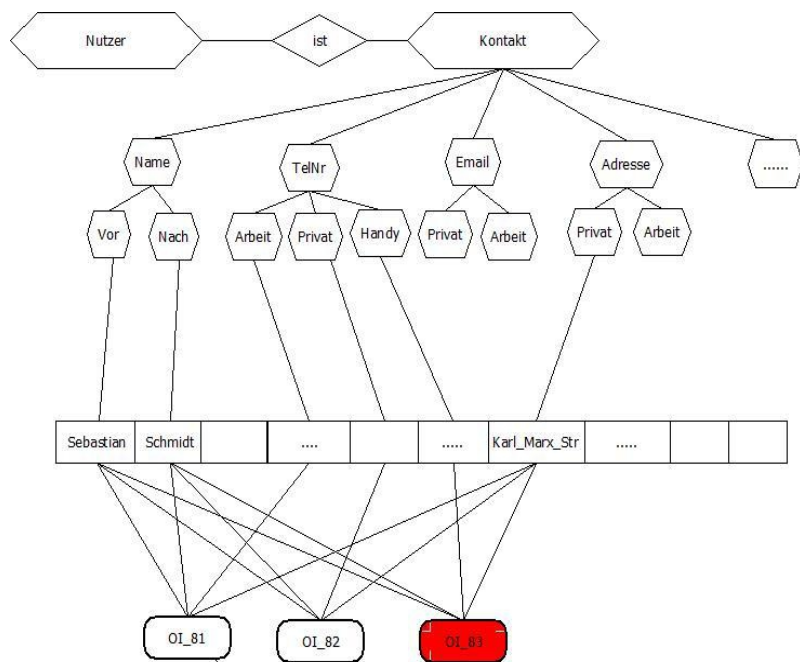
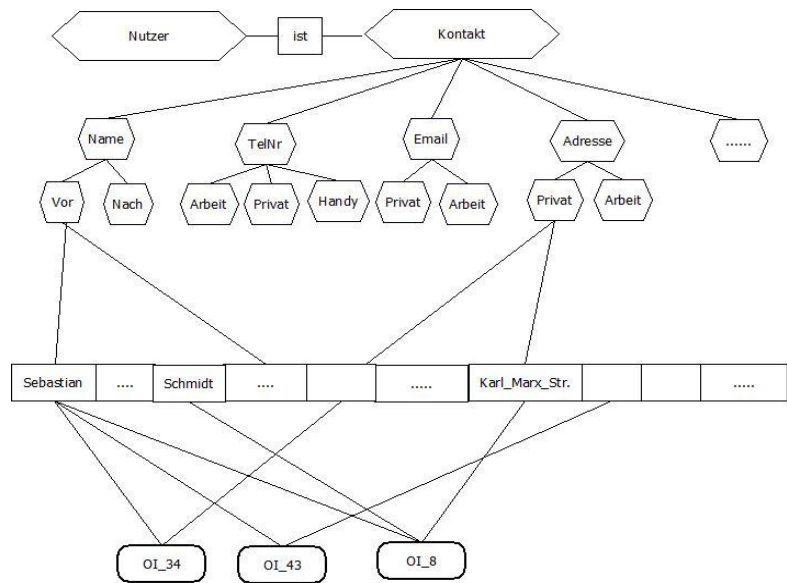
**System:** Meinst du Sebastian Schmidt?

**Nutzer:** ja den meine ich

**System:** Sebastian Schmidt hat verschiedene Nummern, welche soll ich anrufen?

**Nutzer:** seine Handy Nummer.

**System:** Die Handy Nummer von Sebastian Schmidt wird angerufen



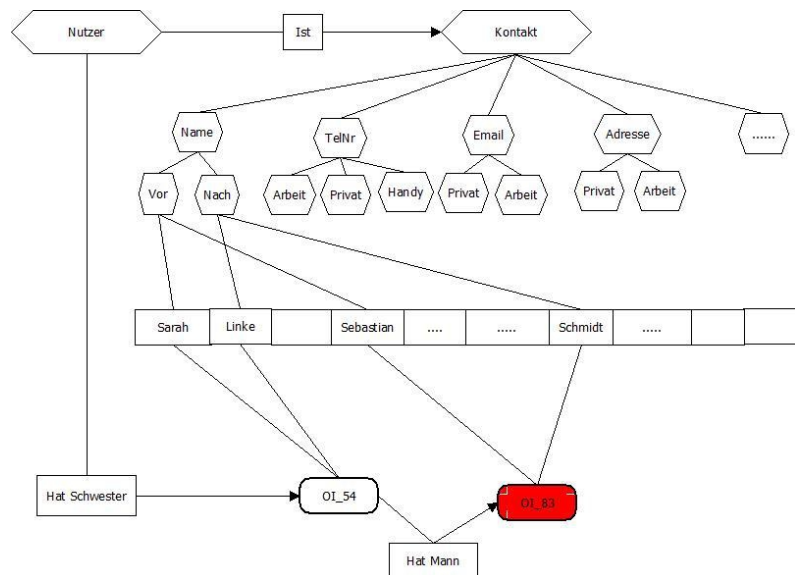
### Szenario 2 (indirekte Identifikation):

Der Nutzer weiß nur, dass die gewünschte Person der Mann seiner Schwester ist und nicht, wie sein Name ist oder seine Email-Adresse oder ähnliches lautet. Das System sollte in diesem Fall direkt durch eine Beziehung zwischen dem Nutzer und den Kontakten oder Beziehungen zwischen den Kontakten die gewünschte Kontaktperson identifizieren. Diese Beziehungen sind nicht im Weltmodell definiert und werden während einer Trainingsphase vorgemerkt und gespeichert. Das System soll

während dieser Trainingsphase merken, dass der Nutzer in seiner Kontaktliste eine Schwester hat und diese Schwester mit einem Mann verheiratet ist, der auch in der Kontaktliste zu finden ist. Ein Dialog könnte im Idealfall ganz kurz und schnell ablaufen:

**Nutzer:** rufe den *Mann* von meiner Schwester an

**System:** die Handynummer von Sebastian Schmidt wird angerufen

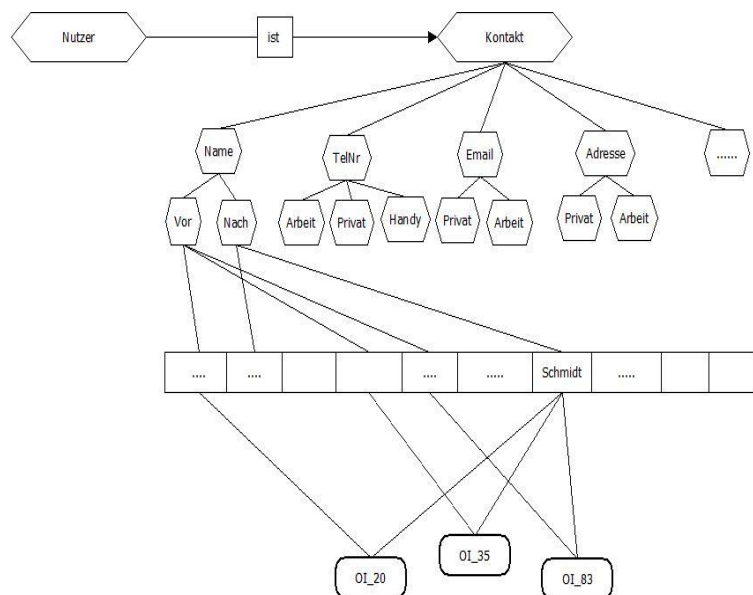


### Szenario 3 (Kombination):

Der Nutzer kennt den Familiennamen und weiß, dass er der Bruder seiner Frau ist. In diesem Fall gibt der Nutzer am Anfang den Nachnamen der gewünschten Person an, aber das System hat in der Kontaktliste mehrere Personen mit diesem Nachnamen. Der Nutzer weißt nur noch, dass diese Person der Bruder seiner Frau ist. An dieser Stelle sollte das System wissen welche Person gemeint wird.

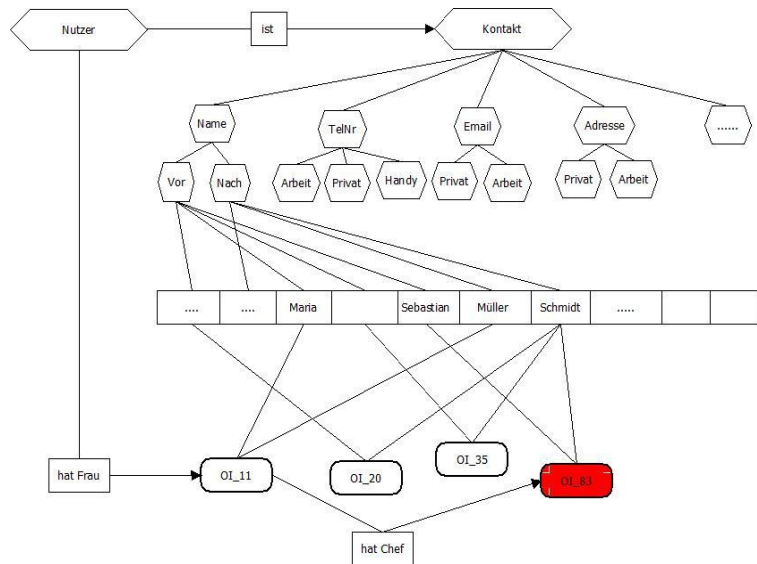
**Nutzer:** Ich möchte Herrn Schmidt anrufen.

**System:** Es gibt mehrere Personen mit dem Namen Schmidt, welche meinst du?



Nutzer: den Chef meiner Frau.

System: der Telefonnummer von Sebastian Schmidt wird angerufen

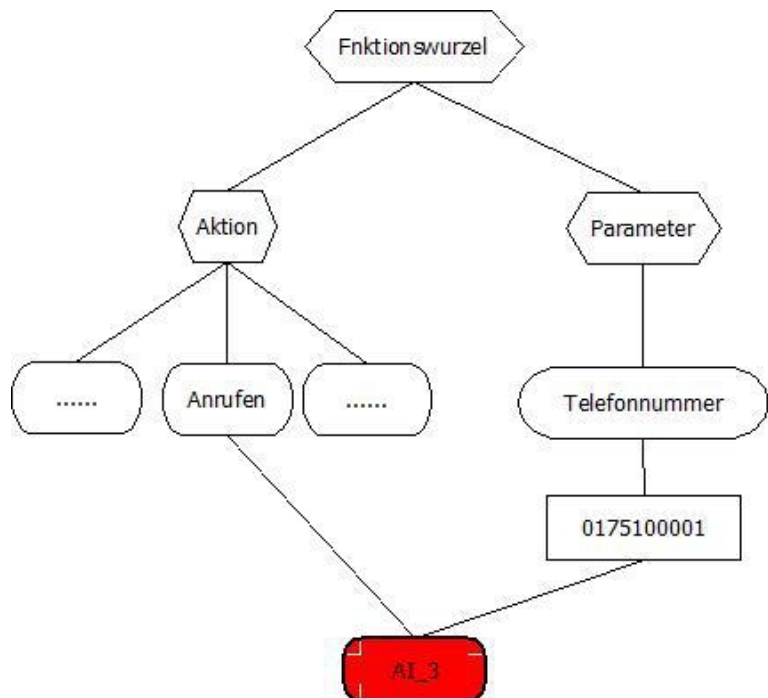


Zur der Erfüllung des Nutzerwunsches gehört auch eine Aktionsidentifikation. Diese soll bei den drei Szenarien, egal wie die Nutzereingaben formuliert sind, gleich sein. Der Nutzer möchte die Aktion **Anrufen** durchführen und dafür ist ein Parameter notwendig um zu bestimmen wie die Aktion durchgeführt wird. Ein möglicher Dialog könnte so ablaufen:

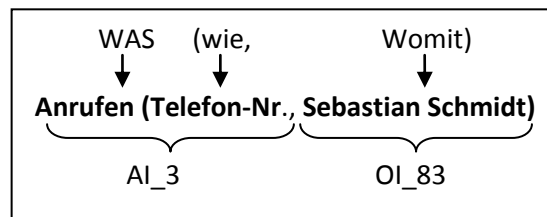
**Nutzer:** Ruf Sebastian an./rufe Herrn Schmidt / ich möchte den man meiner Schwester anrufen

(Das System weißt hier mit Hilfe der Nutzereingaben und der Objekt ID welcher Kontakt-Person gemeint ist)

**System:** die Telefonnummer von Sebastian Schmidt wird angerufen



Mit diesen beiden Komponenten wird die gewünschte Aktion von System durchgeführt und ein Zusammenhang zwischen den beiden Komponenten könnte dargestellt werden(wie in Abbildung 12).



Bei dem Untersuchungsteil in Kapitel 4 werden Online- und Offline Dienste jeweils mit einem ausgewählten Dienst unter die Lupe genommen und die Dialogvarianten in diesem Beispiel überprüft.

### 3. Stand der Technik

Die Interaktion und das Agieren zwischen Mensch und System bzw. Maschine ist schon seit Jahrzehnten ein intensives Forschungsgebiet. Dank der neuen Technologien sind heutzutage verschiedene Anwendungen, die damals bei Sciencefiction- Filmen unvorstellbar waren, verwendbar. Es zeichnet sich ab, dass spracherkennende- und -verstehende Systeme schon bald aus dem betrieblichen Alltag nicht mehr wegzudenken sind.

In den letzten Jahren hat sich der Bereich des automatischen Verstehens von gesprochener Sprache erheblich entwickelt. Insbesondere bei der Mobiltelefonie (Smartphone), die zurzeit nicht auf Telefonieren und Nachrichten schreiben begrenzt ist, sondern viele andere Aufgaben erledigen kann. Aktuell auf dem Markt ist zurzeit das Spracherkennungssystem Siri, das seit Anfang 2010 als App für Smartphones der Firma Apple dient und in das iPhone 4S integriert ist. Diese Software soll per Sprache verschiedene Aufgaben auf dem iPhone steuern.

#### 3.1 Siri und seine Anwendungen

Laut Apple ist Siri eine der intelligentesten persönlichen Assistenten, dem Sie sagen können, welche Aufgaben zu erledigen sind. (Siri versteht, was Sie sagen und meinen. Deshalb ist es auch nicht notwendig, spezielle Befehle zu erlernen und sich an Schlagwörter zu erinnern. Sie können Ihre Anweisungen auf unterschiedliche Art und Weise sagen. Sagen Sie zum Beispiel „Stelle einen Wecker für 6:30 Uhr morgens“ oder „Wecke mich um 6:30 Uhr“. Egal, wie Sie es sagen, Siri versteht es)[14].

Mit dem letzten Software-Update von Apple IOS 6 soll Siri die folgenden Aufgaben per Spracheingabe bedienen:

Anwendung	Aufgaben
Telefon	Personen per Name, Nennung oder Nummer ansagen und anrufen
Nachrichten	Schreiben, vorlesen, verschicken und antworten
Kalender	Termine eintragen, ändern, aufrufen und verschieben
Sport(O)	Ergebnisse, Spielpläne, Spielinformation und Teaminformation anzeigen
Karten	Navigation, Karten und lokale Geschäfte finden und zeigen
Twitter(O)	Posten auf Twitter
Facebook(O)	Status schreiben
Restaurants	Suche nach bestimmten Restaurants
Filme	Suche nach Filminformationen, Kinos und Filmrezensionen vorlesen
Musik	Abspielen von beliebiger Musik oder Wiedergabelisten, überspringen und pausieren
Erinnerungen	Erstellen, aufrufen und bearbeiten
E-Mail(O)	Senden, abrufen und antworten
Wetter	Wetterauskunft
Aktien	Aktienkurse, Information
Uhr	Wecker einstellen und bearbeiten, Weltzeituhr anzeigen und Timer stellen
Kontakte	Nach Informationen fragen, Kontakte suchen, und Beziehungen verbinden
Freunde	Nach Freunden suchen und orten
Notizen	Eintragen, suchen und bearbeiten
Safari	Websuche

Tabelle 3: Dienste die mit Siri arbeiten

Nach dem aktivieren von Siri auf dem iPhone 4s und unter der Voraussetzung der Verfügbarkeit eines Internetzugangs über ein Mobilfunk oder eine WLAN- Verbindung, kann Siri auf zwei Arten gestartet werden:

1. Langes Drücken auf die Home-Taste bis Siri angezeigt wird
2. Das iPhone wie bei einem Telefonanruf an das Ohr halten

In beiden Fällen hört man zwei kurze Töne und auf dem Bildschirm wird „Wie kann ich behilflich sein?“ angezeigt.



Dann sollte laut der Beschreibung von Apple ein Dialog zwischen dem Nutzer und Siri folgen, bei dem man mit ganz normaler gesprochener Sprache sprechen kann. Wie der Dialog durchgeführt wird und was man sagen kann, hat Apple so beschrieben *„Mit Siri können Sie Nachrichten eingeben und senden, Termine festlegen, Telefonnummern wählen, Wegbeschreibungen abrufen, Erinnerungen festhalten, das Internet durchsuchen – und vieles mehr. Und das ganz einfach, indem Sie ganz normal sprechen. Siri versteht, was Sie sagen, und weiß, was Sie meinen. Siri reagiert auf das, was Sie gesagt haben, damit Sie wissen, ob es korrekt verstanden wurde. Und Siri fragt nach, wenn eine Klarstellung oder weitere Informationen erforderlich sind. Das ist einer Unterhaltung mit einer anderen Person sehr ähnlich“* ([14]S.43).

## 4. Untersuchung

In dieser Arbeit wird die Untersuchung der Siri-System-Dienste nur auf die semantische Ebene begrenzt. Bei dieser Gelegenheit soll die Flexibilität und das Verständnis des Systems getestet werden. Hier wird das System an seine Grenze kommen und somit können im Anschluss seine Stärken und Schwächen identifiziert und bewertet werden. Kann Siri halten was ihre Entwickler versprechen? Und wenn ja, wie gut kann es das? Was zeichnet Siri aus und was weniger oder gar nicht? Was soll Siri theoretisch noch können und wie kann man das anhand eines ansprechenden Modelles entwickeln?

Bevor mit der Untersuchung begonnen wird, wird zwischen Online- und Offline-Diensten unterschieden. Alle Dienste die mit dem Appleserver verbunden sind und mit keinem anderen Anbieterserver sind Offline-Dienste. Alle Dienste, die über andere externe Anbieterserver arbeiten, sind Online- Dienste. Für die Offline-Dienste werden wir bei der Untersuchung das Telefonbuch als Beispiel nehmen und für die Online-Dienste wird dann der Yahoo Sport-Dienst- (Sportacular) als Beispiel dienen.

Für jeden Dienst werden alle möglichen Varianten zur Identifikation von Objekten getestet und für jede Variante ein paar Versuche durchgeführt, um fest zu stellen wie gut Siri das kann.

### 4.1 Dialog mit Offline-Applikationen

Für den Dialog mit Offline-Diensten wird als erstes das Telefonbuch untersucht. In dem Kapitel 2.3.4 diente das Telefonbuch ebenso als Beispiel für die Verdeutlichung der Modellierung und des



Dialogablaufes eines Systems im Idealfall. Jetzt wird noch einmal das Telefon getestet, nun aber auf das tatsächliche Können und darauf wie es auf unsere konkreten Wünsche reagiert. Anschließend wird ein ER-Diagramm entwickelt, ganz ähnlich dem von Siri.

### 4.1.1 Das Telefonbuch

Das Telefonbuch auf dem Smartphone von Apple ist eines der häufigsten Dienste, die eng mit Siri zusammenarbeitet. Laut des Benutzerhandbuches des iPhone 4s soll Siri für das Telefonbuch viele Aufgaben erledigen und bei einer Verwirrung wiederholt fragen um die gewünschte Aktion und das Objekt zu identifizieren. Für eine Objekt-Identifikation werden wie in Kapitel 2.3.3 die drei Möglichkeiten von Siri anhand von ein paar Versuchen überprüft und anschließend ein Modell vorgestellt, mit dem Siri möglicherweise im Telefonbuch arbeitet.

#### Versuche Telefonbuch:

##### o Direkte Identifikation

Bei der direkten Identifikation wird der Nutzer seine Wünsche nur mit Attributen, die mit Hilfe von MWR die gewünschte Kontaktperson identifizieren können, eingeben. In verschiedenen Testversuchen im Rahmen direkter Angaben von Attributen versucht der Nutzer gezielt unterschiedliche Äußerungen und ansprechend gezielte Attribute anzugeben um zu schauen wie das System reagiert.

#### Versuch 1: Der Nutzer möchte jemanden mit seinem Vornamen anrufen

Nutzer: Rufe bitte Sebastian an/Sebastian anrufen/Ich möchte Sebastian anrufen.

Siri : Welchen Kontakt meinst du? Sebastian Schmidt, Sebastian Müller, oder Sebastian Linke.

Nutzer : Sebastian Schmidt.

Siri : Welche Telefonnummer für Sebastian Schmidt? Handynummer oder Privatnummer?

Nutzer : Handynummer.

Siri : Die Handynummer von Sebastian Schmidt wird angerufen.



#### Versuch 2: Der Nutzer kennt nur den Vornamen und die Adresse

Nutzer: Rufe bitte Sebastian an/Sebastian anrufen/Ich möchte Sebastian anrufen.

Siri : Welchen Kontakt meinst du? Sebastian Schmidt, Sebastian Müller, oder Sebastian Linke?

Nutzer: Er wohnt in der Universitätsstraße 1.

Siri : Entschuldige, ich verstehe „er wohnt in der Universitätsstraße“ nicht, welchen Kontakt meinst du? Sebastian Schmidt, Sebastian Müller, oder Sebastian Linke? Oder

Siri : Ich kann keine Änderungen an deinen Kontakten vornehmen, Mohammed, du kannst das jedoch in deiner Kontakte-App tun.



### Versuch 3: Der Nutzer möchte die Privat- Adresse von jemandem wissen

Nutzer: Wo wohnt Thomas Siemon?/Wie lautet die Privat Adresse von Thomas Siemon?

Siri : Hier ist die Privat-Adresse von Thomas Siemon. (zeigt alle Adressen)

### Versuch 4: Kontaktdaten zeigen

Nutzer: Zeige mir die Email-Adresse von Sebastian Schmidt.

Siri : Sebastian Schmidt hat mehrere Email-Adressen. ( zeigt alles)

Nutzer: Was ist seine private Email-Adresse?

Siri : Die private Email-Adresse von Sebastian Schmidt ist Sebastian 123@web.de. (zeigt alle E-Mail-Adressen)



#### ○ Indirekte Identifikation

Mit Indirekter Identifikation soll das gewünschte Objekt über Beziehungen identifiziert werden, wie z.B. meine Frau, mein Chef, sein Bruder usw..

### Versuch 5: Nutzer- Daten identifizieren

Nutzer: (Meine) Telefonnummer anrufen/Ruf mich an!

Siri : Welche Telefonnummer von Mohammed Shamsan soll ich anrufen, iPhonenummer oder Geschäftsnummer?

Nutzer: iPhonenummer.

Siri : Die iPhonenummer von Mohammed Shamsan wird angerufen.



### Versuch 6: Kontaktdaten von Nutzer zeigen oder Aufrufen.

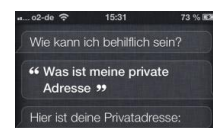
Nutzer: was ist meine Privatadresse

Siri : hier ist Dein Privatadresse(zeigt meine arbeits- und Privatadresse)

### Versuch 7: Zuhause anrufen

Nutzer: Rufe bitte zuhause an?

Siri : Die Privatnummer von Mohammed Shamsan wird angerufen



### Versuch 8: Beziehungen Kontakte mit Nutzern.

Nutzer: Sage meinem Bruder: Ich rufe später an.

Siri : Hier ist deine Nachricht an Ahmed Brother „ich ruf später“, soll ich sie senden?

Nutzer: Ja

Siri : Ok. Sie wurde gesendet.

### Versuch 9

Nutzer: Wann hat meine Frau Geburtstag?

Siri : Franziska Kannengießer hat am 12. Mai Geburtstag.



## Versuch 10

Nutzer: Rufe meinen **Chef** an.

Siri : Die Handynummer von Daniel Vehlo wird angerufen

## Versuch 11: Meine Frau Franziska hat eine Chefin Anja

Nutzer: Rufe die **Chefin** von meiner **Frau** an.

Siri : Mit wem möchtest du sprechen?

Nutzer: Mit der **Chefin** meiner **Frau**.

Siri : Die Handynummer von Daniel Vehlo wird angerufen (*Chef vom Nutzer*).

Nutzer: Nein, ich möchte die **Chefin** von meiner Frau anrufen.

Siri : Ich sehe "Nein, ich möchte die Chefin von meiner Frau anrufen" nicht in deinem Adressbuch.



### o Kombination

Hier wird das Objekt über direkte und indirekte Identifikation überprüft.

## Versuch 12: Daniel ist der Mann von meiner Schwester

Nutzer: Rufe bitte den Daniel an!

Siri : Welchen Kontakt meinst du, Daniel Vehlo oder Daniel Schumacher?

Nutzer: Den Mann von meiner Schwester.

Siri : Ich sehe „Mann“ nicht in deinem Adressbuch.



## Versuch 13: Daniel ist der Vater von Nancy

Nutzer: Den Vater von Nancy anrufen!

Siri : Die Handynummer von Nancy wird angerufen. oder

Siri : Die Handynummer von Sultan Shamsan (Vater des Nutzers) wird angerufen.



## Versuch 14: Anja ist die Chefin von Franzi und Mathias Viehlo ist der Chef vom Nutzer

Nutzer: Ich möchte die E-Mail-Adresse von der Chefin von Franzi.

Siri : Hier ist die E-Mail-Adresse von Daniel Vehlo./oder: Ich weiß nicht; wer deine Managerin ist, aber wenn du es mir sagst, werde ich's mir merken.

## Versuch 15: Nutzer möchte sein eigene Geschäftsadresse an Franziska Kannengießler schicken

Nutzer: Zeige mir meine Geschäftsadresse.

Siri : Hier ist deine Geschäftsadresse (zeigt alle meine Adressen)

Nutzer: Schicke Franzi meine Geschäftsadresse.

Siri : Hier ist deine Nachricht an Franziska Kannengießler „Meine Geschäftsadresse“



## 4.1.2 Analyse

- Direkte ID

Das Siri-System in dem Offline-Dienst Telefonbuch von Apple besitzt die Fähigkeit Kontakte (Objekte) zu identifizieren, allerdings nur über ihre Vor- und Nachnamen. Das heißt dass die Erkennung von Personen ohne diese beiden Bestandteile des Namens in Siri nicht ausführbar ist. (vorausgesetzt, dass mehrere Personen, die den gleichen Vor- bzw. Nachname haben, in der Kontaktliste stehen). Wie man in den Versuchen 1 und 2 bei der direkten Identifikation sehen kann, werden vom Nutzer verschiedene Attribute eingegeben. In dem Versuch 1 hat Siri den gewünschten Kontakt durch die Attribute Vor- und Nachname erkannt. In Versuch 2 werden hingegen nur Vorname und Adresse eingegeben. Siri kann hier das Attribut Adresse nicht erkennen und die gewünschte Kontaktperson war nicht identifizierbar. Nach der ersten Nutzereingabe fragt Siri, wenn die Eingabe noch nicht konkret ist, gezielt nach, welcher Nutzer genau gemeint ist und nennt gleichzeitig alle verfügbaren Kontakte mit Vor- und Nachnamen, die auf die Nutzereingabe zutreffen, so wie in den Versuchen 1 und 2. Der Nutzer muss entweder auf eine der vorgeschlagenen Kontakte tippen oder ihn benennen, sonst ignoriert Siri alle anderen Eingaben bzw. erkennt keine anderen Eingaben. In Versuch 2 versteht Siri nicht, dass der Nutzer die Wohnadresse der gewünschten Person eingegeben hat.

Mittels dieser beiden Versuche konnte festgestellt werden, dass eine direkte Identifikation eines Objektes innerhalb des Telefonbuchs nur über den Namen möglich ist. Das könnte in MWR eventuell wie folgt dargestellt werden( Abbildung 18):

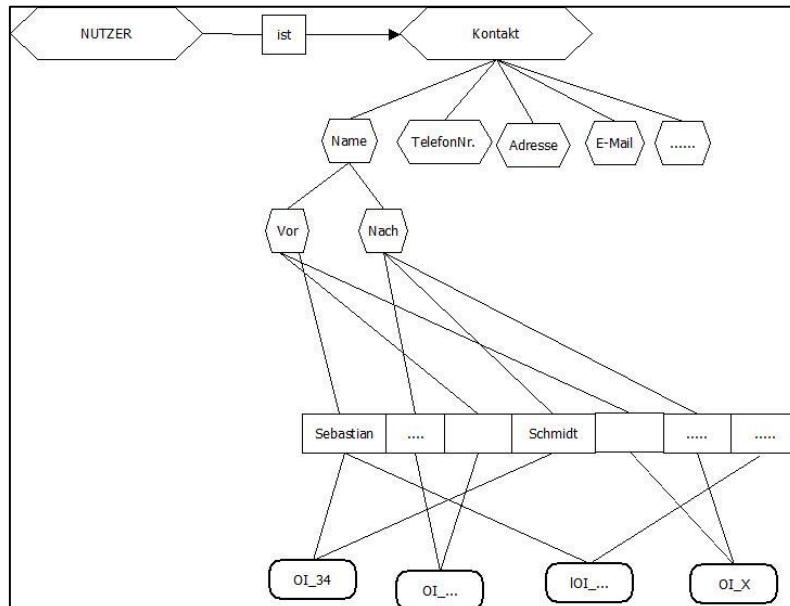


Abbildung 18: MWR für die Objekt-ID in Telefonbuch bei Siri

Nach der Identifikation von Objekten durch Vor- und Nachname oder Spitzname konnten alle anderen gewünschten Attribute aufgerufen werden, wie z.B in Versuch 4.

In Versuch 4 erkennt Siri die gewünschte Person und wird während des Dialogablaufs auch erkennen wer gemeint ist, wenn der Nutzer statt des Namens Pronomen nennt, wie z.B. sein oder mein, ihre usw..

- **Indirekte ID**

Bei der indirekten Objektidentifikation soll Siri über Beziehungen erkennen, wer gemeint ist, ohne die direkte Eingabe von Attributen. Die Versuche 5 - 7 überprüfen die Beziehung Nutzer-Nutzer, die Versuche 8-10 die Nutzer-Kontakt- Beziehungen und der Versuch 10 die Beziehungen Kontakt-Kontakt.

In Versuch 5 verlangt der Nutzer von Siri seine eigene Nummer anzurufen und Siri versteht, wer damit gemeint ist und versucht die Aktion durchzuführen, obwohl das nicht möglich ist. In diesem Fall sollte Siri durch eine Berichtigungstabelle (siehe Kap.2.3.1.3) überprüfen, ob eine Aktions-ID mit einer Objekt-ID zugelassen ist oder nicht. Dies hat Siri in diesem Versuch nicht durchgeführt. Siri erkennt, was der Nutzer mit „mein“ meint, wie z.B. in Versuch5:



(mich)mein Telefonnummer → Telefonnummer vom Nutzer, oder

in Versuch 6: Meine Privatadresse → Privatadresse vom Nutzer



In Versuch 7 hat Siri auch verstanden, was Nutzer mit „Zuhause anrufen“ meint und führte die Aktion Anrufen mit der Privatnummer vom Nutzer aus.

Zuhause → Privatadresse vom Nutzer



Im Telefonbuch von Apple konnte der Nutzer zugehörige Personen zu den Kontakten hinzufügen und Kontakte und Beziehungen miteinander verknüpfen. Der Nutzer hat zum Beispiel seine Frau, Vater, Mutter usw. und Kontakte haben auch Zugehörige die alle in der Kontaktliste zu finden sind. Die Versuche 8- 10 zeigen die Beziehungen zwischen dem Nutzer und seiner Kontaktliste. Beziehungen der Kontakte untereinander wurden mit Siri getestet, hier stellte sich die Frage, inwiefern und wie gut Siri diese Beziehungen identifiziert.



In Versuch 8 verlangt der Nutzer von Siri eine Nachricht an seinen Bruder zu schicken. Siri erkennt den Aktionswunsch (sms senden), obwohl der Nutzer seine Eingabe anderes formuliert hat: „Sag mein Bruder“. Dies kann man durch ein Art Generalisierung von Wortfolgen erreichen, mit der nicht nur genau die Worte oder Sätze, die im Datensystem sind, sondern auch eine gewisse Varianten

davon festgehalten werden kann. Für diesen Versuch kann das Aktionsmodell in MWR dargestellt werden (Abbildung 19).

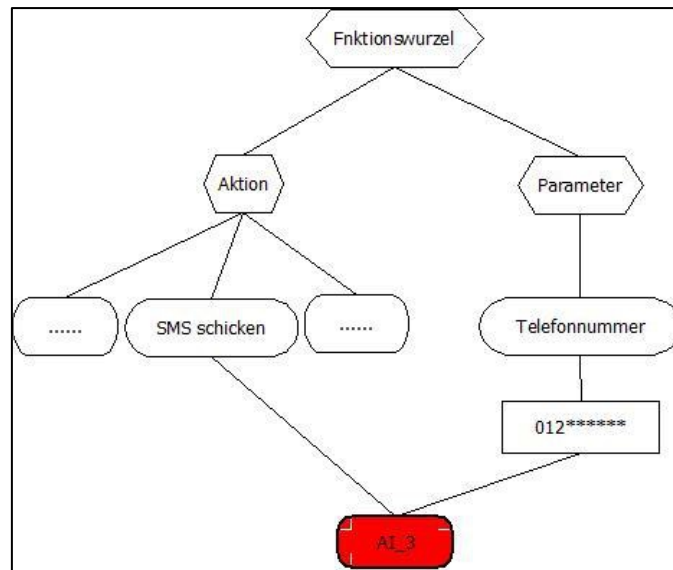


Abbildung 19: Aktion-ID für das Telefonbuch bei Siri

Bei der Objekt-Identifikation hat Siri auch erkannt, was der Nutzer mit „mein Bruder“ meint und konnte die Gewünschte Person identifiziert.



Diese Objekt- Identifikation konnte auch in MWR dargestellt werden (Abbildung 20).

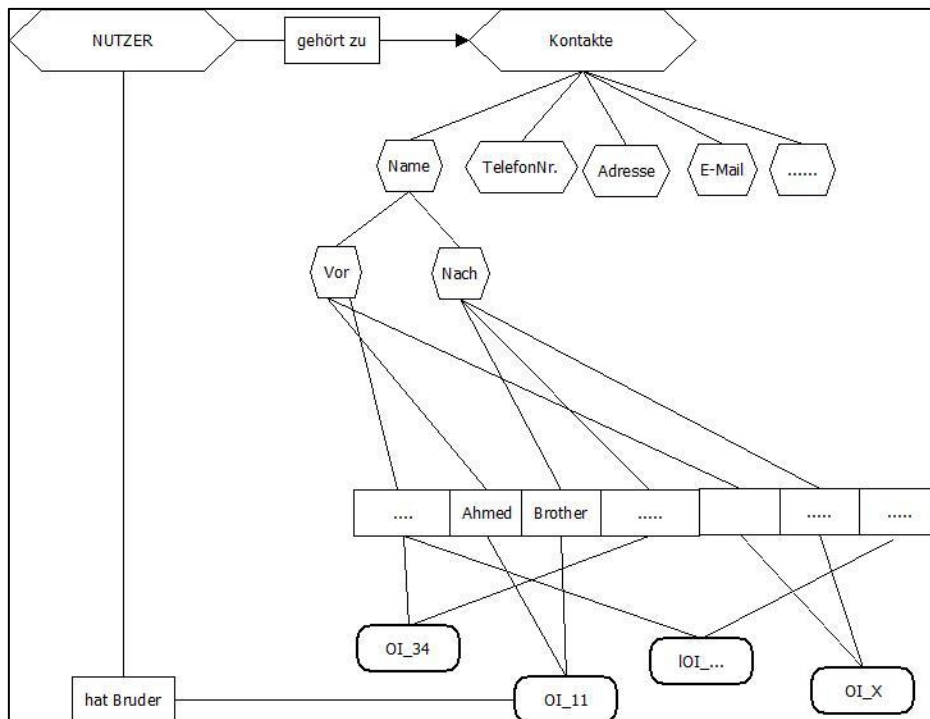


Abbildung 20: MWR bei indirekter Objekt-ID für das Telefonbuch bei Siri

Die gewünschte Aktion wurde hier mit der gewünschten Person verknüpft und die Aufgabe ausgeführt. In den Versuchen 9 und 10 hat Siri auch erkannt was der Nutzer mit „meine Frau“ „und mein Chef“ gemeint wird und führte die entsprechende Aufgabe aus.

Geburtstag meiner Frau: „**meine**“ ist der Bezug zum **Nutzer**, dessen **Frau** am.... **Geburtstag** hat

Meinen Chef anrufen: „**meine**“ ist der Bezug zum **Nutzer**, dessen **Chef**, die **Telefonnummer** .... hat.

In den Versuchen 8-10 sieht man, dass Siri alle Beziehungen, die zum Nutzer gehören (Nutzer-Kontakt) erkennt und identifizieren kann.

In dem Versuch 11 möchte der Nutzer diesmal die Chefin seine Frau anrufen und teilt das Siri mit. Siri wirkt in diesem Fall nach der Aktionsidentifikation (Anrufen) unsicher und fragt den Nutzer noch einmal, mit wem er sprechen möchte. Nach der Bestätigung des Wunsches durch den Nutzer, ruft Siri nicht die Chefin von der Frau sondern den Chef des Nutzers an.



Bei dem Versuch Siri zu korrigieren, in dem man nochmal sagt wer angerufen werden soll, versteht Siri entweder nicht, was der Nutzer möchte, oder versucht wiederholt den Chef des Nutzers anzurufen.

In diesen Versuch erkennt man, dass Siri die angegebenen Beziehungen nur auf den Nutzer bezieht und nicht auf andre Kontakte.

- **Kombination- ID:**

In den Versuchen 12 bis 15 werden die Nutzereingaben bei einem Aktionswunsch direkt und indirekt ausgeführt, also aus Kombination aus den beiden obigen Varianten. In Versuch 12 möchte der Nutzer den Mann seiner Schwester anrufen, nennt aber zuerst nur dessen Vorname. Wie bei der direkten Identifikation findet Siri mehrere Kontakte mit dem gleichen Vornamen und fragt den Nutzer, wer genau gemeint ist. Dieses Mal äußert der Nutzer, dass diese Person der „Ehemann“ oder „Mann“ von seiner Schwester ist. Jetzt erkennt Siri nicht was der Nutzer damit meint und sucht in dem Adressbuch nach dem Wort „Mann“.



Auch in den Versuchen 13 und 14 ist Siri mit den Nutzereingaben überfordert und kann die gewünschte Kontaktperson nicht identifizieren. In Versuch 13 ruft Siri anstelle des Vaters von Nancy entweder Nancy selbst oder den Vater des Nutzers an. im Versuch 14 zeigt Siri entweder die Daten von dem Chef des Nutzers, oder sagt, dass sie nicht weiß, wer die Managerin des Nutzer ist, obwohl das nicht die gewünschte Kontaktperson ist.

In Versuch 15 ist eine Kombinationsidentifikation mit zwei verschiedenen Objekten erforderlich. Der Nutzer möchte hier dass Siri dem einen Objekt „meine Geschäftsadresse“ per Nachricht an das andre Objekt „Kontaktperson“ schickt. Siri erkennt den Aktionswunsch des Nutzers „SMS schicken“ und ebenso, an wen das



geschickt werden soll. Allerdings schickt sie den Satz „ Meine Geschäftsadresse“ und nicht, wie Gewünscht, die Geschäftsadresse vom Nutzer, welche Siri in der Regel bekannt ist.

## 4.2 Dialog mit Online-Diensten

Online Dienste sind im Grunde alle Dienste oder Apps, die ihre Informationen und Daten von ihrem eigenen Dienstserver an das Smartphone weiterleiten können z.B. Wikipedia oder Yahoo und für unseren Fall müssen diese Dienste auch von Siri unterstützt werden. Nach und nach werden mehrere Onlinedienste in Deutschland zur Verfügung stehen, die mit Siri zusammen arbeiten können. Alle Apps in der Tabelle 3.1Kapitel 3), die mit (O) markiert sind, sind die Onlinedienste die nach dem letzten Update vom Betriebssystem von Apple IOS 6 Siri unterstützen.

Als ein relevantes Beispiel für die Online Dienste wird der Yahoo Sport-Dienst untersucht, da Siri mit den anderen Onlinediensten wenig Aktionen durchführen kann. Yahoo Sport hat für das Apple Betriebssystem die App. *Yahoo Sportacular* entwickelt.

### 4.2.1 Sport (Yahoo Sportacular):

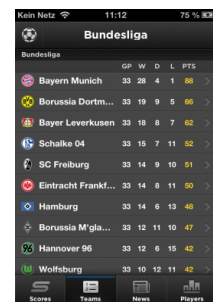
Yahoo Sportacular wird von dem amerikanischen Unternehmen Yahoo Sport für Smartphone wie iPhone oder Samsung entwickelt. Englisch ist die Sprache dieser Anwendung und der Nutzer dieser Anwendung soll laut der Entwickler schnell und einfach alle aktuellen Ergebnisse, Spiel Tabellen und Spielpläne sowie Statistiken und aktuellen Nachrichten bekommen. Diese Anwendung soll nicht nur Fußballbetreffen, sondern auch Baseball, Basketball, Hockey und viele mehr, allerdings sind die meisten diese Sportarten, die diese Anwendung anbietet nur auf die amerikanischen Ligen ausgelegt. Für den Fußball gib es internationale League wie die Italienische, die Spanische oder die Bundesliga und auch die Champions League. Die Struktur dieser Anwendung ist in Ligen unterteilt, wobei jede Liga Mannschaften enthält und für jede Mannschaft Informationen über die Spieler, aktuellen Spiel- Ergebnisse und viel mehr angezeigt wird.

Siri kann diesen Dienst unterstützen und soll laut des Handbuches von Siri folgende Aktionen durchführen:

- Ergebnisse von Spiele anzeigen
- Spielpläne (wo/wann) zeigen
- Spielinformationen zeigen
- Teaminformation zeigen

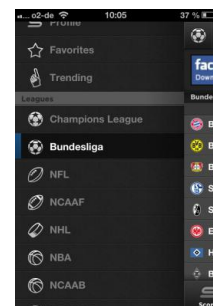
Mit Hilfe dieses Dienstes und anhand von ein paar Versuchen soll nun die Fähigkeit von Siri auf Onlinedienste überprüft werden. Anschließend soll ein passendes ER-Diagramm erstellt und dies anhand der Versuchsergebnisse bewertet werden.

Um einen Überblick über die Struktur dieses Dienstes zu geben, wird zuerst ein passendes ER-Diagramm entwickelt.



The screenshot shows the Bundesliga league table with columns for GP, W, D, L, and PTS. The teams listed are Bayern Munich, Borussia Dortmund, Bayer Leverkusen, Schalke 04, SC Freiburg, Eintracht Frankfurt, Hamburg, Borussia M'gladbach, Hannover 96, and Wolfsburg.

Team	GP	W	D	L	PTS
Bayern Munich	33	28	4	1	88
Borussia Dortmund	33	19	9	5	66
Bayer Leverkusen	33	18	8	7	62
Schalke 04	33	15	7	11	62
SC Freiburg	33	14	9	10	61
Eintracht Frankfurt	33	14	8	11	60
Hamburg	33	14	6	13	48
Borussia M'gladbach	33	12	11	10	47
Hannover 96	33	12	6	15	42
Wolfsburg	33	10	12	11	42





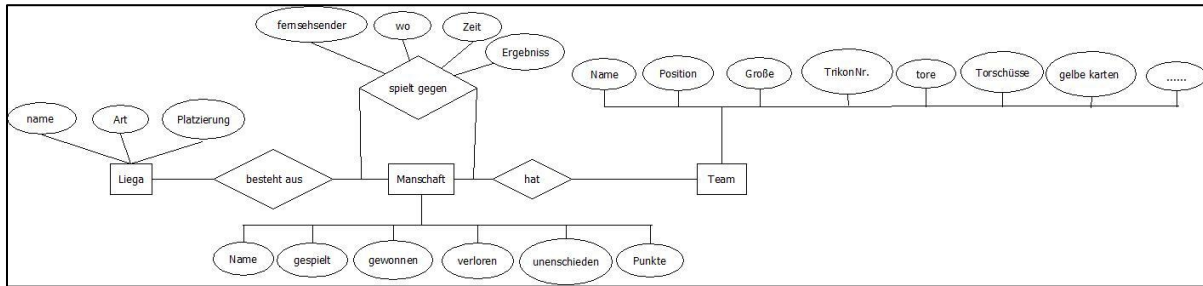


Abbildung 21: Weltmodell für den Onlinedienst Yahoo Sportacular

Wie in der Abbildung 21 hat dieser Dienst 3 Objekte (Liga, Mannschaft und Team), wobei jedes Objekt mit seinen Attributen beschriftet ist.

Die Objekt- und Aktionsidentifikation in diesem Dienst basiert auf den verfügbaren Informationen, die von dem Serverbetreiber des Diensts für Siri zur Verfügung gestellt werden. Siri wird die gefundenen Informationen darstellen. Welche Informationen Siri zeigen kann und welche nicht, das soll anhand der Versuche gezeigt werden.

Bei den Beispielen zum Offline-Dienst hatte man bei der Durchführung der Versuche zwischen direkter und indirekter Identifikation unterschieden, da diese Unterscheidung die Semantik für Siri unter die Lupe nimmt. In diesem Beispiel wird auch versucht, die semantische Ebene von Siri zu untersuchen und die Fähigkeit, indirekte Angaben zu erkennen und zu verstehen. In dem Siri-Handbuch für diesen Dienst gibt es nur Beispiele für direkte Angaben um möglichst ausführliche Informationen bei der ersten Frage oder dem ersten Wunsch zu geben. In den nächsten Test-Versuchen werden sowohl direkte wie auch indirekte Informationen angegeben um zu schauen, wie Siri damit umgeht.

- **Direkte Identifikation**

### Versuch A: Ergebnisse von Spielen

#### A-1

Nutzer: Wie hat Bayern letzte Woche gespielt?

Siri : Die Mannschaft Bayern München hat am letzten Samstag die Mannschaft FC Augsburg mit 3 zu 0 besiegt.

#### A-2

Nutzer: Hat Schalke das letzte Spiel gewonnen?

Siri : Schalke 04 hat am letzten Samstag verloren gegen den VfB Stuttgart gespielt. Das Spiel ging 1 zu 2 aus.

#### A-3

Nutzer: Zeig mir die aktuellen Ergebnisse von der der Champions League(Bundesliga)!

Siri : Leider habe ich keine Informationen über „Champions League“. Tut mir leid! **Oder:**

Siri : (Hier die Ergebnisse der Bundesliga vom letzten Samstag)

A-4

Nutzer: Was sind die Spielergebnisse vom letzten Wochenende?

Siri : Ich verstehe „was sind die Spielergebnisse von letzter Woche“ nicht. Soll ich im Internet danach suchen? (Oder): Ich fürchte ich kann das nicht beantworten.



## Versuch B: Spielpläne

B-1

Nutzer: Wann spielt Borussia Dortmund gegen Bayern München?

Siri : Das Spiel Bayern München- Borussia Mönchengladbach ist übermorgen um 15:30.



B-2

Nutzer: Wann ist das nächste Spiel in der Bundesliga?

Siri : So sieht das in der Bundesliga von übermorgen aus.

B-3

Nutzer: Welche Fußballspiele werden in Fernsehen übertragen?

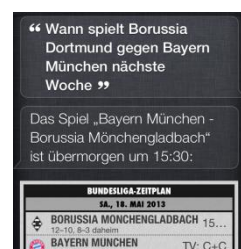
Siri : Hier sind die Ergebnisse der Bundesliga vom letzten Samstag. (zeigt alle Ergebnisse)

Nutzer: Welche Spiele werden nächste Woche im Fernsehen übertragen?

Siri : Welche Liga? MLB,NBA,WNBA,MLS oder NHL?

Nutzer : MLB.

Siri : So sieht es in der MLB von diesem Montag bis 29.Mai 2013 aus. (zeigt alle Spielinformationen der Liga in der genannte Zeit)



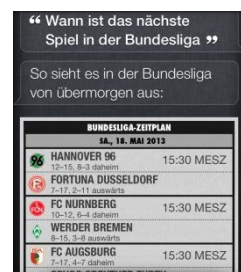
B-4

Nutzer: Welche Sender überträgt das Spiel Dortmund gegen Hoffenheim?

Siri : Entschuldigung, ich habe keine Informationen für diesen Spielplan.

Nutzer: Welche Sender überträgt das Spiel Dortmund gegen Hoffenheim nächste Woche?

Siri : Das Spiel 1899 Hoffenheim – Borussia Dortmund wird auf MD2 gezeigt.



## Versuch C: Spielinformationen

### C-1

Nutzer: Wer hat die meisten Tore in der Champions League geschossen?

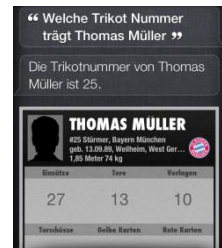
Siri : Leider habe ich keine Informationen zur Champions League. Tut mir Leid.



### C-2

Nutzer: Wie viel Tore hat Dortmund in dieser Saison erzielt?

Siri : Die Mannschaft Borussia Dortmund hat 80 Tore.



### C-3

Nutzer: Welche Fußballmannschaft hat in dieser Saison die meisten gelben Karten erhalten?

Siri : Welche Fußball- Liga?

Nutzer: Bundesliga.

Siri : Tut mir Leid Mohammed, das weiß ich nicht.

Nutzer: Wer hat bis jetzt in dieser Fußballsaison die meisten gelben Karten erhalten?

Siri : Welche Fußball-Liga?

Nutzer: Bundesliga.

Siri : Julian Schuster führt diese Saison mit 11 gelben Karten die Liga an.



### Versuch D: Teaminformation

#### D-1

Nutzer: Welcher Spieler in Bayern trägt die Trikotnummer 25?

Siri : Hier ist die Aufstellung für die Mannschaft Bayern München.



#### D-2

Nutzer: Welche Trikotnummer trägt Thomas Müller?

Siri : Die Trikotnummer von Thomas Müller ist 25.



#### D-3

Nutzer: Wer ist der Torwart von Schalke 04?

Siri : Das hier sind die Torwarte für die Mannschaft Schalke 04. (zeigt alle Torwarte in Mannschaft)

#### D-4

Nutzer: Wer ist der Trainer von Dortmund?

Siri : Keine Ahnung wer gerade Borussia Dortmund trainiert.

- **Indirekte Identifikation**

### **Versuch 1**

Nutzer: In welcher Position spielt Thomas?

Siri : Thomas von welcher Mannschaft? (zeigt alle Spieler Namen von Thomas)

Nutzer: Er spielt bei meiner Favoriten Mannschaft.

Siri : Tut mir leid Mohammed, ich verstehe „er spielt bei meiner Favoriten Mannschaft“ nicht.  
Thomas von welcher Mannschaft?

### **Versuch 2**

Nutzer: Wo spielt der Bruder von Lars Bender?

Siri : Derzeit spielt Lars Bender für die Mannschaft Bayer Leverkusen.

Nutzer: Und wo spielt sein Bruder?

Siri : Lars Bender? Für die Mannschaft Bayer Leverkusen.

Nutzer: Nein, sein Bruder.

Siri : Wie heißt dein Bruder?

## **4.2.2 Analyse**

### **Direkte Identifikation**

in den Versuchen **A** soll Siri dem Nutzer Informationen über Spielergebnisse einer Mannschaft oder Liga mitteilen. Hier werden vom Nutzer verschiedene Fragen gestellt und Siri soll diese Fragen mithilfe des Onlinedienstes beantworten oder die Nutzerfrage hinterfragen, wenn die Frage nicht verstanden wurde oder unübersichtlich war. In Versuch A-1 und A-2 hat Siri die Nutzerwünsche verstanden und zeigte ihm die letzten Spielergebnisse für die Mannschaften FC Bayern und Schalke 04 an. In Versuch A-3 sollte Siri die letzten Spielergebnisse der Champions League bzw. Bundesliga anzeigen. Auch in diesem Versuch hat Siri verstanden, was der Nutzer für Informationen wollte, antwortete aber, dass sie für die Champions League keine Informationen hat, obwohl der Onlinedienst über solche Informationen verfügt. Für die Bundesliga zeigte Siri alle Spielergebnisse vom letzten Spieltag an. In dem Versuch A-4 sollte Siri den Nutzerwunsch hinterfragen, um nach den passenden Informationen suchen zu können, Siri versteht hier allerdings nicht, was der Nutzer möchte und fragt auch nicht nach, was damit gemeint ist. Das Programm sagt, dass es die Frage nicht verstanden hat.

In Versuch **B** überprüft der Nutzer Siris Fähigkeit zur Ermittlung der richtigen Spielpläne mit Hilfe des Onlinedienstes. In B-1 zeigt Siri nicht die gewünschte Information an (statt Bayern gegen Borussia Dortmund zeigt Siri den Spielplan für Bayern gegen Borussia Mönchengladbach). In Versuch B-2 zeigt

Siri auf Wunsch des Nutzers alle kommenden Spiele der Bundesliga für die nächste Woche an. In B-3 soll Siri laut ihres Handbuchs wissen, in welchem Fernsehsender ein Spiel gezeigt wird. Als erstes fragt der Nutzer, wo die Fußballspiele übertragen werden und Siri versteht anscheinend die Frage nicht richtig und zeigt direkt die letzten Ergebnisse der Bundesliga an. Bei dem nächsten Versuch fragt der Nutzer allgemein welche Spiele im Fernsehen gezeigt werden und dieses Mal versteht Siri was damit gemeint ist, fragt nach der gewünschten Liga und zeigt dann alle verfügbaren Informationen für die Liga an. In B-4 gibt der Nutzer als erstes weniger Informationen an, er fragt, wer das Spiel Dortmund gegen Hoffenheim überträgt. Siri antwortet, dass sie das nicht weiß. Beim zweiten Versuch gibt der Nutzer genau die gleiche Information aber mit einem Detail mehr „nächste Woche“. Siri versteht daraufhin den Nutzerwunsch und zeigt die passende Information dazu.

In Versuchen **C** möchte der Nutzer von Siri Informationen über ein bestimmtes Spiel oder eine bestimmte Liga verlangen. Bei C-1 kann Siri keine Auskunft über die gewünschte Liga geben und sagt, dass sie dafür keine passende Informationen hat, obwohl die Champions League in dem Onlinedienst verfügbar ist. In C-2 hingegen geht es um die Bundesliga und deshalb findet Siri hier die passende Antwort. Beim Versuch C-3 stellt der Nutzer die Selbe Frage aber in verschiedene Formulierungen (Sätze), beim ersten Mal (1. Satz) versteht Siri die Frage und verlangt weitere Informationen um die Frage einschränken zu können. Nach der Nutzereingabe sagt Siri allerdings, dass sie die Antwort nicht weiß. Beim 2.Satz/ Versuch reagiert Siri wie beim ersten Satz und fragt nach der genauen Fußballliga. Nach der Nutzereingabe zeigt Siri hier die gewünschten Informationen an.

In Versuchen **D** werden Informationen über das Team und sein Mitglieder verlangt. Bei D-1 zeigt Siri nicht die gewünschte Information, sondern die gesamte Mannschaftsaufstellung. Bei dem Versuch D-2 wird hingegen nur die Information gezeigt, die der Nutzer braucht. In dem Versuch D-3 zeigt Siri nicht den einen Torwart der Startelf an, sondern alle Torwarte der Mannschaft. In D-4 sagt Siri, dass sie nicht weißt, wer die Mannschaft trainiert.

### **Indirekte Identifikation**

Für indirekte Identifikationen wurden nur 2 Versuche durchgeführt. Beim ersten sollte Siri wissen, welche Mannschaft die Lieblingsmannschaft des Nutzers ist, da sie bereits in dem Dienst angegeben wurde. Der Nutzer möchte wissen, in welcher Position ein Spieler spielt, wobei er allerdings nicht den Namen der Mannschaft angibt. Er sagt jedoch, dass die Mannschaft einer seiner Lieblingsmannschaften ist. Beim zweiten Versuch soll Siri den Bruder eines Fußballspielers erkennen, der auch in der Bundesliga spielt und dem Nutzer sagen, wo er spielt. In beiden Versuchen hat Siri nicht erkannt, was der Nutzer möchte.

## 5. Bewertung der Güte der Modellierung für Siri bei den getesteten Diensten

In dem vorherigen Kapitel wurden zwei Dienste mit Siri getestet und je das passende Modell für jeden Dienst dargestellt.

Für den Offlinedienst „Telefonbuch“ von Apple besitzt Siri viele Möglichkeiten um ein Objekt zu identifizieren, nutzt aber nur begrenzte Informationen von der Datenbank des Dienstes.

Bei der **direkten** Angaben von Attributen erkennt Siri den gewünschten Kontakt nur, wenn **Namen** genannt werden z.B Vorname, Nachname oder Spitzname. Andere Attribute wie z.B. Adresse, Email oder Geburtstag sind bei der O-ID nicht relevant und helfen nicht bei der Suche nach der gewünschten Kontaktperson. Bei mehreren Auswahlmöglichkeiten, nach den Nutzerangaben, nennt Siri alle Kontakte der zur Auswahl gehörenden Namen und verlangt, dass der Nutzer den gewünschten Namen nennt oder auf diesen Namen per Finger tippt und erkennt. Daher schließt Siri alle anderen Angaben des Nutzers aus und erkennt keine anderen Angaben außer diese beiden Möglichkeiten an.

Bei **indirekten** Angaben von Informationen durch **Beziehungen** bezieht sich Siri nur auf den Nutzer, wobei der Nutzer und seine Daten zu den Kontakten gehören müssen. Siri weißt, wer der Vater des Nutzers ist, sie weißt allerdings nicht, wer der Vater von anderen Kontakten ist, obwohl diese Daten bereits bei den Kontaktdaten hinterlegt sind.

Siri fragt nach, wenn sie nicht versteht, was der Nutzer möchte und identifiziert bei indirekten Angaben nur die Kontakte die zu dem Nutzer gehören. Wenn jedoch Beziehungen zwischen Kontakten vom Nutzer angegeben sind, erkennt Siri diese entweder nicht oder verbindet die Beziehung mit dem zugehörigen Nutzer, so ruft es zum Beispiel anstatt des Vaters von Sebastian, den Vater des Nutzers an.

Wenn der Nutzer seine Anregungen oder Fragen aus einer **Kombination** (von direkten und indirekten Angaben) an Siri äußert, werden auch nur Objekte identifiziert, die in einer Beziehung zum Nutzer stehen. Siri schließt hier nach der ersten Nutzerangabe auch alle anderen möglichen Nutzerangaben aus und führt einen Dialog aus, in dem es meist selbst bestimmt, was es vom Nutzer hören möchte.

Für die Aktionsidentifikation hatte Siri eine Menge von formulierten Sätzen in einer vorherigen Trainingsphase gewonnen, die dazu führten die Aktionswünsche in verschiedenen Formen zu verstehen. Siri versteht wenn der Nutzer statt „rufe meinen Bruder an“ z.B. „ich möchte mit meinem Bruder sprechen“ sagt. Desweiteren kann der Nutzer „schick Sebastian eine Nachricht“ .... , sowie „Sag Sebastian.....“ oder „sag ihm .....“ sagen. Man könnte auch statt „Meine Privatadresse anrufen“ „Zuhause anrufen“ sagen. Dies passiert aufgrund von Generalisierung von Wortfolgen und Sätzen und ermöglicht es Siri bei verschiedenen Formulierungen die Aktionswünsche durchzuführen[19]. Tabelle 1 zeigt ein paar Sätze und Formulierungen die Siri für Die A\_ID und O\_ID verwendet

Aktion/Objekt	Satz oder Formulierung
Anrufen	<ul style="list-style-type: none"> <li>- Rufe (bitte) &lt;O_ID&gt; an</li> <li>- Ich möchte mit &lt;O_ID&gt; sprechen</li> <li>- Wähle die Nummer von &lt;O_ID&gt;</li> </ul>
SMS schicken	<ul style="list-style-type: none"> <li>- Schicke &lt;O_ID&gt; eine SMS(oder Nachricht)</li> <li>- Sage &lt;O_ID&gt;</li> <li>- Sende &lt;O_ID&gt;</li> <li>- SMS an &lt;O_ID&gt; und &lt;O_ID&gt;</li> </ul>
E-Mail senden	<ul style="list-style-type: none"> <li>- Emaille &lt;O_ID&gt; wegen &lt;Betreff&gt;</li> <li>- Sende ein E-Mail an &lt;O_ID&gt; wegen &lt;Betreff&gt;</li> </ul>
Privatadresse	<ul style="list-style-type: none"> <li>- Zuhause</li> </ul>
Manager	<ul style="list-style-type: none"> <li>- Chef</li> <li>- Boss</li> </ul>
Ehepartner	<ul style="list-style-type: none"> <li>- Frau</li> <li>- Mann</li> </ul>

Tabelle 4: Beispiele für die Generalisierung von Sätzen und Wörtern

Siri besitzt im Telefonbuchdienst keine deutliche Berichtigungstabelle für die Aktions-/Objektidentifikation (siehe Kap. 2.3.2.3). Die letzten Versuche veranschaulichen, dass Siri eine andere Strategie zum Verhindern von falschen Angaben verfolgt. Hier spielen die gegenseitige Steuerung der Identifikation von Objekten und Aktionen und auch die Abhängigkeit bezüglich dessen, was zuerst verstanden wurde, eine große Rolle.

Anhand dieser beiden Komponenten kann Siri einordnen, welche Aktionen zu welchen Objekten passen und daraus wird Siri eine möglichst passende Antwort generieren. Ein Beispiel dafür ist der Versuch 2. Als der Nutzer eine Adresse von der gewünschten Kontaktperson nennt, die er anrufen möchte, dachte Siri, dass der Nutzer seine Adresse ändern möchte. Ein anderes Beispiel: In Versuch 12 versteht Siri die Aktion „Anrufen“, weißt aber nicht was mit „Mann von meiner Schwester“ gemeint ist und versucht in der Kontaktliste die Kontaktperson Mann zu finden. Andere Beispiele sind die Versuche 13 und 14. In Versuch 5 begeht Siri einen Fehler, indem sie versucht die eigene Nummer vom Nutzer anzurufen, was beweist, dass Siri keine definierte Berichtigungstabelle hat.

Beim letzten Versuch bestand das Ziel darin verschiedene Objekte zu identifizieren und sie gleichzeitig miteinander zu verkoppeln. Im Textinhalt einer Nachricht, der an eine Kontaktperson verschickt werden sollte, sollte die Geschäftsadresse vom Nutzer stehen und nicht der Satz „meine Geschäftsadresse“ wie in dem Versuch 15. In diesem Versuch wird Siri nach der Identifikation von einer Aktion (Sag → SMS schreiben) und der ersten Objekte- ID (Franzi → Franziska Kannengießler) alle andere Nutzerangaben nur als reinen Textinhalt für die SMS-Nachricht betrachten.

Insgesamt hat Siri für das Telefonbuch von Apple einen begrenzten Zugang auf die Datenbank für die ersten Identifikationen von Objekten. Siri benötigt für die Identifikation eines Kontaktes entweder einen Namen oder eine Beziehung, die durch eine Verknüpfung zwischen den Kontakten im Telefonbuch entsteht. Diese Beziehungsidentifikation im Siri-System gilt nur für den Nutzer und nicht für alle Kontakte. Durch eine Generalisierung von Wortfolgen im Datenwert ermöglicht Siri es dem Nutzer seine Wünsche und Anregungen Siri auf unterschiedliche Weise mitzuteilen, allerdings muss das in einer vorherigen Trainingsphase in das System integriert werden. Im Vergleich zu Abbildung 17 hat Siri bei dem Telefonbuch-Dienst das folgende Weltmodell (Abbildung 22).

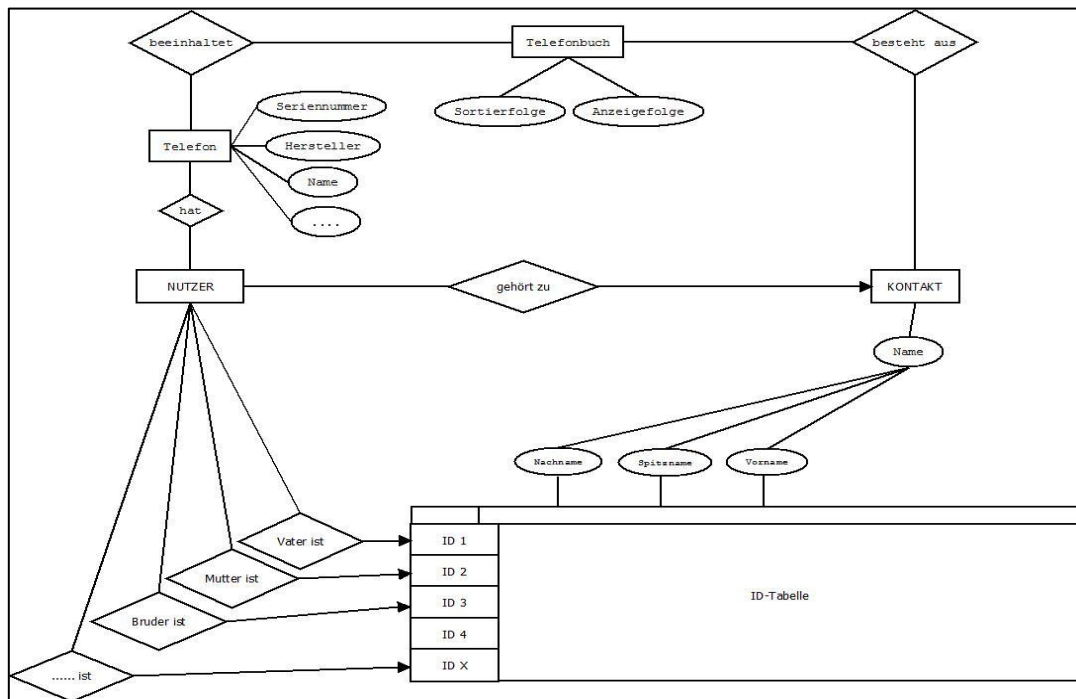


Abbildung 22: Weltmodell für Siri-System für Das Telefonbuch-Dienst

Bei den Online-Diensten hat Siri einen begrenzten Zugriff auf die Datenbank des Dienstes, es hat daher Informationen zur Bundesliga und anderen internationalen und amerikanischen Ligen, aber z.B. keine Informationen zur Champions League (Versuche A-3, C4) und keine vollständigen Informationen über einzelne Mannschaften (Versuch D-4).

Bei direkten Angaben von Attributen werden die gewünschten Objekte erst identifiziert, wenn alle nötigen Informationen vom Nutzer deutlich auf das gewünschte Objekt gerichtet sind und es keine Verwirrungen gibt (siehe Versuch A-1, A-2, B-2, C-2). Am besten sollten die Fragen so ähnlich wie in dem Handbuch von Siri vorgegeben, gestellt werden (siehe Versuch C-3 und D-2). Bei einer Mehrdeutigkeit oder Verwirrung verlangt Siri meistens keine weiteren Informationen um den Nutzerwunsch einzugrenzen, sondern gibt eine falsche Antwort (siehe Versuch B-1, B-3 und D-1). Teilweise gibt das System zu, dass es die Frage nicht versteht (siehe Versuch A-4 und B-4). Bei allen Versuchen verlangt Siri zur Objekt-Identifikation nur die Namen (Liga, Mannschaft, Spieler) alle anderen Informationen spielen für die Identifikationen keine Rolle.

Für indirekte Identifikationen eines Objektes wurden zwei Tests durchgeführt. Beide Tests verdeutlichen, dass diese Online-Dienste keine indirekten Identifikationen von Objekten ermöglichen. Es werden anscheinend zwischen den Objekten keine Beziehungen aufgebaut, sie sind nur durch Attribute identifizierbar.



## 6. Zusammenfassung und Ausblick

Die vorliegende Studienarbeit hat die semantischen Daten und Informationen eines Sprachverarbeitungssystem untersucht und analysiert. Es wird am Beispiel des Siri-Systems gezeigt, wie weit ein aktuelles Sprachsystem semantisch denken kann und welche Schritte es übernimmt, um die nötigen Daten während eines Dialogaufbaus erfassen zu können. Siri ist ein Sprachsystem, welches dem Menschen als Assistent dient und durch die Integration von Offline-/ sowie Online-Diensten dem Nutzer Zeit und Arbeit sparen soll.

Nach den kurzen theoretischen Grundlagen über Sprachsysteme zu Beginn des zweiten Kapitels wurde die Vorgeschichte von Siri vorgestellt und Gründe für die Entwicklung solcher Systeme wurden aufgeführt. Die Entwickler versuchten mit Siri und durch die Modellierung von Aufgaben und das Nutzen von Kontexten (z.B. Ort, Zeit usw.), dass Nutzer- Absichten verstanden werden können. Wie dies funktioniert, wurde in dieser Arbeit untersucht.

Vor der Untersuchung wurde ein mögliches Modell für Siri entworfen und die System-Nutzer-Ebene deutlicher dargestellt. Beide Ebenen wurden einzeln beschrieben und benötigen einen Internetzugang um miteinander arbeiten zu können.

Danach wurde die semantische Modellierung beschrieben und mit Hilfe von ER-Diagrammen ein Weltmodell für einen beliebigen Dienst, der mit Siri zusammen arbeiten kann, entworfen um zu analysieren, wie die Daten in semantischen Ebenen organisiert werden.

Anhand des beliebig gewählten Beispiels Telefonbuch wurde gezeigt, wie ein semantisches System in der Regel arbeiten soll. Mit Hilfe einer Struktur und eines Funktionsmodelles des dargestellten Weltmodelles wurde überprüft, ob die Identifikation von Objekten und Aktionen direkt sowie indirekt erläutert und beschrieben werden.

Die Untersuchung des Siri-Systems im Rahmen dieser Arbeit hat ergeben, dass Siri nur einen begrenzten Zugriff auf die Datenbank des jeweiligen Dienstes hat und nicht alle verfügbaren Informationen, die schon in der Datenbank gespeichert sind, nutzen kann. Aus vorher genannten Gründen kann Siri das semantische Denken nicht immer erfüllen.

Für die direkte Identifikation eines gewünschten Objektes durch die Attributeingaben seitens des Nutzers fordert Siri nur einen Namen zur Identifikation an und schließt alle anderen Attribute aus. Siri weiß z.B. wer Sebastian Schmidt ist, jedoch nicht wer Sebastian, der in der Universitätsstr.10 wohnt, ist. Sowohl für das Offline- Dienst- Beispiel (Telefonbuch), wie auch bei dem Online-Dienst- Beispiel (Yahoo Sportacular) und nach der ersten Nutzeräußerung, kontrolliert Siri den Dialog und bestimmt was es vom Nutzer hören möchte durch eine im Bildschirm gezeigte Wahlliste.

Anhand von indirekten Identifikationen von Objekten durch Beziehungen, erkennt Siri nur die Beziehungen die zum Nutzer gehören und kann alle anderen Informationen von Beziehungen nicht verarbeiten. Siri erkennt z.B. durch eine Verknüpfung eine Beziehung zwischen einem Nutzer und einem Kontakt, so erkennt Siri wer der Vater vom Nutzer ist, kann jedoch nicht darauf zurückgreifen, wer der Vater von der Frau des Nutzers ist, obwohl es sich theoretisch um das gleiche Prinzip handelt. Bei dem Offline-Dienst kann man das deutlich erkennen, da diese Beziehungen vorab in der Datenbank vom Nutzer eingetragen werden. Für Online- Dienste kann Siri nur auf die Daten zugreifen die von diesem Dienstserver angeboten werden.

Gut hat das Siri-System bei der großen Auswahl von Formulierungen der Aktionswünsche durch die Generalisierung und Formulierung von Sätzen und auf Grund der Lernkonzepte abgeschnitten. Mit jeder neuen Eingabe seitens des Nutzers verarbeitet und speichert Siri diese Eingaben auf ihrem Server. Dies ermöglicht eine Wiederverwendung der Daten, ebenfalls für andere Nutzer. Eine erkennbare Berichtigungstabelle hat Siri nicht, sie befolgt ein anderes Konzept um falsche Angaben zu verhindern. Wenn Siri bei dem Nutzerwunsch z.B. eine Aktions-ID verstanden hat, sucht sie die passende Objekt-ID für diese Aktion und umgekehrt.

Mit diesen Testergebnissen ist deutlich geworden, dass Siri einen geführten Dialog bietet, der während eines Dialogablaufes die Kontrolle übernimmt. Aktionen versteht es meist und Objekte werden durch die Dialogkontrolle von Siri nur mit Namen identifizierbar. Wenn Siri nicht sicher ist oder den Nutzerwunsch nicht versteht, fragt es noch einmal nach und wenn das System den Nutzerwunsch erneut nicht versteht (wie z.B. „den Chef von meiner Frau anrufen“), versucht es von den verfügbaren Informationen ( Aktion → „anrufen“, Objekt → „Frau“ und „Chef“) den Nutzerwunsch falsch zu erfüllen in dem es, entweder meine Frau oder meinen Chef anruft. Mit dieser Vorgehensweise wird deutlich, dass Siri die Aufgabe nicht semantisch löst sondern alle verfügbaren Beziehungen nur auf den Nutzer bezieht.

Siri ist ein aktuelles Spracherkennungssystem, das als intelligentes VPA vorgestellt wurde und in das ganz viel Arbeit und Aufwand gesteckt wurde, trotzdem bleibt es bis jetzt eine *Beta*-Version. In Juni 2009 hatte Tom Gruber (CTO & cofounder, Siri) auf der Semantic Technology Conference bei der Vorstellung von Siri festgestellt, dass Siri mit der Kombination mit einem Smartphone die folgende Kriterien erfüllen soll [3] :

- Zugriff auf Globale Netzwerke(Internet)
- Zeitliches und Örtliches Bewusstsein
- Dialogfähiges, kontinuierliches Sprachsystem
- Semantische Daten
- Aufgaben- und Domänen- Server und
- Dienste

Alle diese Kriterien treffen bei dem Siri-System zu. Aufgrund der Menge von Daten und übersichtlicher Datenmodellierung, kann Siri nicht semantisch denken. Es sollen hierbei alle verfügbaren Daten und Attribute zur Identifikation von Objekten im Siri benutzt werden, was mit Hilfe einer Merkmal- Werte- Relation, welche nicht nur Namen semantisch zuordnet, sondern auch alle wichtigen Daten und Attribute, geschafft werden kann. Für eine flexible und unabhängige indirekte Daten- Identifizierung hilft hier eine Beziehungen und Verknüpfungen nicht nur zwischen

dem Nutzer und seinen Daten, sondern auch zwischen den Daten untereinander. Das sollte bei Siri in den nächsten Versionen beachtet werden.

## Literaturverzeichnis

- [1] K.Karnagel, Masterarbeit(2012): semantisch Modellierung von Benutzer-System-Interaktion am Beispiel ein Home Entertainment-Anwendung
- [2] B.H. Juang & Lawrence R. Rabiner (2004): Automatic Speech Recognition – A Brief History of the Technology Development, [Online].  
URL:[http://pdf.aminer.org/000/339/833/intelligent\\_system\\_for\\_automatic\\_recognition\\_and\\_evaluation\\_of\\_speech\\_commands.pdf](http://pdf.aminer.org/000/339/833/intelligent_system_for_automatic_recognition_and_evaluation_of_speech_commands.pdf)  
[Zugriff am 13.06.2013]
- [3] Pfister & Kaufmann (2008): Sprachverarbeitung Grundlagen und Methoden der Sprachsynthese und Spracherkennung, Sprenger Verlag Berlin Heidelberg
- [4] M. Burgstaller, B. Gmeiner (2009) :Sprachbasierte Benutzerschnittstellen, [Online].  
URL:[http://www.pervasive.jku.at/Teaching/\\_2009SS/SeminarSEMMK09/Begleitmaterial/Schriftliche%20Ausarbeitungen/G04-Sprache.pdf](http://www.pervasive.jku.at/Teaching/_2009SS/SeminarSEMMK09/Begleitmaterial/Schriftliche%20Ausarbeitungen/G04-Sprache.pdf)  
[Zugriff am 13.06.2013]
- [5] Manfred Pinkal(2011):Einführung in die Computerlinguistik,Verarbeitung gesprochener Sprache,[Online].  
URL: [http://www.coli.uni-saarland.de/courses/I2CL-06/fohlen/CL-Morphologie\\_I.pdf](http://www.coli.uni-saarland.de/courses/I2CL-06/fohlen/CL-Morphologie_I.pdf)  
[Zugriff am 13.06.2013]
- [6] Carstensen, Kai-Uwe (2010): Computerlinguistik und Sprachtechnologie. Eine Einführung, 3. Aufl. Heidelberg: Spektrum, Akad. Verl.
- [7] Carstensen, Kai-Uwe (2012): Sprachtechnologie ein Überblick , Version 2.1, [Online].  
URL: <http://kai-uwe-carstensen.de/Publikationen/Sprachtechnologie.pdf>  
[Zugriff am 13.06.2013]
- [8] Stephen Euler (2006): Grundkurs Spracherkennung, vom Sprachsignal zum Dialog Grundlagen und Anwendung verstehen, vieweg verl. 1. Auflage Wiesbaden
- [9] Roush, Wade (2010): The Story of Siri, from Birth at SRI to Acquisition by Apple—Virtual Personal Assistants Go Mobile. [Online] .  
URL:<http://www.xconomy.com/san-francisco/2010/06/14/the-story-of-siri-from-birth-at-sri-to-acquisition-by-apple-virtual-personal-assistants-go-mobile/>  
[Zugriff am 13.06.2013]
- [10] David Pogue (2011): iPhone 4S, Das Missing Manual, O’Reilly Verlag, 5. Auflage
- [11] Göde Both(2011): Diplomarbeit, Agency und Geschlecht in Mensch/Maschine-Konfigurationen am Beispiel von Virtual Personal Assistants, [Online].  
URL: <http://edoc.hu-berlin.de/master/both-goede-2011-07-19/PDF/both.pdf>  
[Zugriff am 13.06.2013].
- [12] Balzert, Heide (2004): Webdesign & Web-Ergonomie. Websites professionell gestalten. Herdecke; Dortmund: W3L-Verl.
- [13] Semantic Web(2009):Keynote, the Game Changer, Siri a Virtual Personal Assistant.[Online].

- URL: [http://semanticweb.com/keynote-the-game-changer-siri-a-virtual-personal-assistant\\_b15877](http://semanticweb.com/keynote-the-game-changer-siri-a-virtual-personal-assistant_b15877)
- [Zugriff am 13.06.2013].
- [14] Apple Inc.(2012):Benutzerhandbuch Iphone 4S, Für IOS 5.1- Software. [Online].  
URL: [http://manuals.info.apple.com/de\\_DE/iphone\\_ios5\\_benutzerhandbuch.pdf](http://manuals.info.apple.com/de_DE/iphone_ios5_benutzerhandbuch.pdf)  
[Zugriff am 13.06.2013].
- [15] A. Chery, J. Park, et al. (2005): IRIS: Integrate. Relate. Infer. Share., SRI International. [Online]  
URL: <http://www.ai.sri.com/pubs/files/1207.pdf>  
[Zugriff am 13.06.2013].
- [16] V. K. Chaudhri, A. Cheyer, R. Guili, et al. (2006):A Case Study in Engineering a Knowledge Base for an Intelligent Personal Assistant. [Online].  
URL: <http://www.adam.cheyer.com/papers/ontology-overview-semantic-desktop.pdf>  
[Zugriff am 13.06.2013].
- [17] U. Kastens & H.K Büning (2008): Modellierung Grundlagen und formale Methoden, Hanser Verlag München,2.Auflage.
- [18] Roland Römer(2011): Arbeitspapier, Grundlagen MDP, POMDP
- [19] C. Kölbl, M. Huber, et al.(2013): Konstruktion von UMP-Transduktoren aus Wizard-of-Oz Daten.[Online].  
URL:[https://www.informatik.uniaugsburg.de/lehrstuehle/inf/mitarbeiter/koelbl/publications/esv2013/Konstruktion\\_von\\_UMP-Transduktoren\\_aus\\_Wizard-of-Oz\\_Daten.pdf](https://www.informatik.uniaugsburg.de/lehrstuehle/inf/mitarbeiter/koelbl/publications/esv2013/Konstruktion_von_UMP-Transduktoren_aus_Wizard-of-Oz_Daten.pdf)  
[Zugriff am 13.06.2013].
- [20] Wikipedia (2013):Entity Relationship Modell, ER-Diagramme. [Online].  
URL: <http://de.wikipedia.org/wiki/Entity-Relationship-Modell#ER-Diagramme>  
[Zugriff am 13.06.2013].
- [21] M.Wolf, R. Römer & R. Hoffmann (2012): Hierarchische kognitive dynamische Systeme zur Sprach- und Signalverarbeitung, Dresden, Germany.
- [22] S.Rogger, R.Römer & C. Henschel(2012): Semantische Modellierung eines Virtual Reality Systems mit Gestensteuerung, Cottbus, Germany.
- [23] S.Young (2010): Cognitive User Interfaces, Signal Processing Magazine
- [24] Hans, Jürgen; Gallwitz, Florian; et al.: Spracherkennung und Sprachdialog: Stand der Technik, Einsatzbeispiele und zukünftige Trends . [Online].  
URL: [http://www.sympalog.de/cms/upload/pdf/Artikel\\_Design\\_Elektronik\\_2004a.pdf](http://www.sympalog.de/cms/upload/pdf/Artikel_Design_Elektronik_2004a.pdf)  
[Zugriff am 13.06.2013]
- [25] A.Chery, D. Guzzoni & C. Baur (2007):Modeling Human-Agent Interaction with Active Ontologies.[Online].  
URL: <http://www.ai.sri.com/~nysmith/organizing/sss07/papers/SS04GuzzoniD.pdf>  
[Zugriff am 13.06.2013]

## Abbildungsverzeichnis

Abbildung 1: Auswertungen auf Sprachsignal ([3],S. 25) .....	7
Abbildung 2: Die zwei Stufen des Sprachsynthesystem ([3], S.197) .....	8
Abbildung 3: Transkriptionsstufe der Sprachsynthese([3], S 199) .....	9
Abbildung 4: Phonoakustische Stufe der Sprachsynthese ([3], S 200) .....	9
Abbildung 5: Spracherkennungsprozess ([5], S.8) .....	9
Abbildung 8:Darstellung von Merkmalsvektoren([6], S.5) .....	10
Abbildung 6: Oszillogrammdarstellung einen Schallsignal( [6], S.2).....	10
Abbildung 7: Spektrogrammdarstellung ([6], S.3) .....	10
Abbildung 9:Dialogsystem der Sprachanalyse.....	12
Abbildung 10: Dialogsystem Siri.....	17
Abbildung 11:Siri als Gesamtsystem.....	20
Abbildung 12:Struktur- und Funktionskomponente und ihre Zusammengehörigkeit([1], S.12) .	22
Abbildung 13: Weltmodell für ein Telefonbuch .....	26
Abbildung 14:Aktionsmodell zur Identifikation von Aktionen-ID für das Telefonbuch Beispiel ..	27
Abbildung 15:Objektmodell für Identifikation von Objekt-ID .....	28
Abbildung 16:Modellierung der Entscheidungsprozess ([17],S.2) .....	30
Abbildung 17:Erweiterter Weltmodell mit direkte und indirekte Objekt-ID.....	32
Abbildung 18:MWR für die Objekt-ID in Telefonbuch bei Siri .....	43
Abbildung 19:Aktion-ID für das Telefonbuch bei Siri.....	45
Abbildung 20:MWR bei indirekter Objekt-ID für das Telefonbuch bei Siri .....	45
Abbildung 21:Weltmodell für den Onlinedienst Yahoo Sportacular .....	48
Abbildung 22:Weltmodell für Siri-System für Das Telefonbuch-Dienst .....	55

## Erklärung

Der Verfasser erklärt, dass er die vorliegende Arbeit selbständig, ohne fremde Hilfe und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt hat. Die aus fremden Quellen (einschließlich elektronischer Quellen) direkt oder indirekt übernommenen Gedanken sind ausnahmslos als solche kenntlich gemacht. Die Arbeit ist in gleicher oder ähnlicher Form oder auszugsweise im Rahmen einer anderen Prüfung noch nicht vorgelegt worden.

Ort, Datum

Unterschrift des Verfassers