

SEMANTISCHE MODELLIERUNG VON SYSTEM-BENUTZER-INTERAKTIONEN AM BEISPIEL EINER HOME-ENTERTAINMENT STEUERUNG

Kathleen Karnagel, Ronald Römer, Stephan Rogge und Jens Lindemann

Brandenburgische Technische Universität Cottbus

Kurzfassung: Im vorliegenden Beitrag soll die Interaktion zwischen Mensch und Maschine durch die Funktionsnutzungen einer Home-Entertainment-Anwendung auf semantischer Ebene modelliert werden. Die Arbeit auf der semantischen Ebene ermöglicht es unabhängig vom Kommunikationsmedium ein Modell zu entwerfen. Gleichzeitig verlangt es wohldefinierte Schnittstellen für eine funktionierende Kommunikation zwischen Nutzer und System. Der Umgang mit unsicherer Information kann durch die semantische Modellierung untersucht und verbessert werden. Die Aufgabe ist es das theoretisch existierende System für intuitive Bedienung robust zu entwerfen [1]. Beim Systemaufbau wird zwischen statischer und dynamischer Modellierung unterschieden. Die statische Modellierung ermöglicht neben der Objektidentifikation auf einer Datenbank, die Aktionenidentifikation, welche die Funktionsnutzung des Nutzers widerspiegelt. Ein skizziertes Gesamtsystem modelliert die Interaktion zwischen Nutzer und System. Und ein integriertes Struktur- und Funktionsmodell vereinigt alle statischen Elemente und verknüpft diese durch eine Verknüpfungs- und Assoziationstabelle. Die dynamischen Modellkomponenten, wie Abläufe und Prozesse, sind abhängig vom Informationsstatus und den Nutzereingaben. Damit die semantische Modellierung realitätsnah bleibt, kann ein Experiment unter der Wizard of Oz Methode durchgeführt werden. Aus den daraus gesammelten Datensätzen werden *Utterance Meaning Pairs* (UMPs) gebildet. Die damit verbundenen Bedeutungsfindungen ermöglichen der Home-Entertainment-Anwendung besser mit unscharfen Informationen umgehen zu können. Des Weiteren ermöglichen die Ergebnisse, dass die Systemmodelle noch während ihrer Entstehungsphase verifiziert und gegebenenfalls modifiziert werden können.

1 Einleitung

Die Entwicklung sprachbasierter Geräte, wie Handys oder Fernseher, gewinnt zunehmend an Bedeutung, jedoch ist das Verstehen akustischer Nutzereingaben derzeit noch problembehaftet. Sprachverstehende Geräte müssen lernen robuster mit unsicheren Informationen umzugehen. Vor den interaktiven Systemen sitzen Individuen, welche versuchen sich durch unterschiedlichste Wörter und Befehle auszudrücken. Diese Ungleichheiten müssen in der Entwicklung beachtet werden, um die Geräte benutzerfreundlicher zu machen.

Dieser Beitrag soll diesem Problem entgegentreten und beschäftigt sich mit der semantischen Modellierung einer Home-Entertainment Steuerung. Als Beispiel für eine solche Anwendung soll ein Recorder dienen, welcher durch den Benutzer mit Hilfe von beliebigen Sprach-einheiten gesteuert werden kann. Zu Beginn wird das Gesamtsystem spezifiziert, mit genauerer Betrachtung des strukturellen und funktionellen Aufbaus. Dieser Teil des gesamten Systems ist das integrierte Struktur- und Funktionsmodell. Darin findet sich die Modellierung der Filmdatenbank, der Nutzeraktionen und der dazugehörigen Parameter wieder. Abbildung 1 veranschaulicht den Zusammenhang dieser drei Systemkomponenten. Eine genauere Erläuterung dieser zielrelevanten Problematik der Identifikation von Aktionen (A_ID) und Objekten (O_ID) folgt im Kapitel 3.

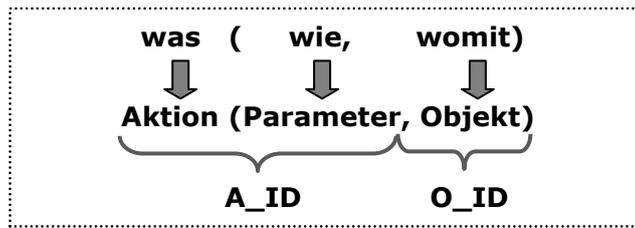


Abbildung 1: Struktur- und Funktionskomponenten und ihre Zusammengehörigkeit

Der Kern dieses Beitrags ist die Erschließung der Interaktion zwischen Nutzer und System. Um diese Kommunikationsebene zu simulieren, wird eine semantische Modellierung für den theoretisch existierenden Recorder entworfen. Mit der Annahme, dass der Recorder bereits auf Spracherkennungsebene funktionstüchtig ist, steht insbesondere die Bedeutungsfindung der vom Nutzer einzugehenden akustischen Einheiten im Mittelpunkt.

2 Spezifikation des Gesamtsystems

Die zu modellierende Home-Entertainment Anwendung hat im Wesentlichen die gleichen Eigenschaften, wie ein handelsüblicher Videorecorder. Der Nutzer soll durch spezielle Zieleingaben die Möglichkeiten haben ein Objekt (Filme, Sendungen etc.) aufzunehmen, zu löschen, wiederzugeben oder mit ihm intern zu arbeiten. Das Besondere an diesem Recorder ist, dass er über akustische Spracheinheiten gesteuert wird. Der Benutzer benötigt keine Fernbedienung um die genannten Funktionen auszuführen. Auf Grund der zu erwartenden unscharfen Informationen seitens des Nutzers, entsteht zwangsläufig eine Interaktion zwischen Mensch und Maschine. Diese technische Kommunikationsweise simuliert die natürliche sprachliche Kommunikation und verschafft dem Nutzer eine komfortablere Bedienung eines technischen Gerätes. Im Prozess der semantischen Modellierung wird davon ausgegangen, dass die Home-Entertainment Anwendung keine Spracherkennungsprobleme hat. Dafür muss ein Weg gefunden werden, die Bedeutung der sprachlichen Einheiten zu ermitteln. Das System besteht aus Struktur- und Funktionselementen. Das Hauptelement des Recorders ist die eigene interne Datenbank, auf der alle Objekte eigens verwaltet werden. Zusätzlich besteht ein Zugang zu einer externen Datenbank. Diese wird von den Fernsehanstalten mit Informationen versorgt und ermöglicht dem Nutzer u. a. auf deren Mediathek zuzugreifen.

2.1 Struktur des Gesamtsystems

Die interne und externe Datenbank (Abbildung 2) sind strukturell gleich aufgebaut. Ihre Modellierung erfolgt mit Hilfe der Entity-Relationship-Methode [2]. Die Objekte werden mit ihren Meta- und Quelldaten darin gespeichert. Durch die vorteilhafte Strukturierung eines ER-Modells bestehen mehrere Verbindungen der Daten. Das ermöglicht, dass Objekte auch auf indirektem Weg ermittelt werden können.

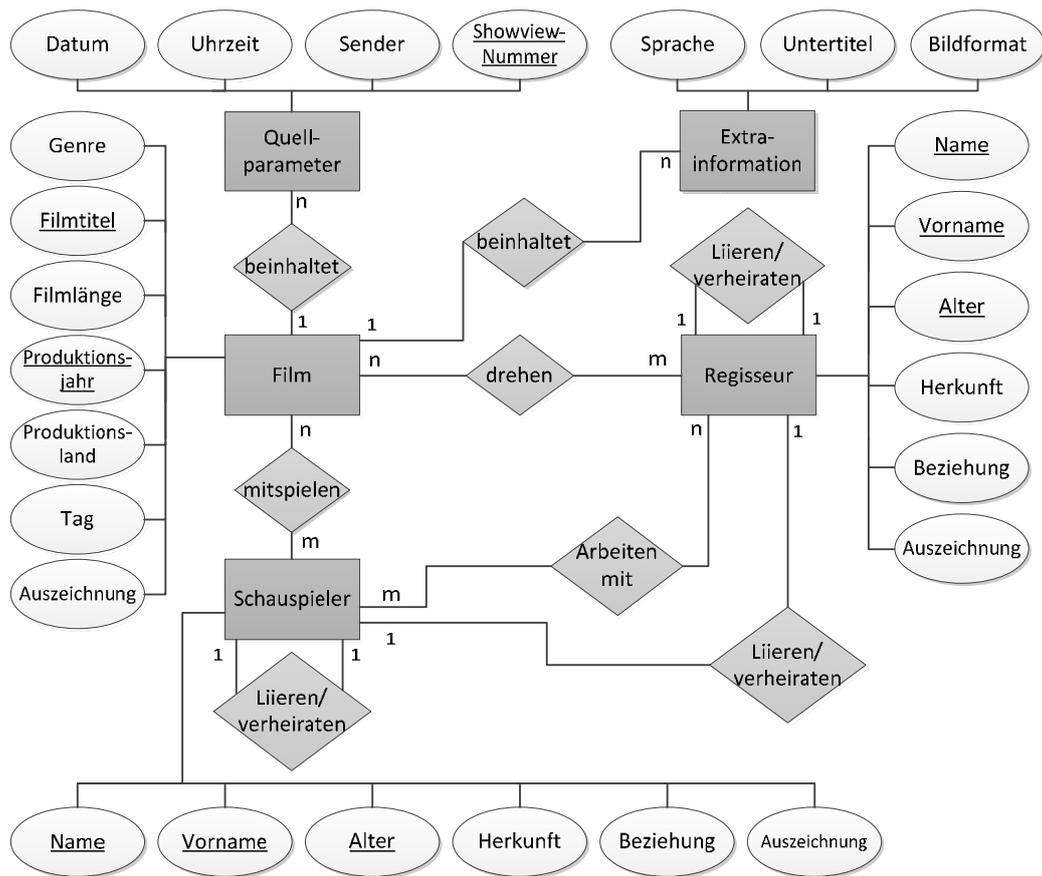


Abbildung 2: Datenbankmodell (nach Chen [3]) der Medienobjekte, wie Filme oder Serien mit Primärschlüssel (Unterstreichung)

2.2 Funktionen des Gesamtsystems

Vom Nutzer wird nicht verlangt, dass er sich elementar mit einem neuen System befassen muss. Ihm stehen folgende bekannte Funktionen zur Verfügung: **Aufnehmen**, **Löschen**, **Wiedergeben** und eine **Bibliotheksfunktion**. Aufnehmen bedeutet, dass ein Objekt über den externen TV-Zugang aufgenommen werden kann, unabhängig davon ob es aktuell im laufenden Programm ausgestrahlt wird oder es sich in einer Mediathek befindet. Soll ein Film auf der internen Datenbank gelöscht werden, geschieht das logisch, nicht physisch. Damit sichert das System die Entscheidung des Nutzers ab. Eine Art Objekt-Historie könnte bei einem Wiederherstellungswunsch wichtige Informationen bereitstellen. Außerdem können von der internen bzw. externen Datenbank Filme und Sendungen wiedergegeben werden. Die Bibliotheksfunktion ermöglicht dem Benutzer aufgenommene Objekte beispielsweise zu Kategorisieren oder zu Listen.

3 Statische Modellierung des Systems

Ziel ist es, die Recorderumgebung auf der semantischen Ebene zu modellieren. In diesem Umfeld gibt es zwei große Hauptkomponenten, den Recorder und den Nutzer, der das System bedient. Damit es zu einer erfolgreichen Kommunikation zwischen diesen beiden Elementen kommen kann müssen Anfragen bzw. Wünsche des Nutzers erfolgreich bearbeitet werden. Zur Verarbeitung der Nutzeranfragen arbeitet das System auf statischen Modellen und Datenbanken. Um die Benutzerfunktionen ausführen zu können, benötigt das Recorder-System immer eine Aktionsidentifikation (A_ID) und eine Objektidentifikation (O_ID). Nur durch die System-Benutzer-Interaktion kann das System beide IDs herausfinden. Abbildung 1 zeigt den Zusammenhang zwischen den Komponenten Aktion und Objekt. Die A_ID setzt sich zusammen aus der Aktion, sprich *was* soll mit einem Objekt passieren und den Parametern, *wie* die Aktion durchgeführt wird. Die O_ID benötigt der Recorder zur Objektidentifizierung, demnach *womit* eine Aktion durchgeführt werden soll.

3.1 Aktionenmodell

Damit das System weiß, welche Parameter für welche Aktion benötigt werden, wurde ein Aktionenmodell (Abbildung 3) entworfen. Eine Aktionsidentifikation ist wichtig, damit das System erkennt, um welche Benutzerfunktion und deren damit verbundenen Parameter es sich handelt.

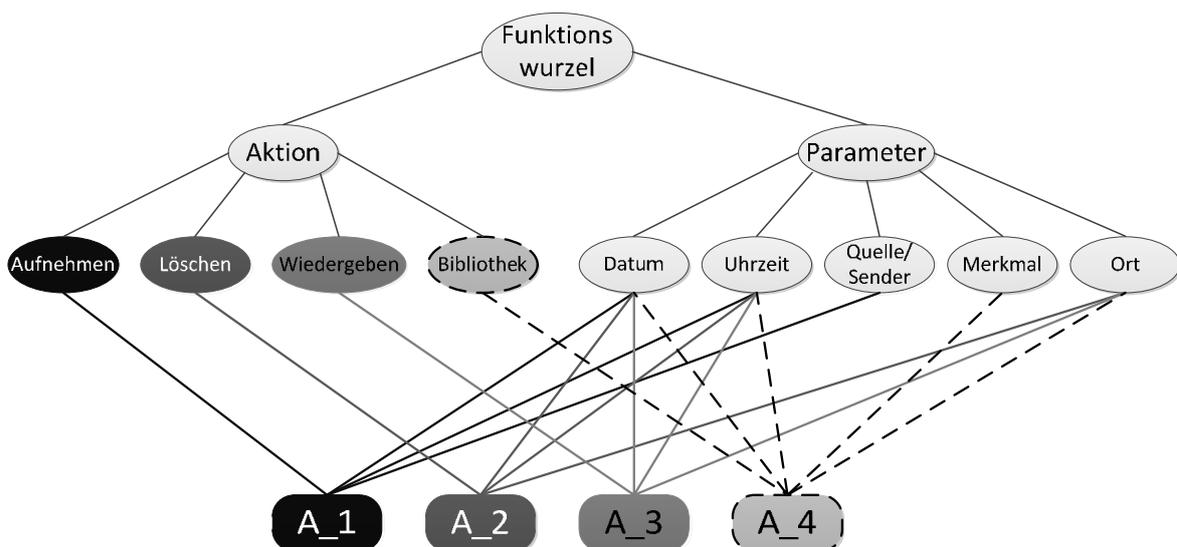


Abbildung 3: Aktionenmodell zur Definition und Identifizierung der entsprechenden Verknüpfungen von Aktionen mit den notwendigen Parametern

Die Abbildung 3 zeigt, dass alle Aktionen mit den Parametern *Datum* und *Uhrzeit* verknüpft sind. Diese Größen benötigen im Aktionsfall Aufnehmen reale Werte, da das System auf die Information der Sendezeit angewiesen ist. Werden in den anderen Fällen keine expliziten Werte angegeben, setzt der Recorder an dieser Stelle einen *Standardwert*. Die *Quelle/Sender* besagt wo das gewünschte Objekt ausgestrahlt wird. Der Parameter *Merkmal* kann den Wert *intern* oder *extern* besitzen. *Intern* lässt auf die vom Nutzer selbst angelegten Merkmale schließen. Das Setzen des Wertes auf *extern*, bedeutet, dass ein internes Objekt nach dem Merkmalsmuster der externen Datenbank abgelegt wird. Der Parameter *Ort* wird bei den Aktionen Löschen, Wiedergeben und Bibliotheksfunktion benötigt. Der gesetzte Wert besagt, ob sich das gewünschte Objekt auf der internen oder externen Datenbank befindet. Um eine Aktion zu identifizieren ist es nicht erforderlich, dass alle nötigen Informationen vom Nutzer stammen. Ist das System infolge hinreichender Informationen durch die Nutzereingaben im

Stünde ein Objekt zu identifizieren kann es anhand der Daten die fehlenden Aktionsparameter auf der entsprechenden Quelldatenbank auslesen.

3.2 Objektmodell

Damit der Recorder die Benutzerfunktion auf dem richtigen Objekt anwendet, muss er den Film oder die Serie auf einer der Datenbanken identifizieren. Die abgespeicherten Metadaten der Objekte werden im ERM gepflegt und aktualisiert. Damit das System das richtige Objekt erkennt wurde ein Objektmodell (Abbildung 4) entwickelt, welches mit den gleichen Objektinformationen wie die Datenbanken ausgestattet ist. In der Abbildung erkennt man die zahlreichen Verbindungen verschiedener Daten zu mehreren Objekten. Das zeigt deutlich, dass teilweise viele Informationen nötig sind, damit ein Film eindeutig identifiziert werden kann.

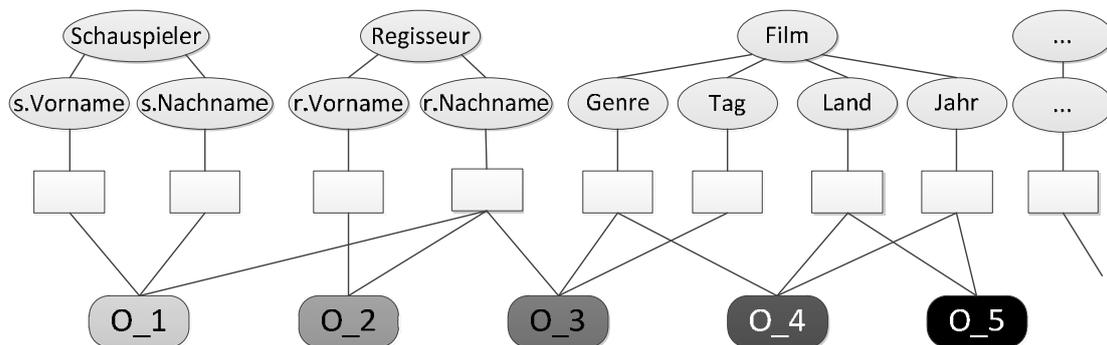


Abbildung 4: Objektmodell zur Identifikation der in der Datenbank befindlichen Objekte. Die Objekte sind anhand ihrer Metadaten auffindbar und führen in richtiger Kombination zum gewünschten Objekt.

Auch bei diesem Modell liegen zwei Möglichkeiten vor, um an die nötigen Parametern zu gelangen. Zum einen kann das System, durch Nutzereingaben über eine Aktionen_ID verfügen und kann mit Hilfe dieser Informationen das entsprechende Objekt auf einer der Datenbanken ausfindig machen. Und zum anderen können alle entscheidenden Objektinformationen vom Nutzer gegeben werden. Ist der Recorder auf das Wissen des Nutzers angewiesen, kann er über Nachfragen, mit Hilfe der Mensch-Maschine-Interaktion, wichtige Informationen sammeln. Die Reihenfolge in der die Aktionen- und Objektidentifikationen vorgenommen werden ist nicht relevant. Es müssen letztendlich beide Komponenten gefunden werden. Fasst man die gefundene Aktion und das Objekt zusammen, bildet sich eine Assoziation. Die IDs werden in eine Assoziationstabelle geschrieben, welche der Recorder wie einen Arbeitsspeicher behandelt. Vollständige Einträge werden abgearbeitet und wieder gelöscht.

3.3 Integriertes Struktur- und Funktionsmodell

Das Gesamtsystem ist in zwei Ebenen unterteilt. Das System, der Recorder, befindet sich auf der System-Ebene und der Nutzer, einschließlich seines Umfeldes, befindet sich auf der User-Ebene. Diese strenge Trennung soll verdeutlichen, dass beide Ebenen ihre eigenen Schwerpunkte aufweisen. Das Gesamtsystem dient u. a. der Veranschaulichung der späteren Experimentierumgebung. Auf der Systemebene befindet sich das integrierte Struktur- und Funktionsmodell (Abbildung 5), welches zur Identifikation der A_ID und der O_ID dient und die Aktionen-, Objekt- und Datenbankmodelle beinhaltet. Alle Komponenten werden durch die Verknüpfungs- und Assoziationstabelle verbunden. Sie definieren erlaubte und nicht erlaubte Verbindungen und zeigen welche Aktion mit welchem Objekt vom Nutzer gewünscht wird.

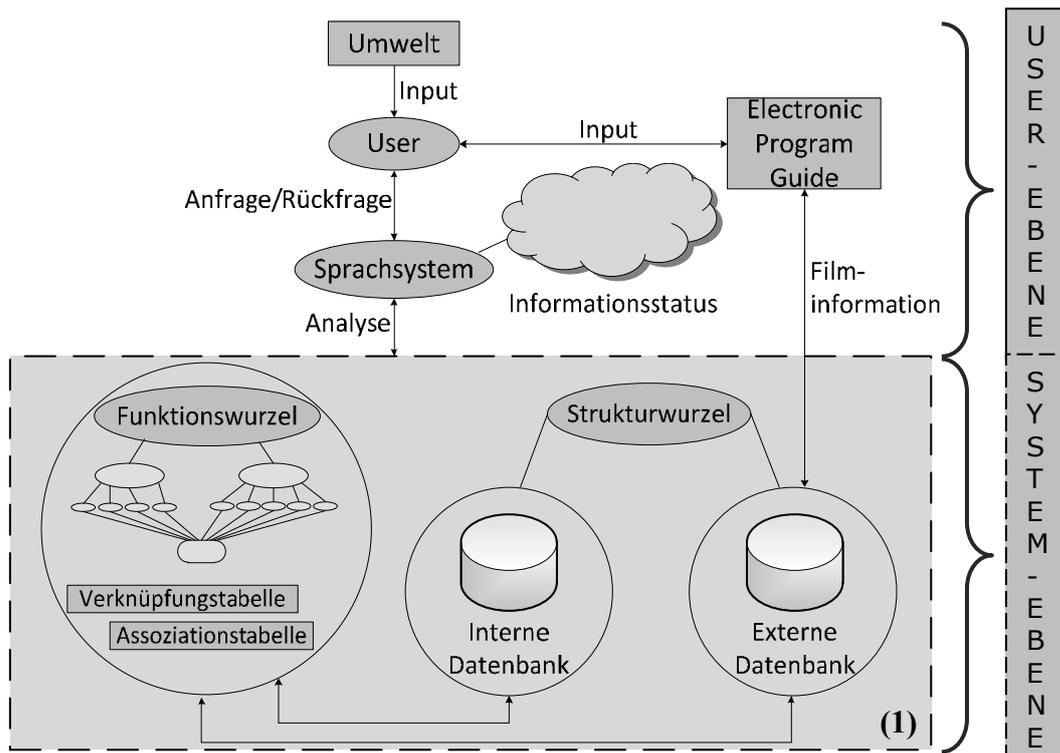


Abbildung 5: Gesamtsystem mit integriertem Struktur- und Funktionsmodell (1)

4 Dynamische Modellierung des Systems

Die dynamische Modellierung soll den Prozess der Aktionen- und Objektfindung während der Mensch-Maschine-Interaktion zeigen. Das Ziel dieser Modellierung ist theoretisch definiert, aber ein Modell zur praktischen Umsetzung ist noch nicht erarbeitet.

Das System wird in Abhängigkeit der Nutzereingaben verschiedene Reaktionen zeigen und diese dem Nutzer mitteilen. Weiterhin ist für die Systemausgaben der aktuelle Informationsstatus entscheidend. Dieser Status verändert sich mit jeder Eingabe des Nutzers. Für diese Zielstellung bieten sich die die umgestellten Funktionsgleichungen f und g an (Abbildung 6).

$$\mathbf{I}' = \mathbf{f} (\mathbf{I}, \mathbf{E})$$

$$\mathbf{A} = \mathbf{g} (\mathbf{I}, \mathbf{E})$$

Abbildung 6: Funktionsgleichungen eines neuen Informationsstatus und der damit verbundenen neuen Ausgabe; I' = neuer Informationsstatus; I = alter Informationsstatus; E = Nutzereingabe; f, g = Funktionen

Während der Mensch-Maschine-Interaktion sammelt das System Informationen und verarbeitet diese. Der Status des dynamischen Systems wird durch diese semantischen Eingaben beeinflusst.

5 Utterance Meaning Pair (UMP)

Bevor Nutzereingaben in das dynamische Modell einfließen können, müssen Äquivalenzklassen dieser akustischen Äußerungen einer Bedeutung zugeordnet werden. Dieses Verfahren wird durch das Bilden von *Utterance Meaning Pairs* realisiert. Dadurch gelingt es dem System trotz unterschiedlicher Wortkodierungen für ein und dieselbe Bedingung die richtige Bedeutung zu definieren. Für eine repräsentative Informationssammlung kann ein Experiment durchgeführt werden. Auf dessen Ergebnissen können anschließend die UMPs gebildet werden. Diese ermittelten Bedeutungen dienen letztendlich als Eingabe für das dynamische Modell und können zu einer Verifikation und Verbesserung des Systems führen. Das Sammeln äquivalenter Begriffe trainiert ein interaktives System und der Umgang mit unscharfer Information wird sicherer [4].

6 Experiment (Wizard of Oz)

Auf Grund der wichtigen Bildung von UMPs muss im Laufe der Modellentwicklung ein praxisorientiertes Experiment durchgeführt werden. Dieses wird einerseits zeigen, wie realitätsnah die Modellierung der Kommunikation zwischen Nutzer und Maschine tatsächlich ist. Andererseits kann durch die Resultate das dynamische Modell verifiziert werden. Da der Recorder nur theoretisch existiert, der Nutzer aber den Eindruck eines funktionierenden Gerätes bekommen soll, kann bei der Durchführung auf die Wizard of Oz Methode zurückgegriffen werden [5]. Dem Probanden wird ein scheinbar funktionstüchtiges Gerät zur Verfügung gestellt, welches allerdings vom Experimentator gesteuert wird. Der Proband erhält vor der Durchführung eine Aufgabenstellung, welche den Umgang mit dem ihm später bereit gestellten Electronic Program Guide (EPG) erklärt. Der EPG ist gezielt mit unvollständigen Informationen einer vorher zusammengestellten Objektsammlung ausgestattet. Die sporadischen Filminformationen werden dem Probanden durch Bilder von Filmausschnitten oder textuellen Hinweisen zuteil. Die vorher aufbereitete Datenbank erleichtert dem Experimentator schnell auf die Nutzereingaben zu reagieren und als System mit dem Nutzer in eine Interaktion zu treten. Das Ziel dieses Experiments ist es einerseits dem Nutzer so wenig Anhaltspunkte wie möglich über das gesuchte Objekt zu geben, aber andererseits eine ausreichende Menge an Hinweisen damit er mit dem Recorder kommunizieren kann. Die Wizard of Oz Methode ermöglicht das sprachverstehende System bereits in seiner Entwicklungsphase zu verbessern.

7 Zusammenfassung

In diesem Beitrag wurde gezeigt, dass die Modellierung eines Systems auf der semantischen Ebene die notwendige Grundlage schafft, eine Mensch-Maschine-Interaktion zu erstellen und zu bearbeiten. Es ist möglich die notwendigen Modelle und Komponenten zu skizzieren und zu verknüpfen. Neben den strukturellen Aspekten eines solchen Systems, müssen außerdem die für den Nutzer zugänglichen Systemfunktionen berücksichtigt werden. Durch statische Modelle kann die Modellierung von Objekten und Aktionen festgelegt werden. Entwürfe wie Datenbanken können anschaulich durch Entity-Relationship-Modelle konstruiert werden. Und das Objekt- und das Aktionenmodell können verständlich in einer hierarchischen Baumstruktur modelliert werden. Die dynamische Modellierung ist das Bindeglied zwischen dem Experiment Wizard of Oz und den UMPs. Die den Prozessablauf beeinflussenden Nutzer-eingaben müssen zuerst in einem Experiment gesammelt werden und aus den gewonnenen Daten die UMPs gebildet werden. Diese Bedeutungen dienen dann als Eingabe im Prozessablauf des dynamischen Modells. Dadurch wird die Struktur des dynamischen Modells weiterentwickelt und verbessert.

8 Literatur

- [1] Young, Steve: *Cognitive User Interfaces*, IEEE Signal Processing Magazin, 2010
- [2] Pernul, Günther; Unland, Rainer: *Datenbanken im Unternehmen – Analyse, Modellbildung und Einsatz*, 2001.
- [3] Chen, Peter: *The Entity Relationship Modell – Toward a unified view of data*, 1976
- [4] Wirsching, Günther; Kölbl, Christian: *Language Modeling with Utterance-Meaning-Pairs*, Technical Report, Institute of Computer Science, University of Augsburg, April 2011.
- [5] Brindöpke, Christel; Johanntokrax, Michaela; Pahde, Arno: *Darf ich dich Marvin nennen? – Instruktionsdialoge in einem Wizard-of-Oz-Szenario*, Universität Bielefeld, 1995.