

# Abbildung von Webformularen auf aktive Ontologien

## Masterarbeit

Wasim Said

Betreut von Martin Blersch und Mathias Landhäußer

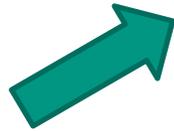
IPD Tichy, Fakultät für Informatik



# Motivation

# Aktive Ontologie

EASIER



Data Entry Form

Project Name

Billing Code

Date

Notes

---

A Web Page

http://

Name

Street

City

Country

email

phone no

---

Registration Details

University:

Institute:

Branch:

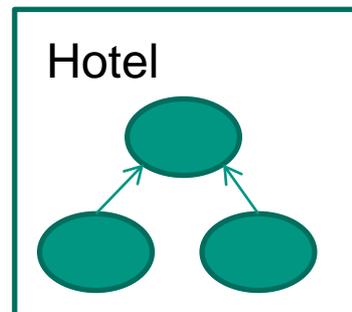
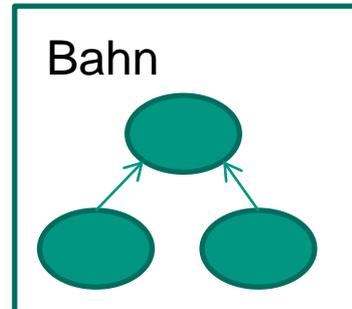
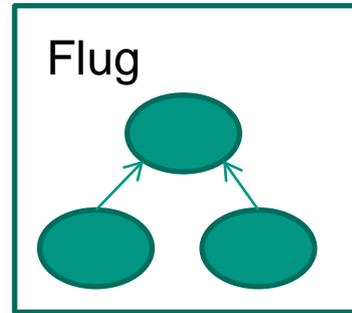
Degree:  Pursuing Completed

Average CPI:  Upto  Th Semester

Experience:  Years

Your Website Or Blog:

1 Step 2



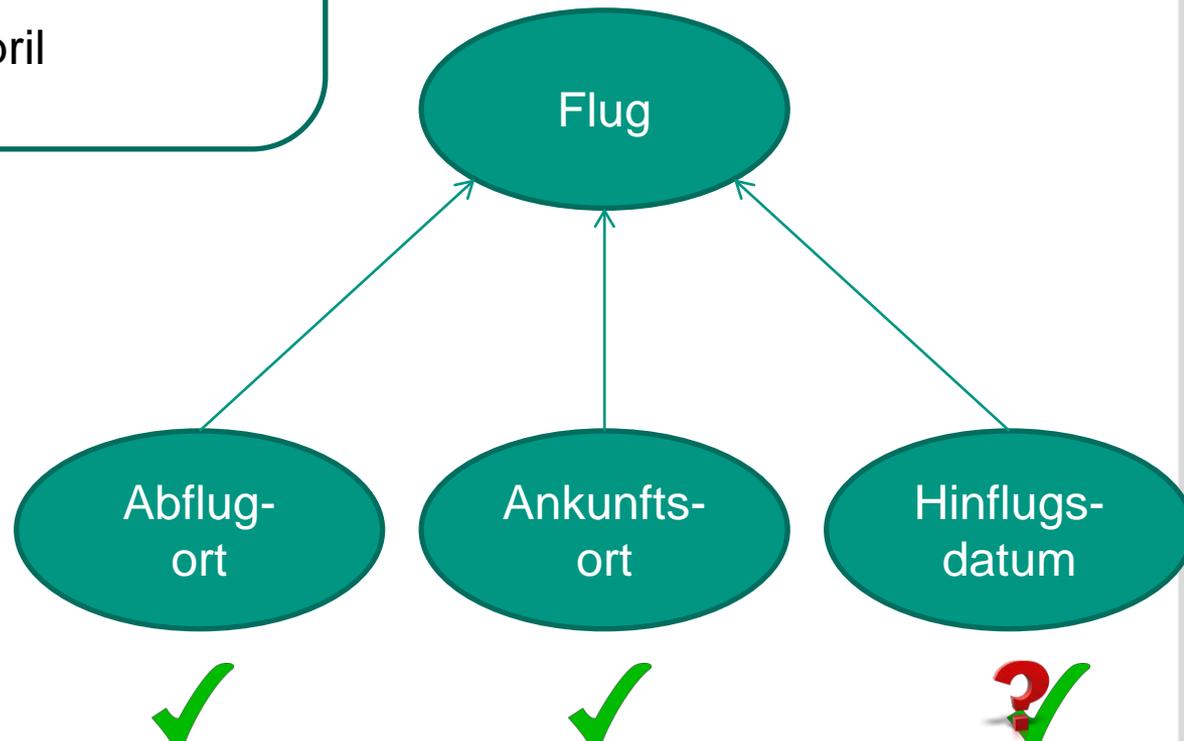
# Motivation

## Dialog Manager

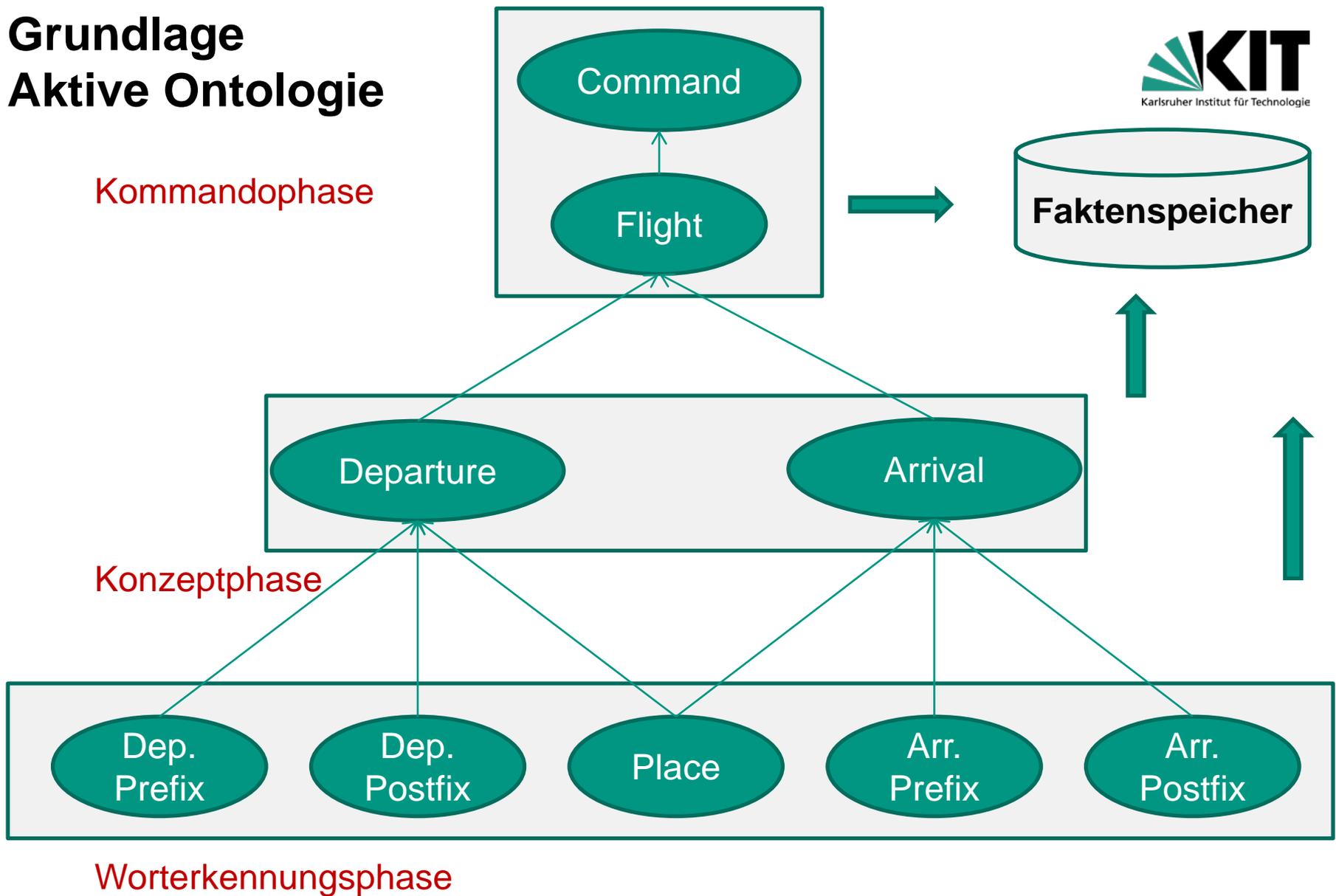
User: Book me a flight from  
Karlsruhe to Berlin

System: When do you want to fly?

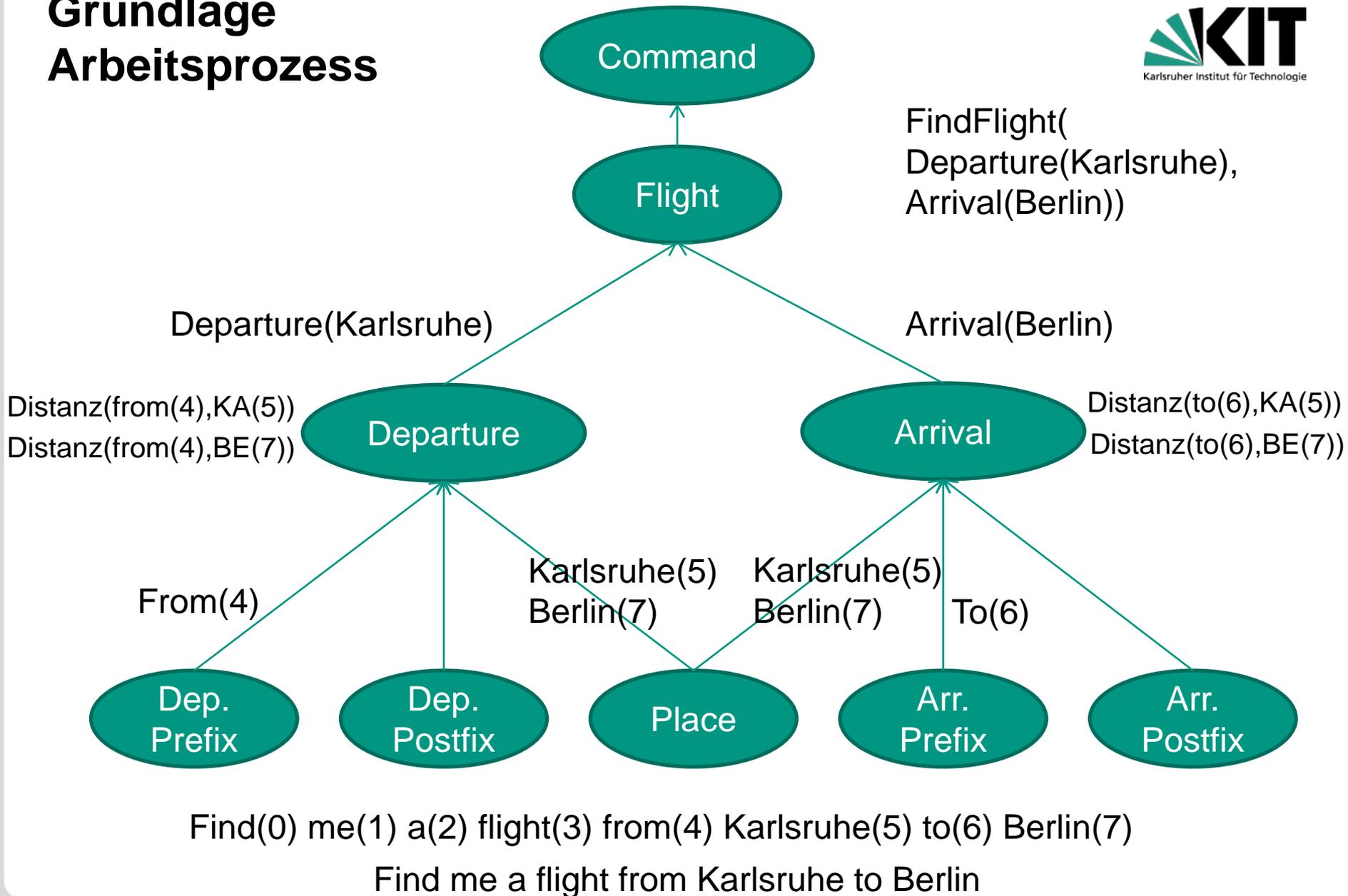
User: On the 15th of April



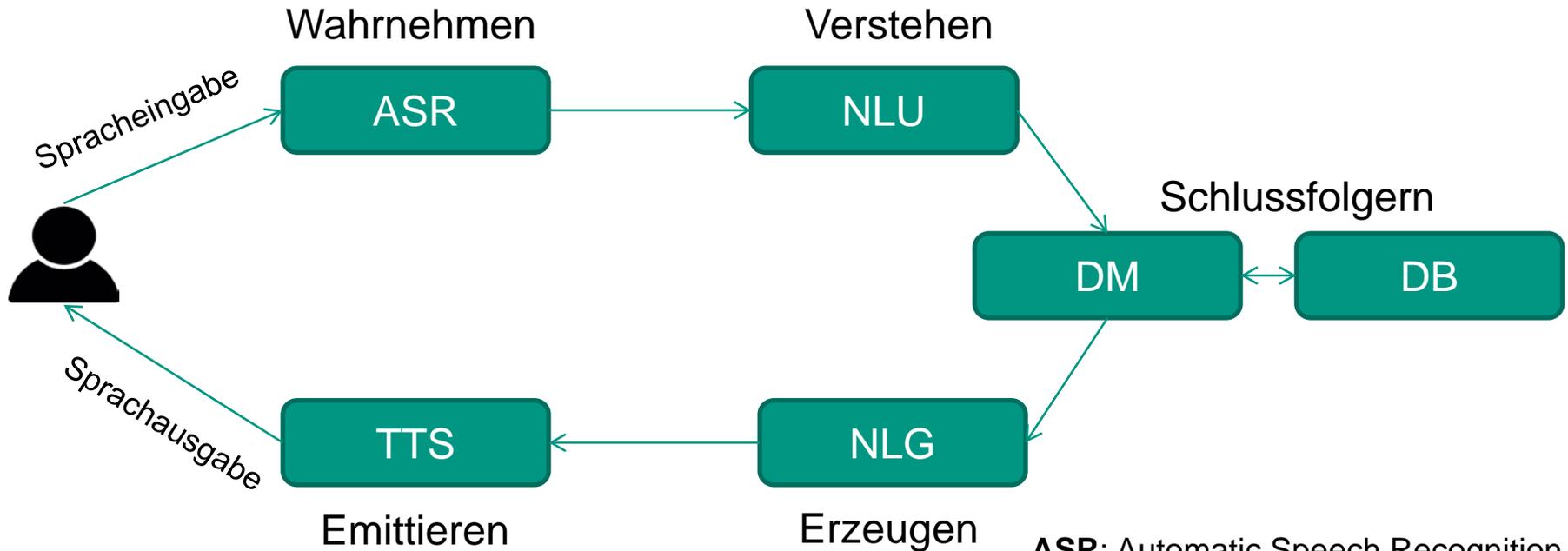
# Grundlage Aktive Ontologie



# Grundlage Arbeitsprozess



# Grundlagen Sprachdialog-Systeme



**ASR:** Automatic Speech Recognition  
**NLU:** Natural Language Understanding  
**DM:** Dialog Manager  
**DB:** Database  
**NLG:** Natural Language Generation  
**TTS:** Text To Speech

# Verwandte Arbeiten

## ■ Erstellung von Ontologien aus den Webformularen

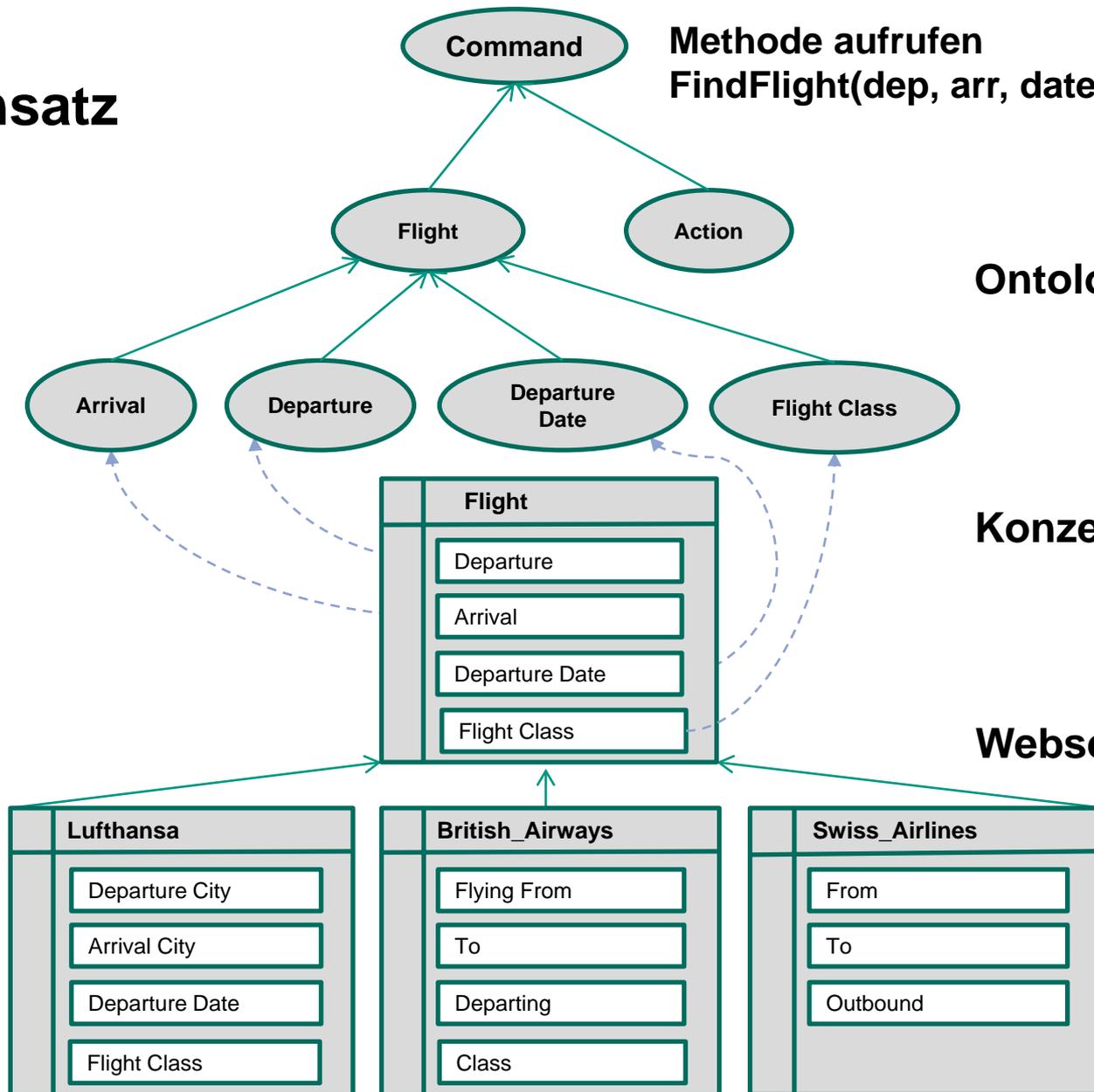
- Faeton: Form analysis and extraction tool for ontology construction [BJNS10]
- Ontobuilder: Fully automatic extraction and consolidation of ontologies from web sources using sequence semantics [RG06]

## ■ Dialog-Manager

- The Philips automatic train timetable information system [AOSS95]
- Partially observable markov decision processes for spoken dialog systems [WY07]

# Ansatz

Methode aufrufen  
FindFlight(dep, arr, date)



Ontologie erstellen

Konzepte identifizieren

Webseiten auswählen

# Ansatz

## Konzepte Identifizieren

Airline	Departure	Arrival	Dep. Date	Flight Class
Lufthansa	Departure City	Arrival City	Dep. Date	Flight Class
British Airways	Flying From	To	Departing	Class
Swiss Airlines	From	To	Outbound	---

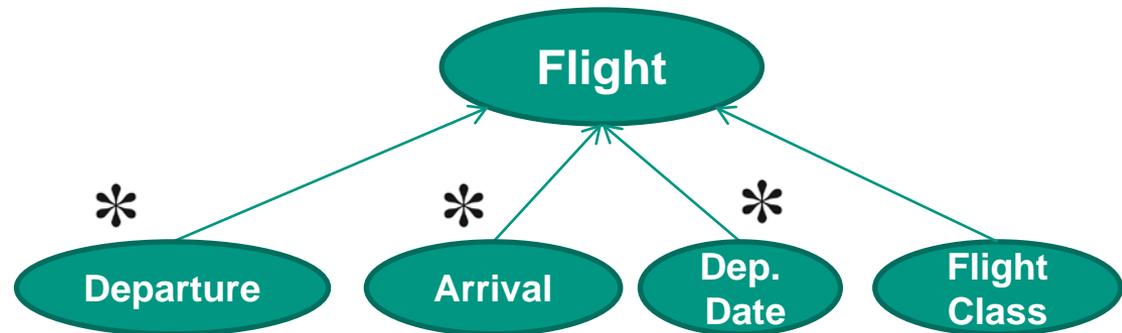
Book a flight \* Mandatory

Departure \*

Arrival \*

Dep. Date \*

Flight Class



### ■ 3 Fragen pro Kategorie

1. Ohne Hilfestellung
2. Bilder von Webformularen
3. Weitere Optionen



### ■ 40 Probanden

- Sprachniveau (20 Muttersprache, 20 Zweitsprache)
- Onlinebuchungserfahrung (24 mit Erfahrung, 16 ohne Erfahrung)
- Alter (von 15 bis 58)
- Bildungsniveau

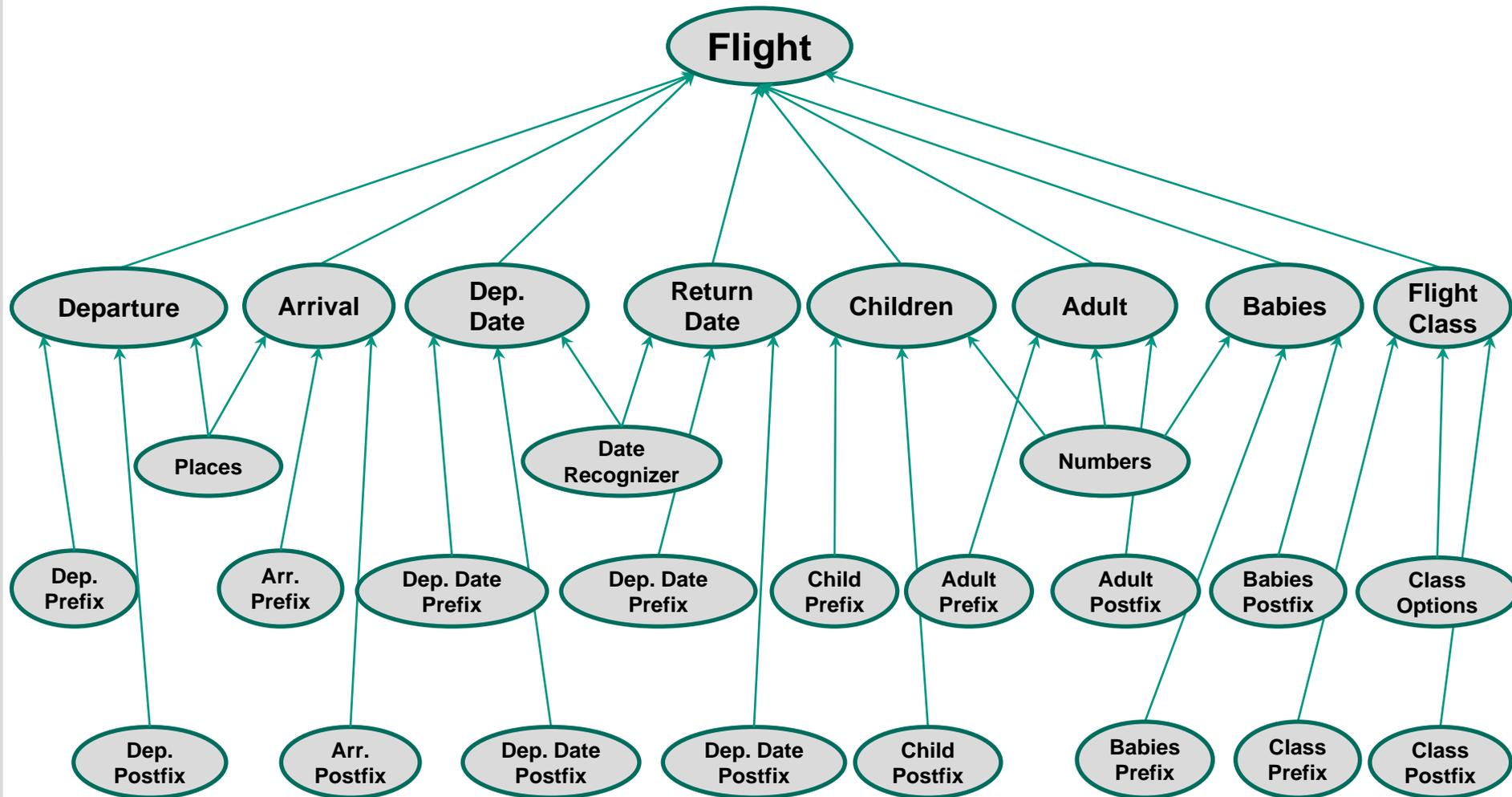
### ■ 50% für die Entwicklung, 50% für die Evaluation

# Ansatz

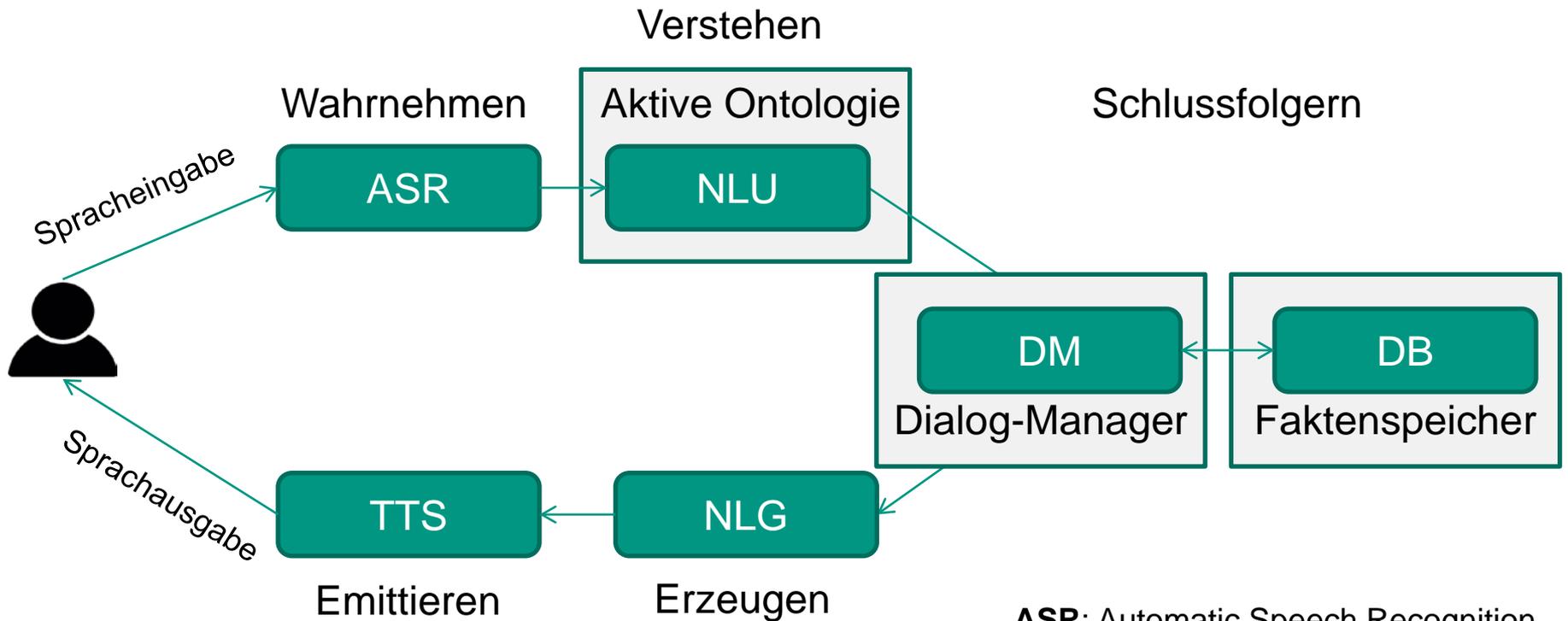
## Erkennung der Konzepte

Konzept	Verwendete Technik	Beispiel
Departure Arrival	Präfixe Postfixe Wörterliste	<b>From</b> Karlsruhe <b>Berlin is the destination</b> <b>Karlsruhe, Berlin</b>
Dep. Date	Präfixe Postfixe reguläre Ausdrücke Bibliothek (SUTime - Stanford)	<b>Fly on</b> 11/12/2015 <b>11/12/2015 is the departure date</b> <b>11/12/2015</b> <b>Next Monday</b>
Flight Class	Präfixe Postfixe Wörterliste	<b>Flight class is</b> economy <b>First class</b> <b>Economy</b>

# Ergebnis Aktive Ontologie „Flug“



# Entwurf Dialog Manager



**ASR:** Automatic Speech Recognition  
**NLU:** Natural Language Understanding  
**DM:** Dialog Manager  
**DB:** Database  
**NLG:** Natural Language Generation  
**TTS:** Text To Speech

# Entwurf Dialog Manager

Book me a flight  
to Manchester

Flight

Departure

Arrival

Dep.  
Date

Flight  
Class

Dialogue

Mandatory

Optional

Fehlende obligatorische  
Eingabe

- 1- Departure
- 2- Departure date

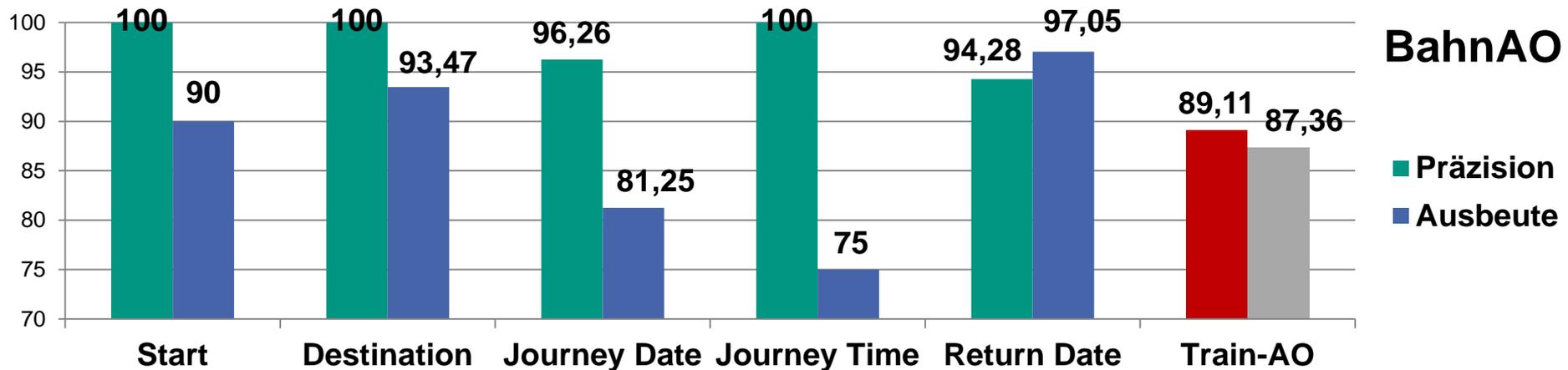
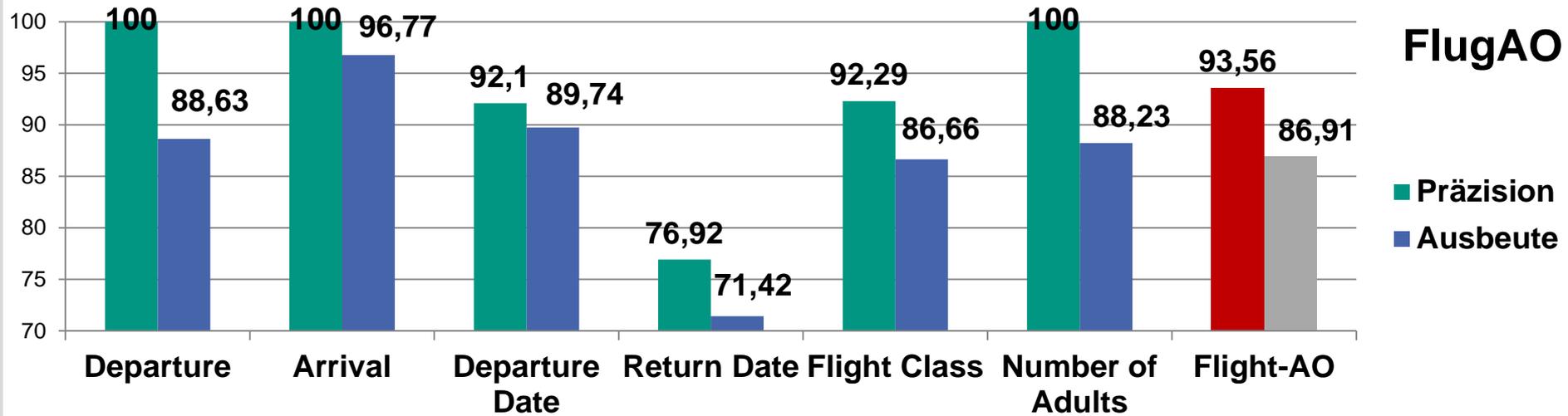
Factstore

Fehlende optionale  
Eingabe

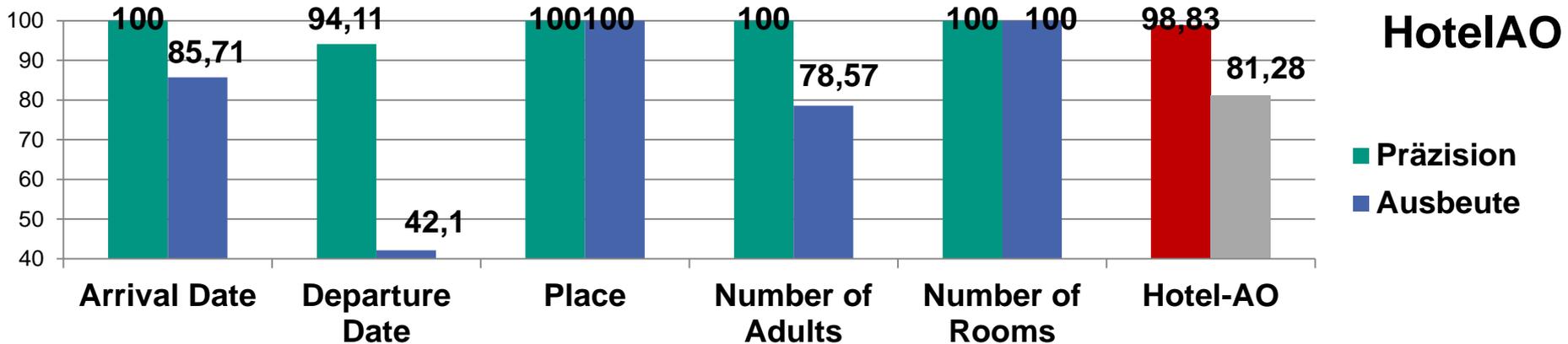
- 1- Flight class

# Demo

# Evaluation



# Evaluation



Eingabe	vollständig	
Erkennung	vollständig	unvollständig
Flug	21 (66%)	11 (34%)
Bahn	15 (68%)	7 (32%)
Hotel	14 (74%)	5 (26%)
Gesamt	50 (65%)	23 (35%)

## ■ **Aktive Ontologien**

- 30 formularbasierte Dienste aus drei verschiedenen Kategorien können verwendet werden

## ■ **EASIER**

- Neue Knotentypen (Datum, Zeit, Ziffern)
- Unterstützung von Ausdrücken die aus mehreren Wörtern bestehen

## ■ **Dialog-Manager**

- Nach fehlende Informationen fragen
- Ergebnisse zeigen

## ■ **Ausblick**

- **Aktive Ontologien**
  - Negation, Referenzen und weitere sprachliche Varianten werden unterstützt
- **Dialog-Manager**
  - Fragen vollautomatisch generieren
  - Aufruf externer Dienste

**Vielen Dank für Ihre  
Aufmerksamkeit!**

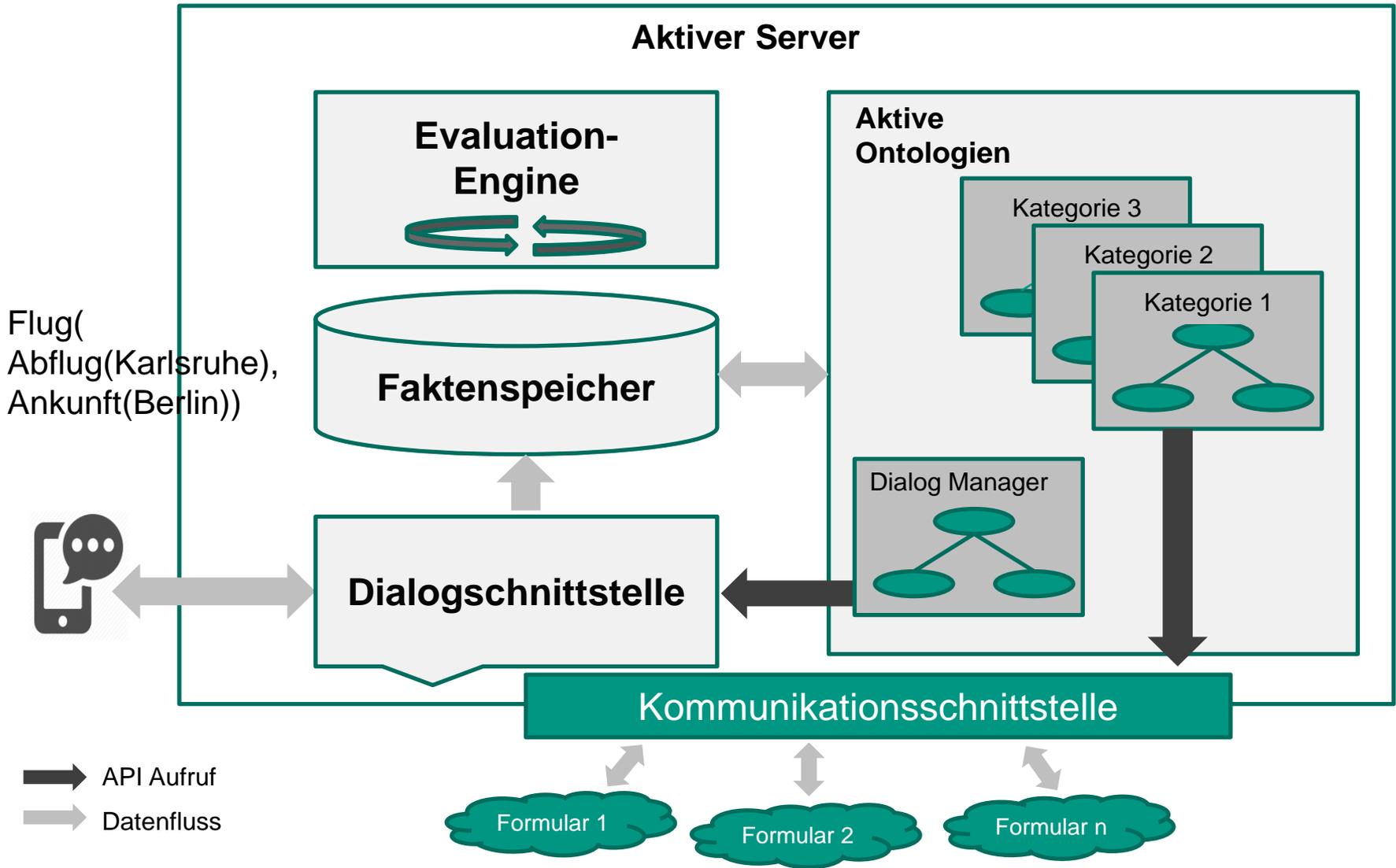
# Literatur

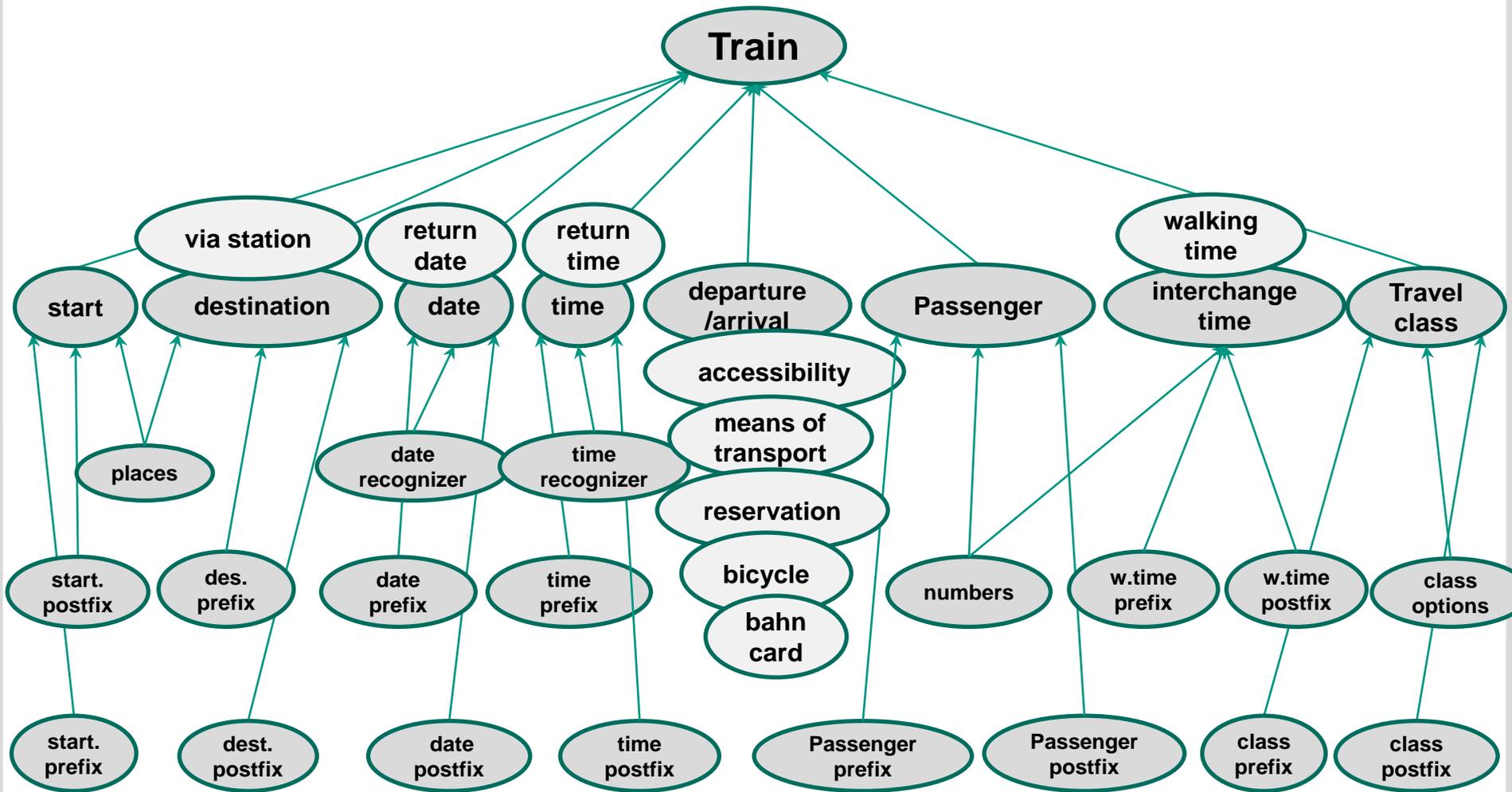
- [Guz08] Guzzoni, Didier: Active: a unified platform for building intelligent applications, Ecole Polytechnique Federale De Lausanne, PhD Thesis, Januar 2008
  
- [BL] Blersch, Martin; Landhäußer, Mathias: EASIER: An Approach to Automatically Generate Active Ontologies for Intelligent Assistants. In: International Symposium on Information Systems and Software Engineering: ISSE 2016
  
- [BJNS10] Berlanga, Rafael ; Jimenez-Ruiz, Ernesto ; Nebot, Victoria ; Sanz, Ismael: Faeton: Form analysis and extraction tool for ontology construction. In: International Journal of Computer Applications in Technology
  
- [RG06] Roitman, Haggai ; Gal, Avigdor: Ontobuilder: Fully automatic extraction and consolidation of ontologies from web sources using sequence semantics. In: Current Trends in Database Technology-EDBT 2006. Springer, 2006, S. 573-576

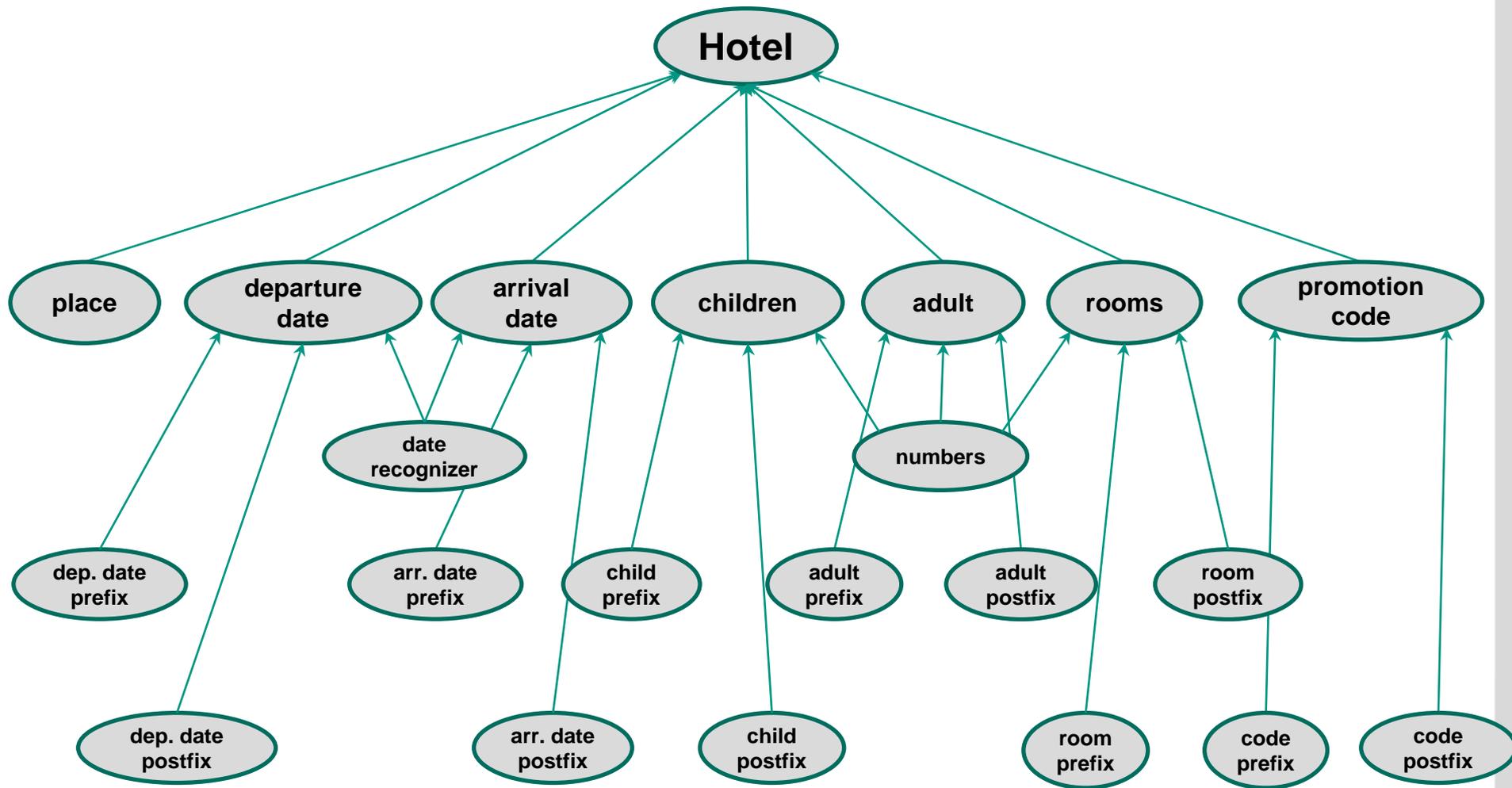
# Literatur

- [DVNR03] Davalcu, H. ; Vadrevu, Srinivas ; Nagarajan, Saravanakumar ; Ramakrishnan, I.V.: OntoMiner: bootstrapping and populating ontologies from domain-specific Web sites. In: IEEE Intelligent Systems 18 (2003), September, Nr. 5, S. 24-33
  
- [AOSS95] Aust, Harald ; Oerder, Martin ; Seide, Frank ; Steinbiss, Volker: The Philips automatic train timetable information system. In: Speech Communication 17 (1995), November, Nr. 3-4, S. 249-262.
  
- [WY07] Williams, Jason D. ; Young, Steve: Partially Observable Markov Decision Processes for Spoken Dialog Systems. In: Comput. Speech Lang. 21 (2007), April, Nr. 2, S. 393-422.

# Ansatz

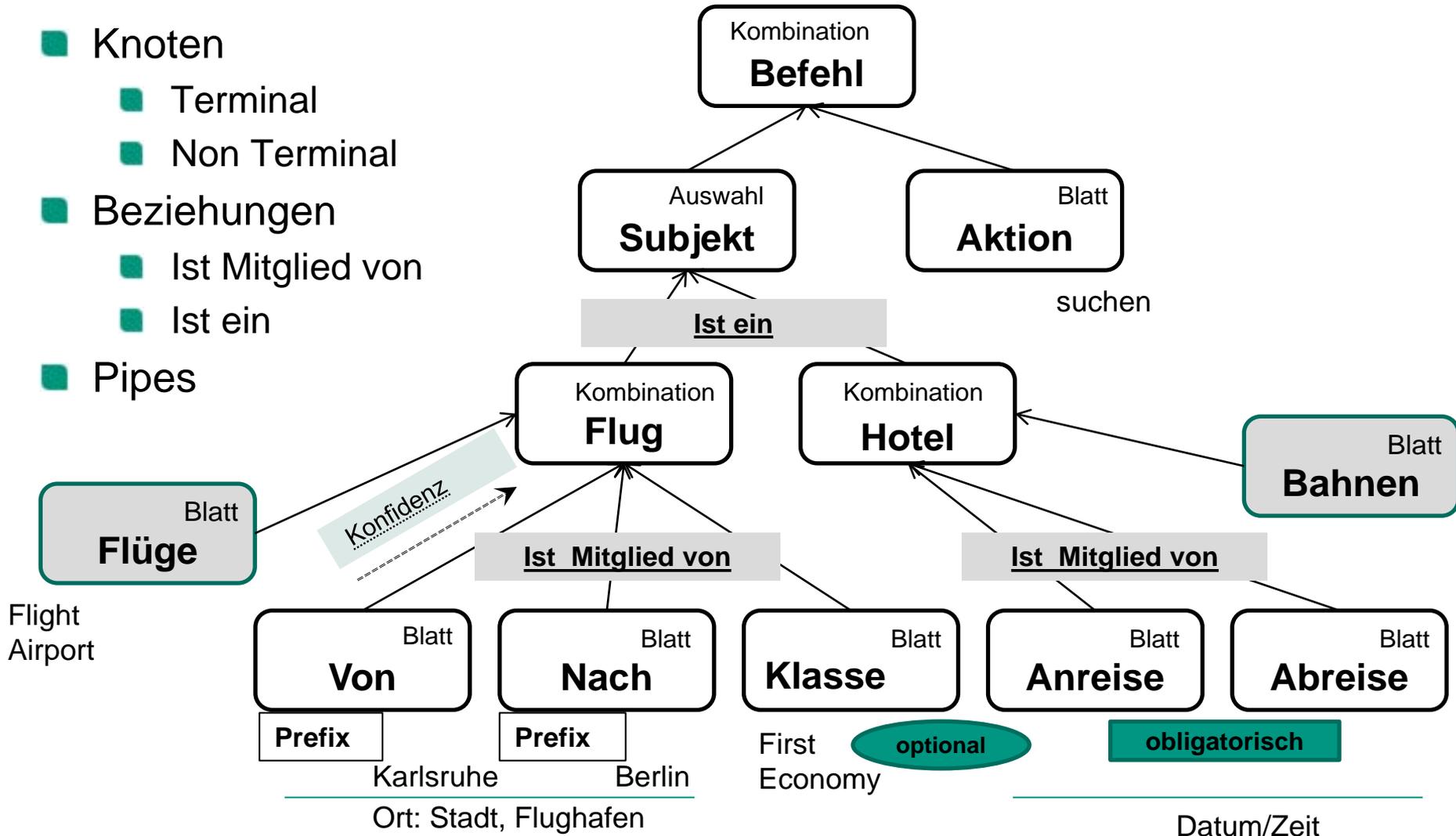






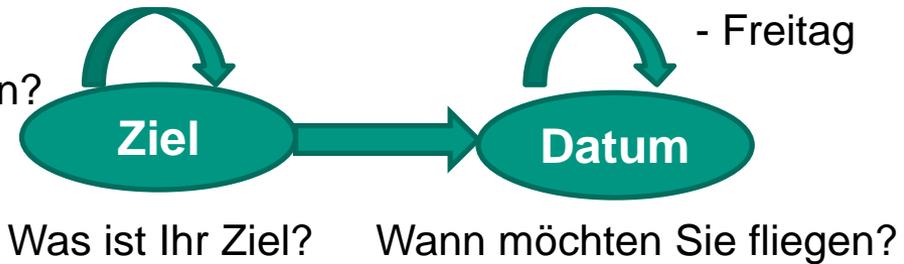
# Grundlagen – Aktive Ontologie

- Knoten
  - Terminal
  - Non Terminal
- Beziehungen
  - Ist Mitglied von
  - Ist ein
- Pipes



## ■ Systeme mit endlichen Automaten

- Berlin
- War es Berlin?
- Ja



## ■ Frame-basierte Systeme

- Was ist Ihr Ziel?
- Berlin am Freitag
- Ich habe die folgenden Flüge



## ■ Agenten-basierte Systeme

- Komplexe Kommunikationen
- Mechanismen für Fehlererkennung und -korrektur
- “Mixed-initiative” Systeme

# Ansatz – Webseite auswählen

Fluggesellschaft	Kontinent	Land	Typ	Bewertung
1- Lufthansa	Europa	Deutschland	Privat	4
2- Air New Zealand	Neuseeland	Neuseeland	Staatlich	4
3- Cathay Pacific Airways	Asien	China	Staatlich	5
4- South African Airlines	Afrika	Südafrika	Staatlich	4
5- British Airways	Europa	Großbritannien	Staatlich	4
6- Ryanair	Europa	Irland	Privat	2
7- Air Canada	Nordamerika	Kanada	Staatlich	4
8- Condor	Europa	Deutschland	Privat	3
9- Emirates	Asien	UAE	Staatlich	4
10- Swiss Airlines	Europa	Schweiz	Staatlich	4

# Ansatz – Nutzeranfrage analysieren

## ■ Probanden

### ■ Sprachniveau: Englisch als Mutter- und Zweitsprache

- “I am headed London and I would like to fly out of Frankfurt”
- “I want to fly from Frankfurt to London”

### ■ Onlinebuchungserfahrung:

- “I would like to use my 25% Bahncard to book a first class train ticket to Berlin”

### ■ Alter: kurzer ungrammatisch Sätze oder lange fehlerfreie Sätze

- “Flight from Stuttgart to Manchester first class 2 adults on 12/01/2016 and back on 20/01/2016”
- “I would like to enquire about flights to Berlin from Frankfurt Airport please. I am off to a concert at September 28<sup>th</sup> until October 4<sup>th</sup>. Could you tell me about the flight availability and prices”

### ■ Ausbildung: Informatiker (Teilnehmer des Praktikums “Dialogmodellierung”)

- “I want to do my Christmas shopping in New York, so I want to fly there”

# Evaluation

Kriterium	Kategorie	Anzahl
Sprachniveau	Englisch als Muttersprache	20
	Englisch als Zweitsprache	20
Onlinebuchungserfahrung	Mit Erfahrung	24
	Ohne Erfahrung	16
Alter	15 - 35 Jahre	24
	36 - 45 Jahre	10
	46 - 55 Jahre	4
	mehr als 55 Jahre	2
Ausbildung	Angestellte	5
	Hochschulabsolventen	7
	Informatiker	7
	Studenten	15
	Schüler	3
	Kein Abitur	3

# Ansatz Dialog Manager

Book me a flight  
to Manchester

Flight

Departure

Arrival

Dep.  
Date

Flight  
class

Dialogue

Mandatory

Optional



Fehlende obligatorische  
Eingabe

- 1- Departure
- 2- Departure date

Factstore

Fehlende optionale  
Eingabe

- 1- Flight class

# Evaluation – Gründe

## ■ undefinierte Ausdrücke

- „Book me flight between Karlsruhe and berlin“
- „I would like to book a ticket on the flight heading towards London“

## ■ undefinierte Formulierungen

- „I do not want the flight class to be first“

## ■ Referenze

- „I want to do my christmas shopping in New York, so I want to fly there“

## ■ Datum: Suntime

- „I want to fly on Saturday first class“
- „I want to book a room in Berlin for two nights“