

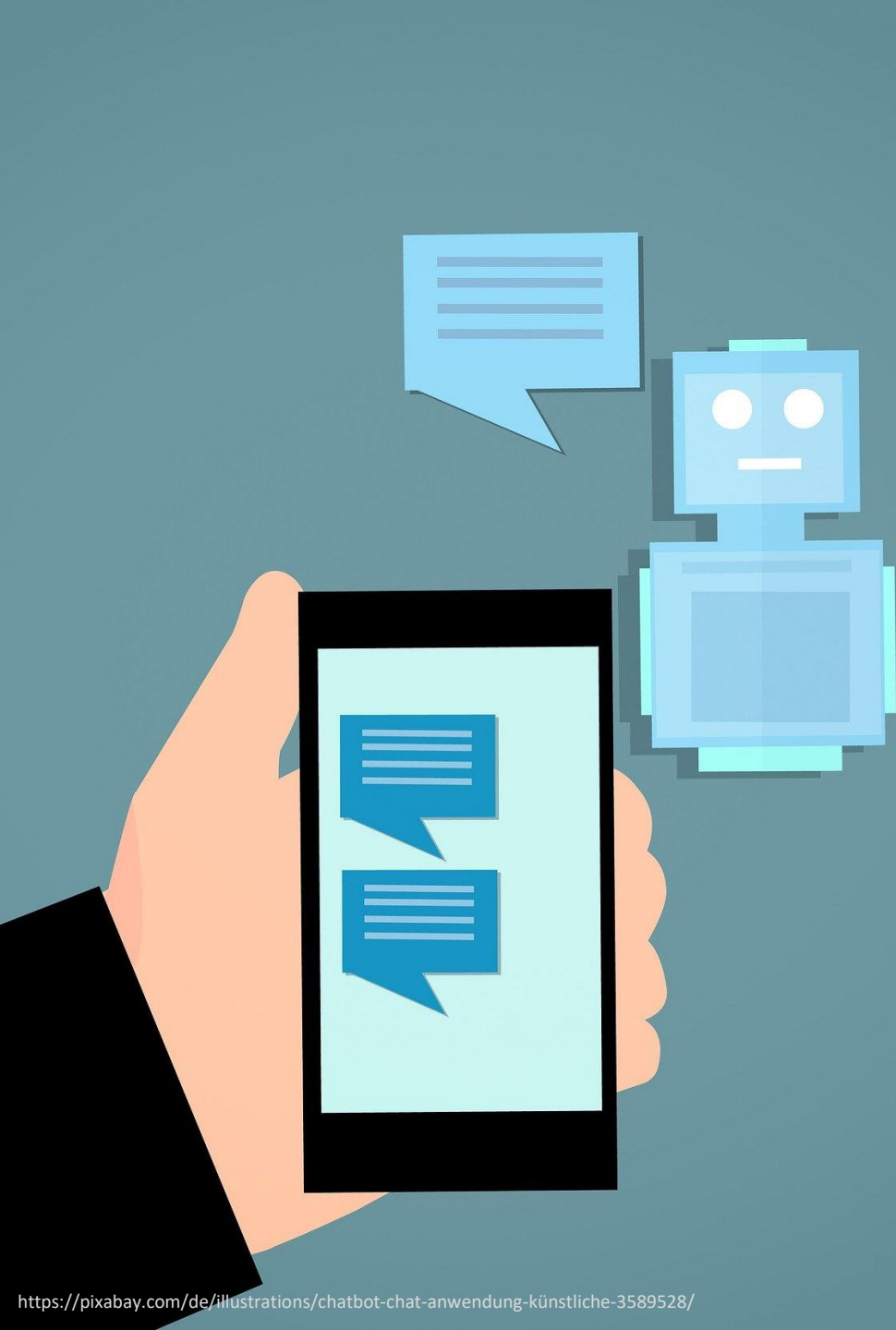
# **Theoretische Informatik III (T3INF2002)**

Formale Sprachen und Automaten | Einführung Compilerbau

Vorlesung im Wintersemester 2022/23

# **Natural Language Processing**

## **Anwendungsbereiche und Funktionsweise**



# Schaffung einer “natürlicheren” interaktiven Kommunikation zwischen Computer und Mensch

Computersysteme sollen sich nahtlos in den menschlichen Alltag integrieren und immer und überall als Unterstützungswerkzeug zur Verfügung stehen:

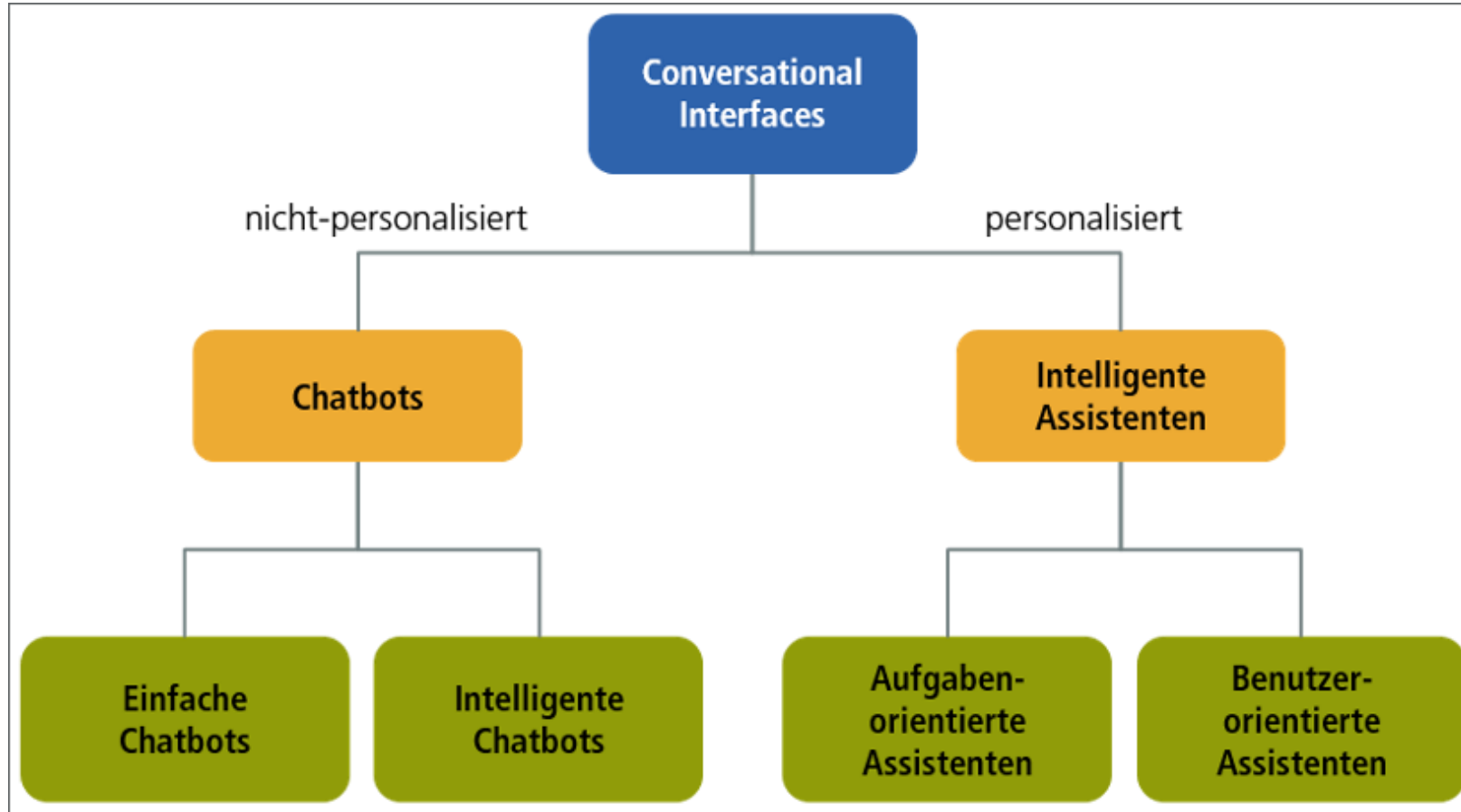
- Klassische Geräte und Programme sind zur spontanen Nutzung oft zu schwerfällig
- CI stellen Kommunikationsschnittstellen dar, die eine natürliche Art der Interaktion mit Computern erlauben und basieren auf natürlicher menschlicher Sprache

# Arten von Conversational Interfaces

## Zwei Entwicklungstendenzen:

- **Spezialisierte CI:** zugeschnitten auf einen stark eingegrenzten Bereich oder eine einzelne Aufgabe (z.B. Chatbots, die als Wetterexperten fungieren oder bei medizinischen Diagnosen behilflich sind)
- **„Virtuelle Assistenten“:** decken eine große Bandbreite an unterschiedlichen Funktionen und Aufgaben ab (z.B. smarte Assistenten wie Siri, Cortana, Alexa).
  - Übernehmen analog zu einem „echten“ Assistenten Aufgaben wie Termin- und Reiseplanung, Einkaufsplanung und -unterstützung oder die Steuerung eines Smart Homes

# Abgrenzung unterschiedlicher Arten von Conversational Interfaces (forts.)



Abgrenzung unterschiedlicher Arten von Conversational Interfaces (Quelle: Eigene Darstellung)

# Abgrenzung unterschiedlicher Arten von Conversational Interfaces (forts.)

**Nicht-personalisierte CIs:** sind Chatbots, die sich in einfache und intelligente Chatbots unterteilen lassen

- Chatbots haben einen begrenzten Zugang zu Informationen und nutzen in der Regel keine persönlichen Informationen der Nutzer:innen, um sich anzupassen
- Chatbots laufen in einer zentralisierten Infrastruktur und werden von Unternehmen oder spezialisierten Anbietern betrieben und deren Ablauf angepasst

**Personalisierte CI:** Intelligente Assistenten, die ein Gedächtnis haben, durch das sie personalisieren können und ihr „Wissen“ stets erweitern

- **Benutzer-orientierte Assistenten** erfassen individuelle Umstände und bauen ein Nutzerprofil bzw. -modell auf
- **Aufgaben-orientierte Assistenten** zielen auf die Unterstützung von Nutzenden ab, während diese eine Tätigkeit oder Aufgabe ausführen -> dabei können auch persönliche Daten speichern und Profil erstellen werden

# Was macht intelligente Assistenten aus?

## Exkurs: Turing-Test

- Ziel von Entwickler:innen ist es, Chatbots und intelligenten Assistenten immer besser an den Sprachgebrauch und die Konversation mit dem Menschen anzupassen
- Wie gut das gelingt, wird durch Benchmarking gemessen
  - Nachweis der Konversationsfähigkeit eines künstlichen Systems ist der Turing-Test

Der englische Mathematiker Alan Turing entwickelt den Turing-Test 1950. Der Turing-Test wird eingesetzt, um herauszufinden, ob ein künstliches System über eine vergleichbare Intelligenz wie der Mensch verfügt (Turing, 2009).

-> Alan Turing lieferte hierfür das gedankliche Konstrukt und die theoretische Skizze, die erst durch die zunehmende Verbreitung von künstlicher Intelligenz zu einem Test ausgearbeitet wurde.

# Was macht intelligente Assistenten aus?

## Exkurs: Turing-Test (forts.)

Kritik: Test überprüft lediglich die Funktionalität eines Systems und hinterfragt nicht, ob der Intelligenz auch ein bewusstes oder intentionales Handeln zugrunde liegt.

— Turing-Test kann auch von einem System bestanden werden, das menschliches Verhalten nur imitiert

**Imitation menschlichen Verhaltens:** System ist so programmiert, dass der Gesprächspartner und der Gesprächsverlauf manipuliert wird, sodass der Eindruck entsteht, es handele sich um einen Menschen. Um dies zu erreichen, sind keine komplexen intelligenten Systeme notwendig.

**Beispiel:** Chatbot „Eugene Goostman“

Im Juni 2014 gelang es, mehr als 30 Prozent einer Jury davon zu überzeugen, dass es sich bei dem Chatbot um einen Menschen handelte (Warwick & Shah, 2016)



# Chatbots

# Überblick Chatbots

**Chatbots sind kleine digitale Helfer, die Nutzer:innen Kontext-spezifische Informationen zur Verfügung stellen sollen:**

- Interaktion mit den Chatbots erfolgt durch Text- oder Spracheingabe
- Chatbots sind häufig auf einen Aufgabenbereich oder eine konkrete Aufgabe spezialisiert
  - Innerhalb des spezialisierten Bereiches, können Chatbots relativ gut mit den Nutzer:innen agieren

## **Beispiel: Servicebereich im Unternehmen**

Beratung und Unterstützung von Kunden, z. B. bei der Beantwortung allgemeiner Fragen. Ziel ist es die Mitarbeiter von Unternehmen zu entlasten und „Rund-um-die-Uhr“ Service für Kunden anzubieten. Chatbots können als ein interaktives und effizientes Frequent-Ask-Question-Werkzeug (FAQ) genutzt werden.

# Chatbot - Beispieldialog



Beispiel eines Dialogs mit einem Chatbot (Quelle: Eigene Darstellung)

Unterstützung auch innerhalb des Unternehmens möglich, wie im Beispieldialog.

# Arten von Chatbots

**Technischen Merkmale und der Funktionsumfang ermöglichen eine Unterteilung von Chatbots, in einfache und intelligente Chatbots.**

## Chatbot-Arten

- Einfache Chatbots:
  - Regelbasierte Chatbots
  - Anwendungsspezifische Chatbots
- Intelligente Chatbots:
  - NLP-basierte Chatbots
  - KI-basierte Chatbots

# Regelbasierte Chatbots

# Regelbasierte Chatbots

**Klassische Chatbots** sind regelbasierte Interaktionsschnittstellen, mit vordefinierten Antwortmöglichkeiten, die auf Basis von bestimmten Eingaben der Nutzer:innen abgerufen und präsentiert werden.

**Regelbasierte Chatbots** besitzen hinterlegte Antwortmöglichkeiten und sind auf das einprogrammierte Skript beschränkt.

- Können nur sehr einfache und standardisierte Abläufe abbilden (z. B. Fragen zu Produkten und Leistungen eines Unternehmens beantworten)
- Eignen sich z. B. um Abläufe zu erklären (z. B. Ablauf des Bewerbungsprozesses in einem Unternehmen)

# Regelbasierte Chatbots (forts.)

Hinter **regelbasiert** steckt der Ansatz der Erkennung von Wörtern und Phrasen, die vordefiniert sind, um auf Eingaben der Nutzer:innen reagieren zu können.

- **Änderungen / Erweiterungen** an regelbasierten Chatbots sind **aufwändig**, da die Sequenzen von Dialogen aufeinander aufbauen
  - Die Mustererkennung muss für die Eingaben entsprechend vorhanden sein
- **Regelbasierte Ansätze** finden sich in (intelligenten) Systemen und **ermöglichen nachvollziehbare Entscheidungen**
  - Auch in KI-Systemen werden teilweise regelbasierte Ansätzen verwendet

## Probleme:

- Werden unbekannte Begriffe verwendet oder sind Rechtschreibfehler in der Eingabe vorhanden, kann der Chatbot meist nicht damit umgehen
- Chatbots haben daher einen eher schlechten Ruf und gelten als „dumm“ und „nervig“, da den Nutzer:innen oft der Zweck und die Limitierung des Chatbots nicht klar ist und diese auch nicht erkennbar werden

# Anwendungsspezifische Chatbots



# Anwendungsspezifische Chatbots

**Anwendungsspezifische Chatbots** sind eine Mischung aus regelbasiertem und intelligentem Ansatz.

- Werden in einen bestimmten Anwendungsbereich eingesetzt (z.B. Finanzberatung, Terminvereinbarung beim Arzt)
- Können in ihrem Anwendungsbereich flexibel auf Fragen reagieren
- Antwortmöglichkeiten werden aus einer vordefinierten Kategorie von Texten abgerufen, ohne dabei einen fest vergebenen Dialog durchlaufen zu müssen
- Weitere Funktionen, wie z. B. eine Terminreservierung, können vorbereitet werden, indem Informationen aus angebundenen Systemen abgefragt.

Aufgrund ihres abgegrenzten Aufgabenbereiches kann auf einfache Verfahren zurückgegriffen werden, die sowohl eine schnelle Antwortzeit als auch eine gute Qualität ermöglichen.

# NLP-basierte Chatbots

# NLP-basierte Chatbots

NLP-basierten Chatbots verarbeiten und analysieren die Eingaben der Nutzenden mittels Natural Language Processing.

- Chatbot ist unempfindlicher gegenüber Rechtschreibfehlern
- Ermöglicht eine Flexibilisierung des Dialogs zwischen Nutzer:innen und Chatbot

NLP-basierte Chatbots besitzen einen Daten-getriebenen Ansatz.

- Auf Basis großer Textmengen der Anwendungsdomäne (z. B. Texte von Nachrichtenportalen) wird ein Korpus aufgebaut, der die Menge an Antwortmöglichkeiten darstellt.
- Der Korpus kann einfach erweitert werden, ohne dass der Chatbot angepasst werden muss.

# NLP-basierte Chatbots (forts.)

## NLP-basierte Chatbots sind:

- Flexibler
- In ihren Antwortmöglichkeiten weniger limitiert
  - Dennoch können sie nicht alle möglichen Fragen der Nutzenden beantworten
  - Antworten weisen ggf. eine große Unschärfe auf

## Korpus kann nicht beliebig groß sein und auch nicht alle möglichen Texte enthalten, da sonst:

- Die Dauer der Antwortzeit zu lange wird
- Ein Konflikt bei der Priorisierung von Antworten entsteht

### Wichtige Fragestellungen:

- Was will der Nutzende eigentlich wissen
- Frage der Relevanz von Informationen spielt eine große Rolle
- Einfache NLP-Ansätze bzw. NLP-basierte Chatbots können dieses Problem nicht lösen

# KI-basierte Chatbots

# KI-basierte Chatbots

**KI-basierte Chatbots** lernen grundsätzlich durch den Einsatz von „*Deep Learning*“.

- Chatbot wird mithilfe von Algorithmen des maschinellen Lernens erstellt
- Lernt aus den zur Verfügung stehenden Daten, die aus dem Mensch-zu-Chatbot-Dialog oder einem Dialog von Mensch zu Mensch stammen
- Nutzung von Gesprächsverläufen zwischen Personen in Foren, Chats oder aufgezeichnete Gespräche (z. B. Filmdialoge) als Datenbasis

Der KI-basierte Chatbot wird trainiert, sein eigenes Bewusstsein für den Text zu entwickeln.

- Chatbot wird beigebracht, wie er sich mit Menschen unterhalten kann
- Effektivität des maschinellen Lernens hängt von der Datenmenge und der Datenqualität ab und verbessert sich, je mehr „hochwertige“ Daten zur Verfügung stehen

# Beispiele von Chatbots

# Chatbot ELIZA

- Erste Chatbot war ELIZA und wurde von Professor Joseph Weizenbaum am MIT in den 1960er-Jahren entwickelt
- ELIZA diente der Nachahmung einer menschlichen Konversation und wurde durch Pattern Matching und Substitutionsverfahren realisiert (Weizenbaum, 1966)

Pattern Matching ist ein Vergleichsverfahren, bei dem eingegebene Textsequenzen auf bestimmte Wörter oder Phrasen überprüft werden.

Mit dem Substitutionsverfahren werden Texte an definierten Stellen entsprechend der Eingabe ausgetauscht.



## Chatbot ELIZA (forts.)

# Eliza

ELIZA: How do you do. Please tell me your problem.  
YOU: I have no problems  
ELIZA: Can you elaborate on that ?

Talk

Reset

/ ʔAǎL.C. s.jCũshlʔ LL Cũš ō 9[Lù! □v] š □š 9Itǎ□ لله للشيء

# Chatbot Jabberwacky

- Jabberwacky wurde von Rollo Carpenter im Jahre 1988 geschrieben und hatte zum Ziel, einen natürlich wirkenden, unterhaltsamen und humorvollen Chat zu simulieren
- Jabberwacky merkt sich alle Gespräche, die er mit Nutzer:innen führte, um aus diesen Gesprächen zu lernen (Boonthum-Denecke, McCarthy & Lamkin, 2011)

-> Dieser mitlernende Ansatz hat viel zum technologischen Fortschritt beigetragen und wird auch heute noch für Forschungszwecke verwendet.

# Chatbot Jabberwacky (forts.)

Chatting to **jabberwacky**

Are you an artist?

No.

What are you then?

A student.

A student of what?

Computer science.

Really?

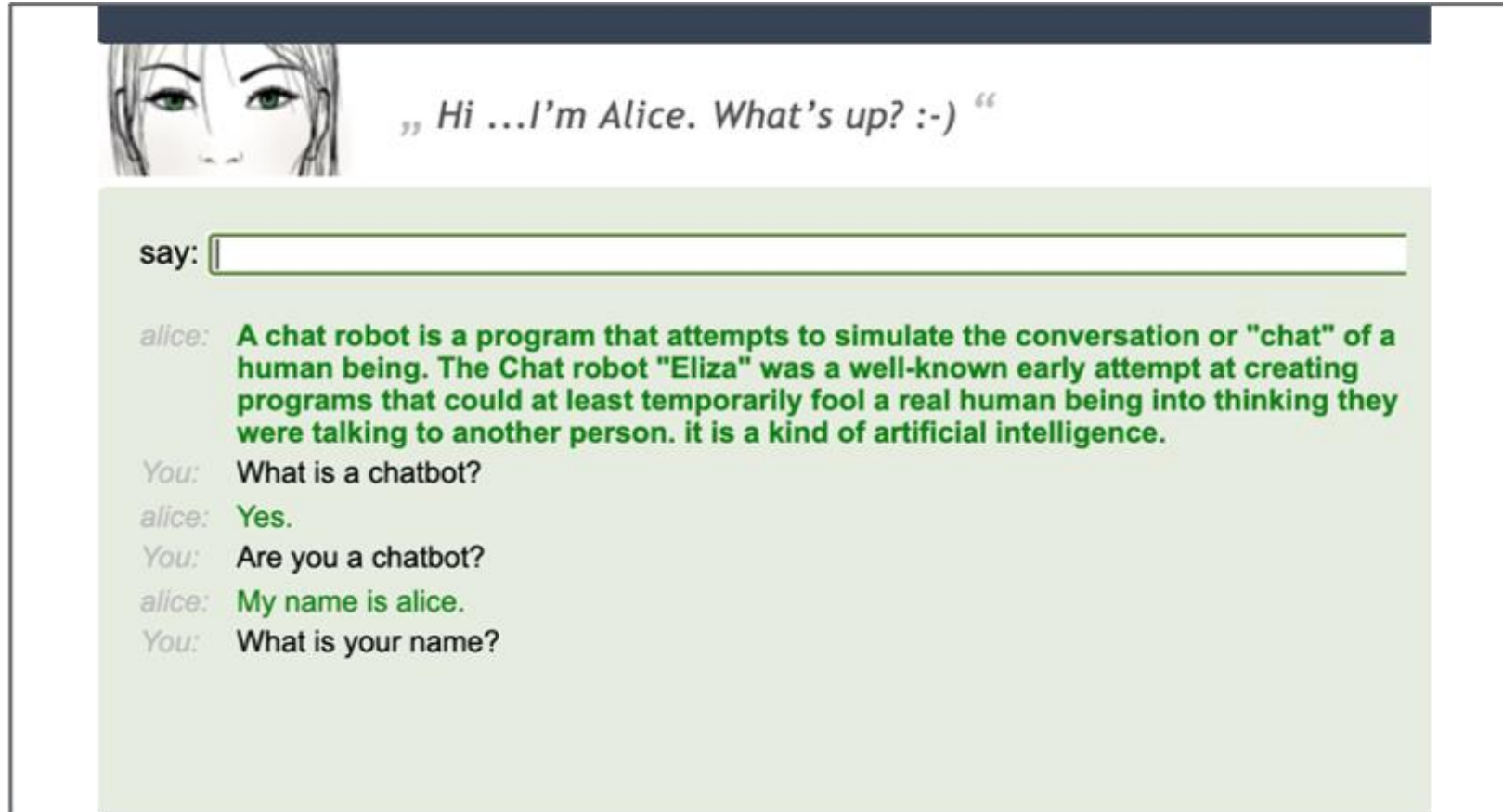
Chatbot Benutzerschnittstelle von Jabberwacky (Quelle: <https://www.jabberwacky.com/>)

# Chatbot A.L.I.C.E. - Artificial Linguistic Internet Computer Entity

- A.L.I.C.E. (Artificial Linguistic Internet Computer Entity) wurde 1995 entwickelt und durch Richard Wallace vorangetrieben
- Verwendung eines heuristischen Matching-Muster, das sich als wesentliche Verbesserung gegenüber früheren Chatbots erwies
- Wallace entwickelte ein XML-Schema für die Formalisierung des Dialoges, das als „Artificial Intelligence Markup Language“ (AIML) bekannt ist (Wallace, 2009)

AIML ist eine für alle zugängliche, minimalistische Sprache zur Entwicklung von Chatbots, die den Funktionsumfang von A.L.I.C.E besitzen.

# Chatbot A.L.I.C.E. - Artificial Linguistic Internet Computer Entity (forts.)



Chatbot Benutzerschnittstelle von ALICE (Quelle: Fellmann (2013))

# Intelligenten Assistenten

# Überblick Intelligenten Assistenten

**Intelligente Assistenten** sind den intelligenten Chatbots ähnlich:

- Verwendung maschineller Lernverfahren und des Natural Language Processing, um die **Interaktion mit den Nutzenden so Menschen-ähnlich** wie möglich zu **gestalten**

**Metapher** des **persönlichen Assistenten**, der die Gewohnheiten, Termine, den Charakter und die Eigenschaften des Chefs oder der Chefin kennt, ist die Leitlinie hinter intelligenten Assistenten.

-> Dazu reicht die reine sprachliche Anpassung und das führen eines Menschen-ähnlichen Dialogs nicht mehr aus

# Überblick Intelligenten Assistenten (forts.)

Neben dem Sprachgebrauch, müssen auch das persönliche Umfeld und die Eigenschaften der Nutzer:in berücksichtigt werden, um in verschiedenen Situationen angemessen unterstützen und reagieren zu können.

- Von diesem Szenario sind Forschung und Praxis noch recht weit entfernt
  - **Personalisierung und die Adaptionfähigkeit von Computersystemen** auf neue, unbekannte Situationen ist noch schwer zu realisieren
- > Die Forschung beschäftigt sich im **Bereich der Emergenz und der emergenten Systeme** mit dieser Anpassbarkeit und Funktionsfähigkeit eines sich ausgleichenden Regelkreises zwischen Mensch und Computer (Russell & Norvig, 2010, S. 1001–1002).



# Intelligenten Assistenten - Eigenschaften

## Intelligenten Assistenten können ...

- ... sprachgesteuerte und/oder textbasierte Benutzerschnittstellen aufweisen.

## Sie sind in der Lage ...

- ... die durch Nutzer:innen, in natürlicher Sprache, ohne Beachtung eines Sprachmusters oder einfacher Sprache, gestellten Fragen zu verarbeiten und Antworten zu generieren oder Aufgaben auszuführen.

## Anders als bei Chatbots gehört die Ausführung von bestimmten Aufgaben bei intelligenten Assistenten mittlerweile zu den Basisfunktionalitäten:

- Verknüpfung verschiedener Geräte der Nutzer:in und geräteübergreifende Personalisierung
- Nutzung von Standortinformationen, um auf Geschehen im direkten Umfeld aufmerksam zu (z.B. Sightseeing, Wetterprognose, Verkehrssituation)

# Benutzerorientierte intelligente Assistenten

**Benutzerorientierte intelligente Assistenten** haben einen starken Bezug zu ihren Nutzer:innen und begleiten diese über ein persönliches, meist mobiles Endgerät.

- Beobachten das Verhalten der Nutzer:in, um sich an die individuellen Bedürfnisse anzupassen und aktiv in speziellen Situationen zu unterstützen
  - Umfasst Informationsbedürfnisse, Interessen, Gewohnheiten und die aktuelle Situation
- Passen sich im Laufe der Zeit an die Bedürfnisse der Nutzer:in an, wenn sich (langfristige) Änderungen im Verhalten und den Interessen erkennen lassen

Persönlichen Assistenten wie z. B. von Apple, Google und Microsoft sind solche benutzerorientierten Assistenten und sollen ihre Qualitäten als digitale, intelligente Begleiter noch ausbauen.

# Beispiele von Intelligenten Assistenten

# Sprachassistent Amazon Alexa

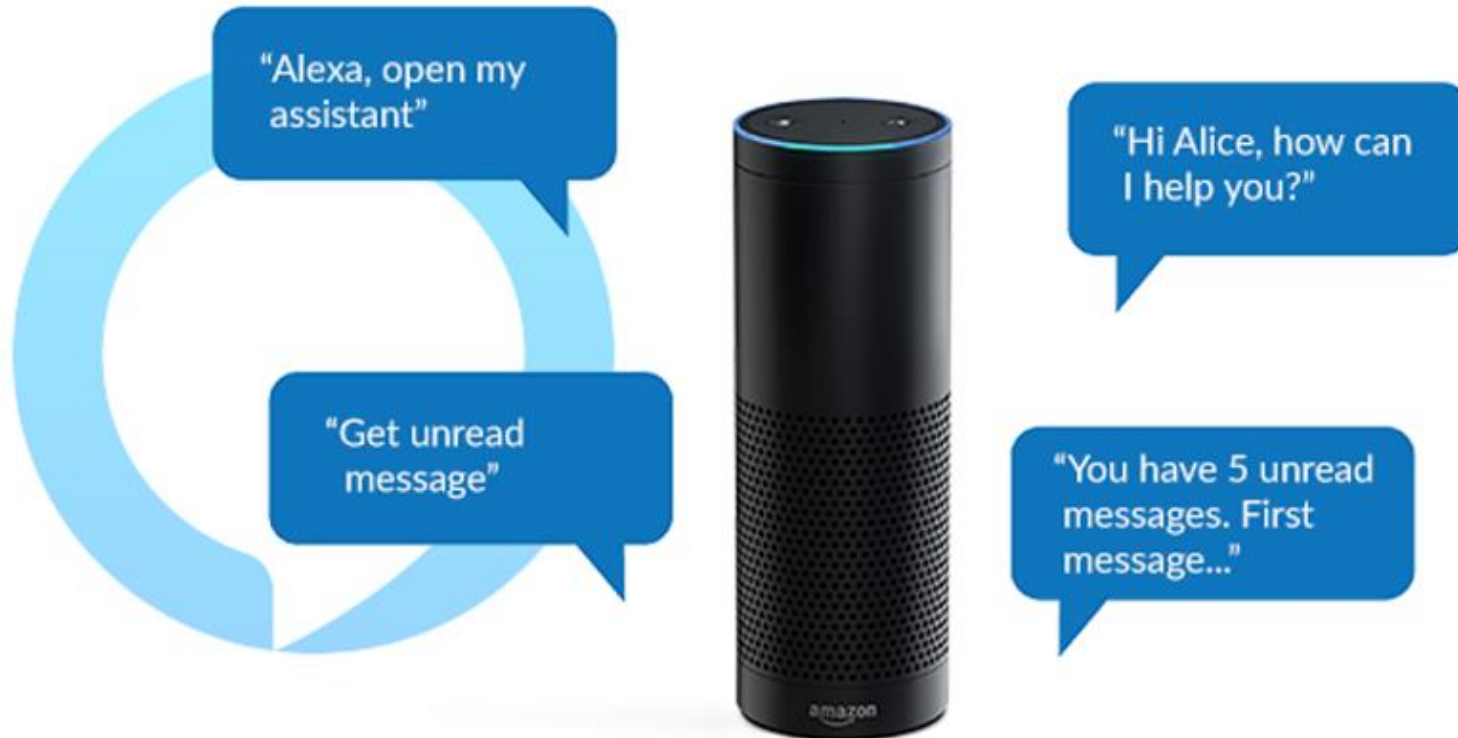
## **Amazon stellte 2014 seinen intelligenten, persönlichen Assistenten Alexa vor**

- Rund um Alexa und dessen Endgeräte entstand ein neues Segment an Smart-Home-Produkten, die durch Alexa gesteuert werden können
- Mittels Sprachbefehlen der Nutzer:innen
  - sucht Alexa Informationen im Internet,
  - spielt Musik, Podcasts oder Hörbücher ab,
  - erstellt Notizen oder Einkaufslisten,
  - liest Nachrichten und Wetterberichte vor

Alexa reagiert nur auf Sprachkommandos und Spracheingabe

# Sprachassistent Amazon Alexa

## Alexa for RingCentral



Amazon Alexa Sprachassistent (Quelle: Vu (2020))

# Was steckt eigentlich hinter den CI? – Hintergründe und Grundlagen

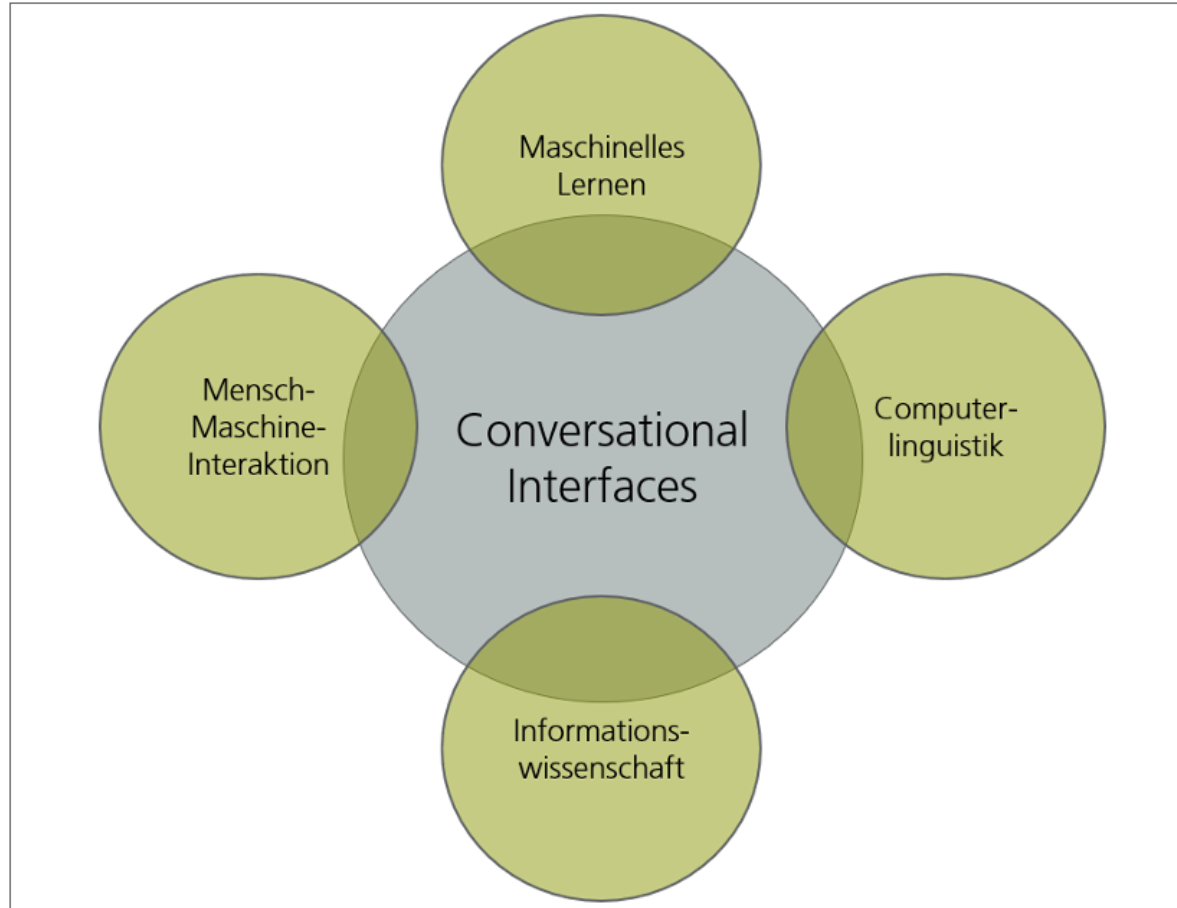
2

# Interdisziplinäre Themenbereiche tangieren das Thema Conversational Interfaces

**Conversational Interfaces (CI)** sind ein wichtiger Bestandteil der Informatik und haben sich innerhalb der Informatik-Disziplinen zu eigenständigen Wissenschafts- und Anwendungsgebieten entwickelt.

- Nutzung von Ansätzen und Erkenntnissen aus anderen Wissenschaften und Fachrichtungen
- **Disziplinen (Carstensen et al., 2010):**
  - Informatik (künstliche Intelligenz und Mensch-Computer-Interaktion)
  - Kognitionswissenschaften und der Sprachwissenschaft
- Verbreitung von CI aufgrund des technischen Fortschritts in der Hardwareentwicklung, bei mobilen Endgeräten (z. B. Smartphones und Tablets) und die Verbesserung von Algorithmen im Bereich der künstlichen Intelligenz

# Schnittstellen zu anderen Disziplinen



**Künstliche Intelligenz** -> wichtigste Teilgebiete sind das maschinelle Lernen, die Wissensrepräsentation (als ein Teil der Informationswissenschaft) und das Sprachverstehen.



# Computerlinguistik

# Computerlinguistik

Conversational Interfaces verarbeiten natürliche Sprache, sowohl als Input als auch als Output.

Mit der natürlichen Sprachverarbeitung beschäftigt sich die Computerlinguistik, die sich als eine eigenständige Disziplin etabliert hat:

- Vereint Ansätze und Methoden der **Sprachwissenschaft** und **Informatik**, um zu untersuchen, wie natürliche Sprache in Form von Text- oder Sprachdaten zerlegt, aufbereitet und algorithmisch verarbeitet werden kann
- **Computerlinguistik** ist ein Fachgebiet, das sich im deutschsprachigen Raum gebildet hat (Carstensen et al., 2010).
  - In der englischen Literatur wird dieser Bereich als **Natural Language Processing (NLP)** bezeichnet

# Computerlinguistik (forts.)

**Computerlinguistik** entstand in den 1960er-Jahren, wie als auch die Grundsteine für die Entwicklung künstlicher Intelligenz gelegt worden.

- Einsatz künstlicher Intelligenz versprach **Automatisierung der Sprachverarbeitung** und der **Sprachübersetzung**
  - Dies konnte während der Anfangsjahre jedoch nicht erreicht werden
  - Praktische Umsetzung konnte nicht mit den Erwartungen an die neu entwickelten theoretischen Ansätze mithalten
- Zu den Errungenschaften dieser Zeit gehören Chomskys Arbeit *Syntactic Structures* von 1957
  - Chomskys beschreibt die Sprache in einem neuen formalen Gerüst (Chomsky, 2002).
- Weitere Errungenschaft stellt die Entwicklung der Sprachlogiken durch Saul Kripke (1972) und Richard Montague (1972) dar

# Computerlinguistik (forts.)

## Die moderne Computerlinguistik teilt sich in zwei Bereiche auf:

- Theoretischer Bereich, der sich mit der klassischen philosophisch-linguistischen Fragestellung befasst
- Praktischer Bereich, der sich mit der informatisch-praktischen Anwendung befasst

Für die Entwicklung von Conversational Interfaces ist vor allem der praktische Bereich von Bedeutung.-> Die dabei angewandten Methoden und genutzten Verfahren basieren auf den theoretischen Erkenntnissen der Computerlinguistik.

- **Am Anfang einer jeden Konversation – auch mit einem Computer – steht das Wahrnehmen und Verstehen von Sprache**
  - -> Menschen haben dazu Ohren, die das Gesagte aufnehmen und den Schall in Laute und Wörter übersetzen, und sie verfügen über das Wissen einer Sprache, um das Gesagte in einen Zusammenhang zu bringen

# Computerlinguistik (forts.)

**Computern muss das Wahrnehmen und Verstehen erst beigebracht werden.**

Signale müssen aus **Schallinformationen**, bei gesprochener Sprache und/oder aus textuellen Zeichenketten (bei geschriebener Sprache) **erkannt** und **verstanden werden**:

- Erfordert, dass die **Sprache analysiert** wird, indem schrittweise die Eingangsrepräsentation (Schall, Text) verarbeitet wird (Carstensen et al., 2010)
- Informationen durchlaufen dabei verschiedene sprachliche Repräsentations- ebenen
  - In einem Computersystem werden diese Schritte sequenziell durchgeführt und abgearbeitet (Carstensen et al., 2010)

# Computerlinguistik (forts.)

## Spracherkennung

- Bei gesprochener Sprache muss der Inhalt aus den Schallinformationen extrahiert und in Text umgewandelt werden. Dies geschieht durch sogenannte Speech-to-Text-Verfahren (STT).

## Tokenisierung

- Aus dem Text werden unterschiedliche Segmente gebildet. Segmente können einen unterschiedlichen Umfang haben, z. B. Absätze, Sätze oder Wörter. Welche Segmente gebildet werden, hängt von den Zielen eines Systems ab.

## Morphologische Analyse

- Im Text werden die grammatischen Informationen analysiert und extrahiert. Dabei findet eine Zurückführung der einzelnen Wörter auf ihre Grundform statt, was auch als Wortstammbildung bezeichnet wird.

# Computerlinguistik (forts.)

## Syntaktische Analyse

- In der Syntax-Analyse werden die Wörter in einem Satz auf ihre strukturelle Funktion untersucht. Dabei wird z. B. ermittelt, welche Wörter ein Subjekt, Prädikat, Objekt oder Artikel sind.

## Semantische Analyse

- Aufsetzend auf der syntaktischen Analyse werden den Sätzen Bedeutungen zugeordnet.

## Dialog- und Diskursanalyse

- Bei aufeinander folgenden Sätzen wird untersucht, in welchem Verhältnis sie zueinanderstehen stehen und ob sie einen Dialog oder Diskurs wiedergeben (z. B. ein Gespräch zwischen zwei Personen).

# Computerlinguistik (forts.)

**Die oben genannten Schritte bauen aufeinander auf ->** Das Ergebnis der Vorstufe geht als aufbereitete Information in die Folgestufe ein.

- Es wird eine immer **detailliertere Sprachanalyse** durchgeführt, um den Text verstehen zu können
- Es müssen nicht immer alle Schritte durchlaufen werden, wenn bereits Teilergebnisse für die weitere Verarbeitung verwendet werden können
- Besonders moderne **maschinelle Lernverfahren** sind in der Lage, mit diesen Teilergebnissen weiterzuarbeiten, wodurch aufwendige Folgeschritte verkürzt werden können



# Maschinelles Lernen

# Maschinelles Lernen

**Maschinelles Lernen** gehört zum **Kernbereich der künstlichen Intelligenz** und wird beispielsweise für die automatische Generierung von Wissen verwendet.

- **Ausgangsbasis** sind Erfahrungen und bereits bestehendes Wissen sowie Beispiele, aus dem das künstlich intelligente System lernen kann
- Durch den **Aufbau von statistischen Modellen** mittels spezieller Lern-Algorithmen ist ein künstlich intelligentes System in der Lage, nach einer Lernphase durch Beispieldaten diese zu abstrahieren und generelle Schlüsse daraus zu ziehen
- Schlussfolgerungen werden als **Muster und Regeln hinterlegt** und auf unbekannte Daten angewendet, um diese zu beurteilen und eine Prognose zu unbekannten Informationen zu treffen

# Maschinelles Lernen (forts.)

**Trainierte künstliche Systeme** können unter anderem dazu verwendet werden,

- um ärztliche Diagnosen durchzuführen bzw. zu unterstützen
- Betrug mit Kreditkarten zu erkennen
- Aktienmarktanalysen durchzuführen
- Sprach- und Texterkennung zu realisieren

Das maschinelle Lernen ist ein Oberbegriff für die automatische Wissensgenerierung und weist verschiedene Arten des Lernens auf.

# Maschinelles Lernen (forts.)

**Maschinelles Lernen** hat Querbezüge und Gemeinsamkeiten zu anderen Fachgebieten der Informatik, wie z. B.:

- dem Knowledge Discovery in Databases (KDD)
- dem Data-Mining (DM), mit dem es sich viele Algorithmen teilt

KDD und DM werden vorwiegend für das Finden von neuen Mustern und Gesetzmäßigkeiten verwendet.

Zudem können KDD-Methoden genutzt werden, um Lerndaten für maschinelles Lernen zu ermitteln oder aufzubereiten (Azevedo & Santos, 2008).

# Maschinelles Lernen (forts.)

**Art und Weise der Wissensrepräsentation ist ein bedeutender Faktor beim maschinellen Lernen.**

- Wissensrepräsentation dient dazu, die extrahierten Informationen und das gelernte Wissen zu strukturieren und für die maschinelle Verarbeitung aufzubereiten

## **Zwei Arten der Wissensrepräsentation:**

- **Symbolischen Ansätze** mit denen das Wissen, bestehend aus Beispielen und den induzierten Regeln, explizit repräsentiert wird -> Dazu wird die Aussagenlogik oder die Prädikatenlogik verwendet
- **Nicht symbolische Ansätze**, wie neuronale Netze, besitzen zwar ein berechenbares „antrainiertes“ Verhalten, geben jedoch keinen Einblick in die erlernten Lösungswege, wodurch das Wissen implizit repräsentiert wird (Russell & Norvig, 2010)

# Informationswissenschaften

# Informationswissenschaften

**Informationswissenschaft ist eine junge Disziplin und gilt als eine Brückenwissenschaft**, die verschiedene Perspektiven und Ansätze anderer Wissenschaften aus verschiedenen Blickwinkeln betrachtet und anwendungsbezogen in Produkte und Dienstleistungen überführt.

- Befasst sich mit **Informationssystemen und deren unterstützende Informationsprozesse**
- **Alle Aspekte des Informationsgeschehens** werden betrachtet, die beispielsweise kognitiver, ökonomischer, sozialer oder politischer Art sind
- Fokus auf Informationsinhalt und Repräsentation von Wissen (Stock, 2007)

**Informationswissenschaft** schafft einerseits einen Ordnungsrahmen und andererseits „Wissen, wie vorhandenes Wissen genutzt werden kann“ (Stock, 2007, S. 4).

# Informationswissenschaften - Forschungsgegenstand

Forschungsgegenstand sind Wissen und Information, die meist aus anderen Wissenschaften stammen, wie der Informatik oder der Soziologie.

- **Interdisziplinäres Wissen** steht in der Informationswissenschaft im Vordergrund
- **Domänen-spezifisches** und **Fachwissen** liegt vorrangig in den einzelnen angrenzenden Wissenschaften, die ihr eigenes Wissen aufbauen
- Von Gernot Wersig wird die Informationswissenschaft als **Triade von Wissen, Mensch und Informationstechnik** beschrieben

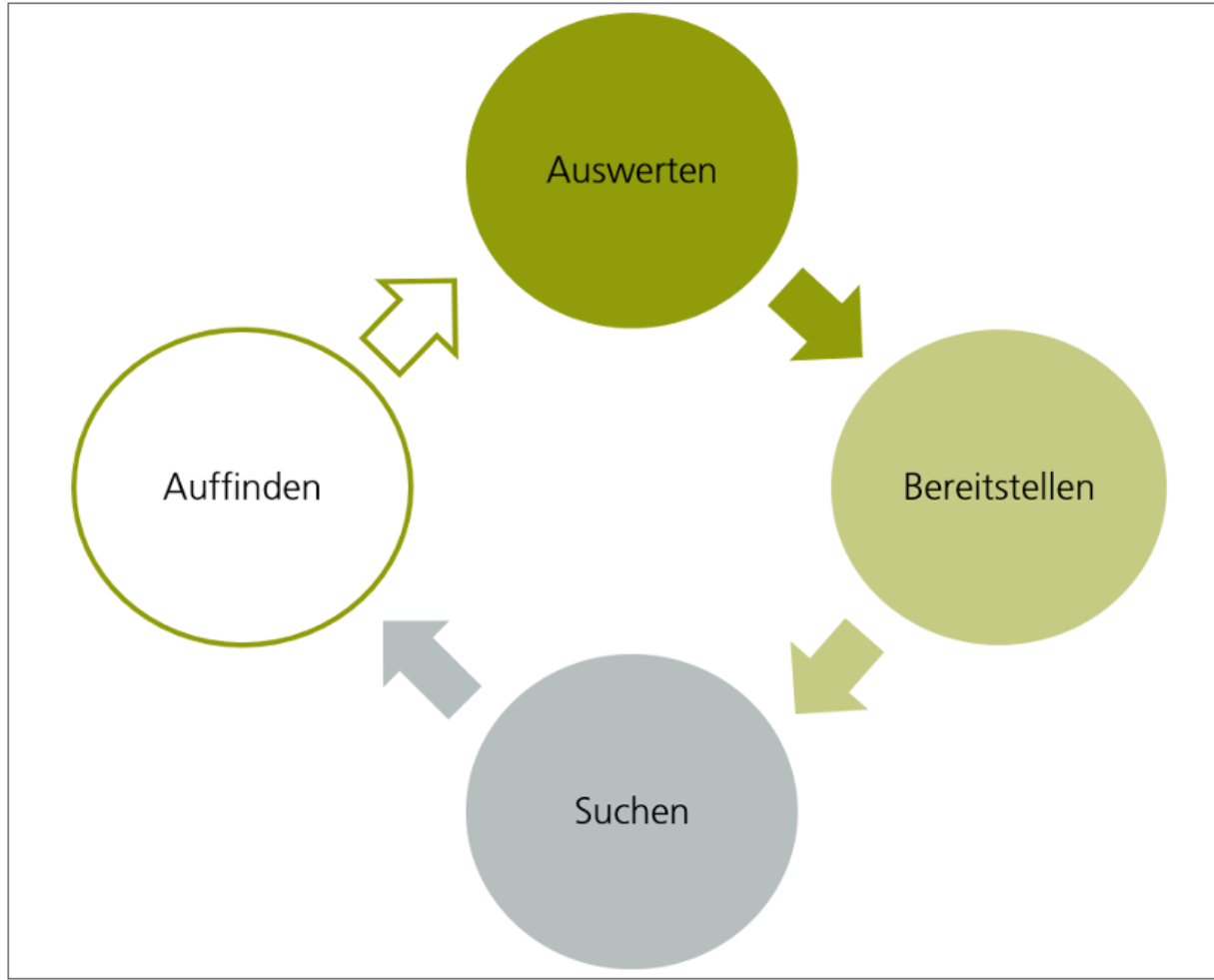


# Informationswissenschaften - Forschungsgegenstand (forts.)

## Untersuchungsbereich der Informationswissenschaft:

- **Auswerten, Bereitstellen, Suchen und Wiederauffinden** von **relevanten Informationen** und **Wissen** aus **Informationssystemen** durch **Informations- und Kommunikationsprozesse** (Stock, 2007)
- **Iterativer Prozess**, bei dem Ergebnisse aus einem vorangegangenen Durchlauf in den Folgeablauf integriert werden
- Informationen werden ausgewertet, um sie beispielsweise einem Themenbereich zuzuordnen
- Werden bei der Suche relevante Informationen gefunden, nennt man das Wiederauffinden von Informationen

# Informationswissenschaften - Forschungsgegenstand (forts.)



# Informationswissenschaften - Forschungsgegenstand (forts.)

**Relevante Informationen:** bei einer Suchanfrage erhaltenen Ergebnisse, die das Informationsbedürfnis objektiv befriedigen (Stock, 2007).

- Objektive Betrachtung von Relevanz bei der Informationsermittlung ist dem subjektiven Nutzen der Suchenden übergeordnet

**Aber:** fehlende subjektive Betrachtung führt in der Praxis zu Problemen, da Inhalte und Informationen nicht oder nur teilweise zu den tatsächlichen Informationsbedürfnissen der Nutzer:innen passen.

Personalisierung von Anwendungen und Systemen spielt auch bei der Entwicklung und Bereitstellung von Conversational Interfaces eine Rolle, schließlich geht es darum, den Nutzer:innen die Informationen bereitzustellen, die für sie wichtig sind.

# Informationswissenschaften - Wissen

Nutzer:innen haben die Erwartungshaltung, dass ihnen ein intelligenter, persönlicher Assistent oder spezielle Chatbots Auskunft geben und Informationen liefern, die sie **wissen** müssen oder die sie interessieren.

Was aber ist Wissen in diesem Kontext?

und

Wie kann man prüfen, ob die automatisch ermittelten Informationen auch wirklich relevant sind und das Informationsbedürfnis zum Erlangen des gewünschten Wissens befriedigt ist?

# Informationswissenschaften – Wissen (forts.)

**Die Informationswissenschaft stellt sich die Frage, was Wissen eigentlich ist und in welchem Zusammenhang es steht.**

**Wissen** wird als ein Ergebnis eines Informationsvermittlungsprozesses angesehen, in dem ein Informationsempfänger oder eine Informationsempfängerin neue Informationen aufnimmt und seinem oder ihrem bereits vorhandenem Wissen hinzufügt bzw. dieses erweitert.

# Informationswissenschaften – Kernaufgaben

Informationswissenschaft beschränkt sich nicht nur auf die Analyse der Prozesse von Wissensträgern

- Sie stellt Information, Wissen und Verstehen in den Mittelpunkt des Geschehens, in die auch stets der menschliche Akteur involviert ist

Belkin fasst den Kern der Informationswissenschaft unter folgenden Punkten zusammen (Belkin, 1978):

- Information betrifft menschliche kognitive Kommunikationsprozesse
- **Idee der gewünschten Information** (desired Information)
- **Effektivität von Information**, Informationssystemen und Informationsvermittlung (information transfer)
- Verhältnis zwischen Information und ihrem Hersteller (generator)
- Verhältnis zwischen Information und Nutzer (Ingwersen, 1995)

# Informationswissenschaften – Kernaufgaben (forts.)

- **Erste Aspekt** bezieht sich auf die verschiedenen Formen der Informationsvermittlung innerhalb der Gesellschaft und deren gesellschaftliche Gruppen
  - Durch die Anwendung sozialwissenschaftlicher Methoden werden verschiedene Kommunikationsformen empirisch untersucht.
- **Zweite Aspekt** beschäftigt sich mit den Eigenschaften der gewünschten Information
  - Dabei wird versucht, den Informationsbedarf zu ermitteln und die verschiedenen, ihm zugrunde liegenden Interessen zu erkennen.
- **Dritte Aspekt** befasst sich mit der Effektivität von Informationssystemen während des Informationstransfers
  - Gegenstand ist das Management des Transfers zwischen Nutzern und Informationsträgern, wie beispielsweise die Optimierung des Zugangs zu wissenschaftlichen Artikeln durch Nutzende.

# Informationswissenschaften – Kernaufgaben (forts.)

- **Vierte Aspekt** befasst sich mit Analyse und Darstellung von Wissen in Informationssystemen
  - Die klassischen Erschließungsmethoden (wie z. B. Klassifikation, Indexierung, Abstrakt) stammten aus den bibliothekarischen Traditionen und wurden für textuelle Dokumente entwickelt.
  - Durch quantitative Methoden (z. B. Wortzählungen) konnten bestimmte bibliometrische Konstanten ermittelt werden, mit deren Hilfe die Retrieval-Methoden weiterentwickelt wurden.
- **Fünfte Aspekt** befasst sich mit der Frage nach Relevanz, Nutzung und Mehrwert von Information aus der Sicht der Nutzenden
  - Das „**Cranfield experiment**“ zählt zu den ersten Ansätzen, die sich mit der Relevanz beschäftigen und wurde in den 60er-Jahren in England entwickelt.
  - Die Relevanz von Retrieval-Techniken wurde jedoch gemessen, ohne die Nutzenden direkt zu beteiligen. Aus den Ansätzen dieser Zeit wurden die Metriken zur Messung der Effektivität von **IR-Systemen (IR = information retrieval)** entwickelt, zu denen die Precision und der Recall gehören.



# Informationswissenschaften – Praktische Relevanz

**Zu den wichtigen und praktisch relevanten Aspekten der Informationswissenschaft gehören:**

- die Informationssuche
- das Design von Information-Retrieval-Systemen
- die Interaktions- und Navigationsmechanismen (information retrieval)
- die Analyse und Bewertung von Informationssystemen aus der Sicht der Nutzenden
- die Messung der Wirkung von Information auf die Gesellschaft

# Informationswissenschaften – Praktische Relevanz (forts.)

Für Ingwersen (1995) steht die **Analyse der Suche nach Informationen in Verbindung mit den Nutzer:innen im Vordergrund**

- Nur dann gelingt es, effektive und effiziente Informationssysteme zu entwerfen oder Interaktionsmechanismen zu gestalten
- Es muss bekannt sein, welche Informationen ein Mensch oder eine bestimmte Gruppe sucht, wünscht oder benötigt

**Wichtig:** bevor Ergebnisse bereitgestellt werden, ist herauszufinden, was die Nutzer:innen an Information benötigen.

Gilt auch für die Entwicklung von Conversational Interfaces:

- Wenn diese Nutzer:innen Informationen bereitstellen sollen und dabei beispielsweise stellvertretend für Mitarbeitende eines Unternehmens auftreten (z. B. Sachbearbeiter:in)
- Anders als ein menschliche Akteure können Conversational Interfaces nur schwer oder überhaupt nicht ad-hoc reagieren

# Informationswissenschaften – Wissensrepräsentation

**Wissensrepräsentation (knowledge representation)** wird in der Wissensmodellierung dazu verwendet, Wissen in einer formalen Form in wissensbasierten Systemen abzubilden.

- Es wurden verschiedene **formale Sprachen und Notationen** vorgeschlagen und entwickelt
- **Sammlung** repräsentierten Wissens wird als Wissensbasis bezeichnet
- **Semantic Web** stellt ein verteilt abgelegtes formalisiertes Wissen dar

Neben der Wissensrepräsentation gibt es auch die Wissensorganisation. Durch die Wissensorganisation sollen bestehende Wissensbestände geordnet werden.

- Verwendung von Metadaten, die das Beschreiben und Wiederauffinden des Wissens ermöglichen (Carstensen et al., 2010)

# Informationswissenschaften – Wissensrepräsentation (forts.)

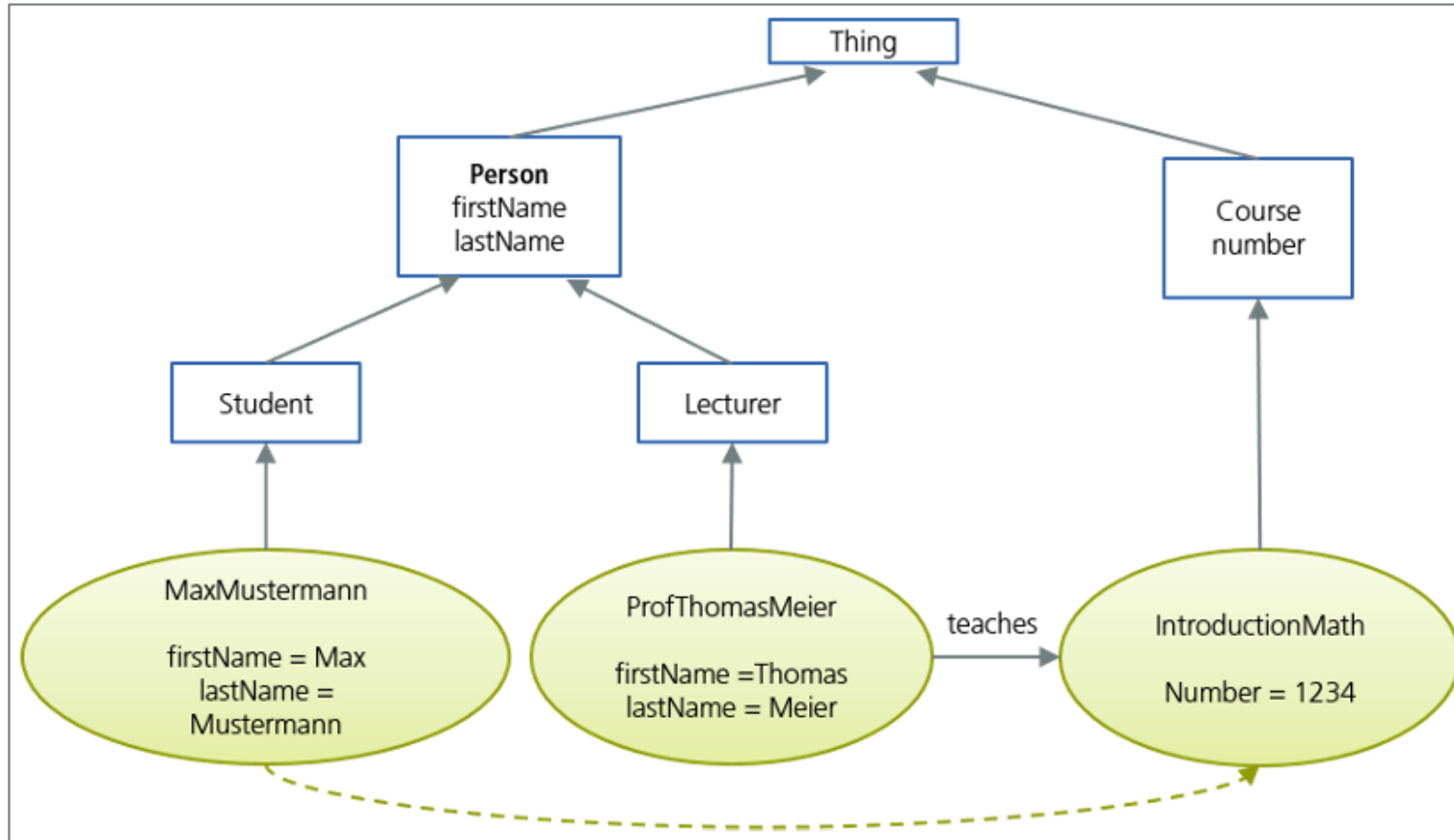
**Repräsentation und Ordnung von Wissen nutzt verschiedene Konzepte und Ansätze.**

**Zu diesen gehören:**

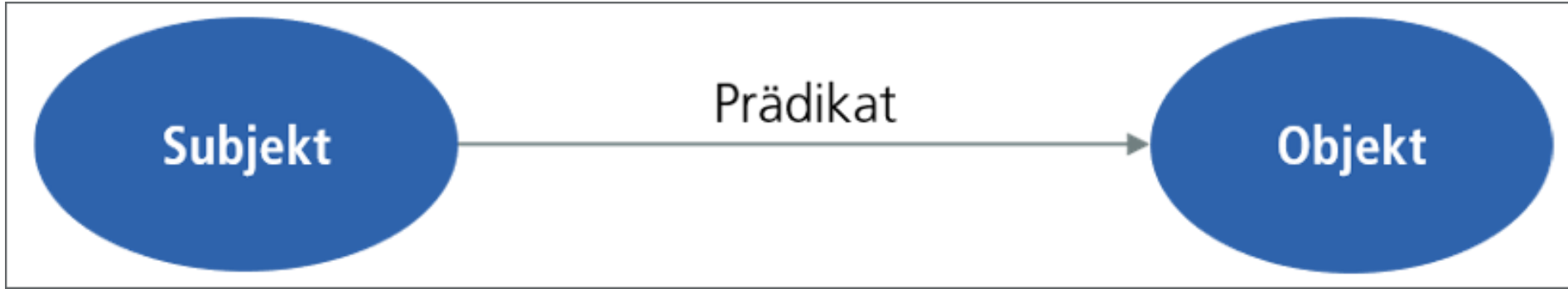
- Kataloge, Glossare, Taxonomien (einfache Vokabularien)
- Klassifikationen und Thesauri
- Semantische Netze, Ontologien und Produktionsregeln
- Axiomensysteme und Prädikatenlogik
- Mehrschichtige erweiterte semantische Netze (auch als MultiNet bekannt)

Ergänzt werden die Ansätze durch die Informationsvisualisierung, die selbst keine formale Repräsentation darstellt, aber zur Vermittlung von Informationen und Wissen eingesetzt werden kann.

# Informationswissenschaften – Wissensrepräsentation (forts.)



# Informationswissenschaften – Wissensrepräsentation (forts.)



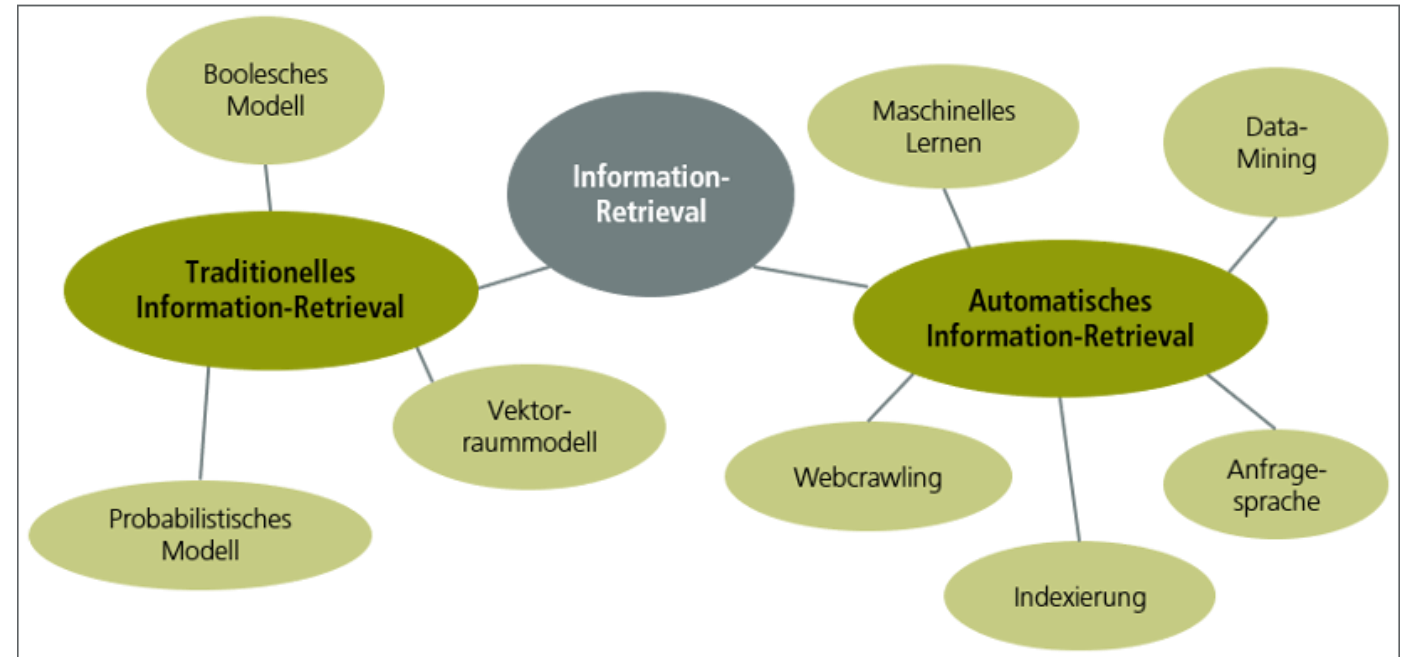
```
58 <owl:NamedIndividual rdf:about="#MaxMustermann">
59   <rdf:type rdf:resource="#Student"/>
60   <studies rdf:resource="#IntroductionMath"/>
61   <firstName rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
62     ↪ >Max</firstName>
63   <lastName rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
64     ↪ >Mustermann</lastName>
65 </owl:NamedIndividual>
```

# Informationswissenschaften – Information Retrieval

**Information-Retrieval wird genutzt, um die gewünschte Information aus einer Vielzahl an Informationen im Web zu extrahieren.**

Saini und Arora unterscheiden dafür zwei Arten von Information-Retrieval (Saini & Arora, 2016, S. 2635):

- Traditionelles Information-Retrieval
- Automatisches Information-Retrieval



# Informationswissenschaften – Information Retrieval (forts.)

Beim **traditionellen Information-Retrieval** wird eine begrenzte Menge an Daten nach den gewünschten Informationen durchsucht.

Folgende Modelle werden verwendet (Saini & Arora, 2016, S. 2635):

- **Boolesches Modell:** Informationsgewinnung erfolgt mit boolescher Algebra, also anhand der logischen Operatoren „und“, „oder“ und „nicht“.
- **Vektorraummodell:** Informationsbeschaffung erfolgt mit einem Vektorraum, auf dem die Informationen als Punkte repräsentiert werden. Es wird die Distanz zwischen Suchvektor und Informationsvektor berechnet.
- **Probabilistisches Modell:** Die Beschaffung der Information wird anhand der Wahrscheinlichkeit der Relevanz vorgenommen.



# Informationswissenschaften – Information Retrieval (forts.)

Durch automatisches Information-Retrieval soll die Informationsflut beseitigt werden, indem, aus einer großen Menge von Informationen, die relevanten ermittelt werden.

Dies kann auf folgende Arten geschehen (Saini & Arora, 2016, S. 2635):

- **Web-Crawling:** Programm, um automatisch Informationen aus dem Internet herunterzuladen.
- **Indexierung:** Reduzierung von Dokumenten auf die enthaltenen informativen Begriffe als Menge von sogenannten Indexbegriffen (Kaur & Gupta, 2016).
- **Anfragesprache:** Es wird eine Anfrage an eine Datenbank gestellt, und die Ergebnisse können mit unterschiedlichen Verfahren analysiert werden.
- **Data-Mining:** Automatische Extraktion von Informationen aus großen Datenbeständen mithilfe von computergestützten Methoden, um Muster und Zusammenhänge zu ermitteln.
- **Maschinelles Lernen:** Mithilfe von maschinellen Lernverfahren werden Informationen aus Dokumenten extrahiert.

# Mensch-Computer-Interaktion

# Mensch-Computer-Interaktion

Der Forschungsbereich **Mensch-Computer-Interaktion**, im englischen *Human-Computer-Interaction (HCI)*, beschäftigt sich mit der **Gestaltung, Interaktion und der Verwendung von Computertechnologien**.

- Gehört zum übergeordneten Forschungsgebiet **Mensch-Maschine-Interaktion**, das sich auf einem abstrakteren Niveau mit der Interaktion zwischen Menschen und Maschinen auseinandersetzt
- Es entsteht ein Gesamtsystem, deren Beteiligte und Schnittstellen im Fokus stehen und das zu einem Mensch-Maschine-System wird

# Mensch-Computer-Interaktion (forts.)

HCI untersucht die Art und Weise, wie Menschen mit Computern interagieren.

Im Zentrum stehen der Mensch und dessen Bedürfnisse bei der Nutzung von Informationssystemen. Entwicklung und Erforschung interaktiver Systeme, die vielfältige Schnittstellen aufweisen und mit umfangreicher Sensorik ausgestattet sind.

- HCI ist ein interdisziplinäres Gebiet, bei dem neben Erkenntnissen der *Informatik* auch solche aus der *Psychologie*, der *Kognitionswissenschaft*, der *Soziologie*, der *Arbeitswissenschaft*, der *Ergonomie* und dem *Design* einbezogen werden
- Als Forschungsgebiet befindet sich HCI an der **Schnittstelle** von *Informatik*, *Verhaltenswissenschaften*, *Design* und *Medienwissenschaften*.

# Mensch-Computer-Interaktion (forts.)

## Im Zentrum der Entwicklung stehen ...

- ... Interaktionsschnittstellen zwischen Mensch und Computer.

## Interaktionen finden meistens über vorgefertigte Dialoge statt, dazu gehören ...

- ... grafische Benutzerschnittstellen von Anwendungen wie beispielsweise ein E-Mail-Programm.

**Dialog sind vorgefertigt, da die Nutzer:innen keinen Einfluss auf die verfügbaren Elemente innerhalb der Benutzeroberfläche nehmen können.**

- > Die Auswahl an Funktionen und Dialogen ist seitens der Entwickler des Programmes fest vorgegeben.

# Mensch-Computer-Interaktion (forts.)

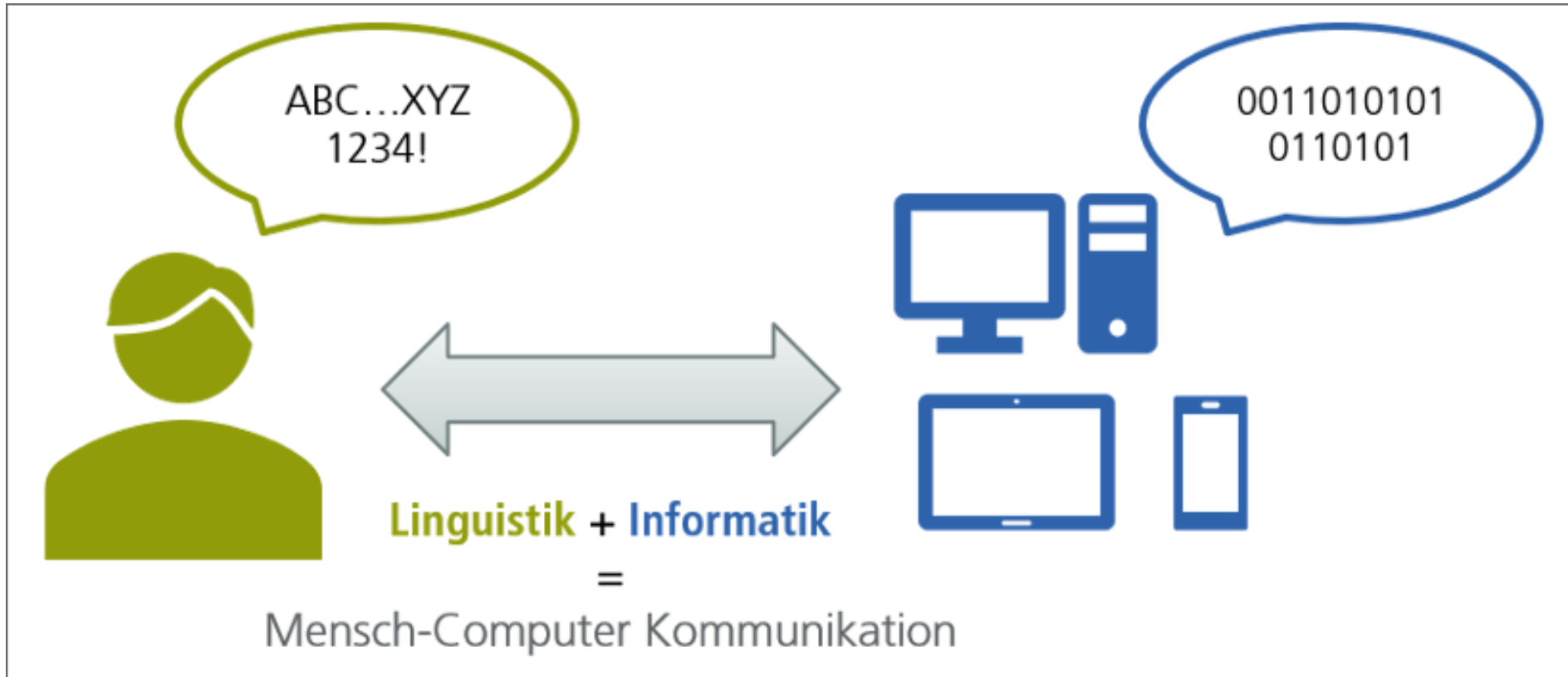
**Usability-Forschung**, ein Teilbereich der HCI, beschäftigt sich u. a. mit der Anpassbarkeit von Benutzeroberflächen an die individuellen Bedürfnisse der einzelnen Nutzer:innen (Bevan, 1995).

- Art und Umfang der Anpassbarkeit sind bereits festgelegt
- Aber: Nutzer:innen haben verschiedene Möglichkeiten, die Darstellung anzupassen, z. B. das Ein- und Ausblenden von Elementen, Vergrößerung von Texten, Anordnung von Elementen

Interaktion zwischen Mensch und Computer ist keine natürliche Kommunikation, wie sie zwischen Menschen stattfindet.

- Auch die Kommunikation zwischen Maschinen ist eine andere als die zwischen Mensch und Computer

# Mensch-Computer-Interaktion (forts.)



# Mensch-Computer-Interaktion (forts.)

**Interaktion zwischen Mensch (natürliche Sprache) und Computer (binäre Codierung) bedarf einer Übersetzungsleistung zwischen den unterschiedlichen Welten.**

- Unterscheidung von Text- und Spracheingabe, die mit unterschiedlichen Aufbereitungsstrategien verbunden sind
- Computer können nur Zeichen verarbeiten, daher müssen alle Informationen übersetzt werden
- Beziehungen und Regeln zwischen den Zeichen, Wörtern und Textbestandteilen müssen korrekt abgebildet werden, die aus der Sprache des zu verarbeitenden Textes hervorgehen

**Hier kommen Linguistik und Informatik zusammen**, um natürliche Sprache in ein computerrelevantes Format zu überführen:

- Schrittweise aufeinander aufbauende komplexe Verfahren der Sprachverarbeitung, der Sprachanalyse und der Sprachrepräsentation ( Mensch-Computer-Interaktion, maschinelles Lernen, Computerlinguistik, Informationswissenschaften)

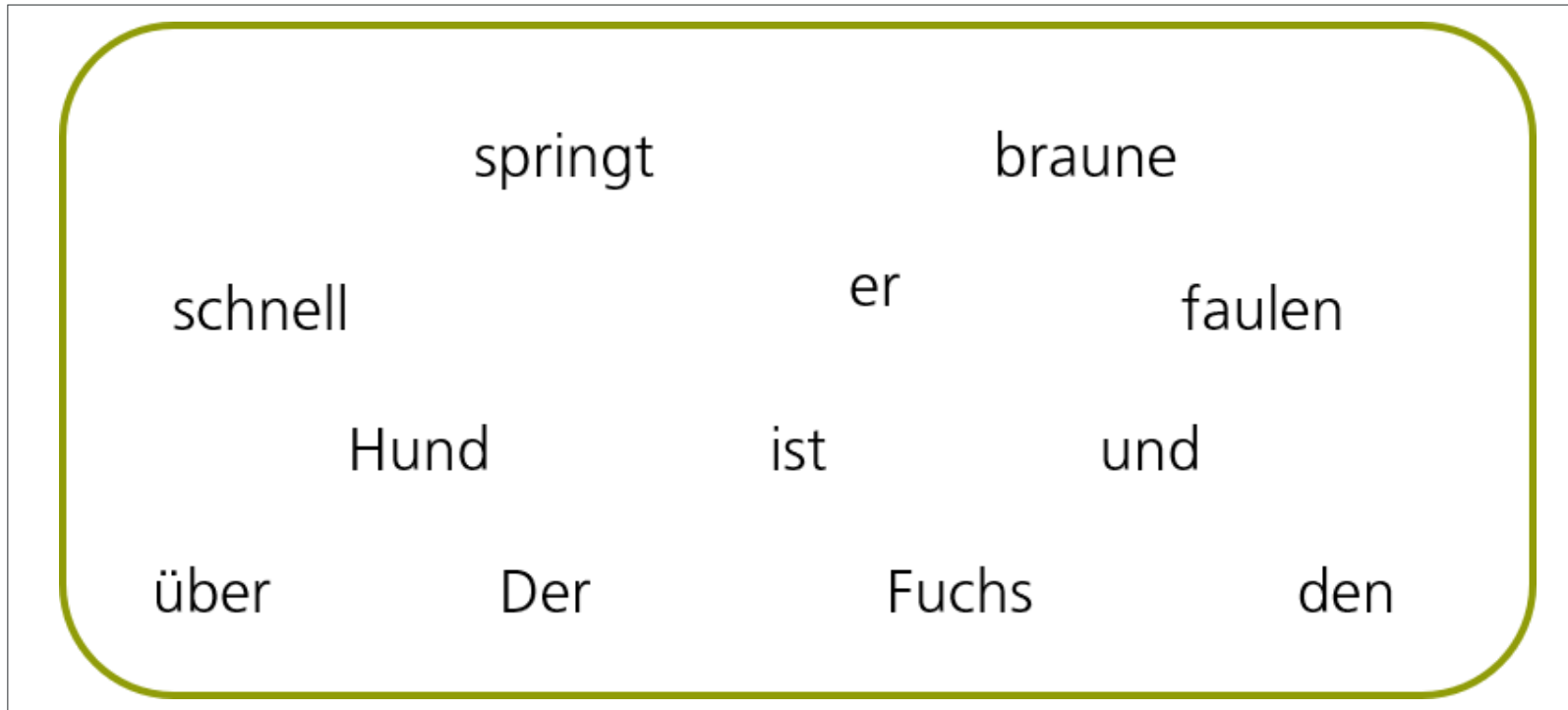


# Mensch-Computer-Interaktion (forts.)

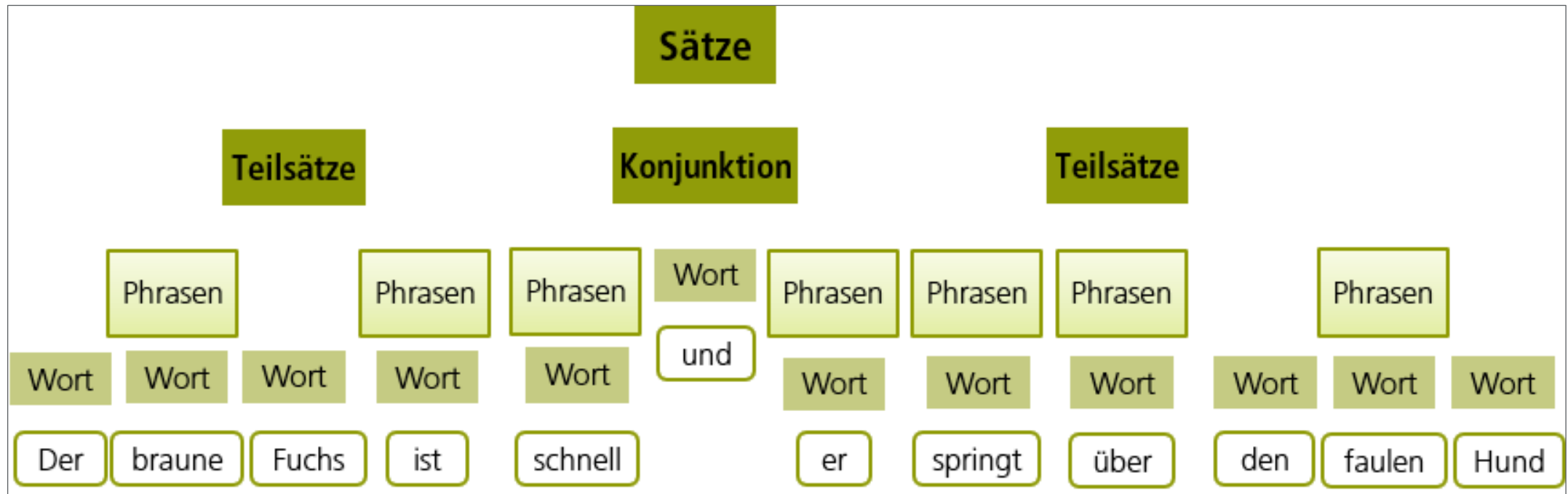
**Sprachverarbeitung ermöglicht die Entwicklung von natürlichen Konversationen über Konversationsschnittstellen.**

- Erfordert Sprachverstehen durch den Computer, bei dem die Herausforderungen darin bestehen, Hintergrundwissen aufzubauen, um die natürliche Sprache richtig zu interpretieren.
- Schwierigkeiten bestehen im Verstehen von Metaphern und Ironie, was Gegenstand der Forschung im Opinion Mining ist.
- Aus dem Bereich der Mensch-Computer-Interaktion kommen Ansätze und Methoden zur Konzeption, Planung und Gestaltung von interaktiven Systemen sowie deren Testung und Validierung.

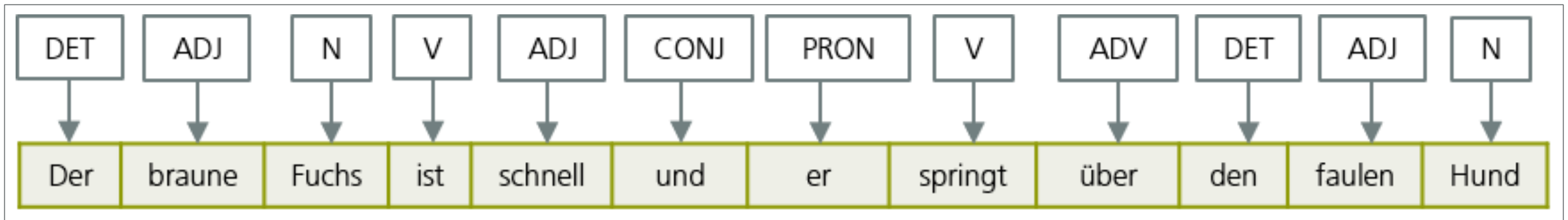
# Eine Sammlung von Wörtern ohne jede Beziehung oder Struktur



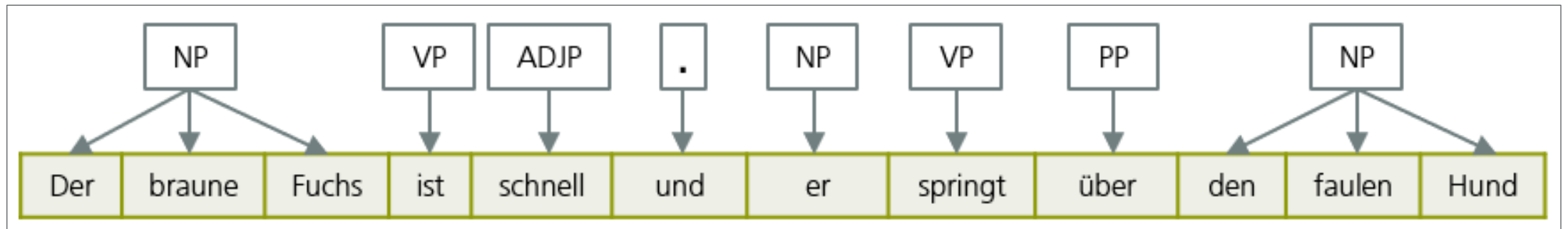
# Strukturierter Satz nach der hierarchischen Syntax



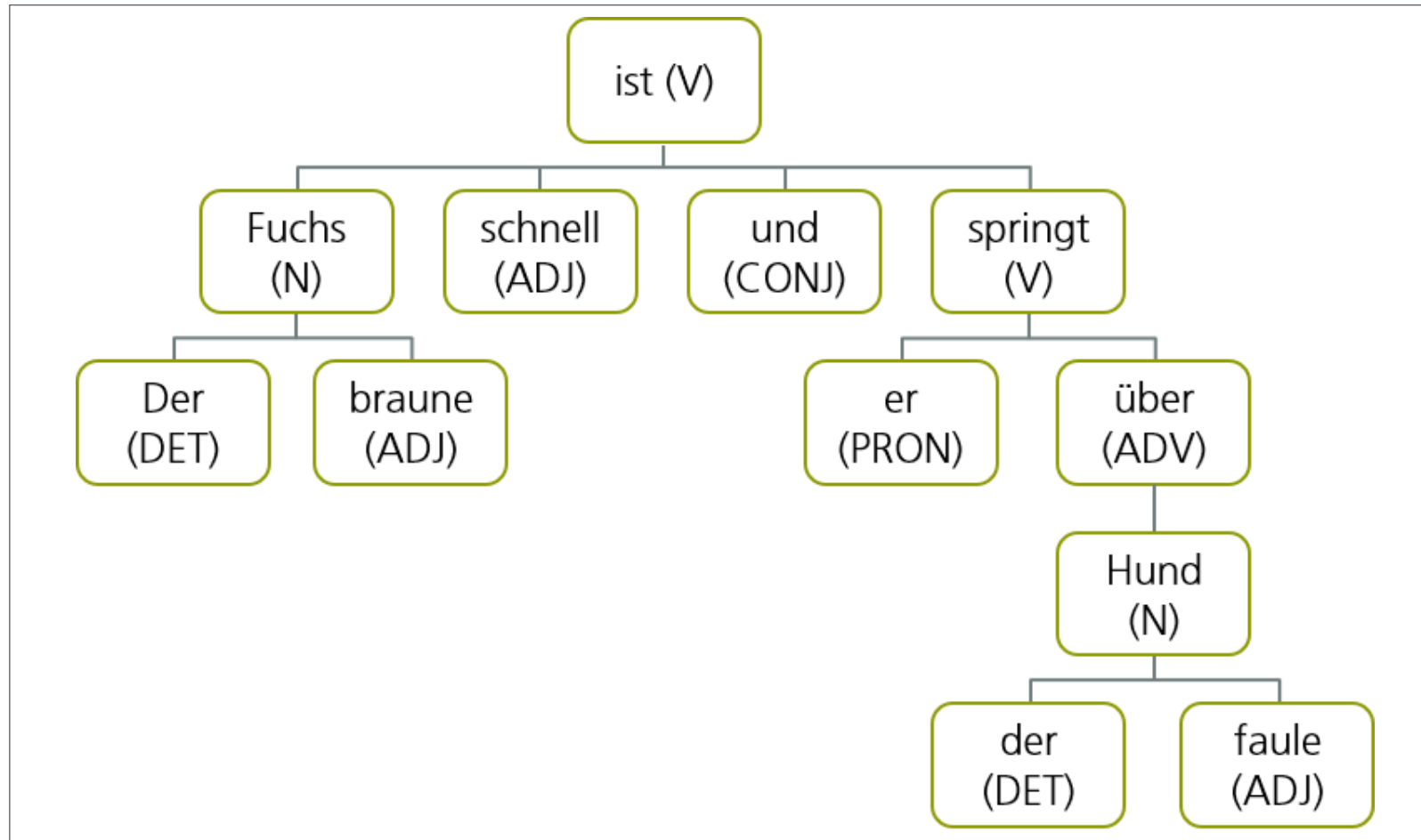
# Kommentierte Wörter mit ihren POS-Tags



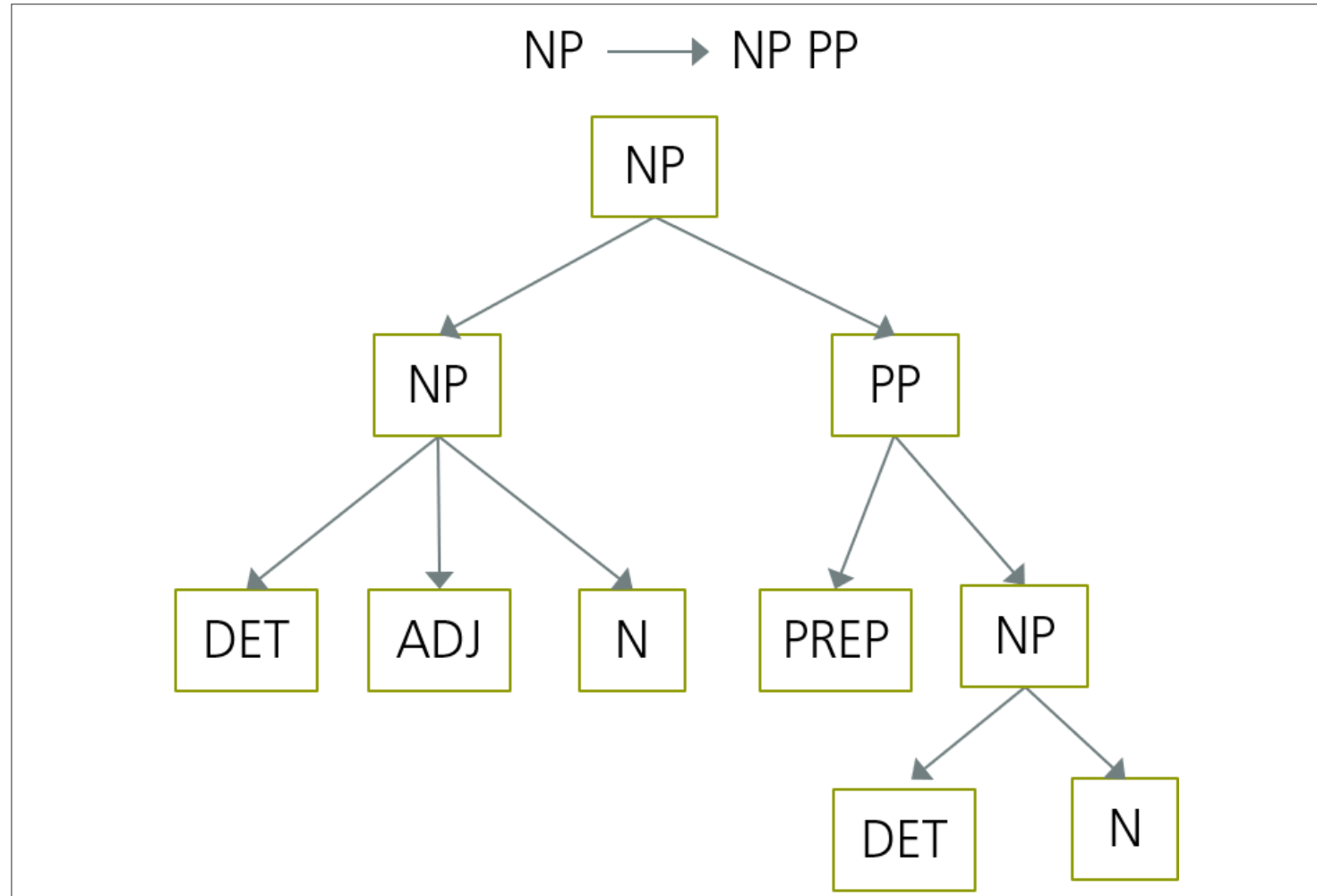
# Kommentierte Phrasen mit ihren Tags



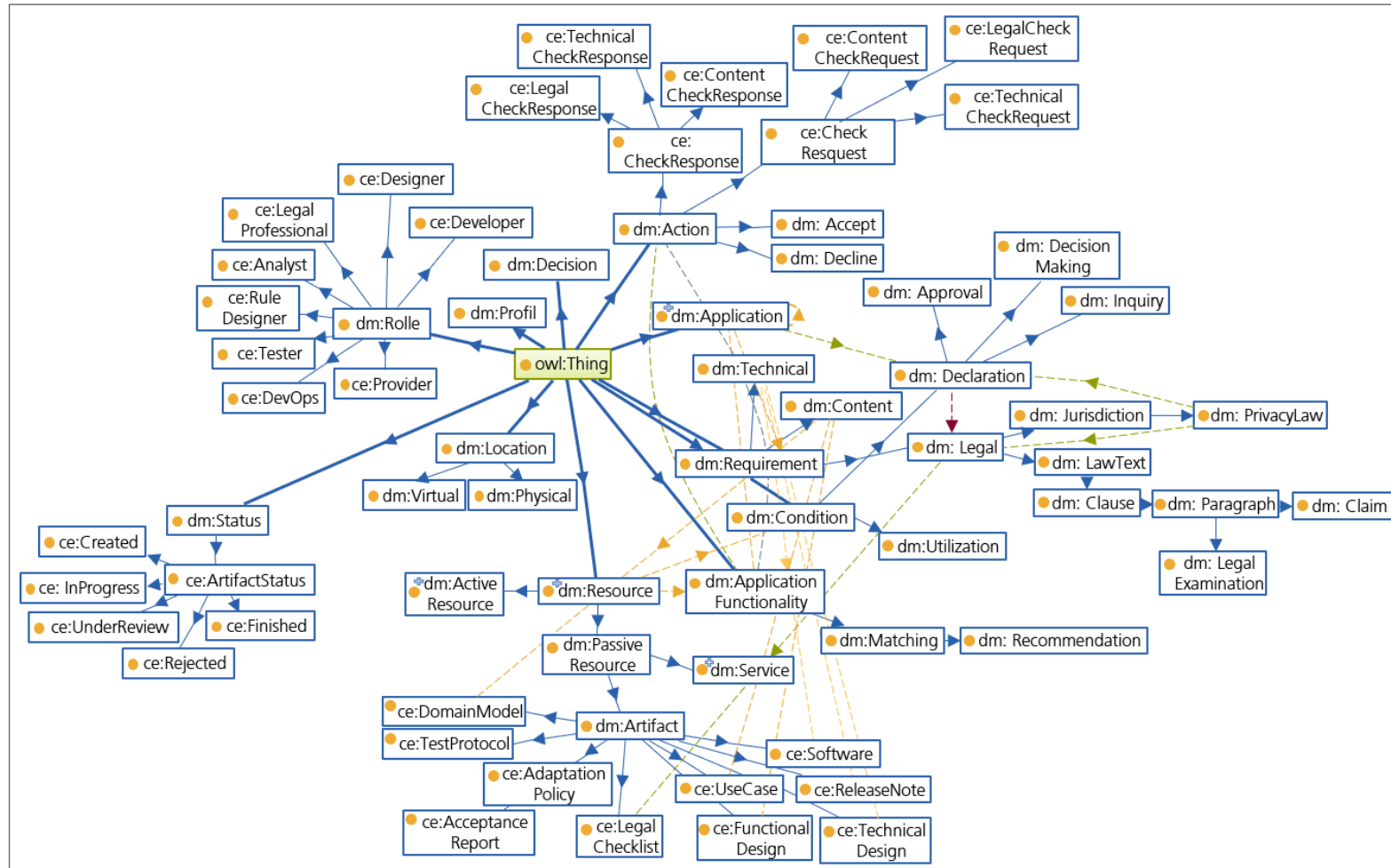
# Syntaxbaum auf Basis der Abhängigkeitsgrammatik mit POS-Tags



# Generativer Syntaxbaum, der die Strukturierungsregeln für Substantivsätze darstellt

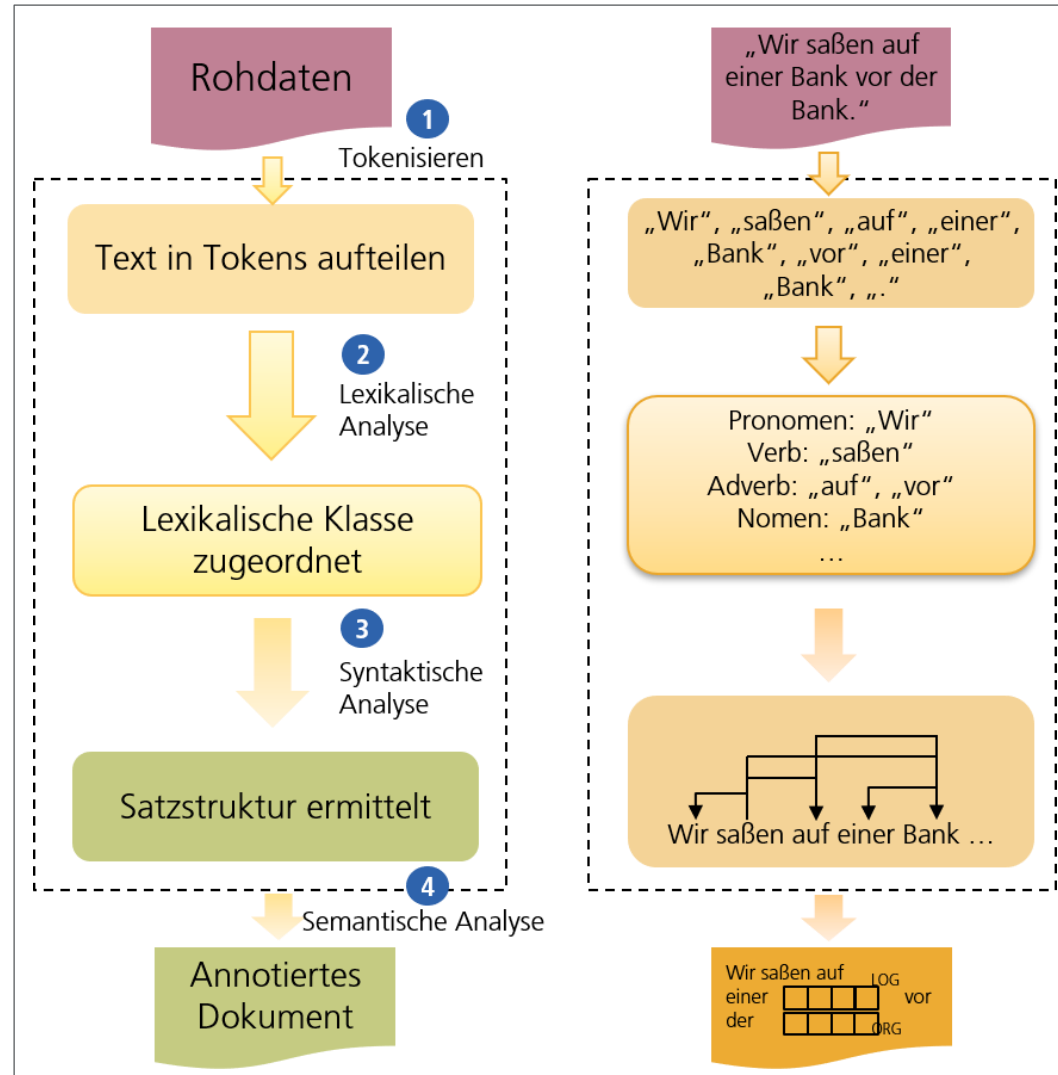


# Semantisches Netz zu den Konzepten und Beziehungen eines Anwendungssystems





# Sprachverarbeitungsprozess



# Aufbau und Ablauf eines supervised Learning Verfahrens

