

# Wissensrepräsentation Vorlesung

Sommersemester 2008

12. Sitzung

Dozent

**Nino Simunic M.A.**

Computerlinguistik, Campus DU

# Übersicht

- **Rückblick, Zusammenhänge**
- **Mysterien**

# Inhalte im abstrakten Überblick

- Künstliche Intelligenz
  - Turing ... Rationales Handeln ... Agenten
- (Symbolische) Wissensrepräsentation
  - Intelligente Systeme (WBS, ...)
  - Wissensarten (Welt- ... wissen)
  - Wissensformen (Regeln, ... Ontologie)
  - Formalismen, Sprachen (S. Netze, ... RDFS)
  - Wissensverarbeitung, Inferenz
- Numerische Verfahren der KI
  - Vorhersage, Klassifizieren
  - Maschinelles Lernen

# Methoden, Gebiete der KI

- Wissensrepräsentation, Wissensverarbeitung
- Symbolische vs. numerische Verfahren
  
- Planen, Suchen
- Diagnostik
- Maschinelles Lernen, Ontology Learning
- NLP
- Intelligente Agenten
- XPS
- Computerspiel(-er)
- Robotik, Mustererkennung
- ...

# WBS/IS: Wichtige Zusammenhänge, Gemeinsamkeiten

Für WBS: Von Regelbasis ... Ontologie

- Trennung zwischen
  - (Deskriptiver) Wissensrepräsentation
  - Wissensverarbeitung
- Formalismus, Sprache je nach Komplexität
  - Thesaurus ... Ontologie / Logik ... OWL
  - Grundvoraussetzung: Effiziente maschinelle Verarbeitung/ Interpretation
  - Natürliche Sprache: Hochgradig informell, ungeeignet
  - Semantisches Netz: Nach Überführung in Logik / OAV
- Wissensakquisition, -auffindung, -generierung
- Knowledge Akquisition Bottleneck

# Konzeptionelles Wissen

- Basis für Formalismen zur Darstellung konzeptionellen Wissens:  
**Taxonomie, (Mehrfach-) Vererbung.**
- **Frames, semantische Netze**
  - Semantische Netze: Inferenz über **graph-basierte Algorithmen**
- Beide Ansätze:
  - Atomare, komplexe Werte, Defaults
  - Wertebereich, Einschränkung
  - Formale Semantik liegt zugrunde
  - Übersetzung und Repräsentation in Sprache → Maschinelle Interpretation

**Viele Ideen umgesetzt in XML Schema, RDF(S), OWL**

# (Deklarative) Wissensarten, -formen der Veranstaltung

- Weltwissen
- Domänenwissen
- Expertenwissen
- Fachwissen

- Regelbasierte Systeme
- (Formale) ISA
- »Thesaurus«

- Regeln, Semantische Netze, Frames, Schemata, Logik(-en), Objekt-Attribut-Wert Tripel
- Prinzipiell : Bilder, natürliche Sprache. **ABER:**  
**Hochgradig ineffiziente masch. Verarbeitung**
- Prozedurales Wissen:  
I.A. schlecht beschreib, wart-, wiederverwendbar

# Numerische Verfahren

## *Warum?*

- Ausdrücken von Unsicherheiten, Ranking, Tendenzen, etc.
- Unvollständige, unsichere Informationen
- Große Datenmengen, »industrielle« Motivation: Robustheit, Genauigkeit, Zeit, Geld

## *Was/Wie/Womit?*

- Stochastik/Statistik/Probabilistik
- Klassifizierung, Vorhersage, ...
- Qualität abhängig vom Modell, Beispielen
- Unüberwachte Verfahren: Wissensarm. I.d.R. auch Induktion von bspw. Regeln aus »Rohmaterial« (im GgStz. zu bspw. annotiertem Trainingsmaterial für POS-Tagger)
- Oft: Glättung

# Vorhersagegenauigkeit bei steigendem $n$

To give an intuition for the increasing power of higher-order  $N$ -grams, we trained a unigram, bigram, trigram, and a quadrigram model on the complete corpus of Shakespeare's works. We then used these four grammars to generate random sentences. In the following examples we treated each punc-

## 1. Unigram approximation to Shakespeare

- (a) To him swallowed confess hear both. Which. Of save on trail for are ay  
device and rote life have
- (b) Every enter now severally so, let
- (c) Hill he late speaks; or! a more to leg less first you enter

## 2. Bigram approximation to Shakespeare

- (a) What means, sir. I confess she? then all sorts, he is trim, captain.
- (b) Why dost stand forth thy canopy, forsooth; he is this palpable hit the King Henry. Live king. Follow.

## 3. Trigram approximation to Shakespeare

- (a) Sweet prince, Falstaff shall die. Harry of Monmouth's grave.
- (b) This shall forbid it should be branded, if renown made it empty.
- (c) What is't that cried?

## 4. Quadrigram approximation to Shakespeare

- (a) King Henry. What! I will go seek the traitor Gloucester. Exeunt some of the watch. A great banquet serv'd in;
- (b) Will you not tell me who I am?
- (c) It cannot be but so.
- (d) Indeed the short and the long. Marry, 'tis a noble Lepidus.
- (e) They say all lovers swear more performance than they are wont to keep obliged faith unforfeited!
- (f) Enter Leonato's brother Antonio, and the rest, but seek the weary beds of people sick.

Anzahl »anzutrainierender«, zu beobachtender  
Wahrscheinlichkeiten, um alle Möglichkeiten abzudecken:

$P(\text{hund} \text{ im} | \text{dem})$

$P(\text{hund} \text{ im} | \text{der})$

$P(\text{hund} \text{ im} | \text{den})$

$P(\text{hund} \text{ im} | \text{vom}) \dots$

$P(\text{hund} \text{ auf} | \text{vom}) \dots$

**Problem:** Genügend  
Text, Trainingsmaterial,  
um alles zu beobachten.  
**Lösung:** Glättung

Model

Parameters

1 st order (bigram model):

$20,000 \times 19,999 = 400 \text{ million}$

2nd order (trigram model):

$20,000^2 \times 19,999 = 8 \text{ trillion}$

3th order (four-gram model):

$20,000^3 \times 19,999 = 1.6 \times 10^{17}$

Table 6.1 Growth in number of parameters for n-gram models.

Größe des  
Vokabulars

More data = Better data

# Vorwärtsverkettung Bottom-up Verfahren

Datenorientiert

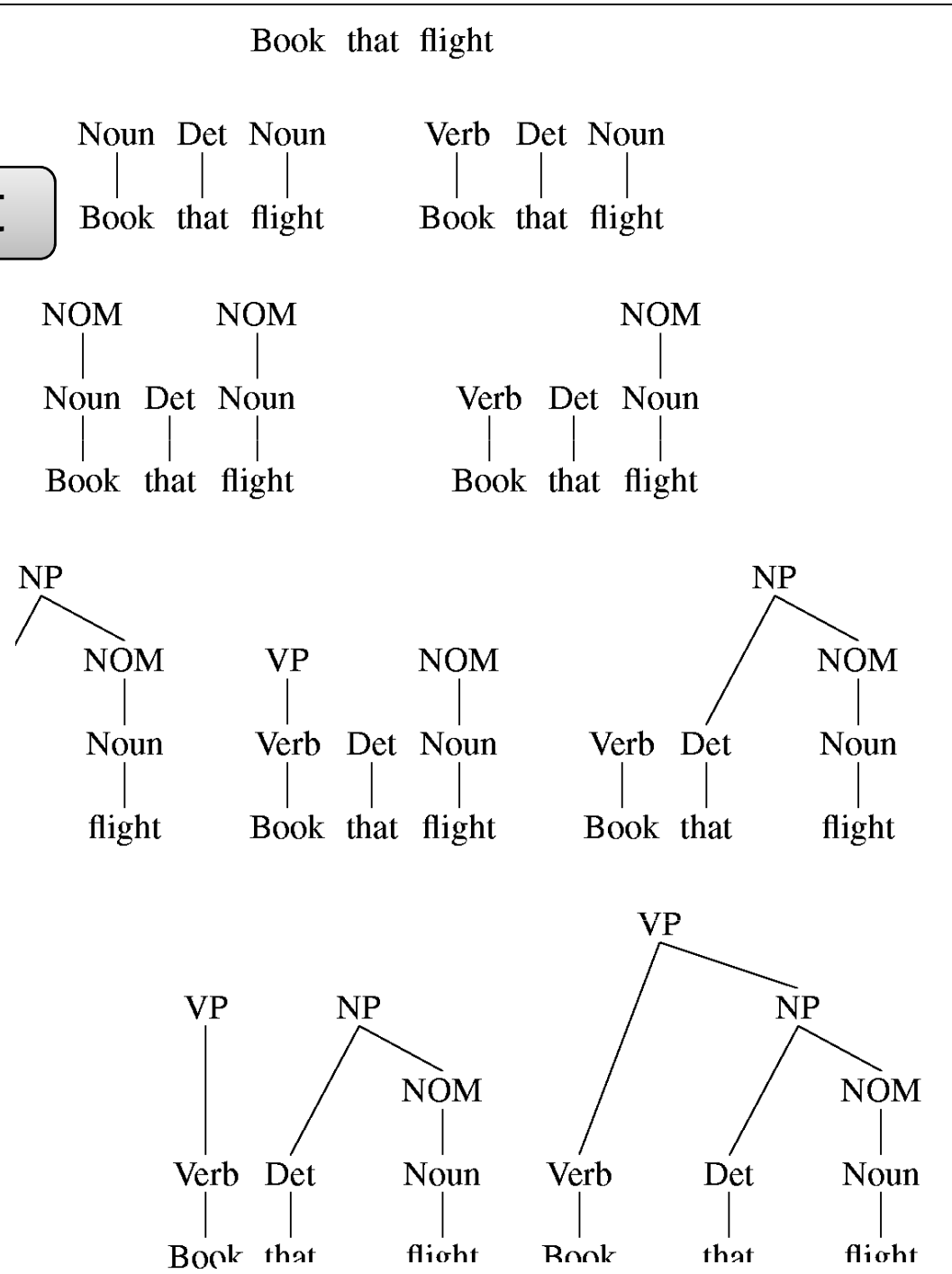
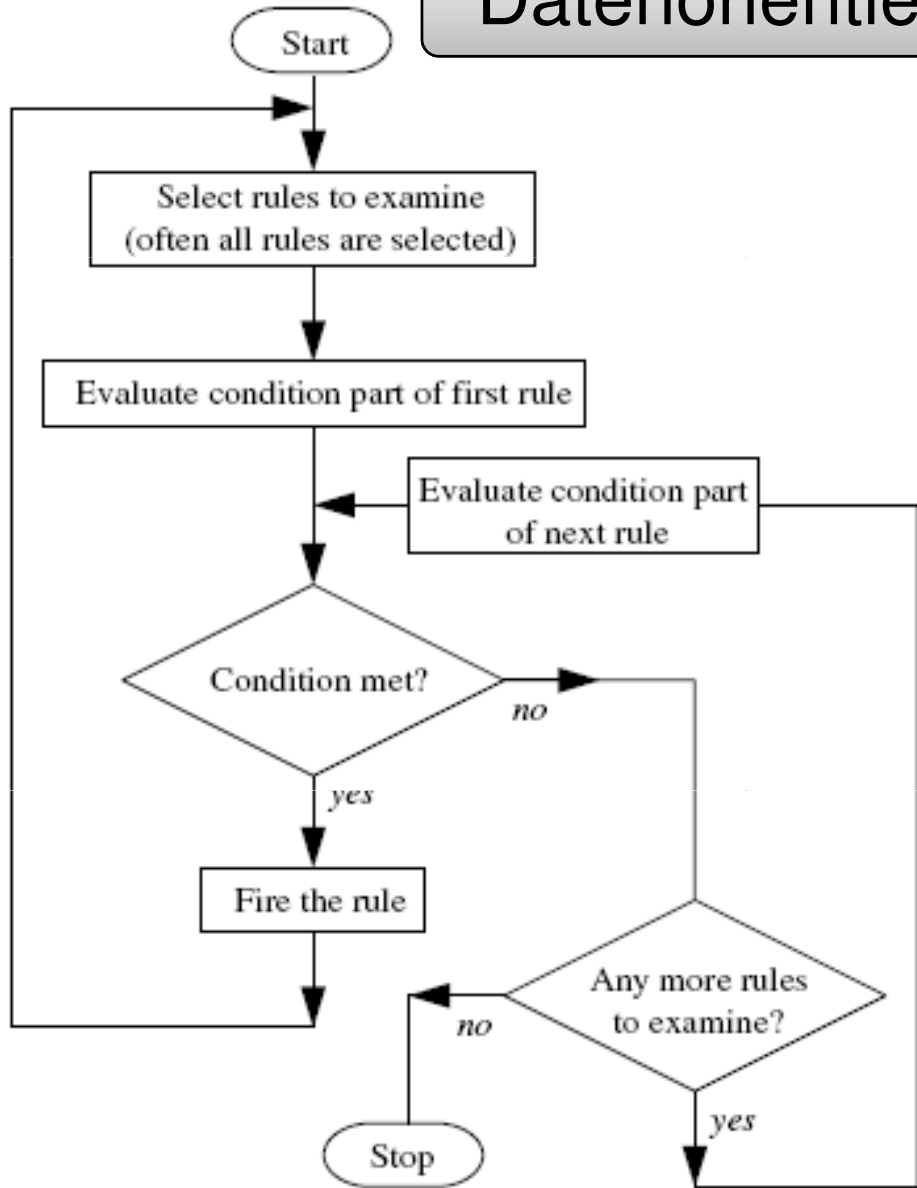
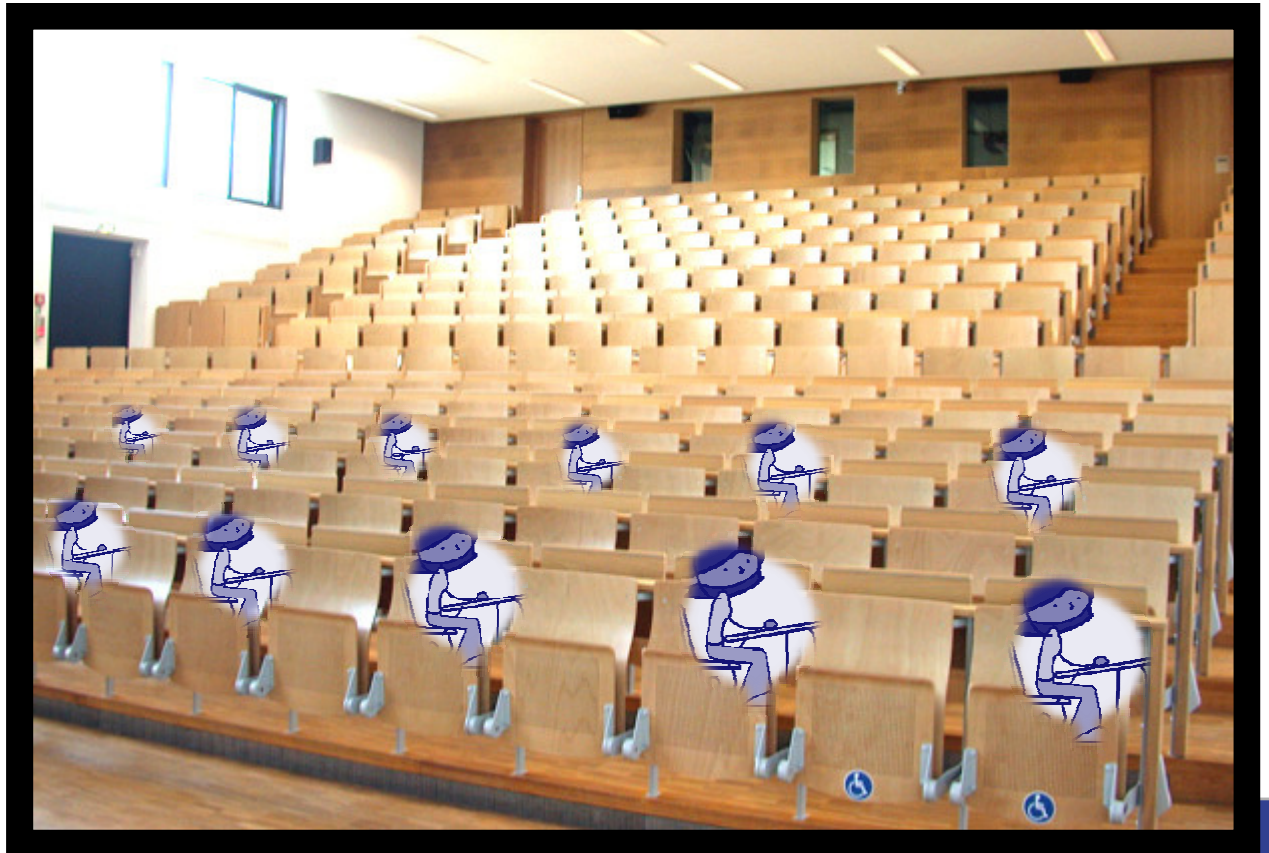


Figure 10.4 An expanding bottom-up search space for the sentence *Book that flight*. This figure does not show the final tier of the search with the correct parse tree (see Figure 10.1). Make sure you understand how that final parse tree follows from the search space in this figure.



# Zur Klausur

- 16.07., LB 107, 14.00 - 15.00 Uhr (90 min.)
- Ausweise mitbringen (Student., Personal.)
- Hilfsmittel: Alles außer Notebook(-artigem, inkl. PDA, Handy)
  - Bitte selbständig arbeiten. Sonst: Täuschungsversuch.
- Sitzordnung:
  - Versetzt (Schachbr.)
  - 1 Reihe frei



Viel Erfolg für die Klausur,  
die weiteren Prüfungen,  
das weitere Studium,  
Lotto, NachbarIn,  
.....

**Tschüss,  
schöne Ferien**

