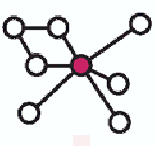


D. Wissensrepräsentation



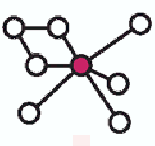
Prinzipien der Wissensrepräsentation



Davis, Shrobe, and Szolovits (1993) discussed **five principles for a knowledge representation**:

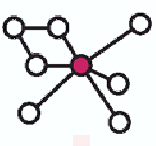
1. A knowledge representation is a **surrogate**,
 2. a **set of ontological commitments**,
 3. a **fragmentary theory of intelligent reasoning**,
 4. a **medium for pragmatically efficient computation**,
 5. and a **medium of human expression**.
- The way a knowledge representation focuses on these principles characterizes its spirit.
 - Each knowledge representation is a trade-off between these principles.

Knowledge representation as surrogate



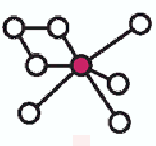
„Knowledge representation is most fundamentally a *surrogate*, a substitute for the thing itself, used to enable an entity to determine consequences by thinking rather than acting. [...] Reasoning is a process that goes on internally [of a person or program], while most things it wishes to reason about exist only externally.“

Knowledge representation as as set ontological commitments



A knowledge representation “is a *set of ontological commitments*, i.e., an answer to the following question: ,In what terms should we think about the world? [...] In selecting any representation, we are [...] making a set of decisions about how and what to see in the world. [...] We (and our reasoning machines) need guidance in deciding what in the world to attend to, and what to ignore.”

Knowledge representation as theory of intelligent reasoning



„The initial conception of a knowledge representation is typically motivated by some insight indicating how people reason intelligently, or by some belief about what it means to reason intelligently at all. [...]

It is a *fragmentary theory of intelligent reasoning*, expressed in terms of three components:

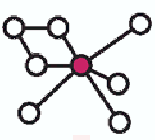
- (i) the representation's fundamental conception of intelligent reasoning;
- (ii) the set of inferences the representation *sanctions*;
- (iii) and the set of inferences it *recommends*.

The authors consider five fields which have provided notions of intelligent reasoning:

- Mathematical Logic (e.g., Prolog)
- Psychology (e.g., frames)
- Biology (e.g., neural networks)
- Statistics (e.g., bayesian networks)
- Economics (e.g., rational agents)

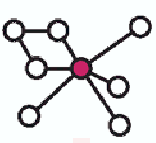
(Siehe drei Folien weiter.)

Knowledge representation as medium for efficient computation



Knowledge representation „is a *medium for pragmatically efficient computation*, i.e., the computational environment in which thinking is accomplished. One contribution to this pragmatic efficiency is supplied by the guidance a representation provides for organizing information so as to facilitate making the recommended inferences.“

Knowledge representation as medium of human expression

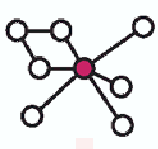


„Knowledge representations are [...] the medium of expression and communication in which we tell the machine (and perhaps one another) about the world. [...]

Knowledge representation is thus a medium of expression and communication for the use by *us*. [...]

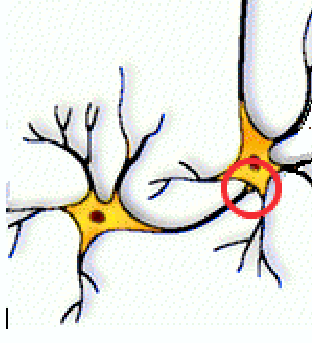
A representation is the language in which we communicate, hence we must be able to speak it without heroic effort.“

Wissen



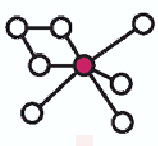
Es gibt verschiedene Wurzeln der Wissensverarbeitung, die zu unterschiedlichen Modellen von Wissen führen:

- Biologie: Vernetzung
→ neuronale Netze
- mathematische Logik: Deduktion
→ Logikkalküle, Prolog
- Statistik: Unsicherheit
→ Fuzzy-Logik, Bayessche Netze
- Psychologie: Begriffe
→ wissensbasierte Systeme
- Ökonomie: Ziele, Bewertungen
→ Agenten



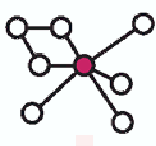
DIN 2330 Seite 11





Neuronale Netze

- Wir werden in dieser Vorlesung den Schwerpunkt auf die begriffliche Wissensrepräsentation legen.
- Vorher aber ein kleiner Ausflug in die neuronalen Netze, um die Verschiedenheit der Ansätze zu beleuchten.
- Neuronale Netze werden detailliert in der Vorlesung „Knowledge Discovery“ besprochen.



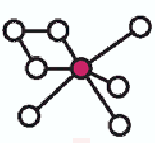
(Künstliche) Neuronale Netze

sind massiv parallel verbundene Netzwerke aus

- einfachen (üblicherweise adaptiven) Elementen in
- hierarchischer Anordnung oder Organisation,

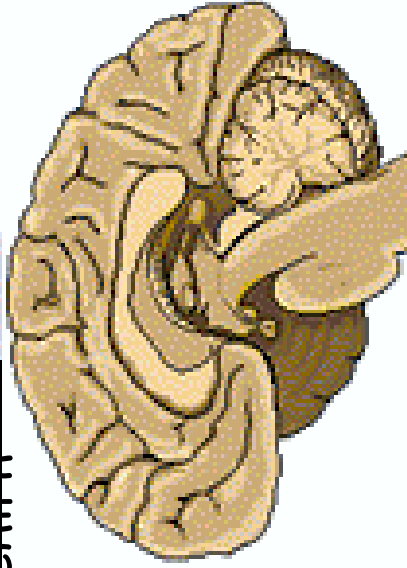
die mit der Welt in der selben Art wie biologische Nervensysteme interagieren sollen. [Kohonen 1984]

Neuronale Netze



Vorbild Biologie

Gehirn

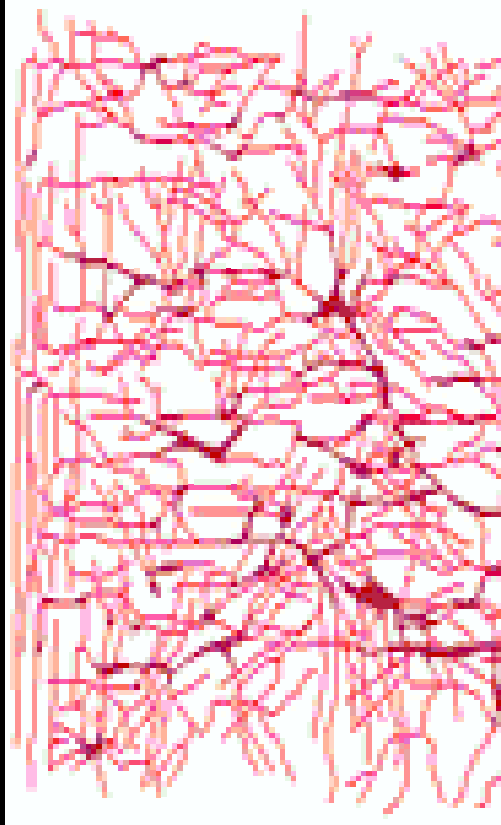


Das Gehirn besteht aus

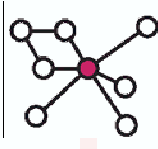
- ca. 10^{11} Neuronen, die mit
 - ca. 10^4 anderen Neuronen
 - ca. 10^{13} Synapsen
- verschaltet sind.

In der Vergangenheit wurden Forscher aus den verschiedensten Fachgebieten durch das biologische Vorbild motiviert.

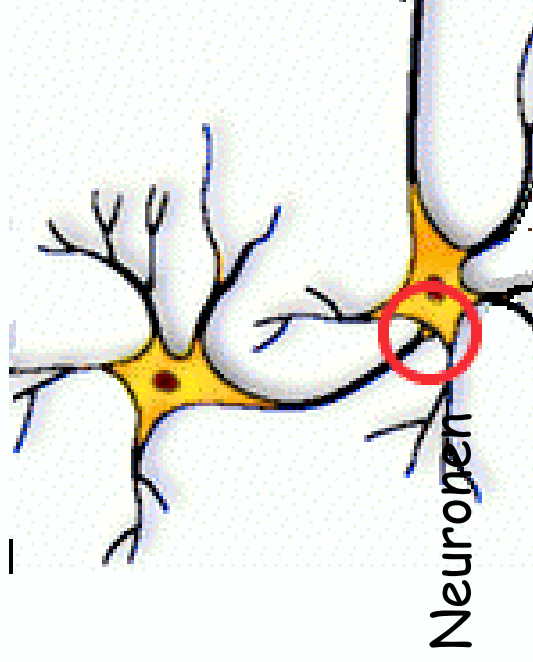
Geflecht aus Neuronen



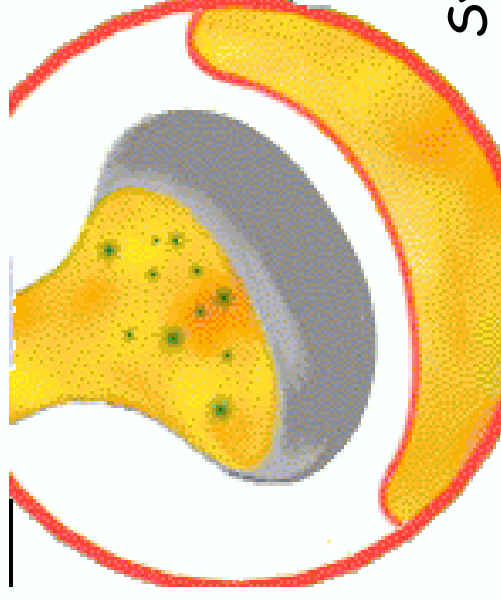
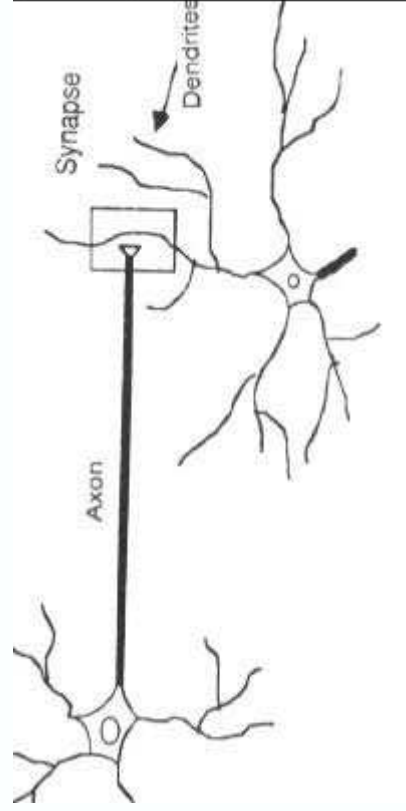
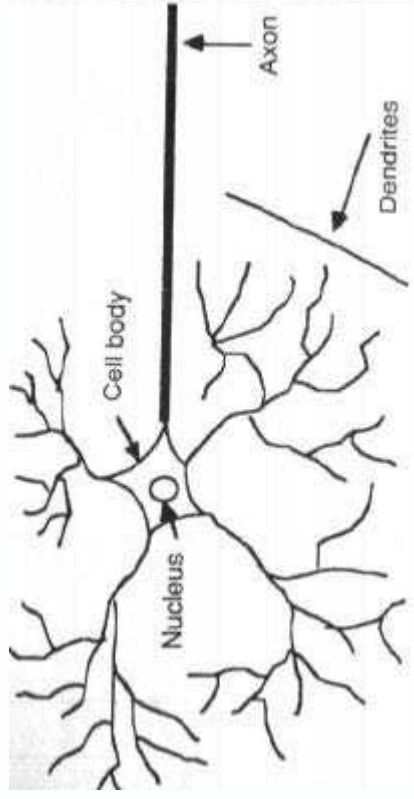
Neuronale Netze



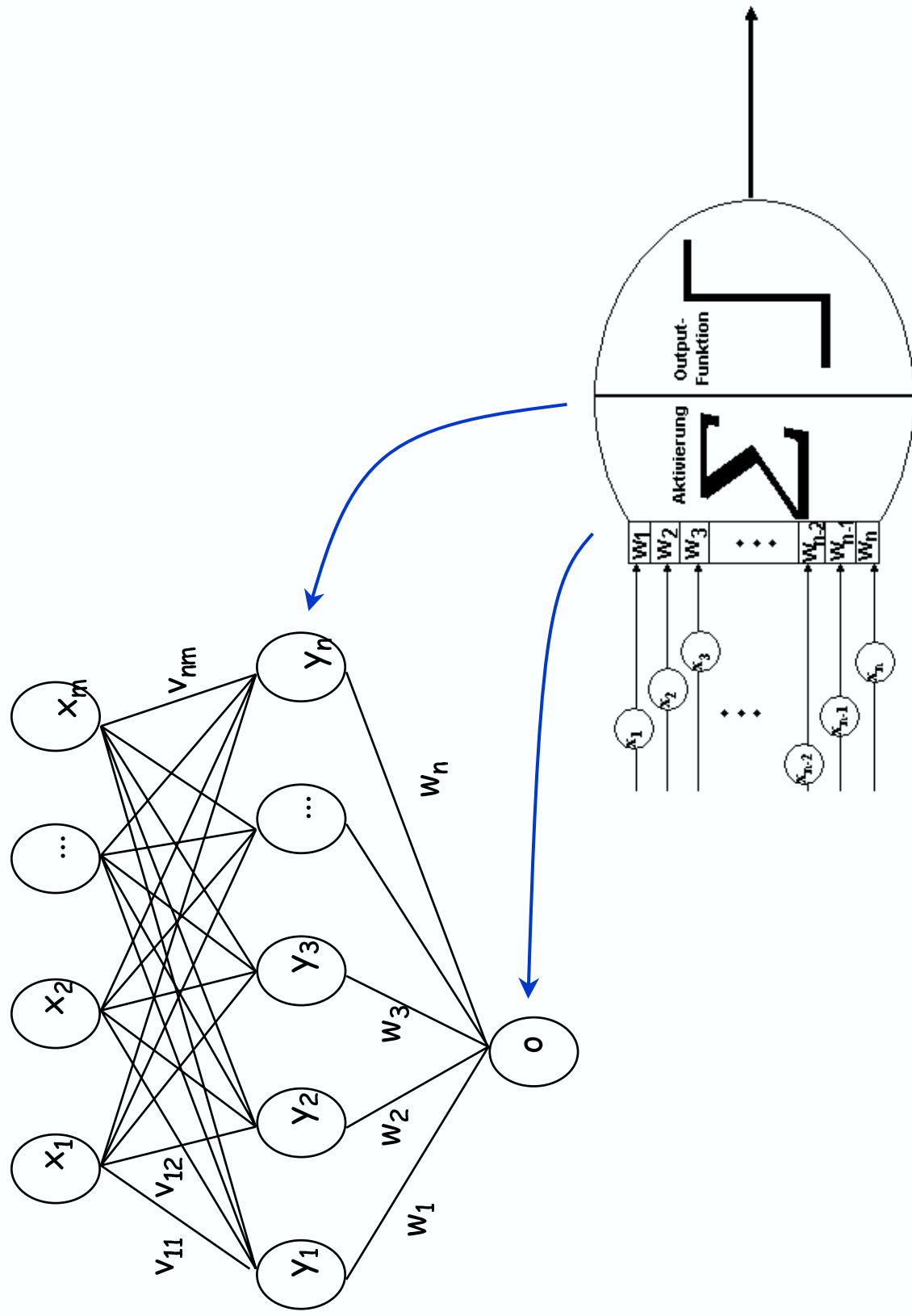
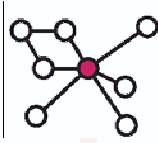
Vorbild Biologie



Neuronen



Synapse



Neuronale Netze

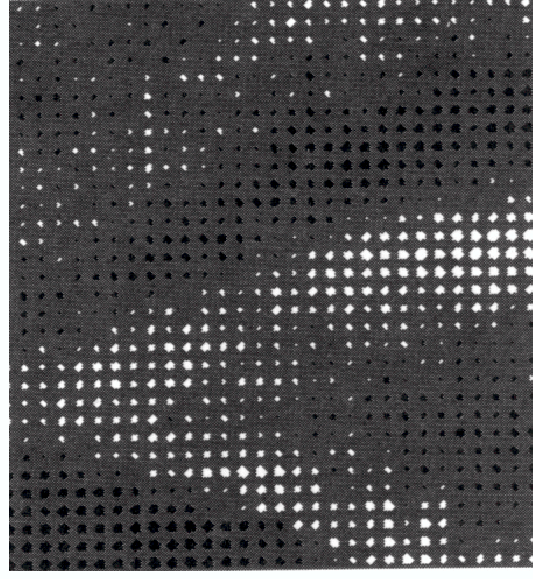
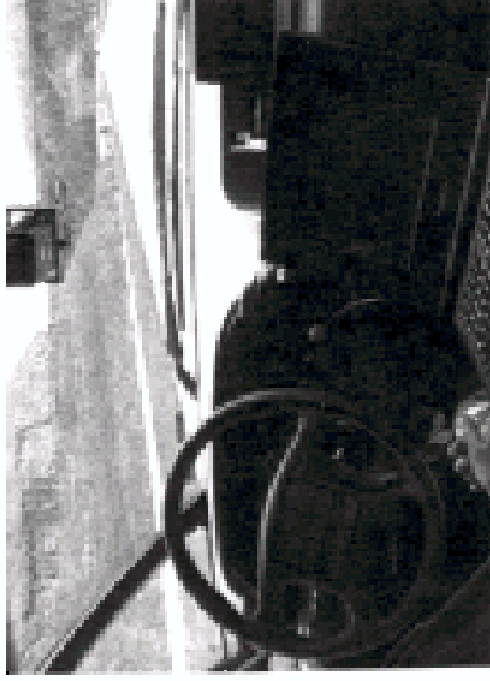
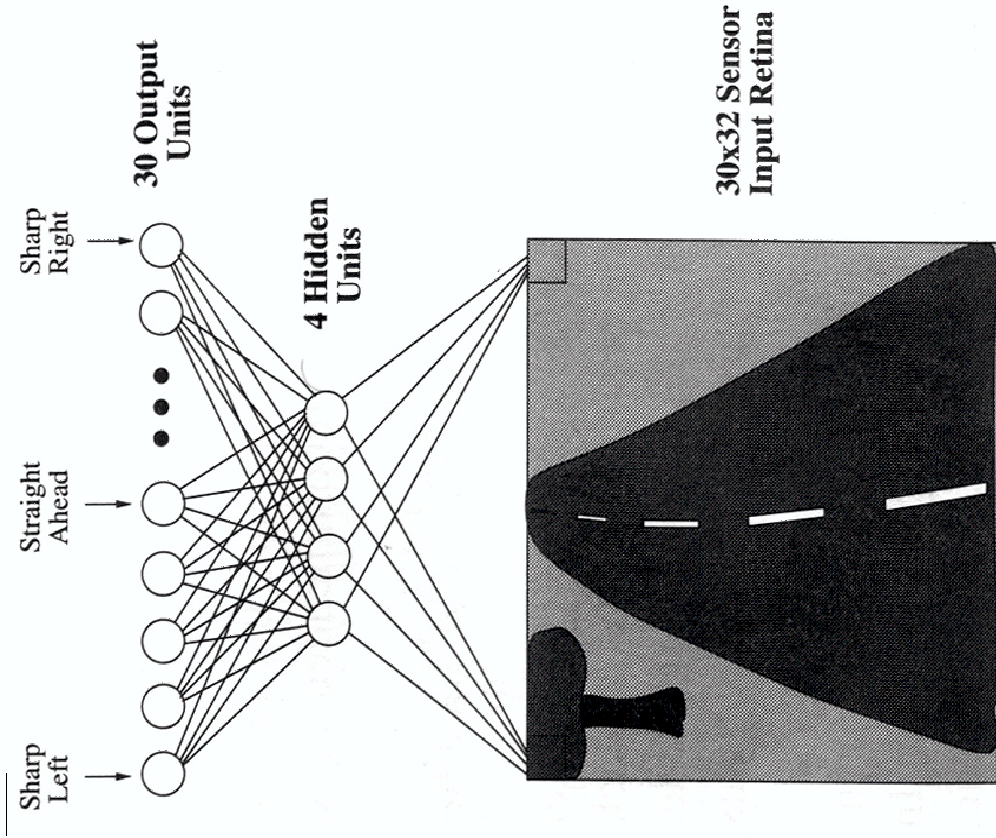
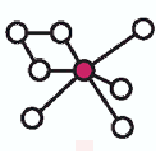
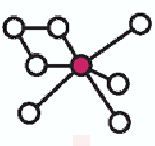


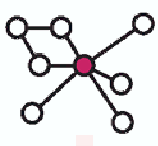
Figure 1: Neural Network learning to steer an autonomous vehicle (Mitchell 1997)

Neuronale Netze



Wofür nutzt man sie?

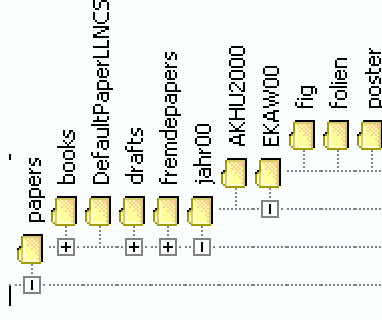
- Forschung:
 - Modellierung & Simulation biologischer neuronaler Netze
 - Funktionsapproximation
 - Speicherung von Informationen
 - ...
- Anwendungen (z.B.):
 - Interpretation von Sensordaten
 - Prozeßsteuerung
 - Medizin
 - Elektronische Nase
 - Schrifterkennung
 - Risikomanagement
 - Zeitreihenanalyse und -prognose
 - Robotersteuerung



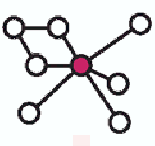
Begriffliches Wissen

- Wir werden in dieser Vorlesung den Schwerpunkt auf die begriffliche Wissensrepräsentation legen.
- Die psychologische Perspektive führt zu begrifflichen Modellen von Wissen:

- Semantische Netze
- Frames
- Begriffliche Graphen
- Formale Begriffsanalyse
- Beschreibungslogiken
- Ontologien/Semantic Web



Begriffliche Wissensrepräsentation



Was ist ein Begriff?

Begriff. Denkeinheit, die aus einer Menge von Gegenständen unter Ermittlung der diesen Gegenständen gemeinsamen Eigenschaften mittels Abstraktion gebildet wird.

Benennung. Aus einem Wort oder mehreren Wörtern bestehende Bezeichnung.

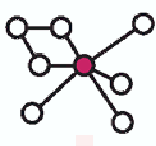
Definition. Begriffsbestimmung mit sprachlichen Mitteln.

[DIN 2342 Teil 1, Begriffe der Terminologielehre - Grundbegriffe, 10.92]

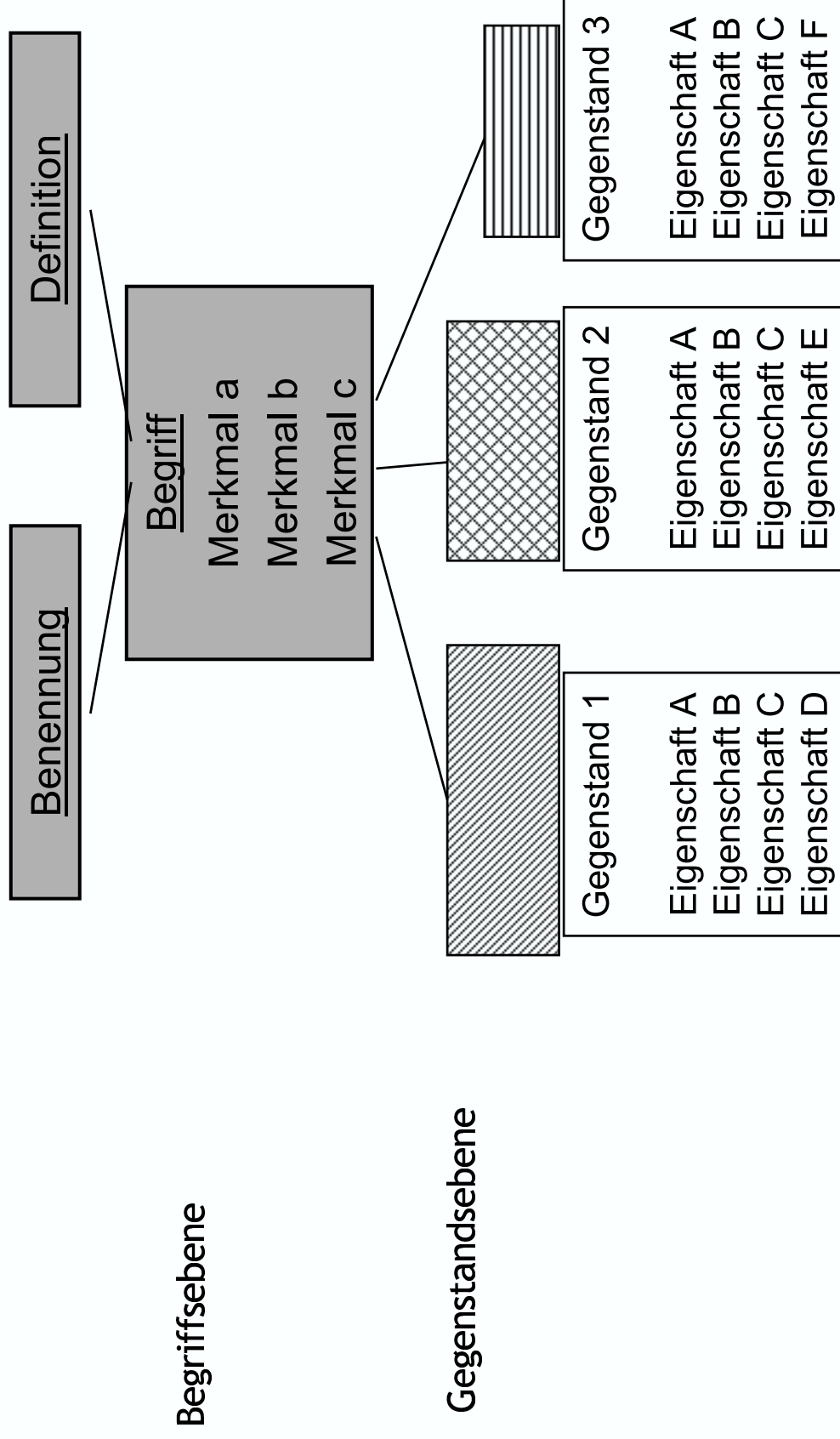
The set of *characteristics* that come together as a unit to form the *concept* is called the *intension*. The *objects* viewed as a set and conceptualized into a *concept* are known as the *extension*. The two, *extension* and *intension*, are interdependent. For example, the *characteristics* making up the *intension* of ‚lead pencil‘ determines the *extension*, those *objects* that qualify as lead pencils and vice versa.

[ISO 704: Terminology Work: Principles and Methods, ISO 2000]

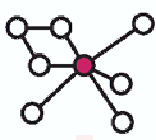
Begriffliche Wissensrepräsentation



DIN 2330: Begriffe und Benennungen. Allgemeine Grundsätze



Begriffliche Wissensrepräsentation

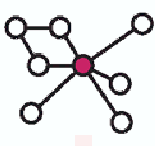


Wir stellen nun kurz ein paar Formalismen zur Repräsentation begrifflichen Wissens vor:

- Natural Language
- Concept hierarchies
- Existential Graphs
- Semantic Networks
- Frames
- Conceptual Graphs
- Description Logics

Letzteren werden wir im Verlauf der Vorlesung wiederbegegnen als formale Grundlage für Wissensrepräsentation im WWW.

(siehe u.a. J. Sowa: Semantic Networks. <http://www.jfsowa.com/pubs/semnet.htm>)



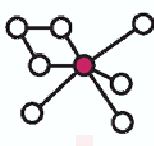
Advantages of natural language:

- Very expressive.
- Easy to understand for humans.

Problems with natural language (for automatic processing):

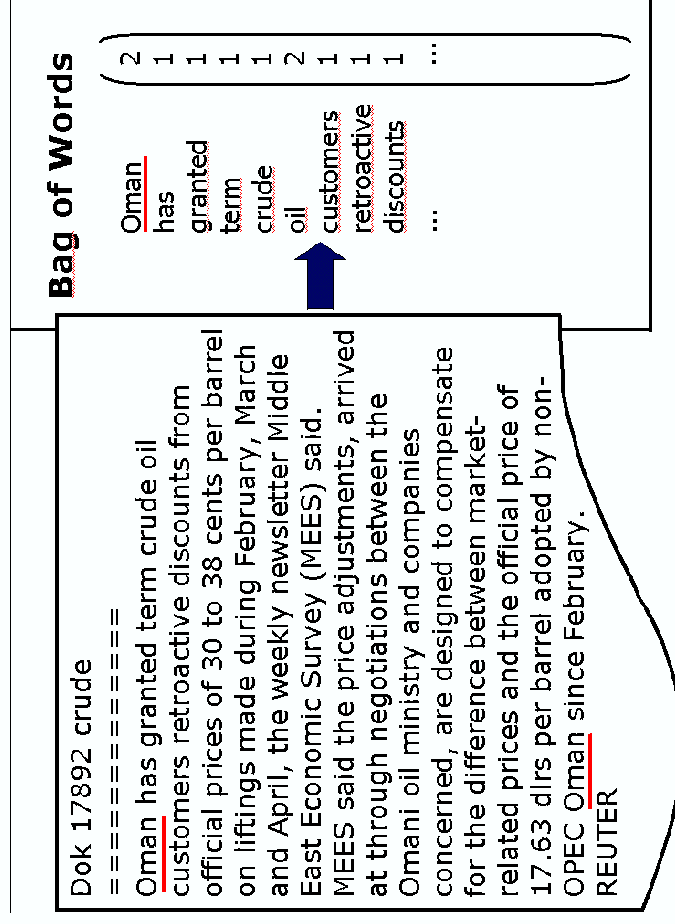
- Natural language is often ambiguous.
- Semantics are not formally defined.
- There is little uniformity in the structure of sentences.

Knowledge Representation by Natural Language



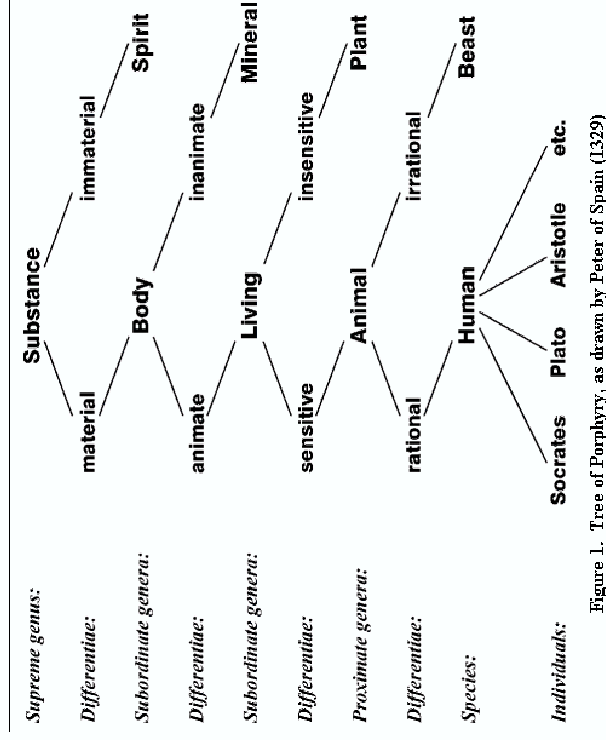
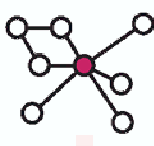
Excursion: Workaround for Information Retrieval and Text Mining:

- “Bag of Words” Model (see lecture on Information Retrieval)
- (Is not considered as ‘Knowledge Representation’ in the narrow sense.)



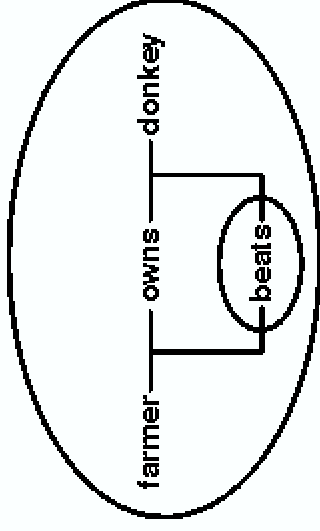
- Advantages:
 - *easy computation* in a vector space
 - “similarity” between documents can be *measured*
- Disadvantage:
 - *internal structure and meaning* of the text gets *lost*

Concept hierarchies



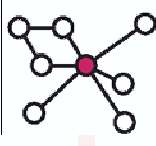
- going back to Aristotle/Porphyry
- Advantages:
 - easy to understand
 - widely used
- Disadvantage:
 - not very expressive (but sufficient for many applications - eg, the file system on your PC)

Existential Graphs

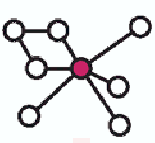


"If a farmer owns a donkey, then he beats it."

- Introduced by Charles S. Peirce in 1897(!).
- Advantages:
 - intuitive reading
 - intuitive calculus
 - formal semantics (but hard to extract from Peirce's manuscripts)
- Disadvantage:
 - non-linear notation difficult to represent in the computer

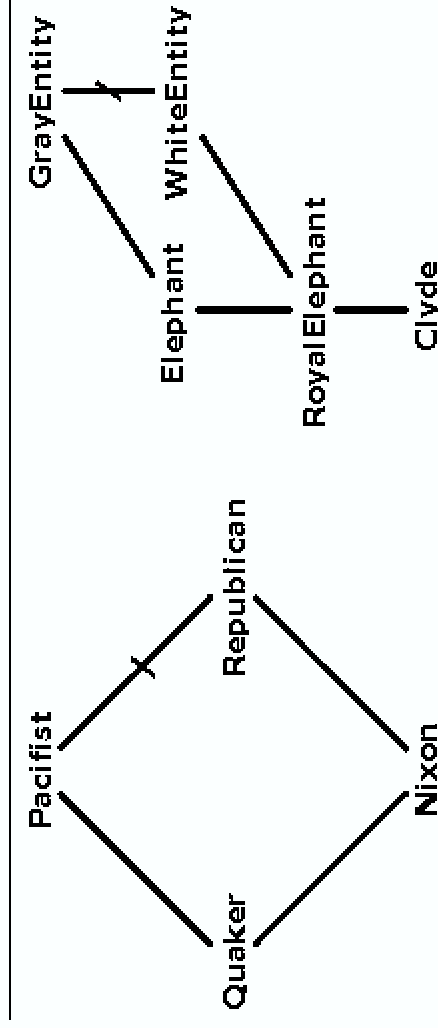
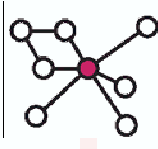


Semantic Networks



- "Semantic Nets" were first invented for computers by Richard H. Richens of the Cambridge Language Research Unit in 1956 as an „interlingua“ for machine translation of natural languages.
- They were developed by Robert F. Simmons in the early 1960s and later featured prominently in the work of M. Ross Quinlan in 1966.
- Did not provide clear, formal semantics. → Different implementations could treat the same data differently!

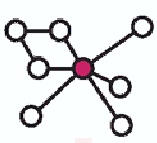
Semantic Networks



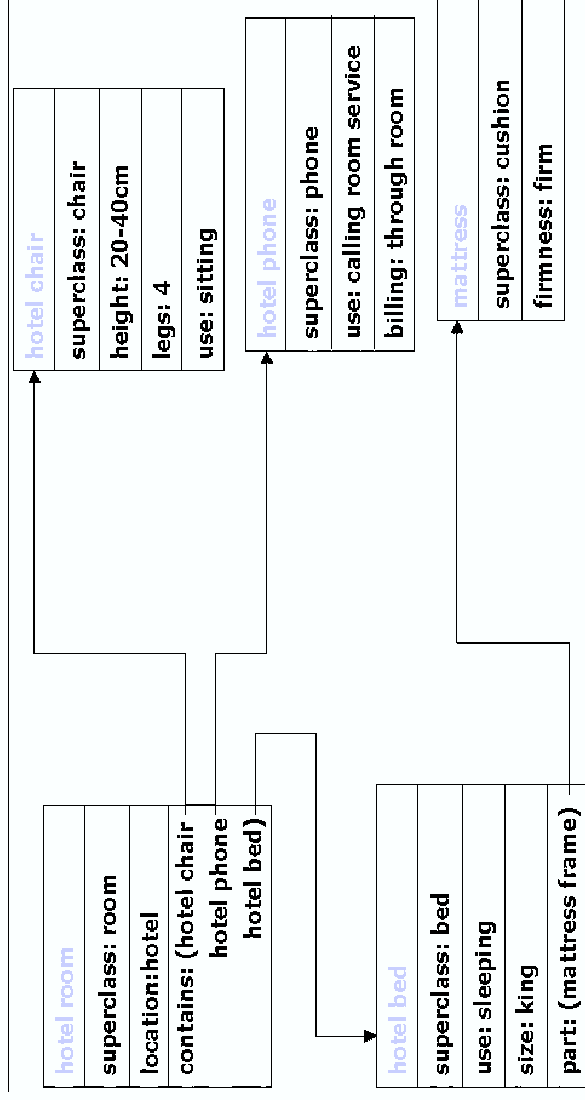
The example above shows default reasoning: Usually, elephants are grey.

- Problem: How are royal elephants treated? → The formalism is ambiguous. Different implementations may treat the same data differently. A formal semantic would answer this question.
- Later, Description Logics were introduced to overcome this problem by introducing such a semantic.

Frames

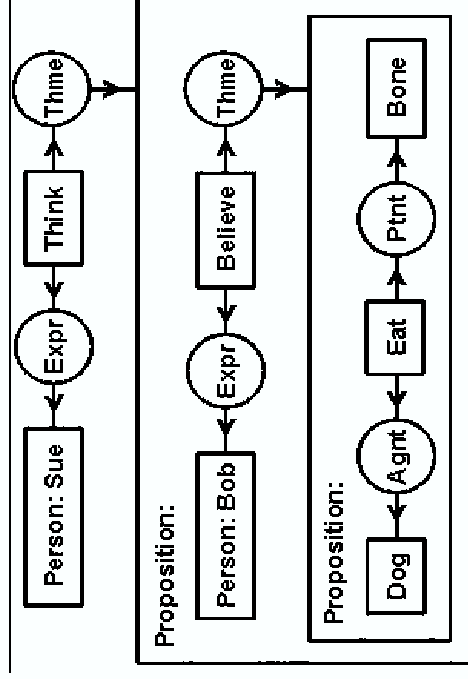
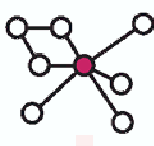


- Much like a semantic network except each node represents prototypical concepts and/or situations.
- Each node has several property *slots* whose values may be specified or *inherited* by default.
- Frames are similar to object-oriented modelling



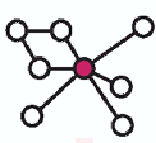
- Advantages:
 - is intuitive for many users
- Disadvantages
 - formal semantics are missing

Conceptual graphs



- Based on Peirce's existential graphs
- Aim at providing a formal semantics, by mapping to predicate logic.
- Advantages:
 - intuitive reading
 - expressive (beyond first order logic, see example above)
 - provides linear notation for easier computation
- Disadvantages:
 - mapping to predicate logic is not (yet) well defined → formal semantics are still missing

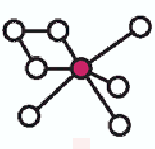
Description Logics



Woman	≡	Person $\sqcap$ Female
Man	≡	Person $\sqcap$ $\neg$ Woman
Mother	≡	Woman $\sqcap$ $\exists$ hasChild.Person
Father	≡	Man $\sqcap$ $\exists$ hasChild.Person
Parent	≡	Father $\sqcup$ Mother
Grandmother	≡	Mother $\sqcap$ $\exists$ hasChild.Parent
MotherWithManyChildren	≡	Mother $\sqcap$ ≥ 3 hasChild
MotherWithoutDaughter	≡	Mother $\sqcap$ $\forall$ hasChild. $\neg$ Woman
Wife	≡	Woman $\sqcap$ $\exists$ hasHusband.Man

- Beschreibungslogiken sind eine Familie von entscheidbaren Fragmenten der Prädikatenlogik.
- Im Gegensatz zu Frames und semantischen Netzen bieten sie eine formale Semantik.
- Erster Vertreter: KL-ONE (1985)
- Heutzutage Grundlage für das Semantic Web (→ Kapitel E & F)

Conclusion



- There are many formalisms for representing conceptual knowledge.
- Depending on the purpose, a different formalism might be the most suitable.
- In this light, we will discuss (in the next section) three formalisms in more details that were established for the (Semantic) Web: XML, RDF, OWL.