STRUCTURING KNOWLEDGE: ABSTRACTION

Abstraction is fundamental!

- *Abstraction* is a logical and constructive procedure.
- Objects are being subsumed under a concept, if they have equal, explicitly denominable properties.

  We arrange (associate) all objects, which have equal property (feature) instantiations, into one class.
- By means of this classification we define an *equivalence relation:*
  - Reflexivity $\forall a.(a,a) \in R$, also: aRa
  - Symmetry $\forall a,b.(a,b) \in R \rightarrow (b,a) \in R$
  - Transitivity $\forall a,b,c.(a,b) \in R \wedge (b,c) \in R \rightarrow (a,c) \in R$

Abstractor

- To express that certain statements are invariant wrt. to an equivalence relation, we use a *hypothetical abstract* object which has only the *common* properties.
- Invariant statements A are expressed in a new way by introducing an *abstractor* α: $A(\alpha x)$ with the *abstract* object αx

Example: The Natural Numbers

- Basic class: counting signs.
  Different number representations (lists of strokes, . . .)
- Abstractor: number (as abstract counting sign)
- Equivalence relation: consisting of as many counting signs

Abstraction Schema

Abstraction is the transition from statements

$$A(a)$$

about objects $a, b, \ldots$ among which an equivalence relation $\sim$ has been established, to statements

$$A(\tilde{a})$$

by means of the *abstraction schema*

$$A(\tilde{x}) \rightleftharpoons \forall y(x \sim y \rightarrow A(y))$$

In this way we introduce talking about *abstract objects* $\tilde{a}, \tilde{b}, \ldots$

If two objects a, b obey the relation $a \sim b$, we say that a and b represent the same abstract object $\tilde{a}$ (or $\tilde{b}$, respectively).

Concepts

Concepts ("Begriffe"): Abstraction over predicators wrt. the equivalence relation of invariance relative to a rule system of terminological agreements $\Rightarrow$ *intension*.

LEIBNIZ' theorem of indiscernibility: logical equality.

States of affairs ("Sachverhalte") are gained from statements by means of abstraction wrt. to equality of contents.
Facts: **True** states of affairs.

(Propositional) Knowledge of one or more persons:
A set of statements which satisfy certain conditions on their content, and the *truth* of which the persons can *justify* by (good) reasons they agree upon.

KNOWLEDGE AND COGNITION

The term *"cognitive"* is introduced as a general distinction of the areas of perception, thinking, and imagination from other mental areas like emotion or will.

Nagao: *knowledge* = *cognition* + *logic* ?!?

Truth??
Justification??

Knowledge acquisition through cognition: perception and recognition of objects.
Without knowledge we cannot re-cognize anything.

Theoretical knowledge: *analytic* vs. *synthetic*

Knowledge and Cognition (2)

Rational theory of cognition: Ability to acquire knowledge through the intellect (ratio). Focussing on the deductive-demonstrative function of knowledge acquisition and construction.

Experience theory of cognition: Ability to acquire knowledge on the basis of experiences we make when we interact with our environment. Focussing on induction, classification and description.

Historical remarks:

LOCKE: Empirism
DESCARTES: Three categories of ideas — those gained from sensual data, from imaginations and from innate ideas
BACON: Descriptive knowledge classification.

Recording and Organizing Knowledge

Science arises with the invention of notation systems: Recording and structuring of knowledge as a prerequisite for the invention of the scientific method (Thales).

A brief look at structuring and organizing existing (scientific) knowledge from a top-down perspective

- Structuring in the large: Encyclopedias and systematics in the sciences
- Derivation of classifications according to content — conceptual structure, e.g. library classification schemata
- Classification of items/instances in terms of properties, e.g. of books in a library: *metadata*
- Schematic descriptions of content: from thesauri to formal ontologies (cf. ch. 4, 9)
- Formal representation languages and inference, e.g. for automatic classification and querying: *LOGIC*

Encyclopedic Structuring of Knowledge

Methodology and **didactic purposes**

- Aristotle's writings: logic, science (physics, astronomy, meteorology, biology, physiology, psychology), ethics, esthetics, metaphysics ("first philosophy")
- Ancient classification in mathematics and science: Pythagoras, Ptolemy, Pliny,. . .
- Cassiodorus (Martianus Capella, Isidor of Sevilla): Artes liberales
  - Trivium – grammar, rhetorics, dialectics
  - Quadrivium – arithmetics, geometry, astronomy, music
- Medieval encyclopedism

- From humanism to rationalism and modern science (Descartes, Bacon)
- First encyclopedias in alphabetical order
- Diderot and d'Alembert: L'Encyclopédie — Enlightenment
- Buffon, Linné: systematics in biology, empirical sciences
- Leibniz' program: universal encyclopedia, characteristica universalis, calculus rationcinator; ars inventoria
- Modern science: global view: Comte,. . . ; disciplinary views
- Non-European cultures: Chinese encyclopedism

Factual Knowledge and its Organization

Systematization and *Classification*

Classification: library and information science
— static view (schema $\leftrightarrow$ medium)

New terms through scientific and technological progress

Administration of knowledge corpora by computers: **Access** — **Search**

- *Primary information*
- *Secondary information, METADATA* — keywords, annotations of primary information

Using Primary, Secondary, and Tertiary Information

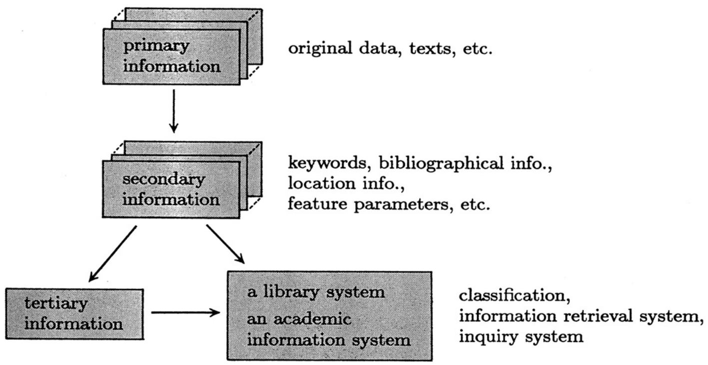


Figure 2.1 Systems for making use of various kinds of information.

Factual Knowledge

Kinds/types of data $\Rightarrow$ "information" (interpretation relation)

Collection and *organization* of data

Classification:
System of concepts represented by a *system* of terms (words) —
THESAURUS

Condensation of data
abstracting, indexing (manual/automatic)

Access methods: search in primary data and metadata

Dewey's Decimal Classification for Libraries

Organisation von Faktenwissen: Dezimalklassifikation (Dewey)

Table 2.1 The 10 classes of the Nippon Decimal Classification

0	general
1	philosophy
2	history
3	social science
4	natural science
5	engineering/technology
6	industry
7	art
8	language
9	literature

Table 2.2 Division of the Nippon Decimal Classification

1(00)	philosophy	class	
12(0)	eastern philosophy	code	
121	Japanese philosophy	item	
121.5	modern	detailed item	
121.52	study of classical literature	more detailed item	small items
121.54	the doctrines of Chu-tzu	more detailed item	

Principles of Decimal Classification

- *Consistency:* each area is divided based on only one classification principle
- *Exclusiveness:* each subclass should not overlap with another
- *Sufficiency:* the classification should not leave anything out
- *Gradualism:* the classification should progress from broader concepts to narrower concepts, and should not leave gaps

Complex classification: combination of *feature structures*
Example: A book about *economy and politics*

Table 2.3 (a) Electrical engineering items

500	technology/engineering	**540**	electrical engineering
510	civil engineering	541	circuit
520	architectural engineering	542	electrical machine
530	mechanical engineering	543	generation of electricity
540	*electrical engineering*	544	transmission, variation, supply of electric power
550	oceanographical/vessel engineering	545	lights
560	metallic engineering/mining engineering	546	electric train
570	chemical engineering	547	communication
580	manufacturing engineering	548	information engineering
590	domestic science	549	electronics

Table 2.3 (b) Information engineering items

548	information engineering Here are categorized only the engineering treatments of computer science. General computer science, software and systems are included in 007.
.2	computers
.3	automatic control, automation, robot
(.4	cybernetics → 007.1)
(.5	human engineering → 501.8)
(.6	bio-engineering → 501.8)
.7	simulation

Table 2.4 Computer science items

007	computer science *Information engieering is categorized in 548. General computer science and software are catego rized here.	.53	indexing
		.54	abstracting
		.57	standardization of information records → 014.3
.1	Information theory (communication and its media, cybernetics, symbols, language and semantics)	.6	data processing: database management system
		.61	system analysis
.2	history/circumstances	.63	computer systems
.3	information and society, information industry	.64	computer programming
		.65	input, accumulation, output
.4	information sources	.68	machine retrieval
.5	documentation, information management	.7	information systems

Organization of Factual Knowledge: Classification Schemata

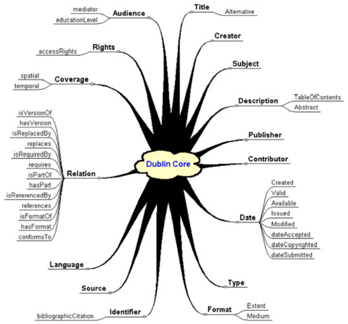
General: Libraries and bibliographic information services

- Library of Congress classification
- Bibliographic classification of German,. . . libraries
- Taxonomies for web pages (Google, Yahoo!) — *automatic*

Domain specific: for most disciplines,
e.g. ACM Computing Classification System

Dublin core

A common core of “metadata elements”, i.e. property types, for resource description



Classifying Items: Feature-based Classification

Example: **Bibliographic entries**
author, title, publisher, place, date, language, format, type, pages, number of figures, . . .

Combination of features:

- coarse grained classification (e.g. disciplines),
- fine grained classification by feature combination

Relations between subjects:
parallel, hierarchical, part–whole, opposition, comparison/ contrast, cause/effect, influence, main subject/ instance, main subject/ points of view, main subject/ purpose, theory/ application

Thesaurus

Thesaurus: Systematik von Termini (Wörtern)

ROGET: Hervorhebung von *Synonymie* vs. Begriffsumfang (enger/ weiter)

Table 2.6 Roget's system

class	section	given code
1. abstract relations	existence	1–8
	relation	9–24
	quantity	25–57
	order	58–83
	number	84–105
	time	106–139
	change	140–152
	causation	153–179
2. space	space in general	180–191
	dimensions	192–239
	form	240–263
	motion	264–315
3. matter	matter in general	316–320
	inorganic matter	321–356
	organic matter	357–449
4. intellect (the exercise of the mind)		
(1) formation of ideas	general	450–454
	precursory conditions and operations	455–466
	materials for reasoning	467–475
	reasoning processes	476–479
	results of reasoning	480–504
	extension of thought	505–513
	creative thought	514–515
(2) communication of ideas	nature of ideas communicated	516–524
	models of communication	525–549
	means of communicating ideas	550–599
5. volition (the exercise of the will)		
(1) individual volition	volition in general	600–619
	prospective volition	620–679
	voluntary action	680–703
	antagonism	704–728
	results of action	729–736
(2) social volition	general social volition	737–759
	special social volition	760–767
	conditional social volition	768–774
	possessive relations	775–819
6. emotion, religion, and morality		
	general	820–826
	personal emotion	837–887
	interpersonal emotion	888–921
	morality	922–975
	religion	976–1000

Thesaurus of Technical Terms:

Structure of the thematic field

Relations between nouns

1. Synonyms, antonyms
2. Broader, narrower terms
3. Similar words
4. Relations: part/whole, order, cause/effect, succession, logical relation, common features
5. Compound words, derivative words

Partial Sample of the INSPEC Thesaurus

Table 2.7 Symbols for expressing the relations among nouns

hierarchical	BT	broader term
	NT	narrower term
generic	BTG	broader term generic
	NTG	narrower term generic
partitive	BTP	broader term partitive
	NTP	narrower term partitive
associative	RT	related term
antonymic	A	
equivalence	USE	use
	UF	used for
	USE+	use in combination

Table 2.8 Partial sample of INSPEC (International Information Services for the Physics and Engineering Communities) thesaurus

integrated circuits			
UF	IC	TT	networks (circuits)
	microcircuits	RT	integrated-circuit manufacture
	microelectronics		integrated-circuit technology
NT	digital-integrated circuits		integrated-circuit testing
	hybrid-integrated circuits		masks
	linear-integrated circuits		modules
	microwave-integrated circuits		radiation hardening (electronics)
	monolithic-integrated circuits		semiconductor devices
	thick film circuits		substrates
	thin film circuits		thick films
BT	networks (circuits)	CC	B2220 B2570
		FC	b2220-s b2570-r
		DI	January 1973

TT: top term(s)
CC: version of the code as used in the INSPEC Magnetic Tape Services
FC: full form of code as input to the INSPEC database
DI: date of entry or input

Thesauri / Semantic Word Nets: (Euro) WordNet

Hierarchical lexicon with representation of sense carrying relations between words — focus on linguistics (word usage), often not as rigid as formal ontologies.

Example:

```
seat  (furniture that is designed for sitting on)
  chair
    armchair
    folding chair
      camp chair
  bench
    park bench
 PARTS:  seat, upholstery
```


Formation of *lexical concepts* through **Synsets** (synonym sets); distinction of readings (seat: 6 senses; sense 3 = furniture. . .)

EuroWordNet: Polylingual extension of WordNet with background ontology “Inter-Lingual Index” (ILI)

Relations in Word Nets

- Lexical relations: **Synonymy**, synset-internal (*teacher, instructor*) and **Antonymy** (*birth / death, to love / to hate, beautiful / ugly*)
- Concept relations, holding for all realizations of a lexical concept within a synset
  - **Hyponymy**, inverse: **hyperonymy**, hierarchy-shaping (*duck* and *bird*)
  - **Meronymy**, inverse: **holonymy**, part-whole relation (*roof* and *house*)
  - **Causation** (*kill* and *die*, *opening* and *open*)
  - **Entailment** (*succeed* and *try*)

Furthermore: **Cross classification** (*banana* as *plant* and *food*)
Subcategorization frame: Syntax-semantics interface

The Semantics of Word Net

. . . lies in its **structure**, depending on the interpretation of the relations:

- synonymy — equivalence relation
- antonymy — complement/negation
- mereonymy — axioms, transitivity

(Euro) WordNet Synsets

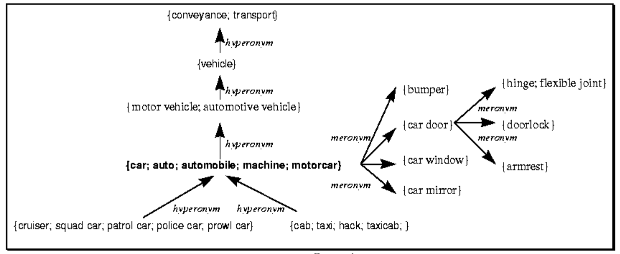


Figure 1: Synsets related to "car" in its first sense in WordNet1.5.

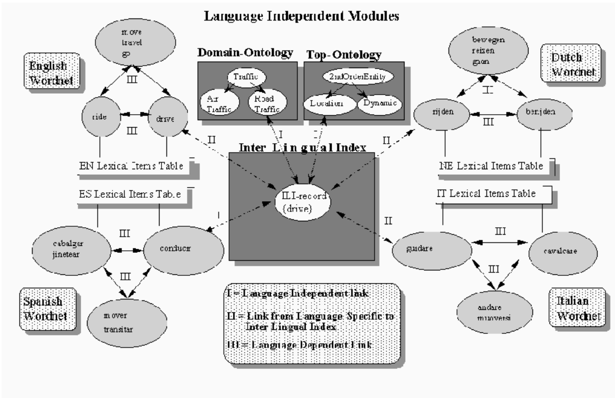


Figure 2. The global architecture of the EuroWordNet database.

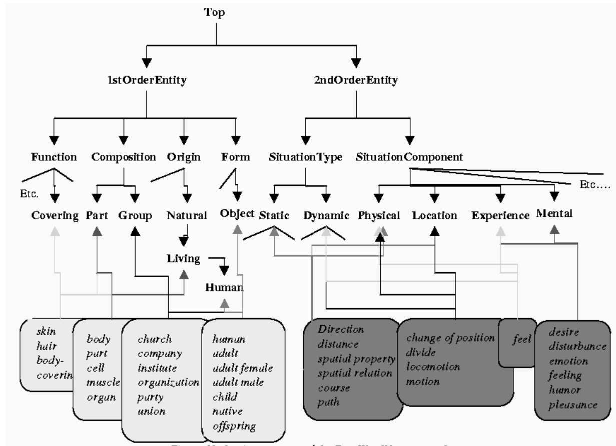


Figure 12: Lattice structure of the EuroWordNet top-ontology

Top^{0}	
1stOrderEntity^{1}	**2ndOrderEntity^{0}**
Origin^{0}	
Natural^{21}	**SituationType^{6}**
Living^{30}	Dynamic^{134}
Plant^{18}	BoundedEvent^{183}
Human^{106}	UnboundedEvent^{48}
Creature^{2}	Static^{28}
Anima^{123}	Property^{61}
Artifact^{144}	Relation^{38}
Form^{0}	
Substance^{32}	**SituationComponent^{0}**
Solid^{63}	Cause^{67}
Liquid^{13}	Agentive^{170}
Gas^{1}	Phenomenal^{17}
Object1^{62}	Stimulating^{25}
Composition^{0}	Communication^{50}
Part^{86}	Condition^{62}
Group^{63}	Existence^{27}
Function^{55}	Experience^{43}
Vehicle^{8}	Location^{76}
Representation^{12}	Manner^{21}
MoneyRepresentation^{10}	Mental^{90}
LanguageRepresentation^{34}	Modal^{10}
ImageRepresentation^{9}	Physical^{140}
Software^{4}	Possession^{23}
Place^{45}	Purpose^{137}
Occupation^{23}	Quantity^{39}
Instrument^{18}	Social^{102}
Garment^{3}	Time^{24}
Furniture^{6}	Usage^{8}
Covering^{8}	
Container^{12}	
Comestible^{32}	
Building^{13}	

Important Issues in KNOWLEDGE REPRESENTATION

(Fikes)

- What knowledge needs to be represented to answer given questions?
- How is incomplete or vague information represented?
- How is qualitative knowledge represented?
- How are assumptions represented and reasoned with?
- How can knowledge be encoded so that it is reusable?
- How can knowledge be reformulated for a given purpose?
- How can effective automatic reasoning be done with large-scale knowledge bases?

- How can computer-interpretable knowledge be extracted from documents?
- How can knowledge from multiple sources be combined and used?

FORMAL REASONING AND COMPUTATIONAL LOGICS

Reasoning

- Computational methods for creating "new" knowledge from existing knowledge
  - Primarily task-specific methods, e.g. planning, scheduling, constraint satisfaction, diagnosis,. . .
  - Methods for managing reasoning, e.g. hybrid reasoning, parallel processing,. . .
- Analysis of reasoning methods: soundness, completeness, required resources
- Methods for creating explanations of reasoning results

Computational Logics: Proof Procedures, Resolution and Unification

(Acknowledgement: Enrico Franconi, Richard Fikes)

In logic, clearly distinguish the definitions of

- the *formal language*
  syntax and semantics ‖ expressive power
- the *reasoning problem*
  decidability ‖ computational complexity
- the *problem solving procedure*
  soundness and completeness ‖ (asymptotic) complexity

The Ideal Computational Logic

- expressive
- with decidable reasoning problems
- with sound and complete reasoning procedures
- with efficient reasoning procedures – possibly sub-optimal

⇒ *Specialized logic-based representation formalisms*

Description Logics — explore the "most" interesting expressive decidable logics with "classical" semantics, equipped with "good" reasoning procedures.

Reasoning: First-Order Logic and Knowledge Representation

Consider First-Order Logic (FOL) as a language for the formal representation of knowledge and reasoning.

REMINDER:

Operations can be defined formally (proof theory) and receive semantic properties (truth theory): $\boxed{\vdash \; \equiv \; \models}$

Well-defined semantics!

Problem: **Undecidability** (Semidecidability)
Even if the decision problem is made solvable by language restrictions, it is in general not solvable within realistic time.

Preliminary Remarks on the Automation of Reasoning

How can we get around undecidability/semidecidability, the fundamental problem for automatic theorem proving in full FOL?

- Use heuristics ⇒ loss of completeness or soundness
- Restrict the language to a decidable fragment

Common inference procedures

- Resolution: Generalization of the propositional resolution calculus to FOL with unification
- Tableau calculi: Generalization of formal logic dialogues to tableaux — in particular for Description Logics

Problems and Algorithms

A **problem** is a “general question” to be answered. It is described by giving

- a general description of all its parameters, and
- a statement of which properties the answer, or *solution*, is required to satisfy.

An *instance* of a problem is obtained by specifying values for all parameters.

An **algorithm** is a finite, effective and determined instruction for solving problems.

An algorithm is said to *solve* a problem Π, if

- it can be applied to any instance I of Π, and
- it is guaranteed to always produce a solution for Π.

Reasoning Procedures

A reasoning procedure is an algorithm trying to solve *specific instances* of a *specific reasoning problem* in a *given logic*.

- Whenever a **sound** reasoning procedure claims to have found a solution for a given instance of a problem, then this is actually a solution.
  - “no wrong inferences are drawn”
  - A sound procedure may fail to find a solution for some instances of the problem, when they actually have one.
- Whenever an instance of a problem has a solution, a **complete** reasoning procedure computes the solution for that instance.
  - “all the correct inferences are drawn”
  - A complete procedure may claim to have found a solution for some instances of the problem, when they do not have one.

Pragmatic Aspect: Use — Basic Operations

- **Tell** (Store): Add a statement to the knowledge base; may include
  - testing for consistency
  - deriving consequences and storing them
- **Untell** (Remove) a statement from the knowledge base — *non-monotonic!*
  - must include removing derived statements: **Reason Maintenance**
- **Ask**
  - whether a statement is entailed (theorem proving)
  - for entailed instances of a statement schema (query answering)
  - for variable values that satisfy constraints (constraint satisfaction, CSP),

in particular
find truth values for variables that satisfy a propositional logic theory (SAT)

- for the effects of performing an action in a state (projection)
- for a plan to achieve a goal (planning)
- for models that explain observations (diagnosis)
- . . .

“Intelligent” Reasoning

- Provides a conception of “intelligent inference”
- Sanctions a set of inferences,
  i.e. what can we infer from what we know?
- Recommends a set of inferences,
  i.e. what ought we to infer from what we know?

A Few Historical Remarks on KR Languages

- Early history (late 1950’s to 1970’s)
  - Research on problem solving and natural language “understanding”
  - Many ad hoc representation schemata
  - “Procedural” vs. “declarative” konwledge controversy
  - No formal semantics
- (Dedicated) Knowledge Representation Languages (1970’s and 1980’s)
  - “Semantic” (associative) networks, “Conceptual graphs” (Sowa)
  - Frames: introducing more structure, object-oriented descriptions with inheritance, prototypes
  - Production rule systems: if-then inference rules, situation-action rules, hybrid procedural-declarative representation

  - Qualitative physics: representation and reasoning with incomplete knowledge, qualitative descriptions
  - FOL and extensions: declarative representations, rigorous theoretical analysis, resolution theorem proving, formal semantics