On the fuzzy use of basic terms in informatics

Werner Zorn

zorn@hpi.uni-potsdam.de

1 “Gundstbegriffe” -most basic terms of informatics

For the “Grundstbegriffe” (“Grundst-“ is the artificial german comparative of Grund-. = basic) equal to core basic terms in my vocabulary I have to thank my 4-year participation in the AA1 working party of DIN from 1976-1980, in which a 10-15 person committee, in 3-4 sessions per year, set about creating a new edition of the DIN 44 300 terminology standard, "Information processing terminology", originating from March 1972, and also brought its work to a conclusion in so far as, in October 1985, the draft of the new DIN 44 300 was published as a Yellow Paper. Whilst this achieved the desired White Paper status in 1988, it was withdrawn in 2001 due to lack of acceptance both by experts and the market, and was replaced by foreign-language standard ISO 2382 (English/French).

At that time, "Grundstbegriffe” were understood to mean that hard core of axiomatic basic terminology, which the remaining terminology can rely on as a foundation by using asterisked (*) references, and without which a stable terminological world cannot be built.

In physics these were the terms: force, mass, acceleration in the field of mechanics, in chemistry the terms: element, compound and molecule, and in electrical engineering the terms: charge, current, voltage and field strength, together with the associated conceptual models.

For the "information processing" scope of DIN 44 300 the following were considered to be core basic terminology within AA1

code, data, information, signal, symbol, knowledge and character.

Here though, it was decided not to define the terms "knowledge" and "information" within the standard, but rather to fall back on what is understood through general colloquial usage.

The fact that endeavours to reach uniform definitions of terms and thereby achieve a common understanding of the basic terminology of information processing (synonymously computer science and informatics) have hitherto failed not only in German but also worldwide, has surely less to do with it not being possible to find such a basic understanding than that, up to now, one could apparently live quite well without this.

To examine the quality of the *state of the art* let us consider the following arbitrary selection of examples of common use of language:

- Connectionless communication
- State of a business process
- Instancing of objects
- Throwing exceptions
- Deploying applications
- Overlapping use of *name, address and identity* or equally of *data, information and knowledge processing*
- Running of programs
- Processing of customers
- Sending to classes
- Everything is an object
- Authentication of persons (reflexive as well)

The above list contains not just examples of imprecise or sloppy use of terminology but indeed cases open to erroneous interpretability, in that static and dynamic, active and passive, hierarchical and equal ranking, element and relation, state and property can no longer be differentiated or are utterly mixed up.

In the analysis of the causes, the observations of two precursors of information technology are helpful, namely:

> *"Information is information, not matter or energy"*, Norbert Wiener, 1961

and

> *"Informatics is the engineering science of the mind"*, Karl Ganzhom, IBM, 1982.

According to the first observation, informatics differs in its essential content from the natural sciences, in which every formation of a model is compared to the experiment, with the demand for sufficiently good and reproducible agreement. According to the second observation, the model world of informatics emerges from the creative mind of the engineer, for which the demand for technical achievability and freedom from contradiction count, but not subject to experimental verifiability of the results beyond the technical artefacts corresponding to the model world.

Taken together, both observations permit a characterisation of informatics as a metaphysical discipline, in that it comes distinctly close to philosophy on the question of its abstract concepts. For example, in his paper "The composition of the real world" [Hart1940], the philosopher Nicolai Hartmann (1882-1950) dealt with the problem of forming categories as the appropriate abstractions in a multidimensional fundamental set. Precisely this is presently driving informatics forward in an extension of object-oriented modelling under the concept "Ontologies" as a highly topical area of research.

Against this background, the definitions of terminology discussed below can be divided methodically into two classes: in casuistic, in which all conceivable cases are enumerated and in abstracting – generalising, in which the generalising definition of insignificant differences ("accidents") abstracts and categorises the essential common aspects ("essences"). As regards method, the second approach is always to be preferred, although by far the more difficult, since it amounts to considering terminology definitions not as separate entries in a specialist dictionary but as parts of a global picture, as complete and consistent as possible, of the respective specialised field.

Causes of the problem stated in the title are analysed in the text below, using selected terminology definitions as an example, which as a possible approach to a solution is contrasted with the understanding of terminology that has emerged at the Hasso- Plattner- Institute (HPI) over the past years. In this, the dimensions taken as a basis are stated, with categorical ordering as by Hartmann in the form of pairs of opposites. First of all though, the informatics-encompassing notion of the model is to be considered.

2 Model as conceptual model

Let us consider the miniature version of a BMW Isetta from the 1950s shown in the picture below, or otherwise referred to as a "toy" car.

The following questions arise: Is the part shown in Fig. 1

a) a "toy" car

b) a model of the car of type BMW Isetta

c) or both?

Figure 1: Miniature version of a BMW Isetta [source: Internet]

Here, those questioned may include those that know about the Isetta (I.), and others who have never seen one (II.), e.g. a "Martian". People under I. will probably answer with c) and those under II. (e.g. Martians) with a).

The contradictoriness of the replies as regards the ability to model the observed part being dependent on the person asked, allows one only to conclude that we are dealing here not with an objective property but with a subjective relationship between the real part and the individual observer with their individual conceptual model. We shall make considerable use of this in defining the term "information".

Before this though, the term "model" is to be defined accordingly:

System model: Abstraction of a system under specific aspects

with abstraction = coarsening of the way of observing and aspect = point of view. The definition expresses, as intended, that an observer adopts certain ways of looking at a system, within which he perceives some factual situations but not others. The system concerned may be real or conceived. In the Hartmann system of 12 categorical pairs of opposites, this comes under the 11th pair of opposites, substratum ⇔ relation.

The comprehension of a model underlying the definition implies that *everything* about which we talk, reflect, teach or communicate is always only our subjective conceptual model of *that* which actually exists. For informatics science this would, for example, mean that lectures about

operating-, database- and communication systems etc.

should actually read:

modelling of operating-, database- and communication systems

or that the rider "modelling of ..." could be dispensed with, and, to avoid erroneous interpretations, should then also be dropped compulsorily once and for all.

Section 8 at the end of the article looks into the question as to whether even more pairs of opposites, such as abstract ⇔ concrete or conceived and real, are not inherent in the above model definition.

3 Information and knowledge

In DIN 44 300 the terms "information" and "knowledge" are adopted as common linguistic property and therefore are not defined separately. A definition that mirrors colloquial usage can be found, for example, in the Wikipedia encyclopaedia and runs:

D.0 [de.wikipedia.org/wiki/Information]

Information (from latin: informare 'to shape, give form to through instruction') is potential or actually existent useful or used knowledge. Essential for the information is recognisability as well as news content based on a particular sample of subject matter and/or energy form in space and/or time: the sample used is significant for an observer within a certain context of meaning and, as a result, alters his internal state in a human context, especially his knowledge. More formally, information is the elimination of uncertainty.

In contrast to this, in the specialist literature one finds numerous different definitions, mostly in the form of explanatory usage, from which the following have been selected arbitrarily:

D.1 Albert Endres in [Endr2003]

Information is data that one is able to interpret

Knowledge: Objects and models that we consider true and useful, since they explain the world within and around us, and allow us to deal with things more rationally

D.2 Gerhard Krüger in [Krü2000]

Information: The meaning that a person can ascribe to data based upon the underlying agreements (conventions)

Caution: The notion of information therefore relates solely to humans! Man and machine can handle data, but only man can extract information from data

Recommendation: As far as possible avoid using the term 'information' in precise forms of expression in the lecture

D.3. Manfred Broy in [Hub2003]

Information is what one calls the abstract content ("semantic content", "semantics") of a statement, description, instruction, message or communication. The external form of portrayal is called representation (concrete form of the message)

Let us first consider the term "information" in D.1-D.3:

As regards method, in all three definitions we are dealing with abstracting-generalising with casuistic underpinning in the case of D.3. In all three definitions, the meaning is the main point of the information (assuming that under interpreting, assignment of a meaning to the interpreted matter is what is meant). D.2 attributes the ability to assign meaning solely to humans, whereas in D.1 and D.3 the possibility of machine interpretation is not excluded. D.1 and D.2 support information based on representation in the form of data. D.3 distinguishes between form and semantic content, which is invariant with respect to the representation. In D.3 a sender-/receiver relationship can be assumed, whereas in D.1 and D.2 the source of information remains undefined. In contrast to the encyclopaedia definition, D.0 does not base any of the definitions D.1-D.3 on "knowledge", but rather on their material/energetic representation in the form of "data" or "messages".

Despite the differences, all of the definitions D.1-D.3 draw a distinction between interpretable form and assignable content, whereby the adjectival suffix "-able" expresses an objective property of the associated noun, irrespective of the subjective ability of an individual to be able to also make use of this property. Put another way: provided somebody in the world is able to assign a meaning to a form, e.g. assign a secret number to a secret place, then this secret number is information.

In addition, a series of outstanding questions remain such as:

Re: D.1: Are interpretable data the sole form of information?

Re: D.2 (taking account of the associated remark and recommendation): Why should the term 'information' be avoided when the form of expression is precise? This question arises particularly with regard to the preceding remarks in section 2 about the model as a subjective conceptual model of a rational individual, with his information about the particular system under consideration.

Re: D.3: Is the statement, description etc. truly still information, if information is the semantic content abstracted from the concrete representation? If no, what difference then exists between information and meaning?

Summarising, the assertion regarding the imprecise use of basic terminology set out in the title of the paper is confirmed with regard to the term "information", from which it is not difficult to derive the hypothesis that the respective lack of clarity is inherited in different ways in terminology associated with "information".

In endeavours to overcome the inconsistencies and lack of clarity found, one will have to allow oneself to be guided methodically by the question:

a) What is the highest abstraction, i.e. the most comprehensive conceptual model?

and this calls for clarification of the following specific questions as regards content:

b1) What possible manifestations of information are there?

b2) What relationship exists between "information" and "knowledge"?

If one begins with question b1), then in view of definitions D.1-D.3 this can be put in concrete form as follows: What manifestations of information are there over and above the representation in the form of data (D.1 and D.2) and messages (D.3)?

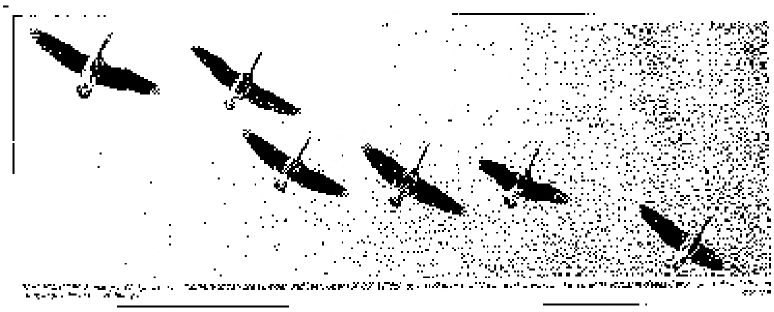
By illustrating the question with the help of Fig. 2 it is evident that:

1. The real world is full of information, whose manifestation we are able to perceive by means of our senses directly or indirectly (i.e. with additional aids)
2. By virtue of our brain we are able to generate new knowledge from existing knowledge and, in some cases, additional perceptions.

With the 2nd observation, wouldn't "knowledge" then simply equate with "information"? The answer is no, but before giving reasons for this, let us illustrate possible manifestations of information by reference to two examples.

If we consider the real world as the primary source of all that is knowable, then we experience this directly by means of perceptible physical quantities, which, in contrast to non-perceptible quantities, are indicated as signals. Thunder and lightning are such signals just as are the flight paths of wild geese in Fig. 2, which represent neither data nor messages. Depending on their source and method of observation, the value progressions of signals may be continuous or discrete in time and space.

We are able to perceive the thunderstorm with thunder and lightning, and it thereby presents us with the information that it is taking place. With further observations we can gain additional information about the precise instant in time, duration, energy released etc. The thunderstorm does not have a meaning that goes beyond this ("Thor is angry", "Thunderstorm in May, April is over (in german: vorbei)"). Correspondingly, the wild geese in Fig. 2 do not send us any message, but provide the



astute observer with "unlimited" information.

Figure 2: IN FORMATION wild geese fly over the Rheiderland between Emden and Leer

[Source: BNN – Badische Neueste Nachrichten 2001, Photo dpa]

For many years Egyptian hieroglyphic scripts could not be interpreted, until finally, in 1799, when the discovery of the Rosetta stone supplied the key to deciphering them – which J.-F. Champollion succeeded in doing in 1822. Were the hieroglyphic scripts information before or only after the discovery of the Rosetta stone? They were, of course, information before the stone's discovery – in contrast to the imprints of chicken's feet found in Assyrian clay tiles, which for a considerable time were similarly taken to be ancient scripts.

The range shown led to the following definition

Information: Knowable

Siegfried Wendt, 1989

which understands "information" as the superset of all

discoverable, ascertainable and conceivable

and therefore includes all existing, as well as any potential further, knowledge [Wendt89].

Let us consider a few more illustrative examples on this point:

the contents of all the books in the Chinese National Library in Beijing

the conception of the world of philosopher Jürgen Habermas (*1929), who is still alive

the diameter of all rocks > 10kg on the dark side of the moon

the biochemical laws of Darwin’s theory of evolution

Accordingly, not information are

the contents of the Library of Alexandria (can no longer be discovered)

the knowledge of the dead (no longer ascertainable) the songs of the planets (no longer conceivable)

the boundlessness of the universe (not yet conceivable)

In contrast to this, reversing the statement by Norbert Wiener, the following are not information and thus not knowable:

Everything material/energetic

as are similarly of different nature:

Everything emotional (perceptible) and transcendental (believable)

where the limits between the knowable and unknowable have moved over the course of human history. Information about what is credible or perceptible can indeed be communicated and thus transferred securely, but not the belief or feeling itself.

The scientific field itself can serve as a further illustration for the definition "Information = Knowable". Its activity – abstracted – consists in nothing other than the continual determination and evaluation of information as knowable. Here, the potentially high degree of compression of information in obtaining knowledge through cognition (= recognise) should be pointed out. This is easily substantiated using Kepler's laws on planetary orbits. This mechanism is clarified further in the following remarks.

The definition of "knowledge" still remains. Point of departure is the difference between "knowable" and "knowledge". Linguistically ...able" expresses the potential (see above), which is intrinsic to the matter under consideration, but is lost once use is made of it. A meal on a plate is "eatable", a meal in the stomach is "eaten". Applied to "knowable" this means that the once known has to lose the potential of being "knowable". This should be taken account of in the definitions and gives rise to the following observations:

I. In contrast to the meal (or any other material/energetic system) "information" possesses the property of being know"able" for any arbitrary number of individuals, without losing this property if any one individual uses up the potential. This unique (!) property of information is found again, for example, in the term "communication", which expresses that the content can be shared by two or more people without one or the other losing out. In other words: information remains "knowable", even if a single individual has extracted "knowledge" from it, provided this information is still available for others. If, in the course of obtaining knowledge, the information utilised were to be lost, e.g. through physical destruction or by being forgotten, then the previously knowable would also no longer exist.

II. With people (or equally with other rational living things) a distinction is drawn between the "knowledge", which is stored in the depths of their brain and the "consciousness", which encompasses that part of knowledge that is present in thinking. The former is already conceived, the latter are the actual thoughts.

If one goes along with this conceptual model, then the knowledge stored in, and retrievable from, the brain can in turn be understood as knowable = information (ascertainable) for the person knowing. With his own thinking he accesses this and possibly generates new knowledge (conceivables). For others the knowledge of a person is equally knowable = information (enquirable), providing the person knowing is prepared and able to communicate his knowledge to the person enquiring.

On the basis of this conceptual model the following definition can be made:

Knowledge: Conceived

The question is how it can enter there from outside (corresponding to the transition from "eatable" to "eaten" through "eat"), which is dealt with in the following section. Before this though, two more concluding remarks should counter any possible misunderstandings or contradictions:

1. "Knowledge" are subjective subsets of all information with the relation ("known by") to an individual. "Knowable" on the other hand, is the essential attribute of all information, known or so far not known by anybody.
2. The usual statement under general colloquial usage:

   "The libraries contain the knowledge of the world"

ought to be reworded as

   "The libraries contain information (knowable) about the world"

which expresses that

a) Knowledge is to be differentiated from presented information

and accordingly

b) The acquisition of knowledge from information is to be achieved by each individual themselves intellectually, so that the relation "known by" can be established .

That this, in turn, is done according to individual conceptual models was explained in section 2. This once again emphasises that everything that we designate as machine knowledge exists always only in our conceptual model, and only becomes actual knowledge when once again it has been thought by a person.

4 From signal to knowledge

"Information" and "knowledge" are categorical terms of our conceptual world, which are contrasted by the material/energetic quantities of the real world (see Fig. 3). The real world is represented by physical quantities, whose manifestation we perceive by means of signals. Here, a signal is defined as follows:

Signal: perceptible value distribution in space and time

The perception of the manifestations of signals may be direct by means of our five senses or indirect with the aid of measuring equipment.

The extraction of knowledge from information as perceived signals is explained with the aid of Fig. 3, for which we can imagine a thunderstorm:

Fig. 3 shows the relationship between the material/energetic real world and the informational human conceptual world, using the example of experiencing a thunderstorm: thunder and lightning are perceptible continual signals in the real world, which as a perceived signal represent information in the human conceptual world, i.e. knowable about the natural event experienced. If the person concerned has never experienced a thunderstorm, then in his signal perceived as information he will not associate any knowledge in the form of a discrete name "thunderstorm", but his stored knowledge will merely consist of "amorphous" recollections of separate prominent claps of thunder and flashes of lightning. As depicted in Fig. 3, he will presumably have additional thoughts about the experience, which may be of discrete form, e.g. "What was experienced seemed to be somewhat dangerous". Also shown in Fig. 3 is a mapping of the continually perceptible signal to a discrete one, e.g. in the form of an assignment of a discrete value 7.5 on a thunder scale that is open upwards. This mapping assumed here within the real world, like that which also takes place in a digital computer, for example, merely removes from the human information processing system one recognition (=abstraction) step, as Figures 4 and 5 below are meant to illustrate.

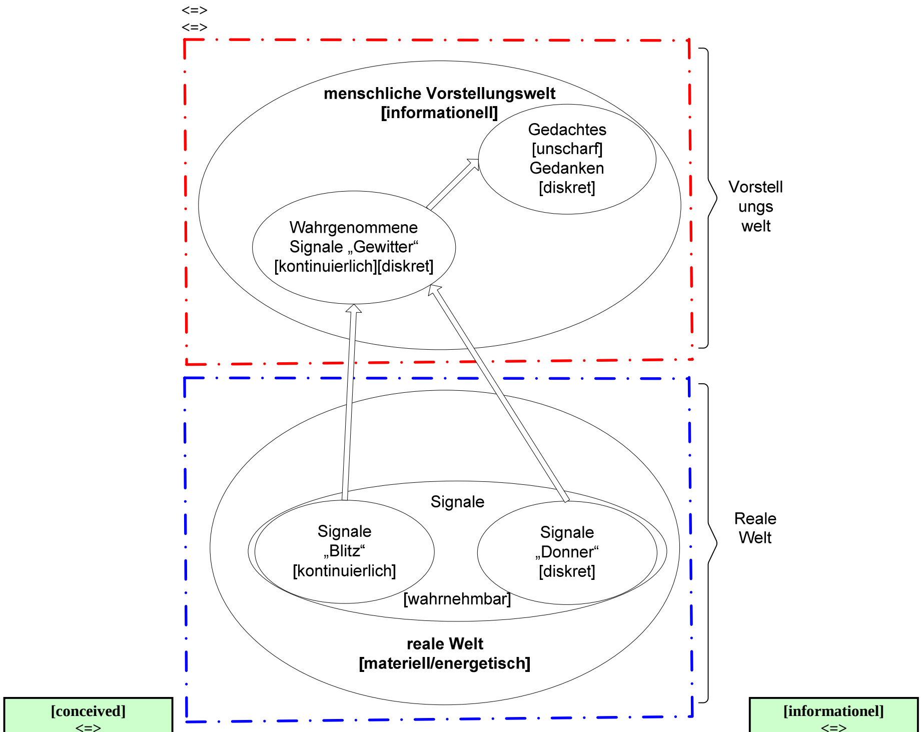
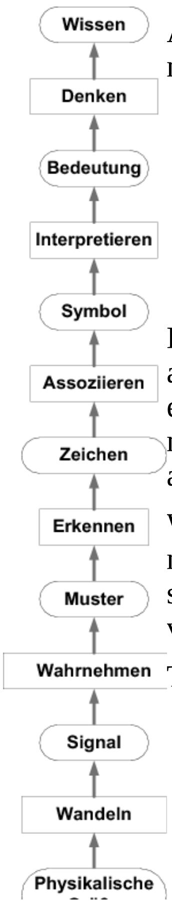


Figure 3: Human conceptual world vs. real world



Against the background of the "Thunderstorm" case study, consider the following definition made in the 44 300 standard:

D.4 DIN AA1 [DIN44300-85] **Signal:** the physical representation of messages(*) or data(*) **digital Signal:** a signal, whose signal parameter(*) represents a message(*) or data(*), which consists only of characters(*).

D.4 totally ignores the physical nature of signals, and instead considers them solely from the aspect of their useful application for representing information. This is rather like seeking to explain electromagnetic waves as nature's own invented modulatable carrier for transporting messages. That signals are themselves information ("knowable" like thunder and lightning) is also ignored.

Whilst characterising the signal as representative does not mention its perceptibility by human or machine, it is implied. Abstraction ensues from the complexity of the processing operations from signal to knowledge and vice-versa, which constitutes an inappropriate coarsening of the way of viewing, as will now be clarified in the text below.

The processing operations that take place on the way from signal to knowledge can be

Figure 4: Mappings of the real world to knowledge
(oval = memory with content, rectangle = operation)

understood as mappings of source and image sets, as illustrated in Fig. 4 below:

The physical quantities as a replica of the real world need to be measurable, so that by measurement and conversion, a perceptible value progression can be generated from this for a particular segment, which we are able to perceive directly or indirectly and thus designate as a signal. The observed real world may, for example, consist of a lump of uranium ore or a fixed agglomeration of copier powder on a piece of paper or a symphony orchestra playing at the moment. We can perceive the first of these audibly by the discrete ticking of a Geiger counter, the second visually as a continuous 2-dimensional grey-scale distribution, and the last one audibly as the superimposition of numerous sound waves. Through perception the signal becomes a model. This mapping may be combined with complex (pre-)processing of the signal, e.g. in the form of rasterisation of grey-scale signals or an instrument-specific filtering of certain frequencies. Through recognition, a pattern is assigned to a character category, e.g. the grey-scale distribution in the form of an originally written two recognised as 2 and the category assigned to the number 2. In colloquial usage, pattern recognition and character recognition are often used synonymously, sometimes with unclear assignment of source and image sets. Naming generally occurs simultaneously on recognition to differentiate the recognised category from possible others. With a recognised character, various categories or even specific objects can be associated (=presented) in the next step, depending on the context, e.g. with the 2 a TV program, school mark, a railway ticket class or a natural number. If the associated carrier is a meaning, we call this a symbol. Interpreting is synonymous with reading symbols or symbolic structures with meaning as a result, where we would have reached the category of the conceived = knowledge, saved in memory from where it can be recalled.

Fig. 4 shows just one path from the signal to knowledge. In fact there are several paths and, in particular, structural mergers at virtually all levels, as Fig. 5 below is meant to illustrate.

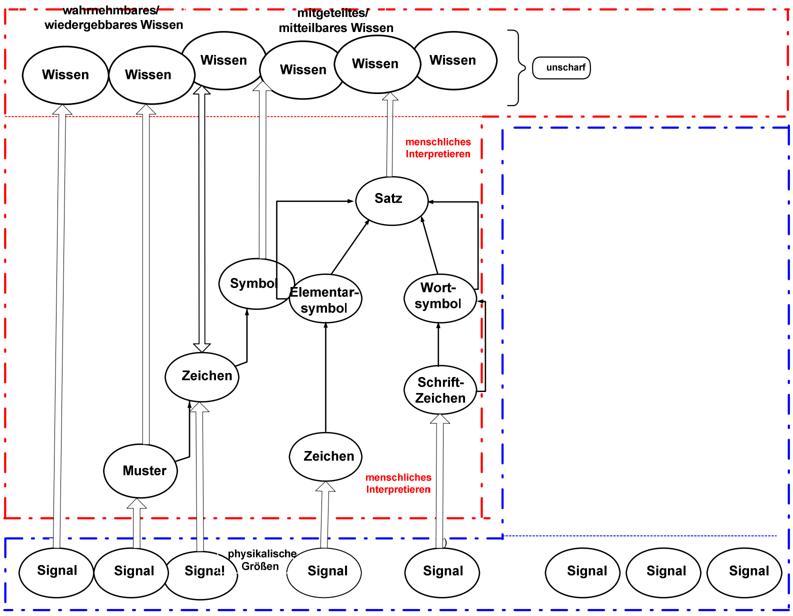


Figure 5: Transitions from signal to knowledge

Here, the heavy arrows signify transitions between the subdomains defined by the pairs of opposites. The large areas bordered by a dot-and-dash line differ in pair of opposites *conceived real* synonym *informational* ⇔ *material/energetic* synonym *human* ⇔ *machine*. The informational section differs by *form* ⇔ *content* synonym *formulable* ⇔ *formulated* synonym *knowledge* ⇔ *formulated information*. The middle area bordered by the dotted rectangle differs from the lower areas in the attributes *discrete* ⇔ *continuous*, and the upper area of knowledge differs from the other areas of the real and informational world by the attributes *formless (=amorphous, structureless)* ⇔ *form-containing* synonym.

Under this categorisation, the terms to be defined can be furnished with the following attributes:

Knowledge(=conceived) [formulable, amorphous]:

- Created (= creative thoughts)
- Interpreted (=interpreted) [formulated, discrete]:
  - Symbol(structure)/Word(symbol)/Sentence (=interpretable)
    - Pattern/Character: [perceived, discrete]
      - Signal in the real world (see below)
- Associated (=associated)
  - Pattern/Character [discrete]
    - Signals in real world (see below)
- Perceived (= seen, heard, touched etc.)
  - Signals in the real world (see below)

Real world [material/energetic]

- Signal [real, continuous, perceptible]

According to this conceptual model one arrives at the following terminology:

- **Knowledge:** Conceived
  - **Sentence:** symbolic structure interpretable as a whole
  - **Word:** character sequence interpretable as a whole (=word symbol)
  - **Symbol:** associated character, interpretable as a whole (-structure)
  - **Compound character:** character structure recognisable as a whole
  - **Characters:** pattern recognisable as a whole
  - **Pattern:** recognisable signal
- **Signal:** perceptible value distribution in space and time

with

Recognise: to already knowingly have perceived something perceived
Associate: associate an idea with something recognised
Interpret: assign meaning to what is presented, generally in the prescribed manner

A good understanding of these terms is conveyed by the work with early childhood speech development [Rei2005].

The presented extraction of knowledge from signals of the real world in the given knowledge steps should not give the impression that this process is schematic and therefore can be formalised. The important bridging function between two formalisable subdomains ought here to include association, in which natural perception combines with an often open imagination.

The observable signals at our interface to the real world are only manifestations, and the knowledge derived from them is far from being a model of the real world. For example, whilst the sinusoidal voltage and current waveshapes within a tuned electrical circuit are a replication of real internal operations, they are not a conceptual model that we would describe as adequate knowledge about such a circuit. Here, it should be clarified once again that forms and formulations at the interface between the real world and our mind are only imaginary representations, which allow us to even think and communicate intelligently about them in the first place.

For comparison, let us recall the definition of knowledge from D.1:

From D.1 Albert Endres in [Endr2003]

Knowledge: Objects and models that we consider true and useful, since they explain the world within and around us, and allow us to deal with things more rationally

The real world considered so far as the original source of all information and thus also of all knowledge does not communicate anything to us. It is simply just there. For that reason we have so far used the term "signal" for perceptible physical quantities but not the term "to signal", since to signal would have a participant as prerequisite, who wants to influence a possible observer, e.g. by communication with the latter. This will now be dealt with in the next section.

5 Communicating knowledge

As normally understood, the communication of knowledge has transmission of information as its prerequisite. The question arises as to how far the hitherto defined terms are compatible with this understanding, and which additional terms are possibly required.

In a sender-/receiver relationship, information as "knowable" means that the sender has "knowledge" at their disposal which constitutes "knowable" for the receiver. The individual contents of communicating knowledge can be specified with the help of the following pairs of terms:

conceived / knowable
presented / presentable
explained / comprehensible or
communicated / experienceable
formulated/ interpretable
employed / associable
formed/ recognisable
created/ perceptible
transferred / receivable

Correspondingly for the operations, the complementary pairs of terms ensue, as depicted along with the contents in the layer model of Fig. 6. For comparison, the layer model in Fig. 4 for acquisition of knowledge from information about the real world is placed alongside.

The number of hierarchical layers, the drawing up of borders between subhierarchies and the terms for operations and contents cannot be mapped adequately by a single abstraction model for the diversity of possible application spheres. However, this is not the objective of the present essay. Rather it is a matter of consist terminology definitions for the respective drawing up of borders between the previously dealt with central basic terms "knowledge" and "information", as well as the "data" and "code" terms still to be defined in the next section.

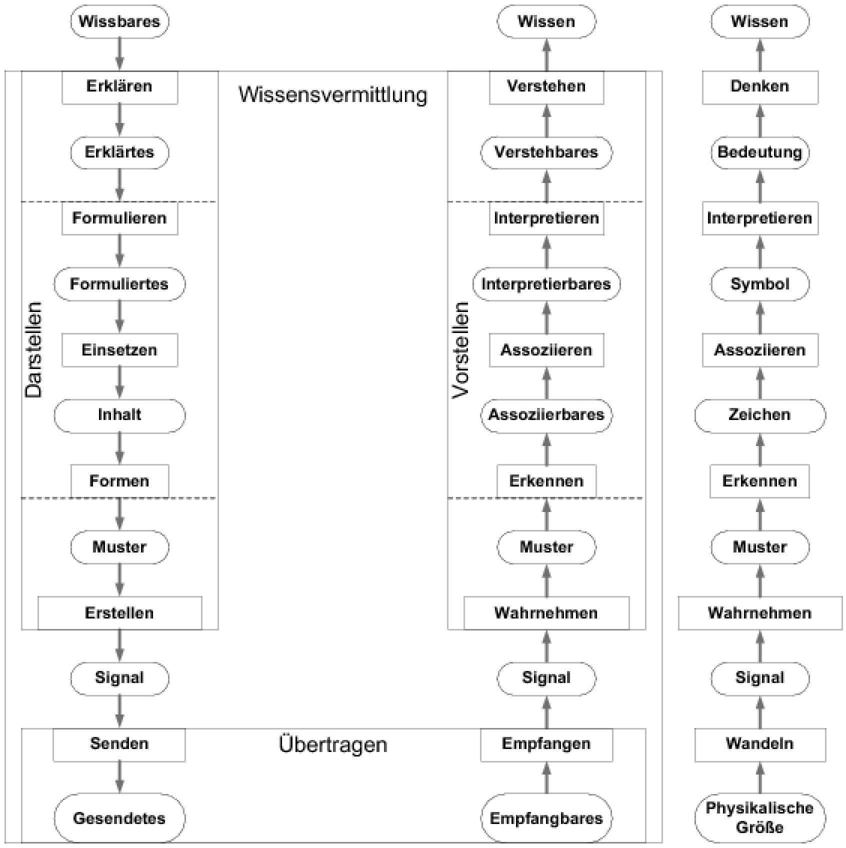


Figure 6: Knowledge conveyance vs. Knowledge acquisition (see Fig. 4)

(oval = memory with contents, rectangle = operations)

Against this background, the following explanations are to be given with regard to Fig. 6: the entire process of conveying knowledge can be roughly divided into the phases:

- Knowledge-side explain/understand [informational, amorphous]
- Linguistic formulating/interpreting [informational, discrete]
- Physical creating/perception [informational, continuous]

summarised in Fig. 6 under

present/understand

with specific demarcation differences in the upper and lower areas, marked by the broken lines. For example, a sentence such as "The moon is made of green cheese" can be interpreted correctly both syntactically and semantically, but its meaning cannot be understood. A painter, for example, applies to canvas the pattern formed in his imagination, whereas a journalist sends off his article to the composing room as a text file. The last phase of

transmission [material/energetic, continuous]

takes place via a physical medium in every case.

According to the conceptual model of Fig. 6 we can ascertain that the enormous amount of what is generally understood to be information forms only a subset of the equally "knowable" very comprehensively defined information, which for differentiation according to Fig. 6 can be designated and defined as follows:

formulated Information: interpretative Knowable

In Fig. 6 this has its source in human knowledge, provided that this "explainable" is also "communicable". What is the situation now with knowledge that is no longer communicable? Is this also still information?

After suffering a stroke, a revered and highly intelligent violinist and lady friend of mine was able to formulate only a single sentence "It is a meadow " and as confirmation that she was correctly understood in what she wanted to express by this, the word "exactly". As previously, she still had great knowledge at her disposal but was only able to communicate this with the greatest of difficulty. In this case, one impinges on the border region of what we are to understand by "communicate" as well as the key question as to whether we want to denote "knowledge" as "information", even if this is not outwardly communicable. The second question can be very easily answered wit "yes", for internally even man takes on the role of a questioner and delves into his memory, just like an outsider in the hope of finding the required information there.

The remaining final question to be posed here is to what extent and, if necessary, how the

information formulated from knowledge

differs from the

information obtained from perception of the real world?

The answer is "In principle not at all", because it is the case with perception that, already at a lower level, formulated information is also produced from the continuous signal in the form of a pattern, e.g. consisting of regularly arranged grey-scale pixels or the discrete amplitude values of an audible oscillation.

After this boundary consideration, we shall now return to the initial concepts of "information" and "knowledge" and state the following ordering that is now possible:

- Information (=knowable) [informational]
  - Knowledge and thoughts[amorphous]
  - Formulated information [discrete]
  - Signals [continuous]
- Non-information ("Not knowable")
  - Real world [material/energetic]
  - Not ascertainable, e.g.
    - Lost knowledge [non existent]
  - Non enquirable, e.g.
    - Unknowns [non existent]
  - Inconceivables: feelings [emotional], belief [spiritual]

Hitherto, the human model has been utilised for forming the terms. In the following section we shall discuss the extent to which these terms can be transferred to machines.

6 Man and machine

Physics has no mission, no less so the inanimate world. Nevertheless, many forms of expression have remained in our colloquial language from prehistoric and mythological times. Equally, technophiles have formed new ones, which make, do, run, reply, set up services, assign and execute jobs etc. on computers and computer-based systems entirely as if they were dealing with human beings. An example that illustrates this especially well is the sphere of humanoid robots.

In the present context, the question arises as to which forms of behaviour and characteristics of people and machines can be described with one and the same terms, since the same kind of matter is involved, and which cannot, because basic differences exist and must be complied with. For the terms considered here, this question will be discussed with reference to Fig. 7 below.

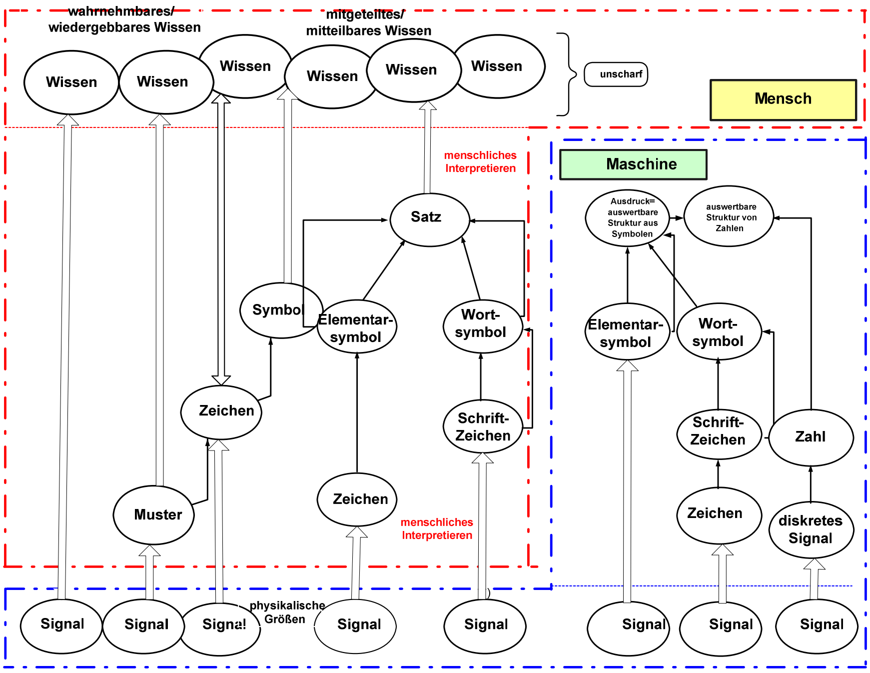


Figure 7: Transitions from signal to knowledge with people and machines

Fig. 7, based on Fig. 4, supplements the categories transferred from the middle region of human signal processing to the region of machine-based signal processing (computers, IT-supported systems, robots and the like).

The interface between man and machine is the real world depicted in the bottom area with the signals perceivable there at the user interface, in the meantime prepared very "human-friendly". Were one to open the computer to trace back the signals at the user interface, one would again find only signals, but this time far less human friendly and not the example, characters and interpretable symbolic structures depicted in Fig. 7. Derived, therefore, is the terminology of our conceptual model, noted down in the right middle region, and were one not to have these due to being a Martian, one would perceive them as nothing other than mysterious signals.

Objection could be raised at this point that the same also applies for humans, since when one opens their head, one finds only signals in their brain and no abstractions at higher levels. This is correct, except that humans are so far the only system that is able to think about its actions and to provide information about what it is actually doing when it does something. Put another way, after initialisation the human requires no outside interpretation in order to know what it is doing.

In contrast to this, a computer is unable to think about itself, and when doing something is able to provide information about what it is supposedly doing, only through predefined messages ("For they know not what they do"- reverse translation of the German movie titel “Denn sie wissen nicht, was sie tun” from “Rebel Without a Cause”). The difference here can be captured in simplified form through the terminological pair of opposites "actually do" and "supposedly do", which expresses what one is "intrinsically" in contrast to what another has "stated". This brings us back to the problem of an adequate conceptual model dealt with in section 2. If we stay with the example of the humanoid robot as a human-like covered construction of chips, motors and rods, then, after what has been said, would this be our conceptual model of a model of a human? With this recursion we would presumably be not only bewildered conceptually but also begin to waver from previous basic notions.
Let us therefore return again to the Isetta toy car: this was a model of an Isetta only in our conceptual model, as a real system, though, a toy car that has nothing in common with a real Isetta. The robot is therefore a real system, which in our conceptual model is the model of a human and in fact also possesses numerous capabilities that come close to those of a human. We can therefore describe this robot as a "human simulator" and use the terminology of the corresponding human area of application to describe its characteristics and behaviour. Why we have described this system as a "human simulator" and not as a "simulation model" will be made clear in section 8 *Abstract versus concrete?*.

We can therefore say, for example, that the hotel service robot is currently taking breakfast to room 17, which the married couple, Mr and Mrs Müller are booked into. Caution is advised if the concepts used in the model world could be more powerful than the capabilities of the real system. If, for example, we say that the robot knows that the Müllers are in room 17, but it has only access to the hotel reservation system, the room coordinates and a calendar plus clock, but has no possibility of authentication, then we should formulate the material circumstances differently. Perhaps this was meant in the recommendation for definition D.2 in section 3.

7 Data and coding

Computers were invented and built to relieve people of time-consuming routine mental work, namely the compilation of tables of logarithms, the control of looms, the evaluation of election results or static computation checks. Here, the subject matter to be processed was supported with the addition commonly used since time immemorial "data ..." for "submitted on...", which then mutated to "data" and extended by the attribute "electronic" evolved to the core term of the abbreviation EDP for machine calculation. What is inherent in the term right from the outset is the property of being "recognisable", which is synonymous with "distinguishable", "countable" and "discrete" [Wendt89]. These are terms and abstractions that can be applied in the same way to objects in the human imagination, such as their representation within discrete processing machines such as digital computers.

In our conceptual model we have ascertained at the uppermost level of observation that information, regardless of its origin in amorphous human knowledge or in continuous real world signals (clarified by the example of extracting and communicating knowledge), can, for the purposes of processing, be mapped onto discrete representations in the domain between:

Amorphous represented knowledge => discrete represented

Continuous signals => discrete represented

In this view, we should not allow ourselves to be led astray by the notion that the mappings from the upper layer of knowledge are the intellectually demanding ones, and those at the lower signal layer, the primitive ones, for which rather than "discrete representation" the term "code" would be appropriate. One can study this taking the example of replication of audible signals in a MP3 representation, automatic announcement of the water level of the Neckar river at Plochingen/ Germany, or many others.

The notion of "discrete representation" as a mapping of both amorphous and continuous in discrete, with humans as with the computer, has led to the following abstraction and definition of terms:

Data: set of discretely represented information

Here, "discrete representation" is an abstraction, which ranges from linguistic formulation at the 'top' level through to intelligent sensing at the 'bottom' level.

In contrast to these generally complex mappings, each of which has an underlying creative mechanism, "coding" is understood as the mechanisable conversion of already discretely represented matter, usually in elementary 1:1 relationships, which led to the following definition:

Coding: content-invariant mapping of data

The notion of coding of continuous signals is thereby replaced with that of a two-step mapping (see also remarks below on "analogue data").

The embedding of these terms in the previous surrounding terminological world is shown in Fig. 8 below.

The direct coding between human and machine is conceived, but in reality occurs via the signal layer.

Owing to the usually imprecise demarcation of the terms "code" and "data", here the term "code" has been contrasted with that of "discrete representation". With his corresponding pair of opposites, Hartmann thus also did not choose the original- and image sets but the pair "substratum ⇔ relation" alias "element ⇔ relation". According to the definition given here, "data" is considered to be the common abstraction of human/intellectual and machine contents, limited to the subset of "discretely represented information", in which a common abstraction "information" as "knowable", "interpretable" "comprehensible" for machines is disregarded, true to the observation about machines made in the previous section "For they know not what they do". Instead of pointing machine processing in the direction of "artificial intelligence", with comparable work the human is seen here in the role of a data processor, who in some cases similarly does not know what he is doing when transferring numbers from left to right.

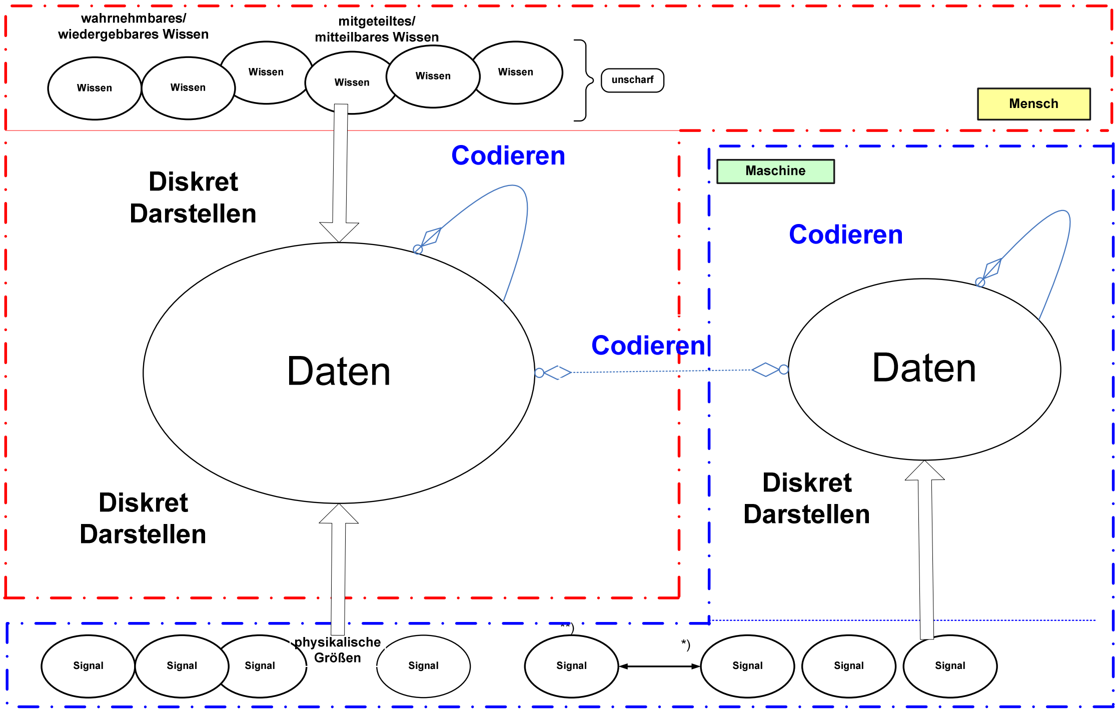


Figure 8: An illustration of "data" and "coding"

After our own definition of terms, let us look again at other definitions:

D.5 DIN 44 300

Data: Characters(*) or continuous functions that represent information on the basis of known or supposed agreements and primarily for processing purposes.

- **digital data:** data(*) consisting solely of characters (*)
- **analogue data:** data(*) consisting solely of continuous functions

From D.1 Albert Endres in [Endr2003]

Information is data that one is able to interpret

D.5 declares data to be information only on condition – similar to D.1- , whereas as it presents itself the figure 2 written on a piece of paper is information = knowable, for anyone seeing it can say "I know there is a 2 written there". Continuous functions as they present themselves are not data at all. The analogue computer considered in this definition displays continuous signals at its output, but can always be mapped to discrete value(ranges) by means of a scale, similar to the way that a clinical thermometer always has a built-in scale without which it would be unusable. To that extent, not only should one not define the term "analogue data" in the standard, but also forbid it.

8 Abstract versus concrete?

The Hartmann system of categories [Hart1940] does not base itself upon the pair of opposites

Abstract ⇔ Concrete

because there are many abstractions for it, and instead uses the pair

Principle ⇔ Concretum.

Therefore, "abstract" continues to be used colloquially, even willingly, as essential for thinking, and "concrete" is taken as essential for the things of the real world. Both notions quickly lead to inconsistencies because man can think concretely in any way whatever: for example, Mozart had a new symphony in his head right down to the last note and needed only to write it down whilst he chatted to his wife Constance.

In connection with the term "model" we have come across the terms: abstraction, system and system model many times and remind ourselves of our own definition:

System model: abstraction of a system from specific aspects

where we understood abstraction to mean a coarsening of the way of observing with "removal" of the unimportant aspects as seen from the respective points of view. At the end of the discourse we raised the question as to whether a humanoid robot should be regarded and designated as a "simulation model" or as a "human simulator", and still have to supply the reasons for choosing the second alternative.

Let us consider Fig. 9 below which is based on the following scenario: rough requirements exist for the functionality of a new system to be developed (e.g. the aforementioned robot). Starting from an initial rough model, this is refined step by step, i.e. "put in concrete terms", until this process has completed its last step of refinement, through which a previous "model specifiable more concretely" has become a "concrete model". A model is termed "concrete" when it is so precisely specified that it can be produced, i.e. "producible" alias "makeable". This modelling process takes place in the human conceptual world, which is characterised by the attributes "informational" and "conceived".

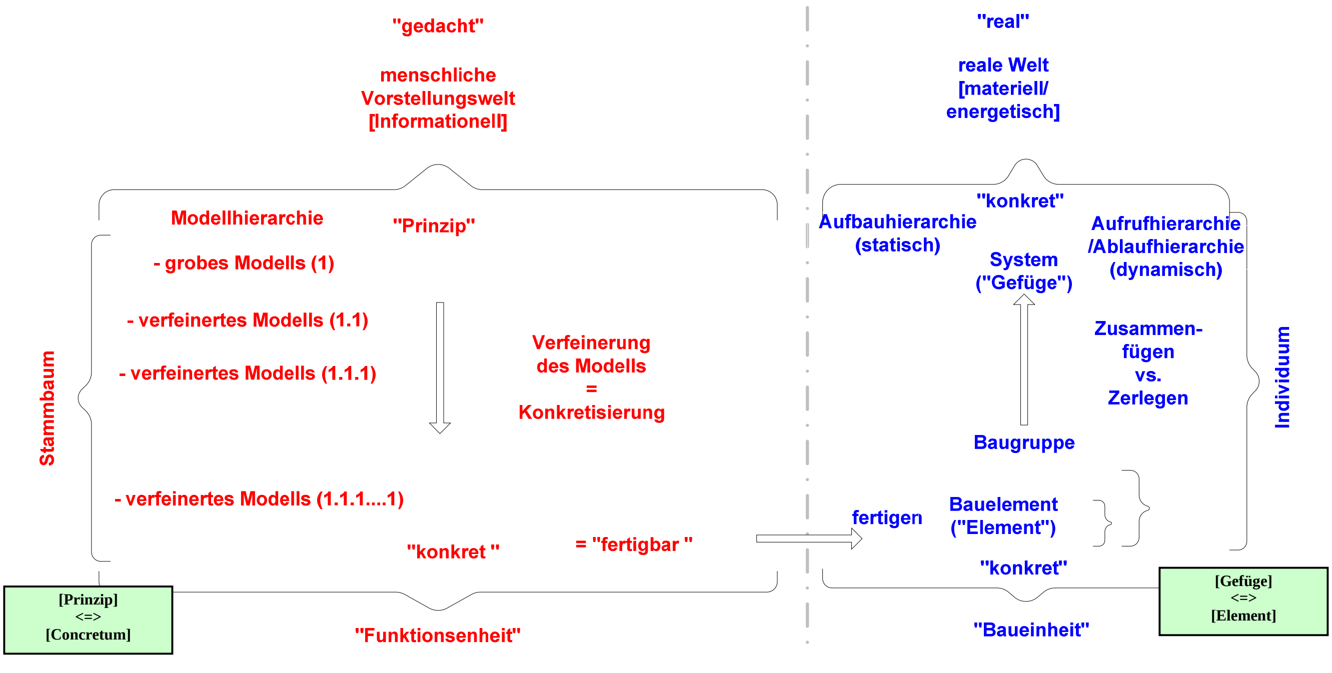


Figure 9: From the rough model to the real system

The production process takes place in the real world, which is characterised by the attributes "material/energetic" and "real". Although the system produced comes into being in individual production stages, it is not ready until the last step has been completed and only then is it actually a system. This process of composition differs from that of decomposition in that the latter is generally understood as a refining of the way of looking at the real system, in which one opens the car bonnet to view the innards and not a mechanical dissection of a car that once worked in its no longer operational separate parts. This is why we also refer to the real system on the right-hand side as "individual" alias "indivisible whole", which loses its essential nature if one takes it apart.

In the left half of Fig. 9 we have a generation (or family) tree of models, and in the right half just an observational unit, namely the emerging or already existing system. On the left we have the conceptual model of a functional unit, and on the right the real building block. For the human conceptual world on the left, Hartmann introduced the already mentioned pair of opposites "principle ⇔ concretum", and for the system emerging or already existing in the real world, the pair of opposites "element ⇔ construction". On the left, a transition takes place from "general" model, with many possible decision paths for refinement, up to specific or "concrete" model, while on the right is the transition from "concrete individual parts" to "concrete system".

Against this background, a definite answer is to be given to the robot-related question: an existing and functioning robot is a real system, which in our conceptual model is a "human simulator". In contrast to this, a "simulation model" would be a third system, namely a real simulation system, which in our conceptual world simulates the robot and not the human, in order to investigate subsequent behaviour in the design phase, for example.

9 Concluding remarks

The text before you, whose descriptions follow the course of the lecture transparencies, turned out to be far longer than the spoken word at the conference, which in this case contradicts the old saying that a picture says more than a 1,000 words. Here, the fundamental problem is evident that it does not suffice to place terminology in a glossary and hope it will be understood correctly. The danger of selective use and reinterpretation is far too great. If one considers terminology standard DIN 44 300, which got bogged down, then the explanatory section that would actually be needed with the present state of the art would be way out of proportion for the imprecise linguistic usage complained about in the title of this paper. To overcome this problem would call for a broadly based discussion, which can merely be suggested here, but not begun.

In retrospect, the title should actually have been formulated positively, since the attempt at precise linguistic usage takes up the major part of the text. The objective was not only to explain the particular terminological structure, but also to point out that even the most elegant structure can waver or be brought down when finally just a single building block does not fit into place. "Analogue data" would have been just this kind of block with the risk of caving in. A further matter not mentioned up until now is particularly worth highlighting: one the one hand, informatics is inundated from the market by anglicisms and, on the other, is dominated within science by foreign words mostly of latin origin. Meanwhile, the rich vocabulary of the German language has increasingly faded into the background, and yet has so much to offer in depth and power of imagination! One needs only consider terms such as Denken, Gedanken, Gedachtes, Gedächtnis or Deuten and Bedeutung, Verstand and Verstehen, Wissen and Bewußtsein, Darstellen, Vorstellen, Herstellen, Erstellen or Kennen and Erkennen, Wahrnehmen, Formen and Formulieren and many others. Anyone of German mother tongue can perform a simple test to see whether, using the corresponding foreign words, he is able to associate the same natural perception as with the aforementioned terminology of his German mother tongue.

Certain things have fallen victim to the fact that the paper was drawn up on the Atlantic coast of France, and, in spite of the Internet, not all sources were to hand, while other matters had to succumb to the fact that, despite the generosity of the publisher and editorial staff, there had to be a final deadline sometime for submission of the paper, which was today, the 1. September 2005. As a result, the proofreading may possibly have missed some errors or inconsistencies.

However, it is assumed that there will still be opportunity for a revised and expanded version, in which all mistakes have been corrected. Any suggestions will be most welcome, for which please accept my thanks in advance.

10 Acknoledgments

I would like to express my thanks to my esteemed colleague and friend Siegfried Wendt, the founding director of HPI, who, with his brilliant abstraction of "information" as "knowable", motivated me to hook into my past involvement with the 44 300 standard, and risk a second attempt on the subject of terminological definitions. Wendt points out a subsequently found confirmation of his concept of information in the case of Carl Friedrich von Weizsäcker in [vWeiz86], where the latter writes: "Information is the measure of an amount of form". We shall also say: Information is the measure of richness of form. Form 'is' neither material nor consciousness, but it is a characteristic of material bodies, and it is knowable for the consciousness. We can say: Material has form, consciousness knows form." With this, a link is also forged to Shannon. Further terminology of the present paper stems from numerous conversations with Wendt, who further motivated me to present the conceptual models that lie behind the terminology, and hopefully also make them understandable. Therefore, with the concept of information in mind, formulated and thus interpretative knowable information has arisen from "knowable", as a possible source of new knowledge on the part of the reader.

11 HPI glossary (extract)

Information: Knowable(*) [informational]

- **Knowable:** ascertainable, conceivable, enquirable
- **Knowledge:** conceived
- **Formulated information:** interpretative knowable(*)
- **Sentence:** symbolic structure interpretable as a whole(*)
- **Symbol:** character interpretable as a whole (*)
- **Symbolic structure:** arrangement of characters interpretable as a whole (*)
- **Character:** pattern recognisable as a whole(*)
- **Pattern:** recognisable signal(*)
- **Data:** set of discretely presented information
- **Coding:** content-invariant mapping of data(*)

Signal: perceptible value progression in space and time [material/energetic]

Literature

[BauFL94] Bauer, F. L.: "Entzifferte Geheimnisse", Springer, 1994

[DIN44300-72] DIN: "Informationsverarbeitung-Begriffe" Beuth, 1972

[DIN44300-85] DIN: "Informationsverarbeitung-Begriffe" (Gelbdruck) Beuth, 10/1985

[Endr2003] Endres, A.: "Die Wissensgesellschaft und ihr Bezug zur Informatik" in GI-Spektrum Band 26, Heft 3, 20.Juni 2003, S. 195-200

[Hart1940] Hartmann, N.: "Der Aufbau der realen Welt" Walter de Gruyter, 1964 (Reprint of the original edition of 1940)

[Hub2003] Hubwieser, P.: "Didaktik der Informatik" (2. Aufl.), Springer, 2003

[Krü2000]Krüger, G.: Folien zur Vorlesung "Telematik" im WS 2000/2001 an der Universität Karlsruhe

[Rei2005] Reimann, B.: "Mutterspracherwerb" in http://www.mutterspracherwerb.de

[SteinRu73] Steinbuch, K., Rupprecht, W.: "Nachrichtentechnik", Springer, 1973

[vWeiz86] von Weizsäcker, C. F.: "Aufbau der Physik", Hanser, 1986

[Wendt89] Wendt, S.: "Nichtphysikalische Grundlagen der Informationstechnik – Interpretierte Formalismen", Springer, 1989

[Zo0405] Zom, W.: Folien zur Vorlesung "Kornmunikationssysteme I" im WS 2004/05, Hasso-Plattner-Institut an der Universität Potsdam