On the Structure of a Global Knowledge Space[1]

Peter Jaenecke

Abstract
A semantic network is defined as the formal basis for a global knowledge space. The fundamental constituents of that space are the knowledge modules. Their structures are explicated in terms of the semantic network terminology, however, they can be viewed also as the semantic for markup languages like topic maps. The modules are identified as elementary units with which a scientific theory can be reconstructed. Theories are represented in the knowledge space as paths. The consequences of such knowledge space are discussed for both epistemology and science.

Keywords: Knowledge module, knowledge network, knowledge space, semantic network, semantic of topic maps, unity of sciences.

> *The totality of all sciences can be compared with the ocean which is continuous and without disruption or partition everywhere, even if the humans conceive partitions and give them names according to their convenience (Leibniz).*[2]

1. Introduction

In scrutinizing of what we call 'scientific knowledge' we will find a lot of things called 'theories' (which mostly do not deserve this qualification), and beyond this we will find a variety of contradictory meanings, hypotheses, short-lived paradigms and so on. We must concede that scientific knowledge consists of more or less isolated knowledge items scattered over a large field of redundant publications. In a former paper a plea was given for constructing real theories,[3] a task which remains still a vital demand. In the meantime, however, the electronic media underwent a rapid rise providing a nearly unbounded storage and retrieval capacity in the world-wide web. Sometimes such a web is seen as the way out of the "knowledge crisis", and, in fact, in many respects the web proves to be a helpful tool for closing the gaps in one's knowledge. But in the long run it is more likely that the "knowledge crisis" will be intensified by the web because of the uncontrolled and incomprehensible variety of offerings. It remains, therefore, a challenge for knowledge organization to provide an instrument with which knowledge can be represented in a systematic and manageable form taking into account both the potential of the electronic media and the systematics of scientific theories. In this paper, the structure of such an instrument is described.

2. Formal Structure of a Knowledge Space

It is assumed that a knowledge space could best be realized by a semantic network. Because in the literature semantic networks are not introduced uniquely, we first give a definition of them and explain features which we will meet again in characterizing the knowledge modules.

---

[1] La representación y organización del conocimiento: Metodologías, modelos y applicaciones. Actas del V Congreso ISKO España, Abril 25 – 27 de 2001, Alcalá de Henares, p. 301 – 312.

[2] LEIBNIZ, Vorausedition to A VI.4, p. 1335.

[3] JAENECKE, Knowledge Organization due to Theory Formation

A *weighted and typed digraph*

$\mathbf{G} = (\mathbf{V}, \mathbf{E})$

consists of a set

$\mathbf{V} = \{\mathbf{V_1}, \mathbf{V_2}, \ldots\}$

of *vertices* and a set

$\mathbf{E} \subseteq \mathbf{W} \times \mathbf{N} \times \mathbf{V} \times \mathbf{V}$

of *edges*. **W** = [0, 1] is the range of weights, and **N** is the set of relation names. A quadruple

$(w, n, v, v') \in \mathbf{E}$

is called *edge from v to v'*; v is called the *tip-* and v' the *tail vertex* of the edge. A sequence of vertices $v_0, \ldots, v_k$ joined uni-directionally with edges of the same name, is called a *path* of the length k which connects the vertices v_0 and v_k with each other.

Definition

A semantic network is a weighted and typed digraph.

Based on the semantic network, the knowledge space is characterized as follows:

Definition

A knowledge space has the structure of a semantic network. Its vertices represent either knowledge modules or names of knowledge modules; its relations represent relationships between two knowledge modules, or between a knowledge module and its name.

The basic unit in the space is the two ary fact which means that A is linked with B by means of the relation R (Figure 1). A and B are vertices of the space, the relation R is a directed, typed and weighted edge. The name of the link is the name of the relation. The number of relations in the space is not limited.

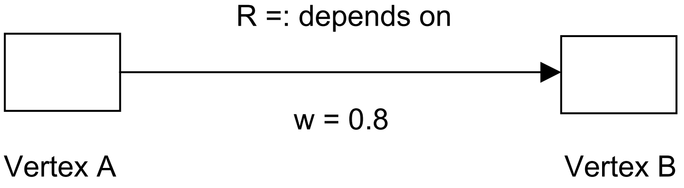

Figure 1: Two ary fact.

The fact consists of vertices A and B which are linked together with the relation 'depends on' and weighted by a factor 0.8; 'depends on' is the name of the relation; the arrowhead points to the tail vertex. Note that A and B are names of the vertices, i.e. the names of what is placed into the boxes.

3. Knowledge Modules

3.1. Rules for Identifying Knowledge Modules

'Knowledge' is an ambiguous concept; its numerous "definitions" are not very helpful for characterizing knowledge modules. Some of the module species are well-known, e.g., definitions, propositions, methods, proofs and so on, however, they are not yet viewed as elementary knowledge units embedded in a knowledge space. We reduce the variety of meanings by the following constraints:

- Knowledge modules are restricted to scientific knowledge, i.e., it must be possible to reconstruct scientific theories by them, and only by means of them.
- Each knowledge module must have a uniquely defined function in the context of a scientific theory.
- Knowledge modules must be elementary units, i.e., the function of a knowledge module must not be achievable by another module, or by a combination of other modules.

3.2. Characterization of Knowledge Modules

There are three abstraction levels in the structure of knowledge modules:

- the module schema,
- the level of module species, and
- the level of intrinsic knowledge models.

3.2.1. Schema of the Knowledge Module

The schema of the knowledge modules includes the general features common for all modules; it can be viewed as a questionnaire having generic and specific feature sections (Table 1).

Name of the module species
Semantic relationship
Role
Quality criterion
[Concatenation operation(s)]
Name of the module species type
Name of the knowledge module
Module core

Table 1: **Schema of a knowledge module.**

The white lines specify generic, the shadowed lines specify specific features. The specific features again are subdivided into type and module name, and into the module core which contains the intrinsic assertion of the knowledge module. Its structure depends on the form of module species in question.

The generic part includes questions for features typical for a module species; it has to be filled out once for each species. The specific part characterizes the structure of a module from that species whose general features are denoted in the generic part. The specific features are subdivided into type and module name, and into the module core which contains the intrinsic assertion of the knowledge module. Its structure depends on the form of the module species in question. The specific part of the module schema has to be filled out individually for each single knowledge module.[4]

3.2.2. Module Species

Defining a module species means (1) instantiate the generic features given in the module schema, and (2) constitute the general form of the knowledge modules of that species (see Table 2 for an example). A species is identified by its *name*. Its *semantic relationship* defines the transmission kind of the meaning; from it follows a substitution rule. The r*ole* of a module results from the purpose it should carry out in a scientific theory. Knowledge needs a demarcation to non-knowledge; the demarcation is done by the *quality criterion*; it consists of conditions which necessarily must be fulfilled. It follows from the nature of the knowledge species that each of them requires an own quality criterion. For some knowledge modules there are *concatenation operations* defined like the sentential connectives AND, OR, NOT etc. for propositions.

Abbreviation	*is a*	Name of the module species
Meaning(short form) = Meaning(long form)	*is a*	Semantic Relationship
Shortening the length of Text	*is a*	Role
Length(short form) < Length (long form) The Term 'short form' must be unique in the space	*is a*	Quality Criterion
		no concatenation operation
Term$_1$	*is a*	Name of the abbreviation type
Term$_2$	*is a*	Name of the knowledge module
Short form	*is a*	Term
$=_{abb}$	*is an*	Equivalence relation
Long form	*is a*	Text

|← Structure of the module species 'abbreviation' →|← Space management information →|

Table 2: Structure and space management information of the knowledge module species 'abbreviation'.

3.2.3. Knowledge Modules

On the lowest abstraction level there are domiciled the knowledge modules themselves, e.g., the definition of 'continuous function' or the regula falsi algorithm. A knowledge module is the instantiation of the generic module form of the species. It is characterized by the *name of the species type*, by its own *module name*, and by the *module core*, which contains the elementary constituents of the module (see Table 3 for an example). The module core varies from species to species; all modules of the same species inherits the general features of the species so that it is not necessary to repeat them by each module.

---

[4] For representing module features, language modules are required like 'term', 'text' etc.; it is assumed here that these language modules as well as the knowledge on their handling are known.

Example: 'Abbreviation QPSK'

Acronym	*is a*	Name of the abbreviation type
Abbreviation QPSK	*is a*	Name of the knowledge module
QPSK	*is a*	Term
$=_{abb}$	*is an*	Equivalence relation
Quaternary Pulse Shape Keying	*is a*	Text

Table 3: Structure and space management information of the knowledge module 'abbreviation QPSK'.

The specifications as outlined in Table 2 and Table 3 are needed especially for the system developer for building the space, but they are laborious when a theory should be reconstructed with them. For the reconstruction "short specifications" like

Acronym 'Abbreviation QPSK':

QPSK $=_{abb}$ Quaternary Pulse Shape Keying

are more practicable. However, the characterizations of a module species as given in Table 2 can be used as a guide for the analyzer of a theory in order to decide, whether a given theory section is a definition, a method etc., i.e., if such a section was identified, e.g., to be an abbreviation, then it must fulfil the features established in Table 2.

Knowledge modules incorporate two antagonistic aspects: compactness and connectivity. Compactness originates from an enclosed thought defined by the features and the role the module plays within a scientific theory. Connectivity makes it possible to integrate them in a space. It entails that they must have also junctions to the "outside world". The junctions are realized

- by the generic/specific terms given in the questionnaire;
- by interlocking a module with other modules by the names which appear in the module in question; e.g. in the module species 'proposition' a proof name must be included which refers to the proof of that proposition;
- by the properties of the mathematical entities like the equivalence relation '$=_{abb}$';
- by additional links, which, e.g., order the knowledge modules in a path for representing a scientific theory;
- by an obligatory 'is a relation'[5] between concepts which serve the space management; for example: an acronym *is a* type of an abbreviation;
- by relations which represent space management information.

---

[5] The 'is a relation' constitutes the relation: 'is an element of a set'; note that 'is a' is identical to 'is an'.

4. Global Knowledge Space

4.1. Unity and Multiplicity

The structure of the knowledge modules was defined in such a way that the modules can be incorporated immediately into a semantic network so that it is justified to say that the space forms a unity within the multiplicity. However, there is realized also a higher-layer unity by the space based on the fact that a single knowledge module shows empty positions like unclear or undefined concepts, which have to be completed by suitable knowledge modules of the same or different species according to given rules. However, the modules used for filling a gap, show possibly themselves gaps which have to be filled, and so on. The result will be a formation of interwoven knowledge modules. A method, for example, describes a procedure with which a defined goal can be reached if used correctly (see Table 4 and Table 5). The reason, why the goal is reached may yet be unknown, i.e., a missing justification forms a gap which has to be closed in the future. Another example is a phenomenological definition (Table 6 and Table 7). It fixes preliminarily a property from which it is known by experience that it must exist, but which cannot be defined exactly at the moment.

Method	*is a*	Name of the module species
Correct execution → achievement of the goal	*is a*	Semantic Relationship
Reaching a well-defined goal		Role
(1) It must be proved that the method leads to the wanted goal if used correctly (2) The requirements for its applicability must be known	*is a*	Quality criterion
Serialization	*is a*	Concatenation operation
Term_1	*is a*	Name of the method type
Term_2	*is a*	Name of the knowledge module
Text_1	*is a*	Problem description
Text_2	*is a*	Requirement(s)
Text_3	*is an*	Approach

Table 4: Structure and space management information of module species 'method'.

<u>Example: 'Regula falsi'</u>

Algorithm	*is a*	Name of a method type
Regula falsi	*is a*	Name of the knowledge module
Computation of a null of a real function f	*is a*	Problem Description
There exists an $x_0 \neq x_1$ so that $f(x_0)\ f(x_1) < 0$	*is a*	Requirement
$x_{m+1} = x_m - f(x_m)/s_m$, where $s_m = [f(x_m) - f(x_{m-1})]/(x_m - x_{m-1})$	*is a*	Approach

Table 5: Structure and space management information of module 'regula falsi'.

Phenomenological definition	*is a*	Name of the module species
Meaning(concept) = Meaning(phenomenon)	*is a*	Semantic relationship
Preliminary fixing of a concept by experience	*is a*	Role
The experience must be generally known, or it must be verifiable empirically	*is*	Quality criterion
		no concatenation operation
Term$_1$	*is a*	Name of the phenomenological definition type
Term$_2$	*is a*	Name of the knowledge module
Text$_1$	*is a*	Experience
Text$_2$	*is a*	Conclusion
Concept	*is a*	Concept to be defined
$=_{pdf}$	*is a*	Equivalence relation
Phenomenon characterized by Text$_3$	*is a*	Phenomenological characteristic

Table 6: Structure and space management information of module species 'phenomenological definition'.

Example: Phenomenological definition 'force'

Empirical definition	*is a*	Name of the phenomenological definition type
Classical force		Name of the knowledge module
Our muscles provide us the qualitative impression that they can be tensed with different intensities	*is a*	Experience
It must be a property in which these different efforts are embodied.	*is a*	Conclusion
Force	*is a*	Concept to be defined
$=_{pdf}$	*is a*	Equivalence relation
property characterized by the fact that muscles can be tensed with different strains	*is a*	Phenomenological characteristic

Table 7: Structure and space management information of module 'classical force'.

Furthermore, each theory, as extensive it may be, must adopt "from the outside" concepts and principles, whose validity can not be proved with the means of that theory itself. The theory also has empty positions which should be closed by other theories, more precisely: by knowledge modules from other theories. These supplementary elements are the filaments, with which a theory is embedded in all the rest of knowledge. Pursuing this approach with consequence, a global knowledge space will result in which the barriers between the individual theories are dissolved. In such a space, individual theories appear now as more or less arbitrarily defined sections, or paths. There exists only one global theory: the knowledge space itself.

4.2. Consistency Constraints

As mentioned above, each module species is characterized by a special quality criterion which ensures the "local" correctness of the space vertices. If the correctness could be proved for each knowledge module, then in a next step it has to be validated, whether all modules are correctly integrated into the space. These tests are syntactic in their nature. Additionally to the quality criteria so-called consistency constraints are needed referring to the quality of the space as a whole. By means of them it can be checked, whether the knowledge expressed in the modules do match. The space should be tested

(1) whether there are junctions which violate the space topology;
(2) whether there are uncomplete knowledge modules or open problems;
(3) whether there are knowledge modules which can be derived from other modules;
(4) whether there are contradictions between knowledge modules.

Test (1) is based on the properties of the relations used in the space. Each relation has certain formal properties like transitivity, even if they are unknown at the time they are introduced. These properties have to be transformed into procedural rules, e.g.: 'a vertex can have an arbitrary number of outgoing relations r, but at the most only one in-going relation r'. Because a space can be incomplete in many ways, there are (2) a variety of tests for incompleteness. Thus, e.g., the properties of a relation can be missing, or there is no proof as yet for a theorem, or a phenomenological definition proves to be insufficient, etc. Depending on the operations defined for each species, (3) inference rules are given about how a knowledge module can be derived from other knowledge modules which must not be necessarily from the same species. Again, a variety of inference rules exists. And (4), the test for contradictions is based on a logical conclusion onto the space. It is an important special case of the deriving test sort, because it tests, whether a contradictory assertion can be derived from the space.

The set of consistency rules is open as follows from the open set of problems, relations and derivation types. All examinations are based on knowledge; more precisely: they are based on knowledge modules which are an inherent part of the knowledge space. These modules are not the subject of the examination, rather they are needed as examination tools. An examination is always "locally", i.e., only a special section of a space is questioned, whereas, in principle, the complete remaining part can enter into the examination procedure. A knowledge space is consistent, if all possible sections of it are consistent.

5. Discussion and Conclusion

The modern electronic storage capacity and the nearly unlimited number of links nurse the dream of a global and world-wide disposable knowledge network, and, indeed, under that name there is a rich offer in the internet today. These networks can be characterized roughly as follows: They are designed (1) as a collection of definitions and assertions of a specific subject domain; based on multi-media tools they form (2) a collection of facts illustrated by pictures, audio and video documents etc., and they are equipped mostly with links to the literature; they are reconstructed (3) from scientific documents by a document description languages, and (4) there is a variety of approaches called 'ontologies'.

Only the latter type of networks bear some similarity to the network suggested in this paper. In the collections mentioned above a systematic order is missing, and networks based on a document description language show a strong redundancy combined with inconsistencies in both the conceptual and the declarative framework; in addition, they do not allow referring and constraint-based validations. There are also differences between ontologies and the knowledge module approach: Ontologies are based on a conceptual framework, whereas the module species are intended as global building blocks for all domains; the semantic of the module species is fixed by their role they play in a scientific theory. As a consequence, knowledge modules will contain closed semantic units like 'requirement', 'approach' (Table 5), or 'experience', 'conclusion', 'characterization' (Table 7). These more extensive units are dictated by the "logic" of a scientific theory, and it seems to be impossible to reduce them to frameworks which are based only on concepts and relations.

Because of its semantic roots, the global knowledge space has also strong impacts to epistemic and scientific conceptions:

- Knowledge is no longer restricted exclusively to propositions in a two-valued logic, i.e., instead of to be true, a knowledge module must fulfil the quality criterion specified for it.
- In the knowledge space, there is no longer a strict separation into different theories, rather, a theory is now a special path in the space. Thus, the universality of the knowledge modules establishes the possibility to dissolve the boundaries between individual theories; in such a way they contribute fundamentally to the unity of sciences.
- Knowledge was proved as yet "locally" within single sciences or subsciences; however, up to what extent can we say that our knowledge fundus is consistent as a whole? Clearly, it is impossible to perform a global consistency check by hand, but a computer - equipped with a knowledge space - could do this work in a systematic way. Without doubt, such a global check would reveal the inconsistency of our current knowledge.
- However, in spite of expected inconsistencies, the space is not useless as a logical system would be, which contains a contradiction. Rather, the space provides tools for identifying weak-points.
- Instead of a "theory dynamic" now there is the space which is in permanently progress. Its dynamic is constituted by new knowledge modules and new links (relations), by closing gaps, and by debugging inconsistencies.
- The knowledge space is a research tool. Installed on a computer, it can be used to go beyond the usual possibilities by deploying the available knowledge as completely as possible. Thus, questionable propositions can be tested on a trial basis whether they are compatible with the knowledge in the space. The space can also serve the understanding of coherence, e.g., by inferring the presuppositions of a given assertion.
- In order to overcome the disadvantages of specialization it is important to create a construction site at which can be worked at the same time, but at different places and without the need for overlooking the complete knowledge.

The work at the knowledge space is still in progress. What is the status? About 20 knowledge species have been identified as yet in analyzing a physical, technical and a psychological theory, namely: classical mechanics, communication theory and a theory on human memory. These theories represent quite different scientific domains, and it is assumed that knowledge modules abstracted from these disciplines will also be useful in other ones. This may be true also for non-scientific knowledge. It is not clear at the moment, whether still more species are required, and whether some of the identified species are redundant. The analysis of theories with the X-ray eyes of a knowledge module hunter proved to be extremely instructive for the author: It is often very hard to assign a special part of text as definition, assertion, explanation, etc.; sometimes one will meet all in one. The theories are built in a not very systematic way; surprisingly that is true also for the classical mechanics. There are unclear concepts, hidden unproved propositions, contradictory concepts, mental leaps a lot: Identifying knowledge species is a cumbersome and laborious task which requires yet a lot of efforts for finishing.

Another unclear point is the realization of such a global space. The topic maps seem to be the best choice,[6] however, as RATH points out,[7] for constraint-based validations, an extension of the ISO standard will be needed. My analysis of scientific theories reveals that one should expect a variety of different validation constraints with a complexity transcending of what is known as yet from literature. It remains a future task to identify appropriate validation constraints and to find universal inference and derivation rules.

Acknowledgements

I wish to thank Alexander Sigel for his helpful hints.

Literature

BOLEY, H.; DECKER, S.; SINTEK, M. (2000). Knowledge Markup Techniques. [URL: http://www.SemanticWeb.org/knowmarktutorial].

JAENECKE, Peter (1997). Knowledge Organization due to Theory Formation, in: Javier García Marco (editor), Organización del Conocimiento en Sistemas de Información y Documentación, Zaragoza 1997, 39-55.

LEIBNIZ, G. W.: Sämtliche Schriften und Briefe = Akademieausgabe 1923ff; zitiert nach A Reihe, Band, Seite. Hier: Vorausedition zu A VI.4, Philosophische Schriften, Münster 1981ff.

PARK, Jack (editor); HUNTING, Sam; BIEZUNSKI, Michel and NEWCOMB, Steven (technical editors) (2001): XML Topic Maps: Creating and Using Topic Maps for the Web. Addison Wesley (to appear September 2001).

RATH, H. (2001). Topic Map Fundamentals for Knowledge Representation, in PARK et al. (2001).

SIGEL, Alexander (2001). Topic Maps in Knowledge Organization, in PARK et al. (2001).

http://www.peterjaenecke.de/wissensordnung.html 27.7.2001/24.1.2007

---

[6] SIGEL; Alexander: Topic Maps in Knowledge Organization.

[7] RATH, Topic Map Fundamentals for Knowledge Representation.