On Navigating in Information Spaces

Dieter Gernert

Technische Universität München, Department of Economics
Arcisstr. 21, D-80333 München (Germany)
Email: t4141ax@mail.lrz-muenchen.de

Abstract

The concept of information is not yet perfectly integrated into our modern science. As a tentative approach the concept of information space is proposed here. An information space consists of elementary semantically meaningful statements (or interpretable structures) together with specific relations between these elements. Relations can either be seen in analogy to the links known from hypertext systems, or they are defined by similarity relations between the elements. Both the elements and the relations can be modified in the course of time, and possible paths through such a space (navigation) are discussed together with their effects on the underlying structures. The mathematics of similarity relations and their context dependence are studied in some detail. The structures proposed here can be modeled on a standard PC; possible applications are sketched.

Keywords: Information space, information dynamics, similarity metric, information retrieval, polymetric space.

1 Are our Fundamental Sciences up to the „Age of Information“?

Every day we hear and read that we are living in the „age of information“. Is this really true in that general form? Are our central and basic fields of science up to this claim? A closer look at physics will reveal that this discipline is still centered around the concepts of matter and energy, whereas information still plays a marginal role and has not yet received full civil rights in physics.

The so-called „information theory“, created mainly by Shannon and Weaver, was originally intended as a theory of information transmission. The floppy and misleading terms „theory of information“ and „information theory“ emerged only later and do not at all reflect the real capacities of that theory. Shannon himself advanced a passionate pleading against the overestimation of the theory essentially originated by himself. More and more - albeit still reluctantly - the idea is gaining broader acceptance that a theory of information which really deserves this name must also account for the semantic and pragmatic aspects of information (for details see Shannon (1956), Gernert

2 A Proposed Change of Perspective

For heuristic purposes a total reversal of perspective is proposed here. We will start from the concept of *„information space"* - rather intuitive in the beginning, but to be given a stronger specification as the work will proceed.

In the ideal case, the result of such an inverted perspective may be a description of our world such that information can be handled in a natural and comfortable manner - whereas the integration of matter and energy into that theory will make trouble. It would be a partial success to find a theory involving „mirror-symmetric complications" when compared with our present situation. This would be an intermediate step towards a unified theory, which in this moment seems to be out of reach (at least in the author's view).

3 Basic Properties of an Information Space

For the present purpose it will be reasonable to start from a rather comprehensive concept of information:

> Information is everything that was, or is, or possibly may be useful or interesting to some receiving system or interpretant.

Such an interpretant may be a human individual, a team, an organization, an animal, or a technical device; the special case of an „autonomous information space" is to follow (Section 4.4).

This tentative definition emphasizes the semantic and pragmatic aspects of information from the very beginning. At least some „minimum semantic contents" (with respect to at least one real or hypothetical interpretant) will be required; only in exotic cases we can have to do with bits and bytes (if e.g. the arrival of „0" or „1" decisively alters a receiving system). On the other hand, the physical representation is considered less important.

Possible types of information covered by that definition include:

1. *„Information on the way":* A sequence of signals just being transmitted in a channel has a chance to reach its addressee, and hence is information in the present sense. It can be valuated under the aspects of telecommunication engineers, and hence the classical Shannon-Weaver theory is included here as a special case. Also any kind of information stored in books, data bases, etc. is subsumed here.
2. The concept of *pragmatic information* defines information by the impact upon the receiving system. Information begins when the channel has ended; information is anything which alters the state or the behaviour of the receiving system. The connection with the definition given above is evident.

3. We also must envisage *potential information (structural information)*, e.g. the information represented by the structure of a crystal, which enables interpretations by crystalographers.

The term „space" as such is not defined in mathematics. Mathematical dictionaries only give references to more specific terms that are formed by combining the word „space" with an adjective or a proper name, e.g. „vector space" or „Banach space". A mathematical dictionary (Naas and Schmid 1961, vol. II, p. 455) presents about 40 such references. An „encyclopedic dictionary of mathematics" gives a circumscription which can hardly be considered a definition: „The term space is used in mathematics for any set when certain types of properties are to be discussed or when it is intended to use some sort of geometrical terminology." (Sneddon 1976, p. 616)

Accordingly, we define an *information space* by a set consisting of *elements* (of a special kind), together with possible *relations* between these element; both the elements and the relations may vary in the course of time.

An element is understood as an irreducible unit which still carries some semantic contents, and particularly has the character of a statement or proposition - no matter whether its physical representation is built up from bits, bytes, characters, words, pixels, or other small parts whatsoever. Examples may be the items of a knowledge base or single mathematical formulas. In both cases there are smaller constituents symbolizing something, but without the quality of a statement.

The mathematical structure of an information space, which is given by the relations existing between the elements, may also depend on the interpretant, and, furthermore, vary with time. Details are the topic of the next sections. Practical applications, which also supply permanent illustrations for most of the mathematical specifications to follow, are given by modern *hypertext* and *information-retrieval systems*, where the user is guided from one element to the next one as governed by a link, or from one element rated as useful to a „similar" one.[1)] These two examples show that information spaces are already „implemented"; but since it is the purpose of this paper to study information spaces under more general aspects, these analogies should not be overemphasized.

---

[1] In information-retrieval systems rather frequently a measure of the similarity between the search statement and an item of the document file is used. Some more advanced systems also make use of the similarity between two items: those items will be additionally displayed to the user which have a sufficient similarity with other items judged as useful before (see e.g. Kowalski, 1997, p. 152-157). This technique, termed „similarity retrieval" („similarity-based retrieval", „similarity search"), is widespread in applications to chemistry and to image data analysis.

4. Mathematical Structure of an Information Space

4.1 Link Structure

Any element of an information space may carry „pointers", each of which leads to a certain other element; as customary we use the term *link*. With the link structure of the space we have arrived at the simplest case of a *navigation in an information space*, which is simply enabled by following sequences of links.

Since a link has a direction, the set of all elements together with all links forms a directed graph, whose mathematical structure will be regarded later. If links can be inserted, altered, and deleted, the link structure will be time-dependent. Links can be given different weights resulting e.g. from ratings assigned to their targets or from the frequency of previous passages. For special applications we may distinguish between two types of links:

1. *Standard links* lead to a predefined target, without checking side conditions.
2. *Non-standard links* may guide a navigator to one or another target depending on parameters - e.g. the user of a hypertext or retrieval system should not be directed to an element already visited before.

4.2 Cluster Structure

As a consequence of the link structure, there are subsets whose elements are closely interrelated (mathematical formulations are possible, e.g. by use of the tools outlined in Section 5); these subsets will be called *clusters*. In the context of information retrieval, terms like „association domains" are used. Clusters can partially overlap, and they can be combined to „clusters of clusters", and so on, until finally a multi-stage part-whole hierarchy is built up. The cluster structure can vary with time, too; for processes of stabilization or decay of clusters see Section 4.4.

4.3 A First Look on Similarity Metrics

The space of our everyday experience is a polymetric space: there are different metrics for pedestrians, for cyclists, etc. Both the optimal path and the time needed to get from a starting-point to a destination are different. Although this is scarcely made explicit in physics, we have multiple metrics also here, as defined e.g. by the propagation of light and of sound within the the same region.

In an information space a metric can be introduced by a *dissimilarity function* or *distance function* $d(x,y)$ which is defined for any two elements x and y and which has the usual properties of a *metric*, as specified by a well-known system of axioms:

M1: *non-negativity:* $d(x,y) \geq 0$
M2: $d(x,y) = 0$ *if and only if* $x = y$

M3: *symmetry:* $d(x,y) = d(y,x)$
M4: *triangle inequality:* $d(x,y) \leq d(x,z) + d(z,x)$

If these axioms are fulfilled then $d(x,y)$ is a *symmetric* metric; if axiom M3 (symmetry) is not valid then $d(x,y)$ is an *asymmetric* metric. The latter case can be relevant here: as a consequence of the link structure, the distance defined by a series of links showing the optimal route from x to y generally will be different from the distance defined by a counterpart leading from y to x.

It is often easier to speak in terms of the *similarity* $s(x,y)$ between x and y - a large distance or dissimilarity corresponds to a low similarity, and vice versa; the mathematical transitions in both directions are trivial.

It is always possible to define a similarity function $s(x,y)$ which accounts for the internal structure of the elements x and y (for details see Section 5). Inevitably, „similarity" is a *perspective notion*; its meaning depends on the interpretant, the historical context, and the purpose pursued with each individual similarity measurement. The procedure which leads to a mathematical formulation of $s(x,y)$ necessarily takes that context into consideration.

Metrics based upon similarity or dissimilarity have proved their utility in modern information-retrieval systems, where the user can receive hints to entries which are „similar" to those rated as useful before, although the underlying similarity definition generally is rather simple and not well adapted to the individual context.[2)]

4.4 Structures Emerging in Dynamical Processes

Navigating in an information space leaves marks and traces. We can distinguish two principal forms of dynamical processes:

1. There is an external user or external observer who travels from one element of the space to another, guided mainly by links, clusters, and similarity measures, and regularly modifying the state of the space by new weights assigned to links, revised definitions of similarity, etc.
2. In the case of an *autonomous information space* there is no intervention from outside. Nevertheless, there are interactions between elements, and the contents of elements, the strengths of links, and the cluster structure can vary with time, as well as the similarity measures that are permanently adapted to new situations.

Some esssential emerging structures and processes connected with them can be characterized as follows:

---

[2] For two fundamental strategies, a similarity measurement based on feature vectors or an underlying structure (described by graphs etc.), see Section 5.2.

1. *Cluster formation:* New clusters or cluster hierarchies can be formed, e.g. through increased weights assigned to elements or through revised similarity measures.
2. *Path formation:* Frequently used routes may be „stabilized“ - where branching is possible, one of the alternative continuations is given a higher preference; shortcuts can be formed.
3. *Representation tuning:* The style of information representation may be adapted to new requirements.
4. *Information tracing:* There is a chance to reconstruct along which way some information has arrived.

Paths and clusters may win or loose relevance in the course of time, and accordingly they can be stabilized or destabilized. A sufficiently complex substructure of an information space can take over the role of an interpretant and hence govern the interpretation of entries stored elsewhere or rule the formulation of similarity measures.

5 The Mathematics of Similarity Metrics

5.1 Internal and External Similarity

In mathematics the term „similarity“ can have two distinct meanings depending on the context. No mentioning of this fact could be found in literature, nor a distinction in terminology. Therefore the new terms

„internal similarity“ and *„external similarity“*

are proposed here. Later on, it will be found that the two disparate meanings can be reconciled under a unifying concept.

Examples for *internal similarity* are given by similar triangles and by similar matrices. Formally, we have a decomposition of a set **A** into a family of non-empty, pairwise disjoint subsets (such that their join is exactly **A**); these subsets are the equivalence classes defined by an *equivalence relation* S with the following properties (for all $a,b,c \in \mathbf{A}$):

IS1: *reflexivity:* $a\mathsf{S}a$

IS2: *symmetry:* $a\mathsf{S}b$ implies $b\mathsf{S}a$

IS3: *transitivity:* $a\mathsf{S}b \wedge b\mathsf{S}c$ implies $a\mathsf{S}c$

By way of contrast, *external similarity* can be illustrated by a series of comparisons between patterns. If $x\mathsf{S}y$ stands for the fact that the two patterns x and y are rated as „similar“, then we can have $a\mathsf{S}b$ and $b\mathsf{S}c$, but $a\mathsf{S}c$ does not necessarily hold. To give another example, $a,b,c,$ ... may denote alternatives for human actions (in mathematical utility theory), such that two of them may be judged as similar or dissimilar with respect to their expected utility. Formally, a distance function $d(x,y)$ is required, and x and y are regarded as similar if the distance between x and y stays below a certain upper bound: $d(x,y) \leq U$ (where U is a fixed positive number).

In the case of external similarity, S is reflexive and symmetric, but *not* a transitive relation. For some triples transitivity may hold, but that is not the general case. The crucial point is the existence of a distance or dissimilarity function $d(x,y)$ defined for every pair $\{x,y\}$.

A synopsis of internal and external similarity may be possible in the following way. We presuppose that exactly one relation of internal similarity, R_i, is defined for a given finite set **A**. Then **A** is decomposed into a finite family of subsets $E_1, E_2, \ldots E_m$, such that each subset is an equivalence class of objects similar with respect to R_i. Next, a distance function $d(x,y)$ must be defined (for all $x, y \in \mathbf{A}$) with the additional property:

If x and x' belong to the same subset E_k
then $d(x,y) = d(x',y)$ (for all $x, x', y \in \mathbf{A}$).

Under this condition the internal similarity defined by R_i and the external similarity induced by $d(x,y)$ are *compatible*; $d(x,y)$ measures the distance or dissimiliarity between equivalence classes, too.

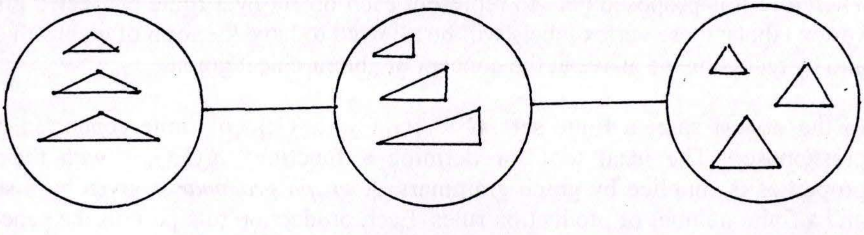

Figure 1: Three different classes of triangles as a demonstration of internal and external similarity

Figure 1 shows three rectangular triangles (in the middle circle), which are similar in the sense of elementary geometry, as representatives of the class of all triangles similar to them (in the same sense), and correspondingly three obtuse and three acute-angled triangles. This particular case is an instance of *internal* similarity. The equivalence classes induced in this way are symbolized by the circles in Figure 1, whereas the connecting straight-lines are to demonstrate *external* similarity: there are smaller and larger distances between equivalence classes (and between two triangles belonging to different classes).

In some cases, an underlying internal similarity together with its equivalence classes may become invisible. Some early authors of graph theory started with a distinction between geometric graphs and abstracts graphs. A geometric graph is a special drawing (where also curved connecting lines are allowed); an abstract graph stands for an equivalence class of isomorphic graphs, such that only the incidence properties count, whereas the kind of drawing becomes irrelevant. In modern texts, generally the concept of abstracts graphs - simply called „graphs" - is tacitly assumed, and „similarity between graphs" always means external similarity.

5.2 Quantification of External Similarity

In the sequel, the term „similarity" always means external similarity. Apparently, „similarity" is closer to human intuition, whereas the terms „distance" or „dissimilarity" are more convenient for mathematical treatment. Given a finite set of objects, it is the present task to set up a distance function such that the distance attributed to any pair of objects will account for their internal structure, too.

In some techniques from information retrieval, but also in cluster analysis, each object is characterized by a finite vector, called „feature vector". It may be difficult to extract such a feature vector from complex real-world objects, and this may imply a significant loss of information. As a rule, the entries of a feature vector represent global and external aspects, whereas the internal structure of an objects is scarcely represented. Therefore it is proposed here to represent each object by a finite connected graph with vertex labels; these vertex labels will be allowed to have the form of graphs themselves, and by recursion we arrive at the concept of „hierarchical graphs".

In the easiest case, a finite set $\mathbf{G} = \{G_1, G_2, \ldots G_n\}$ of finite connected graphs is presupposed. The ideal tool for defining a function $d(G_i, G_k)$ with the required properties is supplied by graph grammars. A *graph grammar* is given by a startgraph and a finite number of production rules. Each production rule permits the generation of a new graph from one of the already existing graphs. This is done by replacing a subgraph of the given graph - where the subgraph fulfills a condition in the production rule - by another graph. For the technical details and for hierarchical graphs only references can be given here.[3)]

For the present application a graph grammar Γ is required which generates at least all graphs in the given set $\mathbf{G}$. If Γ has been fixed, then the requested distance function can be defined by

$$d(G_i, G_k) = \min L(G_i, G_k),$$

where $L(G_i, G_k)$ is the length of a „path" that leads from G_i to G_k by applying

---

[3] See Gernert (1996, 1997), with diagrams and references.

production rules from Γ and the inverse transformations; each step of these kinds contributes 1 to the length L (L corresponds to the number of steps „upward“ and „downward“ in a tree-like diagram representing Γ).

For any given set **G** of graphs a graph grammar Γ can always be found such that at least all graphs in **G** are generated by Γ. But, apart from trivial cases, a graph grammar specified in this way cannot be unique. Rather, there is a multitude of graph grammars which all are suited to represent **G**. The reason behind this fact has been shortly mentioned before (Section 4.3): „similarity“ is a perspective notion, and no definition of similarity or distance can be formulated without a reference to the purpose (or goal etc.) pursued with the individual measurement. Similarity between structures is never a property of the structures alone; rather, it comes into existence through an interpretation by an observer. Fortunately, the method working with graph grammars opens an approach to a mathematical treatment of those factors which go beyond the mere properties of the objects themselves.

5.3 External similarity and dynamical change

Since the elements of an information space can be modified in the course of time, also the equivalence classes defined by internal similarity may undergo changes. Some equivalence classes may be united when certain disparate features of their elements have disappeared; or an equivalence class may be split into several subclasses after significant differences between its elements have emerged. It may depend on the individual implementation what will happen to external similarity as a consequence of such a modification of the internal structure. Remarkable phenomena may be:

1. a persistence of the earlier external similarity for a certain time interval, or, in other words, a time-lag in its updating,
2. a „hereditary character“ of some features of the earlier state that persist in spite of the change,
3. a co-existence of the old and the new external similarity - in the sense of a polymetric space - such that in each situation it may depend upon parameters which of the competing distance measures will become effective.

6 Concluding Remarks

Information spaces in the sense outlined here can be easily modeled on an ordinary PC. Such experimenting, as well as theory-building, may contribute to a better understanding of various processes of information handling. Possible applications refer to information retrieval, data and knowledge bases, information management, Artificial Life, and the theory of self-organizing systems. Fruitful analogies with informational processes in living organisms, but also with processes in physics can be expected.

References

Gernert, Dieter (1996). Pragmatic information as a unifying concept. In: Klaus Kornwachs and Konstantin Jacoby (Eds.), Information - New Questions to a Multidisciplinary Concept. Akademie-Verlag, Berlin, p. 147-162.

Gernert, Dieter (1997). Graph grammars as an analytical tool in physics and biology. BioSystems 43, 179-187.

Kowalski, G. (1997). Information Retrieval: Theory and Implementation. Kluwer, Boston.

Naas, J., Schmid, H.L. (1961). Mathematisches Wörterbuch. Akademie-Verlag, Berlin.

Shannon, Claude E. (1956). The bandwagon. IRE Transactions on Information Theory IT-2, nr. 1, March, p. 3.

Sneddon, I.N. (ed.) (1976). Encyclopedic Dictionary of Mathematics for Engineers and Applied Scientists. Pergamon, Oxford.