FIS2005 - http://www.mdpi.org/fis2005/

Considering an Interdisciplinary Concept of Information for Designing Data Bank Based Information Systems

Federico Flückiger
University of Lugano and
University of Applied Sciences of Southern Switzerland
Innovative Technologies Department, Co-Director eLab
Via Cantonale, Galleria 2, CH-6928 Manno / Switzerland
Tel: +41 58 66 66 588, Fax: +41 58 66 66 571
e-mail: federico.flueckiger@supsi.ch
http://www.elearninglab.org/, http://www.unisi.ch/ & http://www.dti.supsi.ch/
http://mypage.bluewin.ch/federico.flueckiger/

Originally I wanted to start the abstract of my article using text modules written similarly in many other scientific articles:

> Social reality is characterized by an increasing mechanization. Information and communication technologies become more and more present within all areas of human life. Electronic data sources of so called information systems become an ever-increasing importance and its dissemination by the internet make them to the primary information source for different areas of life. They replace more and more reference books of different nature, like phone book, encyclopedia, dictionary, new paper, official gazette, and so on.

With this introduction the subject of the article (the development of an information system) has been motivated using repeatedly information as a not clearly specified concept. This is quite symptomatic for our time where we may live comfortably and working well with a concept of information not really defined. Anyway the progress of technical sciences in general and of computer science in particular was not slowed down at all by that lack of clearness.

However there is some indication that exactly in computer science and parent areas some processes could be easier elaborated using a broadly supported concept of information. Within the development of data bank based information systems we always are going to realize a realistic image on clearly defined worlds to be managed assisted by computer. It is obvious that such processes can be implemented more effectively and convincing using an interdisciplinary concept of information.

At the 1996 FIS conference in Vienna the author of that abstract presented an interdisciplinary approach [4] on definition of information that is usable for designing information systems in a coherent way. It bases on a description of the world composed by structured things surrounded by a semantic and a syntactic closure of relations to other things of the same world.

Introduction

Information is one of those stereotypical, everyday terms, which 'everyone knows' the meaning of. And yet, when asked, everyone finds it difficult to define. It is frequently equated with similar terms such as 'news', 'communication', 'instruction', etc., which themselves are also not clearly defined. This merely dismisses the problem without solving it.

Etymologically the word information is the substantive form of the verb inform which was derived from the Latin 'informare' in the 15th century. While the Latin original meant as much as 'form, give shape to, fashion', the medieval variant was always used in the figurative sense. While 'inform' in the age of Humanism and the Renaissance was a synonym of 'instruct', its field of meaning in the 20th century was expanded to include 'communicate news, provide intelligence'. Most recently the term information has experience a return from the figurative German to its original Latin meaning in the natural sciences, especially in biology (genetic information).

As part of a dissertation [3] the author developed a concept of a unified term information which takes into account the insights of traditional information theories and attempts to eliminate their shortcomings. Beyond this, there are aspects included which, although familiar to researchers in the field, have not found their way into information theory as yet. The resulting strongly interdisciplinary approach of an information theory is sometimes based on known concepts of data bank design, which is why it is generally suitable for the general design of information systems based on data banks. The mathematical equivalents to the statements derived here are documented in another article in the same volume of conference proceedings.

Information Systems and Data banks

Information processing systems were already created in the 50s. Into the 70s a processing-oriented view dominated the development of relevant application systems. The high degree of data redundancy involved caused major consistence problems. Due to the negative experiences development subsequently began of what were known as information systems whose data-oriented (structural-attributive) view was strictly separated from the processing-oriented (functional-cybernetic) view. In such an information system the data are saved according to unitary rules (database) and shielded by a central control program, known as the database management system (DBMS), from direct access by the application programs. This system is what we call the data bank. Independently of this, application programs are developed which only access the data through the DBMS. In so doing we gain two decisive advantages:

1. The data integrity of the information system is assured when the data structures are appropriately designed.
2. The data bank and the application programs are very flexible and the development of each can be designed without a painstaking account of the other's characteristics.

But the prerequisite for this is first and foremost an adequate description of the data in the data bank. This must be created in such a manner that it represents as accurately as possible the section of reality being managed by the data bank. Data bank design thus plays a decisive role in the development of information systems. This usually proves to be a delicate operation, because two worlds collide here which normally do not speak the same language:

- On the one hand, the users who understand little about engineering, but who know the reality to be managed, and who know best the form and manner of how the information system should later function.
- On the other hand, the engineers who must investigate and understand this knowledge in order to convert it into a suitable electronic information system.

It has proven be advantageous to effect this type of interdisciplinary meeting on a conceptual level. On this level the first things discussed are the future information system's relevant documents, objects, ideas, steps, etc., also known as entities (sometimes including objects also known as ontic elements) and their mutual relationships. According to Zehnder [9] and Vetter [8] an entity in this context is an individual and identifying sample of elements (things, persons, or terms) of the real world, or the world of ideas. Hence, an entity can be:

- An individual such as a resident, an employee, a student, a lecturer, etc.
- A real object such as a machine, a building, a product, etc.
- An abstract concept such as a subject area, a lecture, etc.
- An event such as the graduation of a student, posting an entry into an account book, etc.

From the point of view of a system designer an entity is:

- Always a uniquely identifiable unit.
- Always a unit whose existence on an appropriate storage medium must be representable based on an identification characteristic (i.e. a key value), independent of the existence of other entities.
- In most cases an entity for which information will be collected and saved on an appropriate storage medium.

By introducing the term entity, according to Zehnder [9], the world is converted into discrete values in such a manner that the user can still understand it. Entities can be described by attributes, be brought into mutual relationships, and be grouped into sets of entities of the same type. In this way, step by step, a conceptual data structure is created which is then mapped in what is called an entity relationship diagram. This in turn permits the technician relatively simply to derive the technical levels of the physical data structure and the data access structures.

It was revealed that information systems, whose structural-attributive aspects – i.e. the data component – was developed in this manner, are easier to handle, are better accepted by users, and moreover are cheaper to maintain. The entity proved itself during the development of data banks a conceptual elementary building block for the data component of information systems. Difficulties can only arise if under the condition of a strict separation of data and application component, the technical realization of 'active' entities (objects capable of independently triggering functions), is not easy to implement. Moreover the entity term is isolated for the moment and has not been generally integrated into the terminology of informatics.

Unit of Information and Information Element

In [3] I develop a new approach of a unified theory of information with similar insights whereby the focus was not merely on the design of information systems. The new approach was intended to be able to conceptually circumscribe all known information theories. Starting from the known communications systems of Shannon [7] and Nauta [6] a universal communication system was developed (cf. [3] Chapter 4), by means of which the previous determinations of the basic terms of information theory were positioned and adaptively integrated. A presentation of the whole approach to information theory would overstretch the scope of this article. Hence I will limit myself in the following section to a question relevant in connection with information system design, that of how to define the terms unit of information and information element.

This question is discussed heatedly in the standard approaches to information theory. For example, for the supporters of semiotics, as put forward by Morris in [5], the sign or is the unit of information. In the sense that the sign triggers what is known as a semiosis (process of signification) in the interpreting individual, which Nauta in [6] equates with an information event. So, according to this view, in addition to the sign an individual to interpret the sign is also presupposed. The sign is then allocated three levels of relationships, specifically:

- The syntactic, the relationships determining the form of the sign,
- The semantic relationships covering the range of meanings, and
- The pragmatic relations which indicate the effect of the sign on the individual.

In the past, information scientists with a background in semiotics have proposed developing a separate information theory for each relationship level of the sign. These are based on the sign itself as a uniform unit of information, which contains equally syntactic, semantic, and pragmatic information elements. A unified theory to connect all these aspects is not provided for here at present.

There are other difficulties to be surmounted in those information theories drawing on Shannon's communication theory [7]. They basically pose the question, how information can be transmitted over a channel in the form of alternatives. The binary decision where one of only two options is chosen thus suggests itself as the element of information. This definition is also justified by referring to the fact that every n-ary decision can be reduced to a series of binary decisions, which cannot be further reduced. This conception of 'information unit' thus represents a selection from the smallest set that allows a genuine selection, namely a set with two elements. This information unit is often called bit and considered as an acronym of '**b**asic indissoluble **i**nformation uni**t**'.

Unfortunately the term 'bit' is not always used in the same way in the literature. Apart from the version quoted above there is a second variant which considers bit to be the abbreviation of the two words '**b**inary dig**it**'. In that case bit is considered as the smallest possible element for the binary representation of data. Usually the two digits '0' and '1', the character set of the dual system, are used as binary digits. The bit defined in this way is suitable both for a general representation of data and the solution of arithmetic problems.

Thus the term bit has two meanings, on the one hand a binary choice from a binary set, on the other hand a binary digit representing either the values '0' or '1', or adequate code alternatives. In order to distinguish the two definitions, we will call the first alternative bit_1 and the second bit_2. The root of the ambiguity between bit_1 and bit_2 lies in the fact that the alternatives represent different stages of the phenomenon of a binary decision. While bit_1 represents the existence of a binary decision which need not be made, bit_2 represents the result of such a decision in the shape of a fixed value from the set $\{0,1\}$. Thus bit_1 stands for something which may contain either of the two values of $\{0,1\}$, while bit_2 expresses the completed decision by means of one of the two values. Thus bit_1 stands for the carrier of a decision yet to be taken and to be expressed by an element of the type bit_2.

In addition to the just described semiotic and telecommunication technology approaches, the author of this article developed a third variant in [3] (Chapters 3 and 4) to answer the question of [how to define] a unit of information and an information element. The basis is provided by the unified theory of information based on the ontological term of the thing. The thing is always presupposed as a 'something' recognized or postulated by an individual. It is self-evident that such a thing is from the outset 'pragmatic' in every case because it originates in the individual. The thing is always viewed as a whole and as something unique, but it can also consist of other things and itself be a component of other things. The thing normally enters into relationships with other things, the relationships being presupposed to be directed. Those relationships directed towards the thing are attributed a syntactic quality, those away from the thing, a semantic quality. To avoid any confusion with familiar semiotic terms, we call those relationships directed towards the thing d_syntactic, and those directed away from the thing d_semantic. Like the things, the relationships of the individual are also recognized or postulated, which is why the pragmatic is also inherent in them. For this reason, the third pragmatic dimension of relationships need not be provided for using this approach. In summary, the approach of [3] states that the world consists of things recognized or postulated by individuals, and that these things are enveloped in a shell of relationships which is subdivided into a d_syntactic shell and a d_semantic shell.

When we now compare this insight with the first two concepts we can state the following. The semiotic approach can be relatively easily integrated since all semiotic terms have their counterpart in the new concept. In the case of Shannon's concept the situation is somewhat more specific but not much more complicated. In this case Bit1 can be understood as a thing as previously described, which contains a d_semantic shell with merely two elements, specifically the relationships to those two things which are involved in the selection. In this sense Bit1 as a thing sometimes represents the elementary memory location in a computer, whereby both by means of the d_semantic relationships of things referenced by Bit1 represent the values '0' and '1', i.e. their electronic equivalent. If then the case arises that one of the two d_semantic relationships is selected, for example as a value which is stored in one memory location, this relationship refers to a binary number and thus to a thing of the type Bit2. We can draw the following conclusion from these just sketched out statements: in the case of a binary decision, the things of the type Bit1 assume the role of the unit of information which always offer themselves as a potential storage location of a binary decision without anticipating its result. On the other side are relationships referring to Bit2, the information elements in the binary decision. They represent decisions which have been made and thus virtually represent atomic information units.

And what of the general case? The first thought that comes to mind is that the things recognized or postulated by individuals could be the units of information, and the relationships between these things could be the information elements. There is some evidence in favor of the plausibility of this assumption: for one, the observation by semioticians according to which a sign as defined by Morris functions as a unit of information in agreement with the just suggested generalization of the term of unit of information. Moreover, the situation theory of Barwise and Perry in [1] contributes to the support of this thesis. Barwise and Perry describe the world experienced by the individual with situations (things) and restrictions which establish the mutual relationships between the various situations. According to their theory information is always information about something and therefore dependent on restrictions (relationships) between situations. Accordingly there are situations which contain information about other situations, which is marked by the restriction lying between them. Such a restriction is, according to this, always directed from the situation holding the information to that situation which is reference by the restriction as an informational relationship, which agrees ideally with the information theory developed in [3]. The elementary information of the situation theory is thus always an element of the d_semantic shell of that situation which, itself, assumes the role of the unit of information. Because, in turn, the term of the situation (as shown in [3]) merely represents a variant of the more generally defined term of thing.

These statements stake out the field of what can be thought of as unit of information and information element. All the things which can either be recognized or postulated by individuals can assume the role of units of information. Genuine things are not only considered to be those things of the external world and the world of ideas, but also the relationships between such things to the extent which they are the potential objects of scientific research. The information elements, on the other hand, are the relationships between the things assuming the role of the unit of information. In the process it should be remembered that everything either recognized or postulated by some individual can assume the role of a unit of information. Other restrictive characteristics such as the characteristic of truth or factual accuracy, which many researchers postulate as necessary characteristics of information, are reduced by the information theory formulated here to the status of a mere characteristic of a thing as a unit of information, a characteristic which can only make a qualitative statement about the thing. This satisfies the prerequisites for a value-free circumscription of the term information.

Synthesis and Consequences

The aspects of a universal information theory discussed here are restricted for the moment to observations regarding the unit of information and information element. I have excluded from the elucidations important topics such as the communicative aspects in special cases, or the structural-attributive aspects in general. Nevertheless, interesting conclusions can already be drawn:

- The last unit of information is neither the Bit1 as a representative of the smallest set from which a selection is possible, nor is it the sign as the smallest unit of perception. The unit of information is instead the individual thing, either recognized or postulated by an individual. This means that automatically the Bit1 or the sign are also units of information. The information elements finally, which are contained by such a thing, correspond to the relationships which are directed by the thing as a unit of information to other things.

- The entity-relationship concept introduced earlier also agrees very well with these findings, and for the same reason. The entity is then a thing, where it is viewed as something recognized or postulated by an individual and not as something purely external. But this is assured in a superficial manner because the entities determined in connection with the development of data banks are in practice always generated in the engineer – user dialogue, which is why the prerequisite of a thing as described here exists.

In this way the current practice of designing data banks using entities and relationships can be integrated into a higher information theory. More than that: because the general approach presented here to an information theory, alongside the structural-attributive aspects, also sets the functional-cybernetic aspects on the same theoretical foundation, it will be possible in the future when appropriate tools are available, to design information systems as an overall structure without the artificial separation of data component and application component, and also without the likelihood of the loss of data integration. This means that newer types of information systems, such as those based on object-oriented structures, or on knowledge structures of artificial intelligence, can be conceived in a coherent manner. This concerns precisely those concepts in which most recently ontological observations of draft design such as in [3] were proposed and were included in various ways. However, an interactive integration of the various approaches is missing. This is what is achieved by the suggested theoretical reflections and their use as draft concept for all known aspects of informatics.

Bibliography

[1] Barwise, Jon, Perry, John, 'Situations and Attitudes', Cambridge, Mass. and London, England: MIT Press, 1983.

[2] Dretske, Fred Irvin, 'Knowledge and the Flow of Information', Basil Blackwell, Oxford, 1981.

[3] Flückiger, Federico, ‘Contributions Towards a Unified Concept of Information’, Doctoral thesis at Faculty of Science of university of Berne, Switzerland, 1995. (http://mypage.bluewin.ch/federico.flueckiger/Download/Ctrb2uci.pdf)

[4] Flückiger, Federico, ‘Towards a Unified Concept of Information: Presentation of a New Approach’, in: The Quest for a Unified Theory of Information, Edited by Wolfgang Hofkirchner, Gordon and Breach Publishers, ISBN: 90-5700-531-X, Amsterdam, The Netherlands, 1999. (http://mypage.bluewin.ch/federico.flueckiger/Download/Fis96prc.pdf)

[5] Morris, Charles W., 'Signs, Language and Behavior', Englewood Cliffs, New York, 1946.

[6] Nauta, Doede Jr., 'The Meaning of Information', Mouton, The Hague, Paris, 1970.

[7] Shannon, Claude E., Weaver, Warren, 'The Mathematical Theory of Communication', the University of Illinois Press, Urbana, Chicago, London, Fourth Printing of the Paperback Edition, 1969, Original Edition, 1948.

[8] Vetter, Max, 'Aufbau betrieblicher Informationssysteme. mittels pseudo-objektorientierter, konzeptioneller Datenmodellierung', 8., durchgelesene Auflage, B. G. Teubner, Stuttgart, 1998.

[9] Zehnder, Carl August, 'Informationssysteme und Datenbanken', 7., durchgelesene Auflage, VDF Hochschulverlag an der ETH Zürich, 2002.