

DATA vs. INFORMATION

Mihaela MALIȚA¹, Gheorghe M. ȘTEFAN²

mmalita@anselm.edu, gheorghe.stefan@upb.ro

ABSTRACT

Big-Data technologies are about how to extract information from the rough data. The intermediate stage is represented by pre-processed data. But, even from pre-processed data to information the way is long and complex. First of all, we must understand what information means. There is no widely accepted definition for information. By proposing and using a working definition for the concept of information, we present the main techniques involved on the way from the rough data to the useful information. How the information is generated, extracted and how it emerges is shortly introduced. The role played by information in the knowledge process leads to the way information, as structure acting by its meaning, must reconsider the philosophical approach of existence.

KEYWORDS: data, information, machine learning.

Introduction

Over the last hundred years, the rapidly changing technological environment has distorted the relationship between the conceptual space and the technical achievements. Many technical objects are imposed, while the associated concepts are still unsettled. There is a very spectacular play between inventing and catching the resulting reality in well defined concepts. In this process, the market and the corporate space have sometimes more influence than academia. In this context, many notions are long time used before their coherent conceptualization. For example, in our computer dominated world we don't have yet a clear distinction between and the definitions for notions such as *data* and *information*. This fact is not singular in the history of ideas. The most notorious example is the notion of *time*. We are able to measure time with a tremendous accuracy, but we are unable to provide a widely accepted definition for it.

In the next section we propose a working definition for information. The third section deals with various mechanisms for generating information. The fourth section is about how information is extracted from data. How the information manifests on the great chain of being is shortly discussed in the next section. The last section goes beyond what information is and what it does on its way from data to knowledge.

Defining Information

Following the definition provided by Mihai Drăgănescu for the concept of general information [1] we propose a working definition for this very widely used concept.

Definition 1: Information is a structure acting by its meaning.

■

The structure can be: *formal* (based on truth), *non-formal* (based on custom), or *informal* (holistic), while the meaning manifests in three forms: *syntactic*, *semantic*, or as *sense*. Results that we can propose a taxonomy with nine forms of information.

¹ Saint Anselm College, NH; <http://www.anselm.edu/mmalita>

² Politehnica University of Bucharest, RO; <http://users.dcae.pub.ro/~gstefan/>

1. **Formal structure with syntactic meaning (0-type information, 0-TI):** the zero action is provided (it is pure order). Indeed, “syntactic meaning” is a contradiction in terms: the syntactic relation is an internal relation unable to support an external meaning used to trigger an action. Thus, it is about the “zero level” of information, i.e., it represents only a possible support for information. Let us call it *informational structure* [2]. For example, an informational structure is the content of a register file in a processor which is a structure with a syntactic order but with no meaning for the function performed by the processor. The informational structure is the data on which the processing process acts.
2. **Formal structure with semantic meaning (1-TI):** the *program-like action* is performed by the context dependent signification (referential) of the instructions stored in the program memory of a RISC machine. The instruction set architecture of a RISC processor defines the functions performed by a digital structure. The context is the physical structure of the processor able to execute the action coded in each instruction. Only in the context of a specific hardware, the meaning of the formal structure of program is defined and it is able to act. Thus, this is the first level of information. Therefore, a computer is a data processing machine through information, not an information processing machine. The computer, as a physical structure, deals with data under the control of information.
3. **Formal structure with sense (2-TI):** a music score or a verse could be deeply received beyond any possible referential (signification). The “true” art is signification less, but, by an appropriate interpretation, it is able to *act generating sense*. For example:

*Oh, so much silence lies around me that I seem
to hear the moonbeams rapping at my window pane...³*

are extracted from a poem written respecting in full the rules of Romanian language, but the ultimate meaning carried is beyond any worldly referential.



Fig. 1 The last bars ever written by Mozart (in “Lacrymosa” from its Requiem). It is a form which *acts* on the interpreter’s ability to transform it in a *phenomenon*.

³ Lucian Blaga:

*“Atâta liniște-i în jur de-mi pare că aud
Cum se izbesc de geamuri razele de lună.”*

Translated from Romanian to English by Adrian Săhlean.

Let us take another example: few bars from “Lacrymosa” (the last Mozart’s bars from its Requiem): the lyrics (*Lacrimosa dies illa ... judicandum homo reus huic ergo parce Deus...*⁴) and the score (see Fig. 1) are far below the level of the deep meaning offered by the sense carried by an inspired professional interpretation.

4. **Non-formal structure with syntactic meaning (3-TI):** a personalized form without a clear foundation that *acts as a pure ritual* whose meaning is long time forgotten. It is a pure gesture that acts in uniting people around a customary action.
5. **Non-formal structure with semantic meaning (4-TI):** a manifest, a slogan, a mandala, a religious sign ... carrying a cultural or an ideological meaning acting by synchronizing specific communities.
6. **Non-formal structure with sense (5-TI):** Jungian archetypes (a constantly recurring non-formal structure in literature, painting, or mythology), the idea of an ecumenical religion beyond restrictions imposed by any specific religion.
7. **Informal structure with syntactic meaning (6-TI):** sacred ritual supposing a non-mediated connection to the wholeness of existence (various forms of meditation). It can not be validated in community.
8. **Informal structure with semantic meaning (7-TI):** paranormal, transpersonal (encompass wider aspects of humankind, life, psyche or cosmos), altered state of consciousness
9. **Informal structure with sense (8-TI):** the hypothetical pure phenomenological action through intro-openness stated in *Ortophysics* [3].

In various domains the duality data-information manifests: the structure plays the role of data or of information, depending on the context in which it is interpreted. For example, in computer science the memory content of a computer could be data or information. In living entities, the special types of molecules are support structure or provide functional configurations. In our world, the news is usually data, while regulations are information⁵.

The information is *generated* by specific human actions, it could be *extracted* from various environments or it *emerges* from different levels of existence.

Generating Information

The process of generating information manifests mainly on three levels: the formal levels, dominated by the activity of writing programs for computers, the imaginative level of ideologies and artifacts and on a hypothetic holistic level by *intro-openness* [5].

Programs are rigorously organized binary structures designed to control the computational processes in the hardware of a computing machine. The design is done in a formal language and the actual binary form is automatically generated, with more or less efficiency, by the computing tools called *compilers*. Sometimes the program is interpreted by another program as it is, generating the actual binary code which *acts* on the hardware resources of the computing machine. A program written in a high-level programming language is not information. It is the source of information through the compiling or interpreting process.

⁴ *Mournful that day ... guilty man to be judged Lord, have mercy on him ...*

⁵ In the World 3 of Karl Popper [4] we must make the necessary distinction between data and information.

The program is the result of a complex mental activity which starts from a non-formal or even informal specification which is translated in a rigorous formal structure sent as input to a machine (compiler or interpreter). The translation from the non-formal or informal structure of specification to the formal structure of program is not a formal process. The associated mental activity supposes a hard to explain process of formalizing. This mental activity is acquired in a complex learning process involving instruction and training (maybe education).

In fact, as in any transition process, the transition from no-form to form supposes a large spectrum of mental behaviors hard to be caught in few explicit mechanisms.

Ideologies & symbolic artifacts are non-formal constructs developed in specific historic and cultural contexts. They are configured and act by the virtue of self-organizing mechanisms driven by the collective conscious and unconscious mechanisms. They are coherent constructs but do not obey to formal, too rigorous restrictions. Even though partially these coherent constructs can be caught in forms, always a meaningful reminder gets rid of our formalizing work. And, unfortunately, many times in these residuals there are essential aspects. The imaginary, through its three components – intuition, fantasy, will – plays the major role in generating and using this kind of information.

Intro-opening⁶ is the hypothetic process developed *in* the *open* space of the deep existence where the phenomenological information could be modified by our specially trained mind, so as new existential experiments become possible.

Extracting Information from Data

The structures able to act are hidden many times in data distributed on various media or in the continuous flow of signals which flood the environment in which we live. Therefore, there are two main sources to be accessed for extracting information:

- data mining from raw material
- machine learning from selected and annotated (labeled) data

Data mining from raw material is one of the main sources of information. The unstructured data available from various sources is submitted to a structuring process able to provide:

- more organized data without informational properties, namely without abilities to act
- information acting on different levels of reality, mainly supporting various decision processes

Machine learning techniques are the second source of information (when we are unable to explain our expertise) [6]

1. Supervised learning using labeled (annotated) training data (classification, regression)
2. Unsupervised learning using unclassified training data (to find patterns)
3. Reinforcement learning using reward feedback (to maximize performance)

Emerging Information

Information emerges, starting from the phenomenological information, on a desecrated *great chain of being* on the following levels:

⁶ It is the English version of the Romanian term *intro-deschidere*, coined by Mihai Drăgănescu in [5].

-
- **Mineral.** The order of the crystalline structure *acts* being responsible, at least, for the stability of the mineral system.
 - **Plant.** Like all organisms, plants use DNA to pass on their traits. In plants, genetic information *acts* supporting the speciation mechanism. There are experiments, accepted at least partially by the scientific community, showing that plants *react* to information sent them by the state of a conscious mind.
 - **Animal.** Besides the genetic information, emerged at the plant level and developed at the animal stage, information *acts* as an interactive agent in the animal world.
 - **Human.** At this level, the 3rd world of Karl Popper is equivalent with what can be called an info-sphere, a world in itself as part of the human world segregated as part of existence for protecting the evolution of humanity.
 - **Superhuman.** In a world driven by information humans are unable to generate or extract algorithmically. The information extracted or generated using trans- algorithmic techniques are used to design technologically augmented humans. Max Tegmark proposed in his book [7] a vision considering the following stages for human development:
 - *Life 1.0 (biological stage): evolves its hardware and software*
 - *Life 2.0 (cultural stage): evolves its hardware, designs much of its software*
 - *Life 3.0 (technological stage): design its hardware and software* [7]

The *Life 3.0* stage is emerging. The technological improvements started slow, a few millennia ago, when man used a branch to help him walk. But now, the information driven tools starts to be embedded in our body to *act* supporting the way we live.

We conjecture that even the evolution on the *desecrated*⁷ *great chain of being* is made possible by the *action* of various forms of information.

Data – Information – Knowledge

The ultimate action of information is to generate knowledge. Then the *action* environment is our conscious individual or collective mind. Data is unable to generate knowledge.

Definition 2: Knowledge is the mirrored informational image of the deep phenomenological information.

■

The knowledge of existence can be assimilated with the generation of an informational image of the fundamental laws of existence. Can be, because, according to Ian Durham [9], it is not yet clear if:

It from bit

or

Bit from it

⁷ We insist on the desecrated character of the great chain of being for delimiting us from the consecrated Great Chain of Being, “a strict hierarchical structure of all matter and life, thought in medieval Christianity to have been decreed by God” [8].

In [10] is claimed that:

It & bit⁸

could be the most probable relation between what it *is* and what *acts* by its meaning. In this context we are faced with the alternative

IS vs. ACTS

which may substitute the too old debate between

MATTER vs. IDEA

Thus, information, as structure acting by its meaning, becomes the main concept to be used to reconsider the eternal debate unfolded between various dualistic approaches developed through the entire Western philosophy. A concept imposed by the development of a technology – *information* – becomes, in our opinion, the new “star” taking from the old Platonic *idea* the main role in the great story about existence.

References

1. Mihai Drăgănescu, "Information, Heuristics, Creation" in I. Plauder (ed.): *Artificial Intelligence and Information Control Systems of Robots*, Elsevier Publishers B. V. (North - Holland), 1984, pp. 25-28, [Online]. Available: <http://www.racai.ro/media/IHC.pdf>
2. Gheorghe M. Ștefan, *Loops & Complexity in Digital Systems. Lecture Notes on Digital Design in Giga-Gate/Chip Era* (work in endless progress), [Online]. Available: <http://users.dcae.pub.ro/~gstefan/2ndLevel/teachingMaterials/0-BOOK.pdf>
3. Mihai Drăgănescu, *Ortofizica. Încercare asupra lumii și omului din perspectiva științei contemporane*, Editura Științifică, București, 1985.
4. Karl Popper, *Three Worlds*, The Tanner Lecture on Human Values, Delivered at The University of Michigan April 7, 1978, [Online]. Available: https://tannerlectures.utah.edu/_documents/a-to-z/p/popper80.pdf
5. Mihai Drăgănescu, *The Depth of Existence*, [Online]. Available: <http://www.racai.ro/external/static/doe/>
6. Pedro Domingos, *The Master Algorithm: How the Quest for the Ultimate Learning Machine Will Remake Our World*, Basic Books, 2018.
7. Max Tegmark, *Life 3.0: Being Human in the Age of Artificial Intelligence*, Alfred A. Knopf, 2017.
8. Arthur O. Lovejoy, *The Great Chain of Being: A Study of the History of an Idea*, Harvard University Press, 1976.
9. Anthony Aguirre, Brendan Foster and Zeeya Merali (Eds), *It From Bit or Bit From It?* Springer, 2014.
10. Gheorghe M. Ștefan, "Information in the Structural Phenomenology of Mihai Drăgănescu", *Noesis*, 2013-2014, pp. 9-19. [Online]. Available: http://noesis.crifst.ro/wp-content/uploads/revista/2013-2014/2013-2014_1_01.pdf

⁸ Or the *informater* proposed in [5].