Copyright@ Department of Bioinformatics, UIET, CSJM University Kanpur

A Value Added Course on "Data Mining & Data Analytics June 10-June 22 2022

Patron

Honorable Vice Chancellor Dr Vinay Kumar Pathak

CSJMU

&

Dr Brishti Mitra

Director, University Institute of Engineering & Technology

Coordinator

Mamta Sagar

Head, Department of bioinformatics, UIET

Lecture 3 : Information

Value Added Course : Data Mining & Data Analytics

Mamta Sagar

Assistant Professor, Department of Bioinformatics

UIET, CSJMU University, Kanpur

Learning Objective

• Define data, clearly identifying that data has no meaning

• define information and show how data can become information through context and meaning

 • define knowledge and understand that information becomes knowledge when human experience is applied.

- The word information has Latin roots(informo) and Greek (Informatio) which means model, can be considered in two context
- One is Act of molding the mind & the set of communicating knowledge

What is information?

- It is important that students learn the concept of what 'information' is as used in information technology. Information is the result of processing data, usually by computer. This results in facts, which enables the processed data to be used in context and have meaning. Information is data that has meaning.

- If we put Information into an equation it would look like this:

    Data + Meaning = Information

Looking at the examples given for data:

- • 3, 6, 9, 12

- • cat, dog, gerbil, rabbit, cockatoo

- • 161.2, 175.3, 166.4, 164.7, 169.3

- Only when we assign a context or meaning does the data become information. It all

- becomes meaningful when we are told:

- • 3, 6, 9 and 12 are the first four answers in the 3 x table

- • cat, dog, gerbil, rabbit, cockatoo is a list of household pets

- • 161.2, 175.3, 166.4, 164.7, 169.3 are the heights of 15-year-old students

Figure 1-9. Information Theory and the Central Dogma. Information theory applies equally to the replication, transcription, translation, and the overall process of converting nucleotide sequences in DNA to protein.
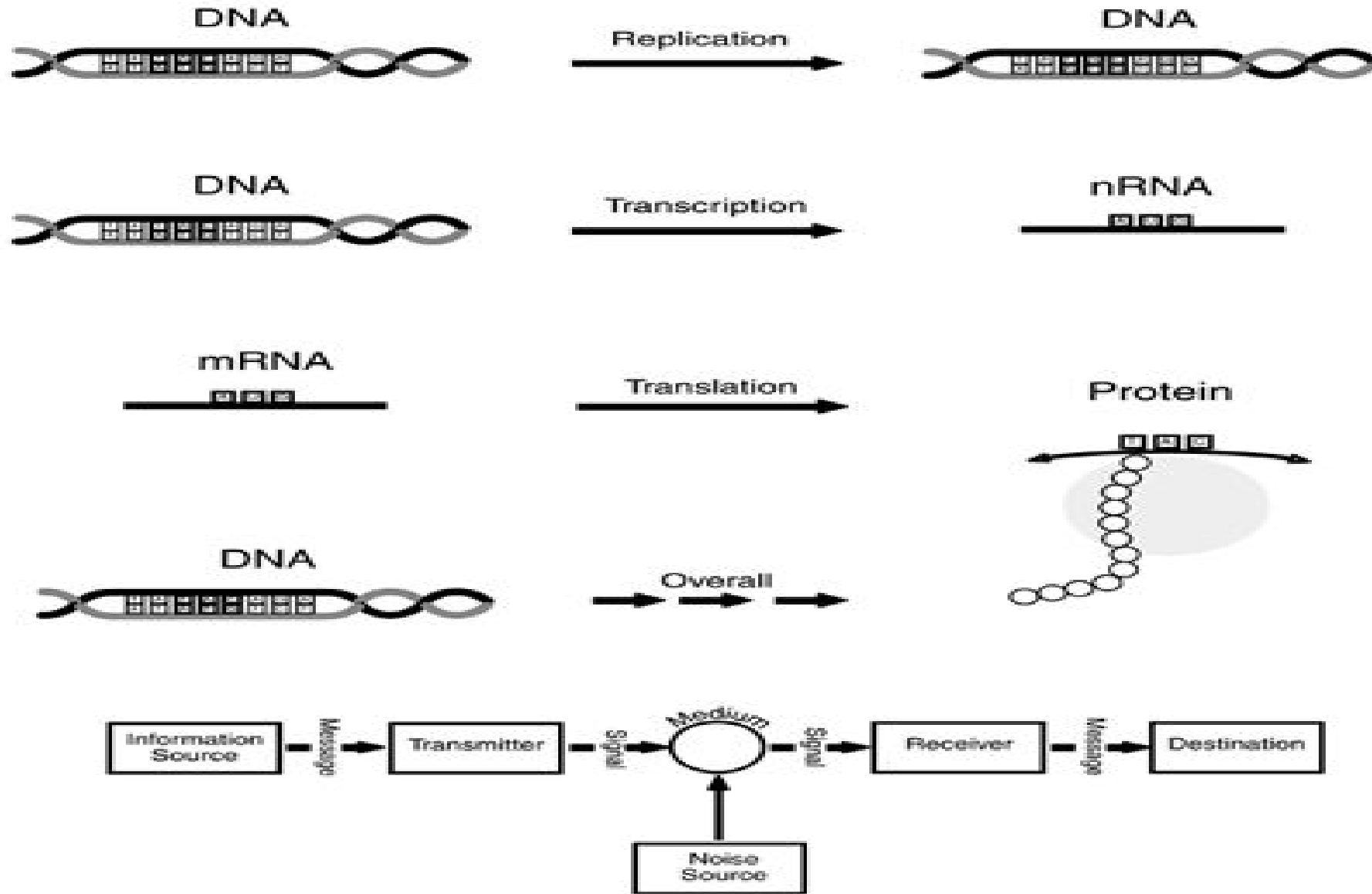
The Concept of information in the Natural Sciences

- Information is prima facie something that flows between a sender and a receiver. But Shannonls definition of information is quantitative concerning possible selections from a repertoire of physical symbols. It is, in fact, as Underwood (2001) remarks, a theory of signal or message, not of information, transmission. Shannon's model of communication (see Figure 8.1) includes six elements: a source, an encoder, a message, a channel, a decoder, and a receiver (Shannon, 1948).
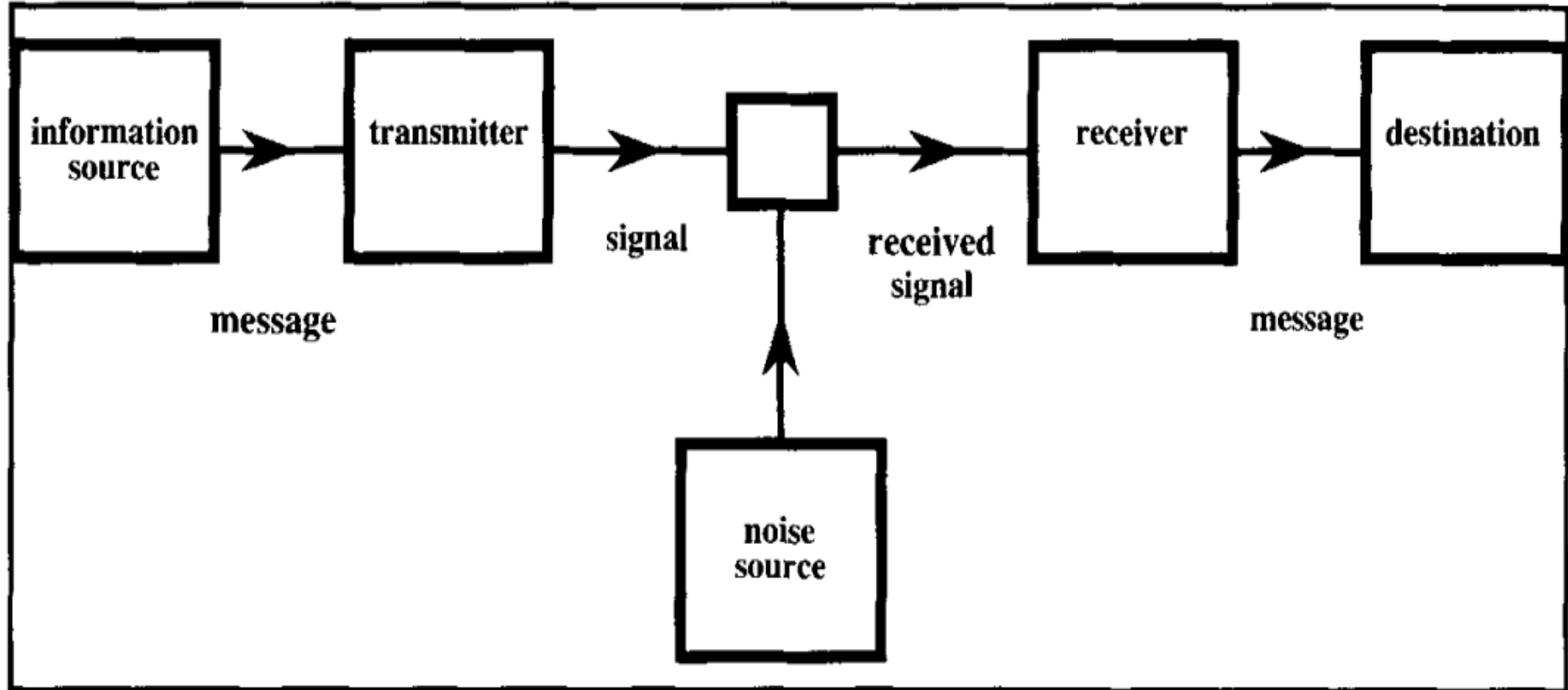
Shannan's Model of Communication



Figure 8.1 Shannon's model of communication.

- Strictly speaking no information could be communicated between a sender and receiver, because this theory is not concerned with the communication of a meaningful message, but rather with the reproduction of a selection process. Shannon correlates information- that is, the number of possible choices in order to create a message- and uncertainty.

- Greater the freedom of choice, the greater the uncertainty; that is, the information.

- This concept of information seems, as Weaver remarks, "disappointing and bizarre-disappointing because it has nothing to do with meaning, and bizarre because it deals not with a single message but rather with the statistical character of a whole ensemble of messages, bizarre also because in these statistical terms the two words information and uncertainty find themselves to be partners" (Shannon & Weaver, 1972, p. 27).

- Information is particularly known from the description of nature at its micro-level of elementary particles and from computer science (bits an qbits), but is also essential in understanding the higher complexity of living organisms as well as the macrostructures such as planets, and galaxies of the Universe. For instance the so-called "Big Bang" and the events that followed, appear to constitute a fine tuned expansion process, in the framework of a very specific set of interrelated physical laws and constants, as it has been unveiled by humans 13.5 billion years later (see for excellent reviews: Davies, 2007, Greene, 2004, Linde, 2004, Görnitz, 2012). In this sense the evolution of our universe can be seen as a dynamic flow of unfolding information and creation of new information.
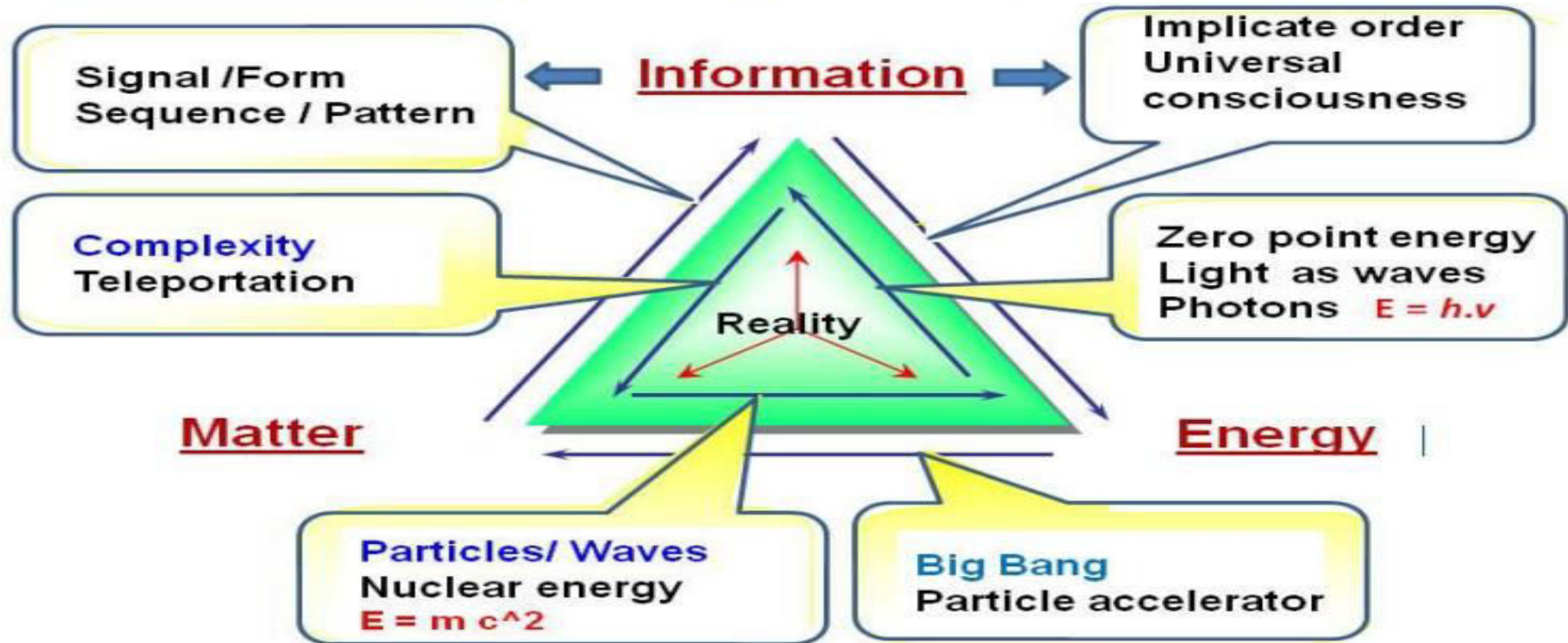
Fig. 1: The fundamental triad of energy/matter/information

how are these three building blocks interrelated?

- This essay is based on the thesis that information is as fundamental as matter and energy in the fabric of reality, in other words: information is physical. Information may even represent a modality of physics that preceded the manifestation of matter (Wheeler 1994, Zeilinger, 2003).

- But how are these three building blocks interrelated? Can information be reduced to energy and vice versa, can energy and matter be defined as modalities of information? Matter and energy were once considered two separate and distinct elements, until Einstein proved they were inter-convertible as indicated in the $E = m c 2$ equation. One may wonder: where the item of information is in this famous equation (see Fig. 2).

- Some see information as a modality of energy and interestingly the constant c in the equation, according to Einstein, was not only indicating the maximal speed of light but at the same time was meant as the maximal speed of information transfer (see Seife, 2006). Umpleby, 2004, published a paper titled Physical Relationships among Matter, Energy and Information, which attempts to connect the three concepts.
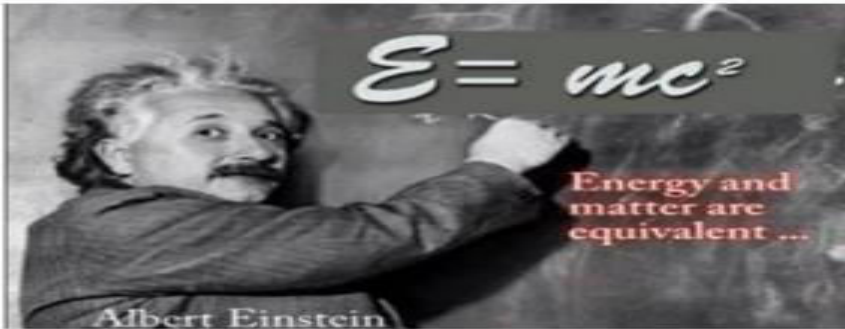
Information and evolution

- The interactions of subatomic waves/particles subsequent to the so called Big Bang, created a dynamic network of quantum information, that finally also enabled the formation of highly complex macromolecular structures. The history of these particular wave/particle interactions, are supposed to be stored in an all pervading quantum field, as it was inherited from the initial information matrix (Zizzi, 2006).

- Each step in the unfolding evolution implied an inherent potential for change and, ultimately, also the ability to generate biological life. The creation of first life was facilitated by processes such as self-organization and autocatalysis, as well as synergistic and symbiotic processes (see Kauffman, 1993, Margoulus, 1998), providing an ever growing, novel, information framework. Further complexity and sophistication in biological evolution was partly realized by genetic mutation, and chromosomal reorganization, combined with the selection pressure of the environment.

What is missing?

The famous equation: What is missing?

$\mathcal{E} = mc^2$

Energy and matter are equivalent ...

Albert Einstein

$E = \cancel{x}a^2$
$E = \cancel{m}b^2$

$E = mc^2$

Without matter, there is nothing; without energy matter is inert; and without information, matter and energy are disorganized, hence useless"

— Anthony Oettinger

Fig. 2: Relating energy to matter, but where is information ?

Cuotation from T.S. Elliot, "The Rock"

- Eliots famous (1969) coutation from T.S. Elliot, "The Rock"
- Where is the Life we have lost in Living?
- Where is the wisdom we have lost in Knowledge?
- Where is the Knowledge we have lost in information?

The Computing Universe: more than an information processor !

Quantum Computation and Quantum Information

MICHAEL A. NIELSEN and ISAAC L. CHUANG
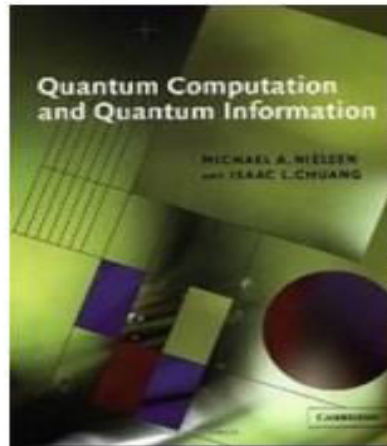
Fredkin   Wolfram   Tegmark   Lloyd   Barrow

Maxwell, Boltzman; Gibbs: Entropy is proportional to the number of bits of information registered by atoms in motion, in the form of polarization, spin, and momentum.
Nielsen and Chuang: atoms and elementary particles can even be programmed to perform digital calculations. Lloyd: The universe is a physical system that is programmed at its microscopic level to perform digital computations.

Fig. 4: The physical information of the Universe.

In a very useful book in this regard with the title: "Decoding the Universe", Charles Seife (2006) explains:

- "What is it that gathers information about the atom and disseminates it into the surrounding environment: it is nature itself that is constantly making measurement on everything. The particles of light and air are nature's probes or measuring devices.

- By observing an object you are simply receiving the information that has already be deposited on those particles. Even if you would remove the earth atmosphere and our sun distant photon's from distant stars are bombarding our planet.

- The universe is teeming with cosmic rays that are composed of photons that were born shortly the big bang. Even without any of those photons, nature would be able to collect information since it creates its own particles at every point of space: on the smallest scales, particles are constantly winkling in and out of existence in the quantum vacuum or zero-point energy field.

- They appear, gather information, disseminate it into the environment and disappear into nothingness from whence they came. These evanescent are the so called vacuum fluctuations occurring throughout the universe, and make it impossible to shield an object completely from Nature's measurements".

2.3.2 When does data become information?

Data on its own has no meaning. It only takes on meaning and becomes information when it is interpreted. Data consists of raw facts and figures. When that data is processed into sets according to context, it provides information.

Data refers to raw input that when processed or arranged makes meaningful output. Information is usually the processed outcome of data. When data is processed into information, it becomes interpretable and gains significance.

In IT, symbols, characters, images, or numbers are data. These are the inputs an IT system needs to process in order to produce a meaningful interpretation. In other words, data in a meaningful form becomes information. Information can be about facts, things, concepts, or anything relevant to the topic concerned. It may provide answers to questions like who, which, when, why, what, and how.

The fundamental character of information

- What Is Information? It arises through interaction !

- Extending the notion of environment or the external world, the following notions of information were given by Gershenson (2010):

- Notion 1: Information is anything that an agent can sense, detect, observe, perceive, infer or anticipate. This notion is in accordance with Wiener (see later), where information is seen as a just-so arrangement, a defined structure, as opposed to randomness and it can be measured in bits.

- Notion 2: An agent is a description of an entity that acts on its environment. Note that agents and their environments are also information, as the can be perceived by other agents. An agent can be an electron, an atom, a molecule, a cell, a human, a computer program, a market, an institution, a society, a city, a country or a planet. Each of these can be described as acting on their environment, simply because they interact with it.

- Notion 3: The environment of an agent consists of all the information interacting with it.

- Notion 4: The ratio of living information of an agent is the amount of active information produced by itself over the amount of active information produced by its environment.

References

- Cambridge International AS & A Level Information Technology 9626 For examination from 2017 Topic 1.1 Data, information and knowledge

- Dirk K F Meijer on 20 May 2015. An Information " what do You Mean?