

## What is information?

### **ICT : Information – Communication – Technology :**

It has been nearly two decades that television, telecommunications and information technology converge. Indeed, it is especially the evolution of the computer technology that has redistributed the cards and helped to blur the boundaries of the two others.

It is common today to declare at a low cost that the lack of understanding between the world of video and IT is deep, that the Broadcast world does not understand IT and vice versa. First, these claims are simplistic and do not reflect the complexity of each of the universe. In addition, both are not able to formulate and structure a new vision of a digital media most appropriate to the needs of collaborative audiovisual production and exploitation.

The concepts of Information and communication technologies, ICT are among the techniques used in the treatment and transmission of “information” via different networks. By extension, this acronym is also a sector of economic activity.

ICTs include a set of resources to handle “information” via computers, programs and networks essential to its creation, storage, management, transmission, archiving, research and discovery.

This definition of the ICT positions the industry as a support of the digital content industry. This differentiation is essential to build a future vision. Indeed, the acronym might suggest that these technologies actually deal with information and communication. In fact, the terms of information and communication are abusive in the real scope of technology.

Essentially, these tools manipulate and transport data's:

Data is a representation of facts, concepts or instructions, for interpretation or processing by humans or by automated tools.

Information and communication are based on other concepts:

- Information is the meaning that humans assign to data by means of conventions. Information requires languages and dictionaries. It can be stored and it is used to communicate in time and space.
- Communication is a process by which living organisms define and share meaning. Communication requires a sender, a message that wraps information and a receiver. It can be remote in time and space. The communication process is complete once the receiver has indicated to the issuer its easy receipt and comprehension of the message. This process requires that the sender and receiver share information as a linguistic and semiotic universe.

In the analysis, ICT does not cover these concepts. Most of the available technologies deals with “information” in the data appearance (signifier) and not in the sense of “processing of knowledge” (properties, relation, signified, ...).

## Linguistic and semantic : Ferdinand de Saussure and Jacques Derrida

At the beginning of the twentieth century, research over language has opened spaces for all people trying to understand the particularly hermetic universe of the meaning. In Europe it is Ferdinand de Saussure who founded modern linguistics and laid the foundations of semiotics.

In his "Cours de Linguistique Générale" (published post-mortem in 1916), he defined the distinction between other language and speech, the arbitrariness of the sign, .... The basic idea of Saussure is that language is a closed system of signs. Any sign is defined in relation to others, for pure difference (negatively), and not by its own characteristics ("positive") and so it constitutes a "system":

The linguist distinguishes two different signs: the signifier and the signified:

- The signifier refers to the graphical representation of a term, a picture, a sound ... What is important in the representation, is the differences that distinguish them from each other. The value of a representation follows from these distinctions. Each language lexicons is constructed with a limited number of characters (phonemes for the language), and syntax that defines the order in which these characters must be organized. In the computer world, the HTML is from all points of view, a signifier!
- The signified is the concept, that is to say, a semantic representation associated with a meaning. This observation leads Saussure also distinguished meaning and value since the existence of different languages necessarily introduces different values for a same meanings. So the signified in a language is a concept defined by the existence or absence of other concepts that are in opposition (enforceable ?).

In all languages carried by humans, there is one to which Jacques Derrida gives primacy over the others: the language (la langue in French !). He sees the system as a metaphysical "logocentrisme". Derrida develops a similar difference to Ferdinand de Saussure, which gives meaning to the significant elements in the form of trace (s). To explicit this meaning he uses the term "differance" (its graphical representation differs by the presence of an "a" instead of an e). In fact, for a same signifier could have many signified (polysemy).

But the "differance" is precisely the movement, the "producer" of these differences: it is the "process" by which different concepts exist.

*WIP : But AV has one specific characteristic: the relation between the signifier and the signified is not arbitrary for a picture ... there is a kind of avatar relation that leads to manage the transition to structured (non-textual) metadata's. For the time being, accurate content adaptation, interactive features in the content, ... will require object tracking and precise positioning of objects and camera's in space and time. More and more xD technology is arriving in the media business !*

## **The cognitive process**

The early days of computing have just helped provide a conceptual arsenal to suggest cognition: the notion of information and information processing.

Cognitive psychology studies the major psychological functions of human beings that are memory, language, intelligence, reasoning, problem solving, perception and attention. Cognition more generally is defined as the set of mental activities and processes relating to knowledge and to the function that makes it.

Although considerable progress has been made since that time, the notion of information system is at the heart of cognitive models: symbolic formalizations (cognition as a system of symbol manipulation), rather connectionist (cognition seen as flow activation in a large neural network), or hybrid (concept of a large neural network that performs functionally a system of symbols).

### **So the concept of information (and Knowledge) is deeply embedded in a cognitive process!**

These analyses provide a major interest to all researchers who are studying theories seeking to establish operational bases for a semiotics accessible to machines. In fact today, text is the only representation format accessible to “interpretation” tools!

## **The semantic web architecture: HTML – XML – RDF – OWL – Logic :**

Most of today's Web content is meant for human consumption, with machines being able to capture and manipulate it only at the syntactic level. The central idea of the Semantic Web initiative is to make the meaning of Web content machine accessible.

This enables the development of sophisticated tools that can provide a much higher functionality in supporting human activities on the Web. The Semantic Web relies on the combination of the following technologies:

- Explicit metadata: they allow Web pages to carry their meaning on their sleeves.
- Ontologies: they describe the main concepts of a domain and their relationships.
- Logical reasoning: it makes it possible to draw conclusions from combining (meta)data with ontologies.

Today HTML is a standard language in which Web pages are written. This language relies on a set of predefined tags, which control the appearance of a Web page (such as bold, italic, numbered and unnumbered lists, line breaks etc.). It is the signifier layer for a linguist as De Saussure !

While XML also relies on tags for marking up Web content, it allows users to define their own tags. In this sense, XML is a domain-independent mark-up meta-language (a language for defining a mark-up language). XML tags do not describe the appearance of Web pages, so XML separates content from formatting, a nice property that is useful for different presentations and views based on the same data. XML is the first layer of the signified for a linguist!

While XML is a universal language for defining mark-up, it does not provide with any means of talking about the semantics (meaning) of data. For example, there is no intended meaning associated with the nesting of tags; it is up to each application to interpret the nesting.

RDF is a framework that has been implemented in XML language for describing resources. Its basic building block is a statement, a triple consisting of an entity (called resource in Web terminology), a property, and a value (which may be another resource). In the Semantic Web design, RDF defines a layer residing on top of XML. As a consequence, RDF has been given a syntax in XML. RDF is domain-independent in that it makes no assumption about a particular domain of use. It is up to the user to define her own terminology in a schema language called RDF Schema. RDF and RDF Schema provide the basic core languages for the Semantic Web.

But there are a number of characteristic user-cases of the Semantic Web that would require more expressiveness. Such extensions include:

- Disjointness of classes
- Boolean combinations of classes
- Cardinality restrictions
- Special characteristics of properties
- Local scope of properties: `rdf:range` defines the range of a property

OWL was designed as a new standard Web ontology language. It is based on top RDF/S, and seeks to find a balance between expressive power and efficient reasoning support.

Reasoning is important because it allows one to:

- check for consistency of an ontology and the knowledge;
- check for unintended relationships between classes and
- automatically classify instances in classes.

The formal foundation of the OWL language is a branch of knowledge representation and reasoning called description logics. (written with the help of ICS/Forth (see [www.ics.forth.gr/isl/swprimer/index2.php](http://www.ics.forth.gr/isl/swprimer/index2.php)) for the semantic web)

So there is no XML vs RDF ... it's only a way to path the semantic universe ...

## **LOD (Linked Open Data)**

In a conventional index, you must define a list of metadata (attributes and values/authority list). Usually this is done in XML, which is quite able to collect various metadata's in this structured language. Unfortunately, these metadata's properties and relationships are only interpretable by humans.

What the audio-visual industry needs is:

Add to the existing contents machine-readable metadata so that information can be analyzed, questioned, reused

Define the relationships shared by these metadata to enable interoperability between applications, but also allow the identification of new relationships by reasoning and machines inferences.

Having identified objects and explicit relations allows the automatic reconciliation of distant information produced by other, facilitating the enrichment, research and information processing. It is therefore essential to design documents that have the information in a form accessible by the machine and keep the links created with the relationship values.

Only at the semantic level it becomes possible to use the computing power of the computer to help users to exploit the information to a greater extent than just reading.

The passage of a "web document" to a Web of "data" means:

- To have typed objects and relations: the Web as a giant distributed database
- Allow the use of knowledge in machine-Web, facilitating research and information processing

The Semantic Web helps the machine in the interpretation of information contained on a web page, for example: mls.jpg is a photo of a person named Marie-Laure Schellings indexed with its original RTBF log!

Advantages of a semantic universe:

- Programs and sites may share information
- Search tools are able to display more relevant information, well based, "relevant" in terms of results!
- The data mashers ("aggregator of knowledge / information) are likely to combine information from different sources to create new and surprising relationships (serendipity search)

All of these resources can be distributed through the Web in a distributed and federated data Graph. And when all resources are described with semantic languages (explicit metadata, relationships), the Web becomes a huge database...

With the help of URI (Unique Resource Identifier), no need for a single common data base, the Web becomes THE data base with a specific search protocol to RDF/OWL: SPARQL

In a traditional database, the schema definition file "data" that defines the physical structure and the meaning of the data stored? This pattern must be defined before it can store data!

In the semantic universe, the pattern is logical (not physical) and "late" (links can be made even after the definition of instances!).

Finally it is the ontology that provides the common semantics for resources on the Semantic Web! Ontologies are developed with increasing levels of semantic expressiveness as RDFS (RDF Schema) and OWL (Web Ontology Language).

## **Evolution of the Semantic Web:**

Linking Open Data (LOD) is not a standard, but a project to link all the semantic databases in the form of "Linked Data", and to obtain a large network, composed by growing database. The Linked Data databases are simply written in RDF/OWL!

Thus, the challenge of LOD is to link all these databases, via HTTP links so that they can be part of the LOD community.

The universe of interconnected data (Linked Data) uses the Web to link semantic databases (RDF, ..) that were not connected, or to simplify the level of data links connected by other methods.

Specifically, Wikipedia defines **linked Data** as a method of publishing structured data so that it can be interlinked and become more useful. It builds upon standard Web technologies such as HTTP and URIs, but rather than using them to serve web pages for human readers, it extends them to share information in a way that can be read automatically by computers. This enables data from different sources to be connected and queried.

<http://linkeddata.org> : the web site provides source and indicators for the resources of the Linked Data (community of interconnected data):

Roger Roberts  
RTBF/Titan & MediaMap+ project  
December 2011