Knowledge Representation from Classification Schema to Semantic Web (I)

Dr. Silvia-Adriana Tomescu
"Carol I" Central University Library of Bucharest
E-mail: silviatomescu@yahoo.co.uk

Dr. Silvia-Adriana Tomescu is librarian at "Carol I" Central University Library of Bucharest, Political Science Library. She completed the doctoral studies in Information Science with a dissertation on distance learning in higher education: „The use of modern means of communication as teaching, learning and training in distance higher education. Case study: librarians' 'raining." Her main study fields of research are: eLearning librarianship, higher education, academic writing and related to current position cultural diplomacy and the Romanian exile. She published translations and studies in library science and communication.

Abstract

In this essay we aim to investigate knowledge as approach of describing possible worlds through classification schema, taxonomies, ontologies and semantic web. We focus on the historical background and the methods of culture and civilization representation. In this regard, we studied the ancient concern to classify knowledge, from the biblical period when the Tree Metaphor concentrated the essence of knowledge, to the Francis Bacon classification and then Paul Otlet and we analysed the languages used in the scientific fields and then in the information science filed, emphasizing on the improvements of the ICT: hypertext and semantic web. We paid a special attention to the knowledge construction through math language and exchange standards. The reason of the approach comes from the logic and philosophic base of the knowledge representation that underline the idea that only properly structured scientific domains ensure the progress of the society.

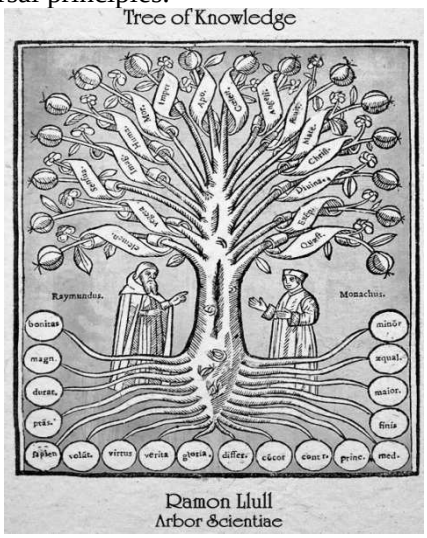
Keywords: knowledge; classification; representation language; semantic web

I. Knowledge representation
1.1 A Historical Perspective

In this research we propose to create a historic path of classification of knowledge in order to understand how this phenomenon contributed to the evolution of science. The interests of intellectuals for organisation of knowledge belonged to the ancient times, and reflect the special position it had, to preserve the values. The extension of any classification system includes the idea of information sharing. The metaphor of *tree* was frequently used to represent human knowledge. In this regard, *Jesse Tree* illustrates the origin of Jesus "*I am the Root and the Offspring of David, and the Bright and Morning Star*"; it is believed that symbolize the basis of genealogies. (1) A reflection of this can be found on the exterior wall of the Voroneț Monastery the Tree of Life. Moreover, this central tree, which covers the whole area of philosophy from cosmos to humanity, is, at the same, time a *Life Tree*. (2)

Porphyry, the Neo-Platonic philosopher, played a significant role in the evolution of thinking. He translated, corrected and revised the writings of his master Plotinus. (3) He wrote in Greek the introduction to Aristotle's Categories – Isagoge which had a full success and led to its translation in Arabic, Hebrew and Latin. (4) Porphyry enumerates five predicates: *genus, species, difference, property, and accident* that determine how to define the categories. Genus and species are complementary to Porphyry (species subordinates to the genus down to human being, considered secondary substance).

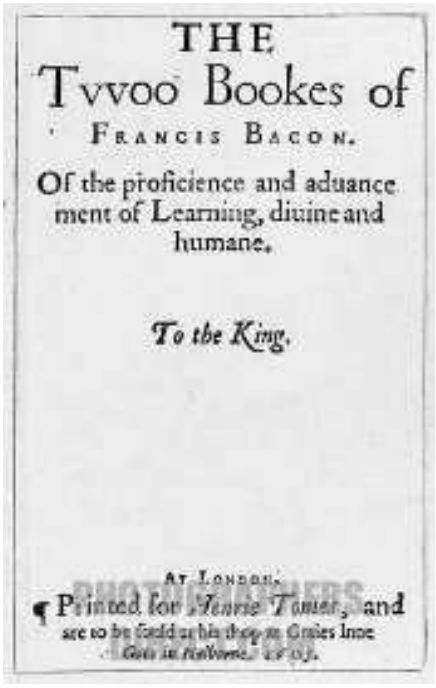
The metaphor of the *Tree of knowledge* of Raymundus Lullus worked also as classificatory representation of sciences. Raymundus Lullus sought to establish a unit of knowledge based on universal principles.



Lullus proposed a unique method of ordering knowledge; in this respect, sought to base a language to express the universe in a *combinatoric of sentences and signs.* He promoted religious beliefs through a unique logic and syntax, limited to nine letters. Lullus focuses on the main topics as Man, God and World, a whole universal system. The diversity of letters and signs combinations express the *algebraic language* that better shaped the combinatory logic of Leibnitz. Aristotle's Categories are also condensations of reality, huge umbrellas which focuses on the spiritual level of world. The author has divided the world into ten categories: *substance, quantity, quality, relation, place, time, situation, possession, action, passion* actually desiring to express entities through Logos.

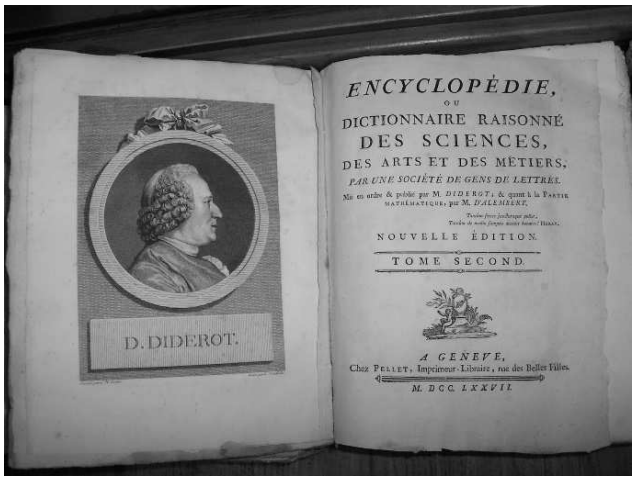
The author tried to establish a hypothetical frame of interrogations about the universe using categories. The more general is the concept, the fewer attributes it has. We agree Constantin Noica's opinion according to which Aristotle's Categories is a *treaty of fundamental questions*. (5)

In 1605 **Francis Bacon** developed his masterpiece „*Of the Advancement of Learning and Proficiency, Divine and Human*" trying to place knowledge around human understanding. He believed that knowledge acquisition is achieved with reason – revelation connection, thus establishing main classes: History, Philosophy and Poetry, each of them associating the psychological function it produces, namely *memory, logic and imagination*.


THE
Tvvoo Bookes of
FRANCIS BACON.
Of the proficience and aduance ment of Learning, diuine and humane.
To the King.
AT LONDON,
Printed for Henrie Tomes, and

Nicolas de Condorcet built a classification system based on decimal divisions. He designed five categories: *objects, methods, ways, uses of knowledge, and ways of accumulating knowledge*. Each of the five classes comprises in turn into ten concepts. His categorical system is considered the core of the classification theory for its exhaustive and exclusive variables between classes.

In 1870, **William Torrey Harris** made a catalogue of St. Louis Public School Library based on Bacon's classification, structured in reverse order, considering it was more logical, and therefore more relevant. Harris developed a classification system in which the influence of Hegelian philosophy is obvious. The draft of this classification appeared in the Journal of Speculative Philosophy published in 1870. (6) Harris considers that on the groundwork of any classification of sciences, there is a system of thinking and, in agreement with Hegel, that philosophy itself is an independent science. Thus the system that Harris proposes was structured as follows: *Philosophy, Theology, Social and Political Sciences, Natural Sciences, Arts, History.* An earlier attempt to represent knowledge in a schematic manner was *Encyclopaedia, or a Systematic Dictionary of the*

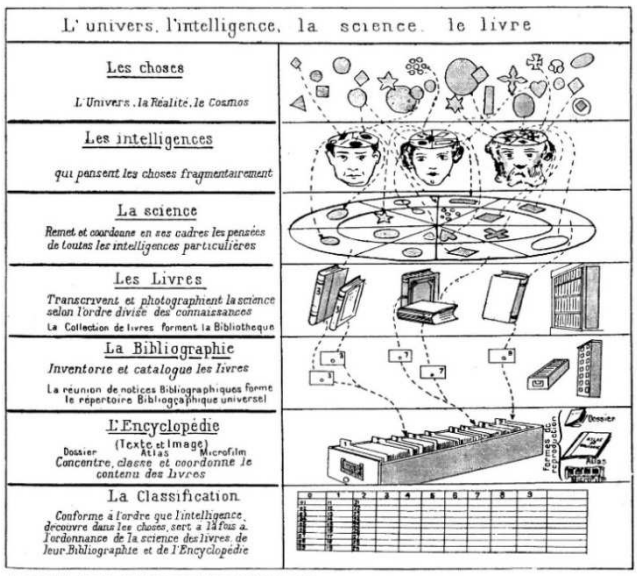


Sciences, Arts, and Crafts collective work, appeared in Paris in 1751, and coordinated by **Denis Diderot** and **Jean d'Alembert**. The information is alphabetically structured, the classified professions linked to the sections they related to, as well as all the information about the scientific fields. The success entitled the period *encyclopaedism.* (7)

To **Andre Marie Ampere** we owe the use of decimal notation of concepts. The idea was taken up by **Melville Dewey** in his *Classification and Subject Index*. The idea was to create a standardized system of shelf reference in libraries.

Auguste Compte in his positive philosophy course argues that History identifies three stages: “theological” explained through myths, "metaphysical” analysed through abstract entities, and “positive” analyses the relations that interconnects phenomena. Compte categorizes the science into *Mathematics, Astronomy, Physics, Chemistry, Biology, Social Sciences*. He considers that a good structuring of knowledge leads to scientific progress and the evolution of *education*. (8)

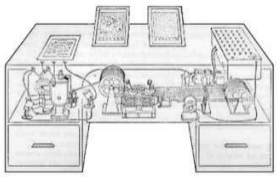
Paul Otlet designed a system of human knowledge classification, entitled *Mundaneum.* (9) The approach involved the annotation and networking of all the information noted on cards; the project started in 1895 and involved a huge team effort, and around 1977 contained almost 16 million records. Otlet’s method sets the *monographic principle* according to which the text splits into fragments with various references noted on standard cards, similar in function to *nodes* and *links* in the hypertext and later conceptual maps. The technique was based on the bibliographic description method to easily identify the information. The thematic organization of the system follows the corrected version of the Universal Decimal Classification developed by



Paul Otlet, Traité de Documentation (1934), p.41

Melville Dewey, whose extended formula appeared in the midst of 1904-1907.

Otlet's ideas explain the architecture of Memex - a plan of Vannevar Bush (who described in his article *As We May Think*, the mechanical machine that he imagined, built on microfilm rapid selector, materialized in the late 1930s).



His statement was the clearest evidence of the usefulness of Memex: *"A Memex is a device in which an individual stores all his books, records, and communications, and which is mechanized so that it may be consulted with exceeding speed and flexibility. It is an enlarged intimate supplement to his memory."* (10)

Bush considered that method of indexing the documents recorded by Memex has provided effective networking and allow quick retrieval of information. The architecture assumed by Bush should have led to a form of encyclopaedic representation of universal knowledge.

II. Knowledge representation languages

2.1 The Universal Decimal Classification

Knowledge representation languages are domains construction instruments with a solid philosophical essence that starts from the idea that the humanity progress depends on a good scientific organization. Knowledge representation languages are the nucleus of lexicons and dictionaries structure, which organizes the notions according to language rules for an efficient search and retrieval; in addition, associates *semantic categories* in order to refine the search. The classification is based on a truthful philosophical foundation centered on the human being. In order to progress, people need systematized information, well-ordered and processed.

The idea of *order* shapes the progress. Information retrieval should provide the best meaning of the term sought after. This way, the indexing follows the reflection of a specific context (the reality that surround a sign) from which the language game is part. So, the *language game* and *linguistic context* must be conceptual reflected. The classification specialist with a distinctive knowledge sets the direction and *meaning* because *the meaning of a word is different from the word itself,* is *related to its use* is *multidimensional* and *related to the context,* thus results from the *use of language*. The meaning of words depends on their use in language; a word has different meanings because the *language games* are different. Its significance differs on its role in different *forms of life* in different *communities* and *discursive practices*. Through language games

we have access to concept as Wittgenstein says in his *Philosophical Investigations. Language games* develop to link to specific *forms of life* and used in the same discursive practices, have the same meanings.

The Universal Decimal Classification (UDC) is a hierarchical multilingual *schema of concepts* organized in discipline-based classes that differ in *attributes* and *differences*. Is an artificial language because assign a *mathematical code* to each concept. These are called *indices* (they refer to the object that it denotes) (11) and enable the expansion from general to particular. Under this language, knowledge is divided into 10 classes, each class subdivided into its logic structure: 0-General, 1-Philosophy, 2-Religion, 3-Social Sciences Theology, 5-Mathematics, Natural sciences, Applied sciences, 6-Technical Medicine 7-Art. Architecture. Music. Games. Sport, 8-Literature, Linguistics, 9-Geography, History. A UDC *sign* is composed of a *general UDC number* which denotes the basic notion with *common auxiliary* number (expresses different relations between subjects) and special auxiliary numbers (concepts related to materials, relations). *Common auxiliary numbers* apply to all classes and are combined using mathematical signs (+ plus sign, denoting coordination; /stroke sign, denoting consecutive extension; : - colon sign, denoting simple relation between concepts; [ ] square brackets, denoting subgrouping). UDC system includes auxiliary signs of place (indicate the geographical region), race (denoting human race), form (the document form-manual, encyclopedia), language (the document language), time (the century, the decade of time). For example a UDC number for *Political Science University Handbook* is - 32(075.8). With the support of this auxiliary all the ambiguities in natural language fade away. But there is a semiotics of decimal classification; the subject *power* indicates the math fields, expressed through 511.132 or political science field, expressed through 342.5. We may notice that “power” has different meanings, it belongs to different discursive domains because the *language games* are different; we may talk about *power* in math science, in physics, in political science. In knowledge representation we are preoccupied by the functions of language, by the concept’s significance as domains’ basis.

The uniqueness of the expression is reflected by the different codification of the two meanings of the concept. Our intention is to underline the relation between the arithmetic languages with the info-documentary disciplines.

Solomon Marcus emphasizes the cultural and transdisciplinary value of mathematical language "… *through social interaction with humanities, mathematics and computers acquire for a diverse public a cultural value”*. (12) Mathematical language in relation to methods of organizing knowledge contributes to spectacular advances. It can provide information about user behavior, specific terms, and fields of knowledge. In order to improve the information architecture, specialists use conceptual modelling.

References

(1) *BibleGateWay*. [Online]. [Accessed 12 October 2014] Available at: http://www.biblegateway.com/passage/?search=Apocalipsa+22,Revelation+22&version=RMNN;KJ21;CEV;NIV

(2) *The Tree of Life*. [Online]. [Accessed 12 October 2014] Available at: http://www.newworldencyclopedia.org/entry/Tree_of_Life_(Judeo-Christian)

(3) Porphiry. In: *Stanford Encyclopedia of Philosophy*. [Online]. [Accessed 12 October 2014] Available at: http://plato.stanford.edu/entries/porphyry/

(4) *Enciclopaedia Universalis*. Corpus 18: Pheniciens-Proclus. Paris: Encyclopedia Universalis Paris SA., 1990, p. 741-742.

(5) ARISTOTEL. *Categorii*. Bucureşti: Humanitas, 1994, p. 126.

(6) BROWN, James D. *Manual of Library Classification and Shelf Arrangement*. London: Library Supply Company, 1898, p. 63.

(7) *Enciclopedia de filosofie şi ştiinţe umane*. Bucureşti: De Agostini, 2004, p. 85, 170, 233, 280-281.

(8) SHIELDS, Charles W. *The order of Science*. New York: Charles Scribner's Sons, 1882, p. 8.

(9) *Mundaneum: Centre de l'archives et space d'exposition*. 1882. [Online]. [Accessed 12 October 2014] Available at: http://www.mundaneum.be/

(10) BUSH, Vannevar. As we may think. In: *The Atlantic Monthly*, iunie, 1945. [Online]. [Accessed 12 October 2014] Available at: http://www.theatlantic.com/past/docs/unbound/flashbks/computer/bushf.htm

(11) PIERCE, Charles S. *Semnificaţie şi acţiune*. Bucureşti: Editura Humanitas, 1990, p. 277.

(12) MARCUS, Solomon. *Singuratatea matematicianului*. 2008 [Online]. [Accessed 12 October 2014] Available at: http://www.academiaromana.ro/discursuri_receptie/d2008/SMarcus-discurs.doc