



A SHANNON'S THEORY OF KNOWLEDGE

Benoît Le Blanc

*Institut de Cognitique, Université de Bordeaux 2, 33076 Bordeaux cedex,
France, leblanc@idc.u-bordeaux2.fr*

Jean-Louis Ermine

*Institut National des Télécommunications, 91011 Evry cedex, France, jean-
louis.ermine@int-evry.fr*

In : *Creating Collaborative Advantage through Knowledge and Innovation*, World Scientific Publishing Company Pte Ltd, Feb 2007, ISBN 978-981-270-451-1, pp 51-68

Our so-called “Information Society”, is gradually changing into a “Knowledge Society”. The Shannon’s Theory provided a formal and mathematical framework for information. It was very fruitful for avoiding ambiguity on the concept of information. This paper proposes a transposition of this theory for knowledge. From the three axes of a formal model designed for knowledge engineering (information, sense and context), three quantitative measures are proposed to get a measure of the quantity of knowledge of a system. This notion permits to consider applications as the cognitive measure of a web site, of a knowledge community (community of practice...).

1. Introduction

There is now a tremendous focus and lot of energies spent around the notion of knowledge. Of course, meanings around this “buzz word” are extremely various, but it is meaningful of a fundamental change of our societies. We can distinguish several meanings around the concept of knowledge.

- Knowledge Society

This is a concept popularised by nations, international organisations... (Anonymous, 2000). [ICT] (Norris, 2004) (Department of Economic and Social Affairs [DESA], 2005). (Mansell & Wehn, 1998; Corniou, 2002; Anonymous, 2001).

- Knowledge Based Economy

This is an new theory in economics developing the notion of the knowledge as an economic good; Precursors are Nelsons (1959), Kenneth Arrow (1962) Herbert Simon (1982), and especially F. Machlup (1984). The modern theory is exposed in the pioneer book of D. Foray (2004) and experienced in Organisation for Economic Co-operation and Development [OECD] publication (Anonymous, 2004). See also in France the Commissariat Général au Plan [CGP] (Anonymous, 2003).

- Knowledge Management (KM)

This is a now very increasing domain that becomes complex, see for instance (Bollinger, 2001).

- Knowledge Engineering

Knowledge Engineering is design methodology for knowledge based systems (Schreiber, 1999; Studer, 1998)(Dieng et al., 2000).

- Information vs. Knowledge

We may conclude that in the concepts of Knowledge Society, Knowledge Economy, Knowledge Management, Knowledge Engineering, ICT is always strongly present. But ICT is dedicated to information processing.

For more than 50 years, information is a very well known and very definite object, notably by the theory of Shannon that gives an operative mathematical definition that solved the ambiguousness problem concerning its nature (Shannon, 1949). What about knowledge? How can one say that ICT processes knowledge? Can we solve the ambiguousness between information and knowledge, and bring a formal answer, even though necessarily partial?

The Shannon's theory of information is not only a very powerful technical tool, but also a very powerful metaphoric tool, (when used scarcely) (Moles, 1975; Bounoux, 1993; Eco, 1972). An extension of the Shannon's theory to knowledge would give a formal relationship between knowledge and information, while providing, in the same way, the fertile metaphors, if they are used with discernment. We propose a sketch of what could be a Shannon's theory of knowledge.

2. Knowledge corpus

○ A formal model of Knowledge Corpus in KM

This model (called AIK model) is an attempt to provide sound basis for the definition of the knowledge capital of organised system. It is described in (Ermine, 2005). This is a mathematical formalism, based on set and morphism theory, general enough to include most of the basic well known concepts of Knowledge Management (Le Moigne, 1990; Morin, 1986; Nonaka & Takeuchi, 1995; Wenger, 1998; Drucker, 1959; etc.). The formal model is summarised in the diagram of Figure 1.

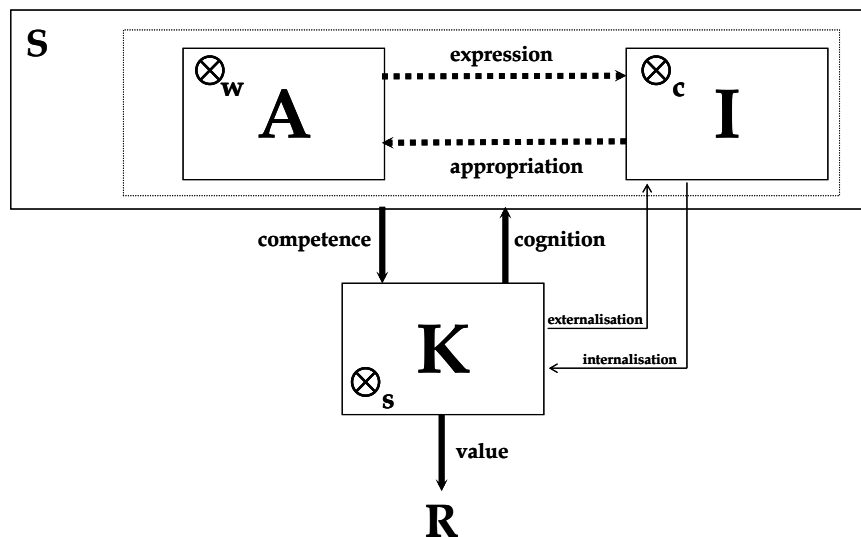


Figure 1: A formal model for KMS

S is the Knowledge Management System composed of the systems of knowledge workers networks A, and the Information System I. K is the Knowledge Capital. \otimes_w is the Wenger's operator of knowledge community aggregation, \otimes_c and \otimes_s are the combination and socialisation Nonaka's operators, there are also the externalisation and internalisation Nonaka's functions. Competence and Cognition are the cognitive functions of the KMS, according to Edgar Morin's theory of knowledge. Expression and appropriation are for knowledge workers, in their relationship with information system.

The value function is a real valuation function that permits to evaluate the added value brought by the KMS.

In the AIK model the function value is essential. It may be expressed as the added value in term of “knowledge capital”; it is the hypothesis of knowledge economy (OECD, 2004). as a strategic assessment function, which gives a grade on a scale of “criticality” (or of risk assessment of the risk) for the knowledge capital (Aubertin, 2006).

The proposed approach is quite different. It starts from the elementary idea, for example, that a data base is made to collect data, and that the value function grows if the quantity of information accumulated in this basis grows. However one knows, thanks to the theory of information, to give a meaningful measure (otherwise applicable) of the quantity of information, calculated in bytes for example. Can one have a similar measure for a knowledge capital? We try here to give some answers to this question.

- A formal model for Knowledge Corpus in Knowledge Engineering

The Knowledge Macroscopic is a tool to structure the knowledge capital of an organised system. That is a kind of knowledge theory that involves a lot of different aspects that have been studied on knowledge and information through times. It has been fully described and justified in (Ermine, 1996, 2002, 2003).

It is based on the “semiotic hypothesis”, considering that knowledge is perceived as a sign, which includes *information* (what is the form, encoded or perceived, of the sign sent to my perception?), *sense* (what semantic representation is generated by the information in my mind?), and *context* (what environment is influencing the sense generated from the information?). Knowledge is information that makes sense in a given context. If the notion of information is clear, referring to Shannon’s theory, the notion of sense (signification or semantic), and the notion of context are far less clear. Then we have to give “measures” for what is supposed to be sense or context. It is not a simple task. We will give some first propositions in that direction.

2.3. Definitions and notations

According to semiotic theory, as quoted above, knowledge can be split into three parts, corresponding to the three points of views of information, sense and context:

I is the Information space, S_e the semantic (sense) space, and C_o the space of context. We have:

$$K = I \times S_e \times C_o$$

$$\forall k \in K, \exists i \in I, s \in S_e, c \in C_o : k = (i,s,c)$$

Then, for $k \in K$, we can define three real valued functions:

$$\text{Val}_I : I \longrightarrow \mathbf{R}$$

$$\text{Val}_S : S_e \longrightarrow \mathbf{R}$$

$$\text{Val}_C : C_o \longrightarrow \mathbf{R}$$

When valuating respectively the value of information, the semantic value, and the contextual value of knowledge, by composition we have the global value of knowledge by:

$$\text{Val}(k) = F(\text{Val}_I(k), \text{Val}_S(k), \text{Val}_C(k))$$

We generalise that definition of knowledge value to a knowledge set, we will call "Knowledge Corpus":

Definition: A Knowledge Corpus H is an element of $\mathcal{P}(K)$ (subsets of K). And by extension, we define:

$$\text{Val}_X(H) = \int_{k \in H} \text{Val}(k) d\mu$$

for a measure μ such that t(he integral is convergent).

Then we can develop a "limited expansion" of the Val function, $\text{Val} = F(\text{Val}_I, \text{Val}_S, \text{Val}_C)$ that we will limit to first order (with an unknown operator \otimes to be defined):

$$\text{Val}(H) = \text{Val}_I(H) \otimes \text{Val}_S(H) \otimes \text{Val}_C(H) + o(H)$$

We will simplify by vanishing the residual term, and we will suppose that:

$$\text{Val}(H) = \text{Val}_I(H) \otimes \text{Val}_S(H) \otimes \text{Val}_C(H)$$

Intuitively, this means to suppose that information, sense and context may vary independently: one can imagine a knowledge corpus that has a

lot of information and few sense (a telephonic directory, for example), a lot of sense and little information (a proverb, for example), a lot of sense but little context –to the sense context of usage, cf. *Infra-* (an intimate diary, or some blogs, for example), little sense and a lot of context (a tenacious and rife rumour, for example) etc.

3. The measure of the quantity of information of a corpus

The Shannon's theory of information permits to define what a quantity of information is (Shannon, 1948; Shannon & Weaver, 1949). This theory is a probabilistic point of view on information produced by a system.

During the communication process, the receptor is waiting of a certain message. Let's take the case of a traffic light. When a person looks at this light, he already has an idea of messages transmitted by this light. A priori, he is unaware of what message is precisely going to be transmitted. However, thanks to his experience, he expects to receive some messages with different probabilities.

If we consider that all messages that can be transmitted potentially by the traffic light had the same probability, the probability of each of them would $P = 1/8$; the 8 possible cases being simultaneous switch on of 0 lamp (1 case), 1 lamp (3 cases), 2 lamps (3 cases) or 3 lamps (1 case). As this value 8 also can be also written as 2^3 and that it is a binary coding (lamps are either on or off), the quantity of information associated to a traffic light, without other precision, is 3. Intuitively it means that there are three pieces of information, every lamp being considered independent of the two others.

In the real word, it is of course different. Only four cases occur (let's take a virtual example in France, it is different in other countries): the red lamp on (45% of cases), the orange lamp on (9,5% of cases), of the green lamp on (45% of cases), or no lamp on (0,5% of cases). This example reflects the reality of information systems of for which the equally probable is a configuration... very unlikely! The quantity of Q information of a message m with a probability of occurrence P is given by the formula (called entropy formula):

$$Q(m) = -P \cdot \log_2(P)$$

In that case, the calculation gives: $Q(m) = 1,4$

One sees on this example that there is a one to one relation between the uncertainty, in terms of probability, of a receptor, relatively to a determined message, and the quantity of information contained in this message. The notion of quantity of information is replaced then advantageously by the notion of entropy that is a mean information quantity in the sense of probabilities, calculated on the set of messages. So if p_i is the probability of occurrence of the message m_i , the entropy is by definition the mathematical variance:

$$H = -p_1 \log(p_1) - p_2 \log(p_2) - \dots$$

This notion requires an important commentary. The more entropy is low, the more informative is the system. In fact the more a message is unlikely, the more it is informative (the message of the assassination of a president is more informative than the message of the fact that there is no snow in Paris in summer!). Then entropy takes the same signification than information, as the possibility of choice for information source, or mean of occurrence probabilities of a set of messages.

In information theory, the introduction of the entropy function was a considerable innovation that was incredibly fruitful. In fact, Shannon introduced this function for simple reasons of regularity, as soon as we have a distribution of probabilities on a set of events (Shannon & Weaver, 1949).

The usual Shannon's entropy is the requested function for $Val_1(H)$ for a knowledge corpus H .

4. The measure of the quantity of sense of a corpus

4.1. Definitions and notations

With the "bit" ("binary" unit or "binary digit"), we have defined an elementary information unit.

The elementary unit of sense, in linguistics, is called "seme", it is often represented by one significant term, framed of signs /, in order to distinguishes it from the common word (for example /beautiful /, /feminine /, /white / etc...). The semantics of a knowledge corpus is represented therefore by the semes that it contains, but also by semantic

links that link these semes. It is that that one calls a semantic network. Works on semantic networks are plethoric, in cognitive sciences, in linguistics or in artificial intelligence. One can define an elementary unit of sense therefore like a set of two semes linked by semantic link. The sense of a corpus K is defined by combinations of elementary units of sense.

The semes of K constitute a finite set of elements (elements or significant terms) S .

Let Ind a function, called indexation function:

$$\text{Ind} : \mathcal{P}(K) \longrightarrow \mathcal{P}(S)$$

that associates to every knowledge corpus H , element of $\mathcal{P}(K)$, a set of semes, element of $\mathcal{P}(S)$.

A semantic graph is a set (V,E) where V is a subset of elements in S , called nodes of the graph, and E a subset of elements of $V \times V$, called vertices or links.

A graph, (we consider only finite graph, with node numbered from 1 to n) is defined by its incidence matrix $P = [p_{i,j}]$: This is a square matrix (n,n) , such that $p_{i,j} = 0$ if there is no link between the node i and j , and $p_{i,j} = 1$ otherwise.

A path of length n in a graph is a sequence (s_0, \dots, s_n) , such that (s_i, s_{i+1}) is a vertex of the graph for every i ; s_0 is the origin of the path, s_n is the end.

A graph is called connected if any two nodes may be linked with a path.

We denote $p_{i,j}(n)$ the number of path of length n that starts from the node i and ends at node j . This is also the coefficient at line i and column j of the matrix P^n : $p_{i,j}(n) = (P^n)_{i,j}$.

A function of semantic graph construction is a function:

$$\Gamma : \mathcal{P}(K) \longrightarrow S \times S^2$$

that associates to every knowledge corpus H a semantic graph $\Gamma(H)$, such that $\text{Pr}_1(\Gamma(H)) = \text{Ind}(H)$. (This is a semantic graph where all the nodes constitute the set of semes indexing H , $\text{Ind}(H)$).

4.2. Quantitative characterisation of a semantic graph: Gurevich entropy

The semantic graph of a knowledge corpus characterizes “the semantic path”, the “semantic random walks” that are possible in the corpus. Hence, the topology of the graph characterizes the semantic complexity of the corpus.

Very similar to the information theory, there exists a very developed theory that characterises the random walks in the graph (hence the semantic random walks in the semantic graphs), this is the theory of graph entropy (Simonyi, 1995).

We will give an approach developed by Gurevich (1969, as cited in Ruette, 2001).

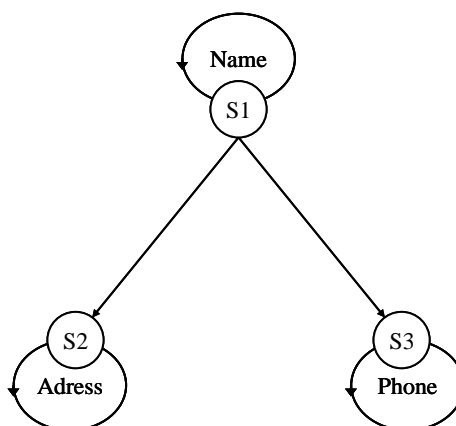


Figure. 2 – Semantic graph of a phone book

Let G a graph, P its incidence matrix. Let recall that $p_{i,j}(n) = (P^n)_{i,j}$.

Gurevich entropy is defined as:

$$h(G) = \limsup_{n \rightarrow +\infty} \frac{1}{n} \log(p_{i,j}(n)) = \limsup_{n \rightarrow +\infty} \frac{1}{n} \log((P^n)_{i,j})$$

For a finite connected graph, this number does not depend on i or j . It represents the exponential growth rate of the number of path with fixed extremities.

To better understand that notion of graph entropy, let's take a very simple example: a phone book. The semantics of a phone book very simple, and given by the graph of figure 2.

Its incidence matrix is:

We can easily calculate for the incidence matrix P that $P^n = (I+A)^n = I + nA$, hence $p_{i,j}(n) = 0, 1$ or n , then $h(G) = 0$, the entropy of that graph is null. Intuitively, it is true that the semantics of a phone book is very poor!

If we try to add a few semantics more, by designing an inversed phone book, where you can find the name from the phone number, we obtain a new semantic graph . The graph is only augmented with a inverse link from the S3 node to the S1 node. The new calculation shows that $P^n = 2^{n-1} (A + A^2)$ Hence $p_{i,j}(n) = 2^{n-1}$ or 0, and then:

$$h(G) = \lim_{n \rightarrow +\infty} (n-1/n) \log(2) = \log(2) = 1$$

The semantics of the new phone book has increased of one bit of sense!

The Gurevich entropy is the requested function for $Val_S(H)$ for a knowledge corpus H .

5. Measure of the context of usage of a knowledge corpus

The third part of our knowledge unit is about the measure of its use. We consider this measure of the context as the use that makes people concerned in a knowledge corpus, but also in relations that settle between these people. This idea considers that a pertinent knowledge will be shared between its holder and his knowledgeable neighbourhood or that conversely a weakly distributed knowledge corresponds to something obsolete, uninteresting or inappropriate. We are conscious to overlook, by this approach, all powerful knowledge that remains kept secret. On the other hand, our model may perfectly apply to knowledge that one tries to distribute, via publication media, paper or electronic, via the media for verbal communication, etc. and that won't have any echoes if they don't cause a lot of interest in knowledgeable communities.

The context of use of a knowledge corpus or its diffusion potential of diffusion is studied here through the existing acquaintance network

between elements, holders and users of knowledge, individuals, groups, or systems. This kind of network corresponds to the social networks studied in psycho-sociology and more lately in graph theory.

- Social networks

The social networks called “small worlds” were initiated by the American psycho-sociologist Stanley Milgram. (Milgram, 1967). He postulated then that every person is linked to any other individual while achieving an average of six jumps, materializing thus the theory of “six degrees of separation”, invented by the Hungarian writer Frigyes Karinthy in a novel of 1929 entitled “Chains”. These experiences have been refuted lately by J. Kleinfeld (2002). Nevertheless, with his very simple and attractive protocol, Milgram shows that it is possible to find an experimental measure of the distance between two random people, by counting the number of necessary mediators to establish a chain between them.

This idea has been studied for specific populations, as for the one of mathematician researchers in mathematical or Hollywood actors. In this case the connectivity between two people corresponds to the realization of a common task: the writing of a scientific article or the apparition in a same movie. See for instance the web site of the Erdős Number Project is: <http://www.oakland.edu/enp/>. or: <http://smallworld.columbia.edu/>.

- Hierarchical Small World networks

The underlying model for all these social networks is a graph where each individual (the nodes) is considered regarding his connections (links). From this graph structure, it is possible to calculate some formal values (Hayes, 2000).

According to a similar approach, there are recent studies on other kinds of networks: subway stations, phone connections, flight connections, but also neurons in simple organisms, diffusion of epidemics, Web pages or specific sites etc. In short, all observable network type, produced naturally or generated by a human activity, may be a subject for this kind of measures (Gaume, 2004). It is remarkable that all these studies assign to all these very various graphs the similar properties. One designed a specific class for these graphs: the Hierarchical Small-World Networks [HSWN] (Albert & Barabási, 2002). For example, Adamic and al. advanced in 1999 that the Web is of that

type (Adamic et al., 2000) and that average 10 clicks only separated any pair of pages, very weak values compared to billions of pages that compose the Web.

To study HSWNs, Watts and Strogatz (1998) proposed to compare them on the one hand to the random graphs¹ and on the other hand to regular graphs². The two studied parameters are the distance between nodes of the network (global diameter of the graph) and the level of nodes clustering (local coefficient of clustering).

Watts and Strogatz proved that small-world networks have the surprising property to be locally dense (as the regular graphs) and to have a relatively short path length (as the random graphs). A small-world graph is thus halfway between a random graph and a regular graph, thus combining properties of local regularity and global disorder.

- The scale-free networks

While working on the more general problem of the construction and the organization of networks, Barabási and Bonabeau (2003) added a third relative metrics to the hierarchical distribution of links on these networks. Indeed the growth observed on certain kind of small-world networks (as the Web) shows a property of preferential attachment that cannot be modelled by the simple mean of path lengths, and clustering of nodes. This new property considers the fact that a new node will stand preferentially connected to nodes that are already greatly connected. This model of accumulated advantage produces a network where most of nodes have few links and some nodes have a lot. The distribution of this nodal connectivity (the probability to have k neighbours) is described by a power law, whereas for the random graphs this probability is described by a Poisson's law. According to Barabasi the coefficient of this power law is a strong characteristic of the network. For the small-world networks, this coefficient is always a number between 2 and 3.

¹ An random graph is a graph with N nodes connected by n links, chosen randomly with probability p among the between uncertainly among the $N(N - 1)/2$ possible links. This graph has $p \cdot N(N - 1)/2$ links randomly distributed

² the regular graphs have their nodes distributed regularly and each node is linked to its i nearest neighbours, " i " being here the degree (distributed regularly) of connection of every node

5.4. Measure of the context of a knowledge corpus

The scale-free networks seem to be a good candidate to model the social networks that use a given knowledge corpus. It is on the basis of this distribution of connection degrees within a network of users that we propose to construct the measure of the context of a knowledge corpus.

We start from the graph formed by the users of a knowledge corpus, and take as connection between two users the representation of an existing link (an e-mail exchange for example). The obtained network is the network of the users of the considered knowledge corpus. According to the above theory, it is a scale-free network. It is then possible to characterize this graph with the distribution law of distribution of the connectivity degree of the different nodes.

If $p(k)$ is the probability for a node of having k neighbours, the usual Poisson's law gives:

$$p(k) \sim g \exp(-k).$$

For a scale-free network, that law is simplified. The power law gives:

$$p(k) \sim g k^a.$$

Then we can easily define and calculate entropy for the scale-free network with the formula:

$$\text{Val}_C(H) = - \sum_k p(k) \log(p(k))$$

This entropy is a good characteristic of the network of the users of the knowledge corpus. We will take it as the measure of the context of this corpus.

This entropy is the requested function for $\text{Val}_C(H)$ for a knowledge corpus H .

6. Perspectives

We now have a proposition for calculating the entropy of a knowledge corpus, by calculating respectively the information entropy (Shannon's entropy), the semantic entropy (Gurevich entropy of the semantic graph of the corpus), and the context entropy (scale-free characteristic of the

user network of the corpus). The combination of those three entropies is not yet clear, and is currently under research, to define global knowledge entropy of the corpus.

The definition of a measure of the quantity of knowledge of a corpus is not only a theoretical objective. Defining a quantity of knowledge of a corpus can bring to numerous innovations, as by example:

- “Scoring” for information retrieval. Search engines, from a research using key words, classify corpora found according to their relevance. This relevance can be calculated on the content (occurrence of terms, for example), on the contextual value of the site (number of connections, for example). The finer quantification of the content in semantic term, or indicators on its context of usage, would permit to give a classification far more interesting.

- Improving the content of a document. Information theory permits to find an optimal coding of an informational corpus. Analogically, if one has a measure of the quantity of knowledge of a corpus, of a document, for example, it may lead to rewrite the original document, so that it would be better understood regarding the signification of reference.

- Supervising knowledge communities. The context of a knowledge corpus is essentially valued by its context of usage. Refined indicators users communities for a knowledge corpus permits to better know these communities and to facilitate their improvement.

These some examples show all the potential that one may have with the notion of measure of the quantity of knowledge of a corpus.

References

- Adamic, L.A., Huberman, B.A., Barabási, A.L., Albert, R., Jeong, H., & Bianconi, G. (2000). Power law distribution of the world wide web. *Science*, 287, 2115a.
- Albert, R., & Barabási, A.L. (2002). Statistical mechanics of complex networks. *Review of modern physics*, 74, 47-97.
- Anonymous. (2000). *Conclusions de la Présidence Conseil européen de Lisbonne 23 et 24 mars 2000*. Public report DOC/00/8, European Community, Brxelle, Belgique, from <http://europa.eu.int>.
- Anonymous. (2001). La société du savoir. *Sciences Humaines Hors-Série*, 32.
- Anonymous. (2003). *La France dans l'économie du savoir : pour une dynamique collective*. Public report, Commissariat Général au Plan [CGP], La Documentation Française, Paris, France.

- Anonymous. (2004). *Gestion des connaissances : mesurer la gestion des connaissances dans le secteur commercial, premiers résultats*. Public report, OCDE, Paris, France.
- Arrow, K. (1962). Economic Welfare and the Allocation of resource of Inventions. In Nelson (Ed.), *The Rate and Direction of Inventive Activity: Economic and Social Factors*. US: Princeton University Press.
- Aubertin, G. (2006). Knowledge mapping: a strategic entry point to knowledge management. In *Trends in Applied Knowledge Management*. US: Hermes Penton, to appear.
- Barabasi, A.L., & Bonabeau, E. (2003). Scale-Free Networks. *Scientific American*, 288(5).
- Bollinger, A.S., & Smith, R.D. (2001). Managing Organizational Knowledge As a Strategic Asset. *Journal of Knowledge Management*, 5(1).
- Bougnoux, D. (1993). *Sciences de l'information et de la communication*. Paris, F: Larousse.
- Corniou, J.P. (2002). *La société de la connaissance : nouvel enjeu pour les organisations*. Paris, F: Hermès science publications.
- Department of Economic and Social Affairs [DESA]. (2005). *Understanding Knowledge Societies*. Public report ST/ESA/PAD/SER.E/66, United Nations, New-York, US.
- Dieng, R., Corby, O., Giboin, A., Goliebowska, J., Matta, N., & Ribière, M. (2000). *Méthodes et outils pour la gestion des connaissances*. Paris, F: Dunod.
- Drucker, P. (1959). *Landmarks of Tomorrow*. New-York, US: Harper and Brothers.
- Eco, U. (1972). *La structure absente, Introduction à la recherche sémiotique*. Paris, F: Mercure de France.
- Ermine, J-L. (1996). *Les systèmes de connaissances*. Paris, F: Hermès sciences publications.
- Ermine, J-L. (2002). *Initiation à la méthode MASK (CD-ROM)*. Troyes, F: UTT Press.
- Ermine, J-L. (2003). *La gestion des connaissances*. Paris, F: Hermès sciences publications.
- Ermine, J-L. (2005). A theoretical and formal model for a Knowledge Management System. In Dan Remenyi (Ed.), *Second International Conference on Intellectual Capital and Knowledge Management*, 187-199, Dubai.
- Foray, D. (2004). *The Economics of Knowledge*. Cambridge, US: The MIT press.
- Gaume, B. (2004). Balades aléatoires dans les Petits Mondes lexicaux. *I3 Information, Interaction, Intelligence*, 4(2).
- Gurevich, B.M. (1969). Topological entropy of enumerable Markov chains (Russian). *Dokl. Akad. Nauk SSSR*, 187, 715–718. English translation *Soviet Math. Dokl.*, 10(4), 911–915.
- Hayes, B. (2000). Graph Theory in Practice. *American Scientist*, 88(1), 9-13.
- Kleinfeld, J.S. (2002). Could it be a big world after all? The Six Degrees of Separation Myth. *Society*, 39(61).
- Le Moigne, J-L. (1990). *La modélisation des systèmes complexes*. Paris, F: Afcet Systèmes, Dunod.

- Machlup, F. (1984). *Knowledge, its creation, distribution and economic significance, vol III*. US: Princeton University Press.
- Mansell, R., & Wehn, U. (1998). *Knowledge Societies Information Technology for Sustainable Development*. Oxford, UK: Oxford University Press.
- Milgram, S. (1967). The small world problem. *Psychol. Today*, 1, 60–67.
- Moles, A. (1975). *Préface à "Théorie mathématique de la communication" par C. Shannon et W. Weaver*. Paris, F: Retz-CEPL.
- Morin, E. (1986). *La méthode, 3. La connaissance de la connaissance*. Paris, F: Editions du seuil.
- Nelson, R.R. (1959). The Simple Economics of Basic Scientific Research. *Journal of Political Economy*.
- Nonaka, I., & Takeuchi H. (1995). *The Knowledge-Creating Company*. Oxford, UK: Oxford University Press.
- Norris, P. (2004). *Building Knowledge Societies*. Public report, UNESCO, New-York, US.
- Ruette, S. (2001). *Chaos en dynamique topologique, en particulier sur l'intervalle, mesures d'entropie maximale*. Doctoral dissertation, Université Aix-Marseille II, Marseille, France.
- Schreiber, G., Akkermans, H., Anjewierden, A., de Hoog, R., Shadbolt, N., de Velde, W.V., & Wielinga, B. (1999). *Knowledge Engineering and Management: The CommonKADS Methodology*. Cambridge, US: The MIT Press.
- Shannon, C.E. (1948). A Mathematical Theory of Communication. *The Bell System Technical Journal*, 27, 379–423, 623–656.
- Shannon, C.E., & Weaver, W. (1949). *The Mathematical Theory of Information*. Urbana, US: University of Illinois Press.
- Simon, H. (1982). Models of Bounded Rationality. *Behavioural Economics and Business Organisation*, 2.
- Simonyi, G. (1995). Graph entropy, a survey, *Combinatorial Optimization. DIMACS, Series on Discrete Math. and Computer Science*, 20, 399-441 (W. Cook, L. Lovasz, P. Seymour eds.), AMS.
- Studer, R., Benjamins, R.V., and Fensel, D. (1998). Knowledge Engineering: Principles and Methods. *Data & Knowledge Engineering*, 25, 161-197.
- Watts, D.J., & Strogatz, S.H. (1998). Collective dynamics of "small-world" networks. *Nature*, 393, 440- 442.
- Wenger, E. (1998). *Communities of Practice : Learning, Meaning and Identity*. New York, US: Cambridge University Press.