

Measuring Knowledge: A Quantitative Approach to Knowledge Theory

Fred Y. Ye^{1,2}

¹School of Information Management, Nanjing University, Nanjing, China

²Jiangsu Key Laboratory of Data Engineering and Knowledge Service, Nanjing, China

Email address:

yyc@nju.edu.cn

To cite this article:

Fred Y. Ye. Measuring Knowledge: A Quantitative Approach to Knowledge Theory. *International Journal of Data Science and Analysis*. Vol. 2, No. 2, 2016, pp. 32-35. doi: 10.11648/j.ijdsa.20160202.13

Received: October 6, 2016; **Accepted:** December 2, 2016; **Published:** December 30, 2016

Abstract: By transferring the DIKW hierarchy to the concept of chain, namely data – information – knowledge – wisdom, the knowledge measure is set up as the logarithm of information, while the information is the logarithm of data, so that knowledge metrics are naturally introduced and the mechanism of Brookes' basic equation of information science is revealed. When knowledge is classified as explicit knowledge and tacit knowledge, qualitative SECI model is changed to quantitative triangle functions on explicit knowledge and tacit knowledge, where the former is measured by the logarithm of data and the latter is measured by the negative entropy of language. The author suggests to treat the unit of knowledge as kit, correspondingly, data as bit and information as byte.

Keywords: Data, Information, Knowledge, Knowledge Metrics, Knowledge Theory

1. Introduction

While knowledge management has increasingly become a hot area, knowledge becomes a scientific key. Although knowledge is well known as a popular concept, it is never defined scientifically and strictly. In Brookes' theory (Brookes, 1980-1981), Popper's scientific philosophy was applied and a basic equation linking knowledge to information was set up. However, there was no theoretical interpretation of interactive mechanism between information and knowledge, though Shannon information theory was well established (Shannon, 1948).

Meanwhile, as a qualitative model, Nonaka et al. (2000)'s SECI model (i.e. Socialization - Externalization - Combination - Internalization) is an influential model for describing organizational knowledge creation. It contains four modes of knowledge conversion between explicit knowledge and tacit knowledge, where S denotes Socialization that describes the dimension of tacit to tacit transfer, E means Externalization that describes the dimension of tacit to explicit transfer, C indicates Combination that describes the dimension of explicit to explicit transfer, and I is Internalization that describes the dimension of explicit to tacit transfer. However, there were

no quantitative relations in the SETI model.

Scientifically, in order to quantitative studies of knowledge, it is necessary to define knowledge and measure knowledge quantitatively. Referring the DIKW hierarchy model (Rowley, 2007; originated by H. Cleveland via M. Zeleny), I try to develop a mathematical method for measuring knowledge, along with my earlier studies (Ye, 1999, 2011).

2. DIKW Chain and Knowledge Metrics

Following the DIKW hierarchy model (Rowley, 2007), Data (D), Information (I), Knowledge (K), and Wisdom (W) together construct a pyramid structure, as shown as the Figure 1.

Although there is argument on DIKW hierarchy (Fricke, 2009), DIKW hierarchy can be transformed to a logic chain, as shown in the Figure 2.

Now the DIKW chain can be quantitatively processed. Let us introduce a median variable *i* as physical information and *J* as subjective information, according to Ye (1999), the logic of quantitative DIKW chain becomes: the objective data transfer into physical information (*i*) via natural transmission, so the *i* is checkable by physical instruments; the physical information (*i*) transfers into objective information (*I*) via

social transmission, thus the I is acceptable by subject; the objective information (I) transforms into subjective information (J) via subject absorption, so that the J bears subject value judgment. Then the subjective information (J) transforms into knowledge (K) via structurization and systemization. The overall applications of knowledge become wisdom (W). This is a transmission chain from objective to subjective side. D and I belong to physical objective field, while K and W belong to cognitive subjective field. It is an objective process from D to I, while it is a subjective process from K to W. There is a key transformation between I and K, where the subjective value judgment is generated.

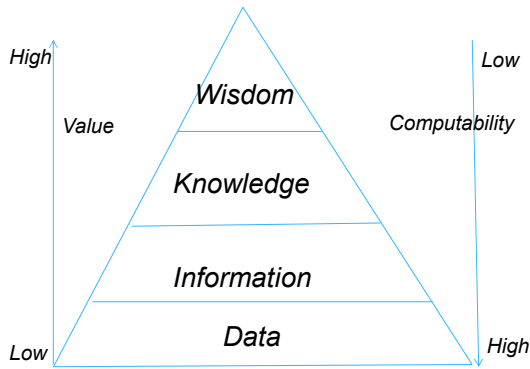


Figure 1. DIKW hierarchy.

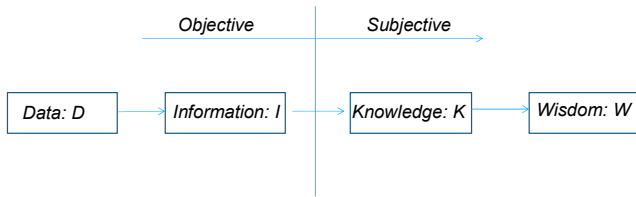


Figure 2. DIKW Chain.

Referring to Shannon information theory (Shannon, 1948), the physical information (i) is defined to be the logarithm of data (D) as

$$i = d \log D \quad (1)$$

in which d is the transformation coefficient from data to information.

From physical information (i) to objective information (I), a transmission channel is in need, where the information motion obeys the wave-heat equations (Ye, 1999, 2011). If there are information compression ratio p and lose ratio q in the transmission channel, there is $I=pqi$. Let $pq=b$, thus we have

$$I = bi \quad (2)$$

Transforming the objective information (I) to the subjective information (J) is a key. Following Brookes' information theory (Brookes, 1980-1981) and the principle of logarithmic perspective, with introducing value coefficient $v \in [0, 1]$ (matching Rescher's model), it is derived as

$$J = \log I^v = v \log I \quad (3)$$

Since valuable information increases knowledge (Ye, 1999), the unit knowledge increment should be proportional to the subjective information, yielding

$$\frac{dK}{dI} = kJ = k \ln I^v \quad (4)$$

where k is knowledge transformation coefficient of information. Therefore, knowledge (K) should be the integral

$$K = k \int J dI = k \int v \ln I^v dI = kvI(\ln I - 1) + K_0 \quad (5)$$

in which K_0 is the integral constant, representing original knowledge.

This is just the Brookes' basic equation of information science

$$K = K_0 + \Delta K; \Delta K = kvI(\ln I - 1) \quad (6)$$

where $\Delta K = kv(\ln I - 1)$ is the increment of knowledge. The process gives the mechanism of Brookes' equation.

As for the measurement units, since D uses bit and I does byte, it is suggested to apply 'kit' to be the unit of knowledge. When $k=1$, it means that 1 byte information can be transferred into 1 kit knowledge.

When an intelligent agent (j) has its knowledge elements as $a_{ik} \in [0, 1]$, there are knowledge vectors

$$A_j = (a_{i1}, a_{i2}, a_{i3}, a_{i4}, a_{i5}, a_{i6}, a_{i7}, a_{i8} \dots) \quad (7)$$

Defining its knowledge matrix as

$$K_j = A_j^T A_j = (r_{ik}) \quad (8)$$

where $r_{ik} = \sum_j a_{ij} a_{jk}$ are intelligent correlatives, the wisdom

(W) can be measured by intelligent grade as the trace of knowledge matrix

$$W_j = \text{tr}(K_j) \quad (9)$$

Therefore, it can be seen that this is a feasible foundation for a unified quantitative model linking information to knowledge based on the logic chain data-information-knowledge-wisdom.

3. Measuring Explicit Knowledge and Tacit Knowledge

Both explicit knowledge and tacit knowledge can be measured in the above quantitative way. However, there are differences between explicit knowledge and tacit knowledge.

In general, while explicit knowledge is codified, tacit knowledge is uncoded. Using the iceberg metaphor, explicit knowledge looks like the peak and tacit knowledge sinks under the sea, by 2/8 rule. So, it is estimated that

explicit knowledge can occupy about 20% and tacit knowledge 80% of the total knowledge volume.

In the SECI model, S (Socialization) is realized by sharing tacit knowledge face-to-face or sharing experiences in a community, which typically occurs in a traditional apprenticeship. E (Externalization) is realized by reading documents, where tacit knowledge is made explicit (knowledge is codified) to enable sharing. C (Combination) is realized by combining different types of explicit knowledge, which is collected and integrated from inside or

outside of organization and then combined, edited or processed to form new knowledge. I (Internalization) is realized by an individual, who learns explicit knowledge and will asset to organization then store as tacit knowledge. These four modes of knowledge dimensions with conversion repeats themselves in a spiral way, as shown as the Figure 3, where the sides of the square represent explicit and tacit knowledge while the inside cross lines mark individual and organization, respectively.

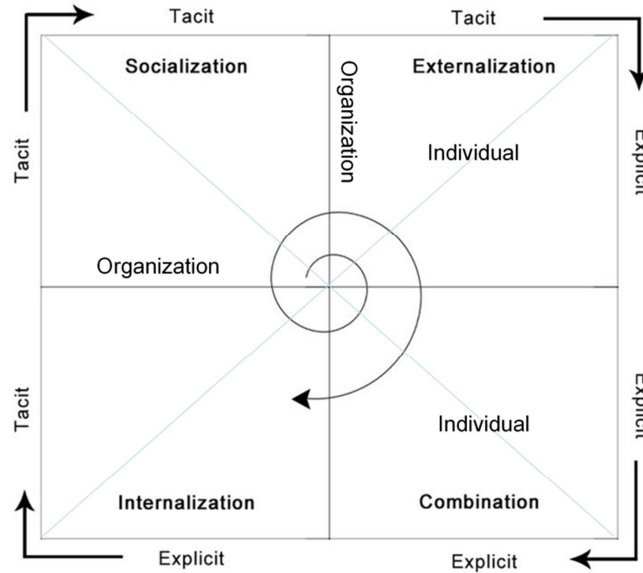


Figure 3. SECI model of knowledge conversion.

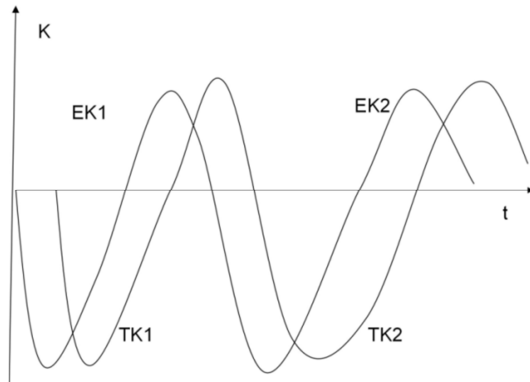


Figure 4. Modeling explicit knowledge and tacit knowledge.

There are two advantages of the SECI model: 1) it appreciates the dynamic nature of knowledge and knowledge creation, and 2) it provides a simple framework for management of the relevant processes. However, there are also disadvantages, in which the key issue concentrates on lack of quantitative analysis. Another issue is based on the study of Japanese organizations that heavily rely on tacit knowledge (Nonaka & von Krogh, 2009). To overcome the disadvantages, it is suggested to introduce a quantitative mathematical pattern as shown in the Figure 4.

In the Figure 4, EK1 and EK2 denote different explicit

knowledge while TK1 and TK2 do tacit knowledge. K means knowledge and t indicates time. A triangle function can be applied to simulate the process as follows:

$$K = a(t) \sin(EK_i) + (1-a(t)) \cos(TK_i) \quad (10)$$

where $a(t)$ is explicit ratio changed by time, $(1-a)$ means tacit ratio and j denotes sum of all items. If $a=0.2$, $(1-a)=0.8$.

When all knowledge becomes explicit knowledge, $a=1$, so that the explicit knowledge is measured by

$$K = \sin(EK_i) \quad (11)$$

When all knowledge changes to tacit knowledge, $a=0$, so that the tacit knowledge is measured by

$$K = \cos(TK_i) \quad (12)$$

4. Analysis and Discussion

The explicit knowledge may express order while the tacit knowledge shows chaos, so that the knowledge from information and data generated explicit knowledge and the entropy caused tacit knowledge. Thus, the explicit knowledge is proportional to the logarithmic logarithm of data by combining Eq. (1) and Eq. (5)

$$EK = c \log \log D \quad (13)$$

where c denotes a proportional constant.

Then the tacit knowledge is proportional to the negative entropy of language (S), which equals the logarithm of words (W)

$$TK = -S = -e \log W \quad (14)$$

where e denotes a constant.

The formula (14) looks like Boltzmann formula. This is another metaphor that physics merges into knowledge system. However, the knowledge system has its unique characteristics when the Eqs. (13) and (14) are inserted into Eq. (10)

$$K = ac \sin(\log \log D_j) + (1-a)e \cos(\log W_j) \quad (15)$$

where j denotes sum of all items.

Another important issue concerns information absorption while information transforms into knowledge. Although we could set k as knowledge transformation coefficient of information in Eq. (4) and calculate the absorption rate into k together, we can also set a specialized absorption rate separated from transformation coefficient if we need to discuss information absorption in the transformation process from information to knowledge. However, as the information absorption highly relies on existing subjective knowledge, absorption rate is a personally parameter and may not be an objective measurement.

Since knowledge measurement is a comprehensive and difficult issue in scientific metrics and knowledge management, few quantitative models are developed based on strict mathematical methodology. Here we try to set up a mathematical model.

Another related issue concerns innovation measurement in knowledge management, on which we mention Triple Helix (TH) proposed by Etzkowitz and Leydesdorff (1995). After it came, Triple Helix has quickly and widely affected the academic world, particularly activates innovation measurement. TH model has also set up quantitative mechanism of interaction among innovative entities and the dynamics with quantitative measures of innovation, based on information theory. Therefore, it will link to knowledge via information. After the Emergence of a Triple Helix as University-Industry-Government (Leydesdorff & Etzkowitz, 1996), the TH model guides us to study University-Industry-Government Relations as well as their actions and functions in innovation. Later, the Triple Helix as a model for innovation studies is emphasized that academic, industrial, and governmental institutions interact at both national and international levels (Leydesdorff & Etzkowitz, 1998). The TH model will also enlighten to study knowledge measurement.

5. Conclusion

In this article, a mathematical theory of knowledge is suggested to include two parts: the first part links the chain of data-information-knowledge-wisdom, and the second part links explicit knowledge and tacit knowledge. The

knowledge is estimated by the logarithm of information and information by the logarithm of data, while the conversion between the explicit knowledge and tacit knowledge are simulated by triangle functions.

It is necessary to develop quantitative metrics in knowledge theory, and the combination of qualitative concepts and quantitative measurement is important for knowledge research. The preliminary exploration here introduces a potential development of knowledge metrics in the future.

References

- [1] Brookes, B. C. (1980-1981). The Foundations of Information Science. *Journal of Information Science*, 1980, 2 (3-4), 125-133 (Part I); 1980, 2 (5), 209-221 (Part II); 1980, 2 (6), 269-275 (Part III); 1981, 3 (1), 3-12 (Part IV).
- [2] Etzkowitz H, Leydesdorff L (1995). The Triple Helix--University-Industry-Government Relations: A Laboratory for Knowledge-Based Economic Development. *EASST Review*, 14 (1), 11-19.
- [3] Frické, M. (2009). The knowledge pyramid: a critique of the DIKW hierarchy. *Journal of Information Science*, 35 (2), 131-142.
- [4] Leydesdorff L, Etzkowitz H. (1996). Emergence of a Triple Helix of University-Industry-Government Relations. *Science and Public Policy*, 23 (3), 279-286.
- [5] Leydesdorff L, Etzkowitz H. (1998). The Triple Helix as a model for innovation studies. *Science and Public Policy*, 25 (3), 195-203.
- [6] Mattmann, C. A. (2013). A vision for data science. *Nature*, 493 (7433), 473-475.
- [7] Nonaka, I., Toyama, R. and Konno, N. (2000). SECI, Ba, and leadership: a unified model of dynamic knowledge creation. *Long Range Planning*, 33, 5-34.
- [8] Nonaka, I. and von Krogh, G. (2009). Tacit Knowledge and Knowledge Conversion: Controversy and Advancement in Organizational Knowledge Creation Theory. *Organization Science*. 20 (3), 635-652.
- [9] Rowley, J. (2007). The wisdom hierarchy: representations of the DIKW hierarchy [J]. *Journal of Information Science*, 33 (2), 163-180.
- [10] Shannon, C. E. (1948). A mathematical theory of communication. *The Bell System Technical Journal*, 27 (3-4), 379-423, 623-656.
- [11] Ye, Y. (1999). An analytical construction on the fundamental theory of information science and technology. *Journal of Scientific and Technical Information Society of China*, 18 (2), 160-166.
- [12] Ye, F. Y. (2011). A Theoretical Approach to the Unification of Informetric Models by Wave-heat Equations. *Journal of the American Society for Information Science and technology*, 62 (6), 1208-1211.
- [13] Ye, Y and Ma, F.-C. (2015). Data science: its emergence and linking with information science. *Journal of the China Society for Scientific and Technical Information*, 34 (6): 575-580.