A method to systematize keywords- Explicit keywords and implicit keywords -

Georgiev, Georgi V.
Yamada, Kaori
Taura, Toshiharu
Nagai, Yukari

A method to systematize keywords

- Explicit keywords and implicit keywords -

Georgi V. Georgiev (Kobe Univ.), Kaori Yamada (Kobe Univ.), Toshiharu Taura (Kobe Univ.), and Yukari Nagai (JAIST)

The purpose of this paper is to propose a method to systematize keywords of multiple articles by using the associative concept dictionary.

In this method, we focus on the keywords behind the expressed keywords that are related to them (hereafter called "implicit keywords"), so that keywords can be identified in a broad and in-depth manner. In order to obtain the implicit keywords, we focus on the relationship between knowledge and memory. We don't believe that "abstraction" is suitable for finding implicit keywords, since most information is lost in the abstraction process. Instead, we focus on "association", since association is considered as a representation process that expresses broad and in-depth feelings, which are related to explicit keywords.

Free association is presumed to reflect how words and concepts are structured and interrelated in the mind (Deese 1965). The background model behind the process of free association examines it as a memory task. Free association is used to study implicit and explicit memory (Nelson *et al*. 2000, p. 887). Research reveals that free association predicts memory recall better than ratings of word similarity (Nelson *et al*. 2004, p. 406). Moreover, the free association procedures can be applied to provide an implicit assessment of the effects of language learning or marketing (Nelson *et al*. 2004, p. 403). By considering the abovementioned features of free association, in our previous research we applied this notion in design (Georgiev *et al.* 2012). In this method, we extend the notion further and apply it for keyword analysis. Hereafter, the unadorned term "association" implies "free association."

Definitions:

- Explicit keyword k_i: a word or phrase that is significant in an article
- Single keyword k_{ij}: a word as an element of explicit keyword k_i
- Implicit keyword k_{ijk}: a word that is not directly expressed but implicitly underlies the single keyword k_{ij}

The procedure of analysis used in this study consisted of the following:

- Procedure 1: Extract explicit keywords from the articles.
- Procedure 2: Find implicit keywords from the explicit keywords.
- Procedure 3: Conduct cluster analysis using both the implicit and explicit keywords.

Procedure 1 Extraction of explicit keywords

We carefully read all the articles and extracted explicit keywords.

Procedure 2 Finding implicit keywords

The associative concept dictionary, which contains the relations between words in terms of association, consists of two types of words–stimulus and response. For our analysis, the single keyword k_{ij} is examined as a response word, and the stimulus words are identified as the set of implicit keywords k_{ijk} (Figure 1). Multiple implicit keywords could exist for each of the single keywords (k_{ij}).

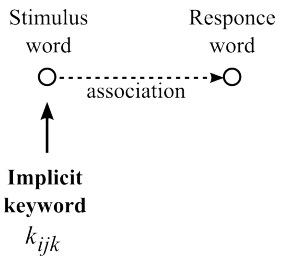

Figure 1. Identification of implicit keywords based on associative concept dictionary.

The associative concept dictionary by Nelson *et al*. (hereafter, called *the dictionary*) used in this analysis was collected from more than 6,000 participants who produced nearly three quarters of a million responses to 5,019 stimulus words (Nelson *et al.* 2004). However, it has the following two limitations. First, the terms in *the dictionary* were collected in the general (daily) sense and not in any technical sense; therefore, terms were not collected with specific implications, even for the *design creativity and innovation* research field. Second, *the dictionary* has limited generalizability across cultures and regions. By considering these limitations, the implicit keywords that are obtained by using *the dictionary* are not the results of this analysis; rather, they are used only to relate the explicit keywords to other explicit keywords.

The steps for finding implicit keywords are as follows:

Step 1: The set of explicit keyword (words/phrases) k_i (e.g., the keyword "deep knowledge") is listed as a set of the single word k_{ij} (e.g., "deep" and "knowledge").

Step 2: The single words "design," "creativity," and "innovation" are omitted.

Step 3: *The dictionary* is used to find implicit keywords k_{ijk}.

Procedure 3 *Cluster analysis*

Two types of matrixes are created–Matrix $R_1(r^1{}_{ij})$, representing relatedness between the

keywords, and Matrix $R_2(r^2{}_{ij})$, representing relatedness based on belonging to the same article.

The detailed analysis consisted of the following steps:

Step 1: Matrix $R_1(r^1{}_{ij})$ is obtained by taking implicit keyword–implicit keyword matching, implicit keyword–single keyword matching, and single keyword–ingle keyword matching into account:

- Implicit keyword (k_{ikm})–implicit keyword (k_{jln}) matching between the two keywords (k_{ik}, k_{jl}) is determined by introducing the coincidence variable C_{ijklmn} as follows (Figure 2(a)):

$$C_{ijklmn} = \begin{cases} 0.1 & \text{if } k_{ikm} = k_{jln} \\ 0 & \text{if } k_{ikm} \neq k_{jln} \end{cases}$$

- Implicit keyword (k_{ikm})–single keyword (k_{jl}) matching is determined by introducing the coincidence variable C_{ijklm} as follows (Figure 2(b)):

$$C_{ijklm} = \begin{cases} 0.1 & \text{if } k_{ikm} = k_{jl} \\ 0 & \text{if } k_{ikm} \neq k_{jl} \end{cases}$$

- Single keyword (k_{ik})–single keyword (k_{jl}) matching is determined by introducing the coincidence variable C_{ijkl} as follows (Figure 2(c)):

$$C_{ijkl} = \begin{cases} 0.1 & \text{if } k_{ik} = k_{jl} \\ 0 & \text{if } k_{ik} \neq k_{jl} \end{cases}$$

- Three types of matches are summed as follows.

$$r_{ij}^{1\prime} = \sum_k \sum_l \left(\sum_m \sum_n C_{ij\,klmn} + \sum_m C_{ij\,klm} + \sum_n C_{j\,ilkn} + C_{ij\,kl} \right)$$

Matrix $R_1(r^1{}_{ij})$ is determined as:

$$r_{ij}^1 = r_{ji}^1 = \begin{cases} r_{ij}^{1\prime} \text{ if } r_{ij}^{1\prime} \le Th \\ Th \text{ otherwise} \end{cases}$$

Th: threshold (in this analysis, Th is determined as 0.5 by considering the distribution of $r_{ij}^{1\prime}$)

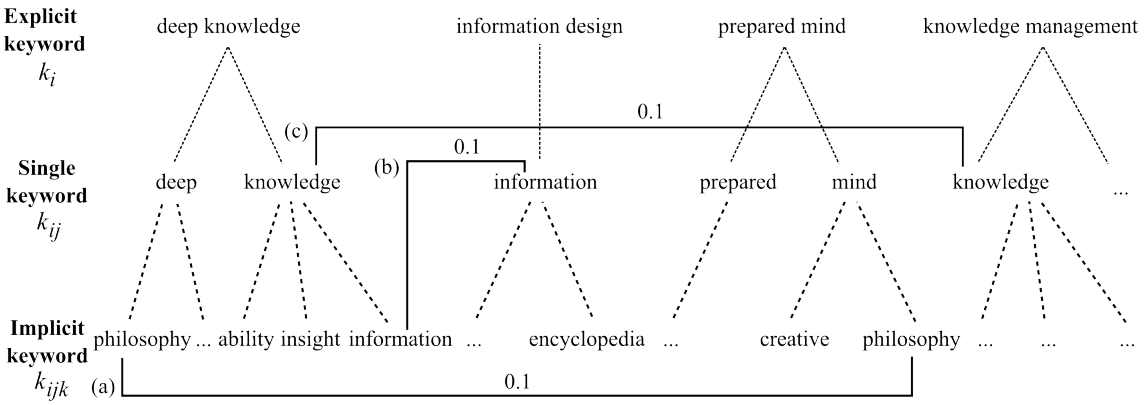

Figure 2. Three types of matching between the keywords.

Step 2: Matrix $R_2(r^2{}_{ij})$ is obtained according to the following procedures.

- For every pair of explicit keywords (k_i, k_j), variables u and v are determined as follows.

  u: the number of articles that involve both explicit keywords

  v: the number of articles that involve only one of the two explicit keywords

- Matrix $R_2(r^2{}_{ij})$ is determined as:

$$r_{ij}^2 = r_{ji}^2 = \frac{u}{u+v}$$

Step 3: Matrix $R(r_{ij})$ is obtained as follows.

$$r_{ij} = w_1 r_{ij}^1 + w_2 r_{ij}^2$$

Here, w_1 and w_2 are weights.

In this analysis, w_1 is determined as 1.4; w_2 is determined as 0.3.

Step 4: The Relatedness matrix $R(r_{ij})$ is employed in cluster analysis. The Ward's method of clustering is used.

References

Georgiev, G. V., Nagai, Y., and Taura, T., 2012. Analysis of the Formation of User Impressions Upon Tactile Interaction with Product Design Materials. *In*: I. Horváth, Z. Rusák, A. Albers, and M. Behrendt, eds. *Proceedings of 9th Tools and Methods of Competitive Engineering TMCE 2012*, Delft: Faculty of Industrial Design Engineering, Delft University of Technology, 515–526.

Nelson, D.L., McEvoy, C.L., and Dennis, S., 2000. What is free association and what does it measure? *Memory & Cognition*, 28 (6), 887–899.

Nelson, D.L., McEvoy, C.L., and Schreiber, T.A., 2004. The University of South Florida free association, rhyme, and word fragment norms. *Behavior Research Methods, Instruments, & Computers*, 36 (3), 402–407.

Nelson, D.L., McEvoy, C.L., and Schreiber, T.A., 2012. *Supplemental Materials AppendixA.xls*. Available from: http://www.springerlink.com/content/u81l55373371rn04/supplementals/

Cramer, P., 1968. *Word association*. New York: Academic Press.

Deese, J., 1965. *The structure of associations in language and thought*. Baltimore: The Johns Hopkins Press.