KFTKGA: KNOWLEDGE FLOW TRACER AND GROWTH ANALYZER FOR COMMUNITY LEARNING

S.M.F.D. Syed Mustapha, Charles A.H. Chong

Abstract: The advances in building learning technology now have to emphasize on the aspect of the individual learning besides the popular focus on the technology per se. Unlike the common research where a great deal has been on finding ways to build, manage, classify, categorize and search knowledge on the server, there is an interest in our work to look at the knowledge development at the individual's learning. We build the technology that resides behind the knowledge sharing platform where learning and sharing activities of an individual take place. The system that we built, KFTGA (Knowledge Flow Tracer and Growth Analyzer), demonstrates the capability of identifying the topics and subjects that an individual is engaged with during the knowledge sharing session and measuring the knowledge growth of the individual learning on a specific subject on a given time space.

Keywords: Knowledge tracing, Knowledge measuring, SECI model

ACM Classification Keywords: H.1.2 User-Machine System

Conference: The paper is selected from International Conference "Intelligent Information and Engineering Systems" INFOS 2008, Varna, Bulgaria, June-July 2008

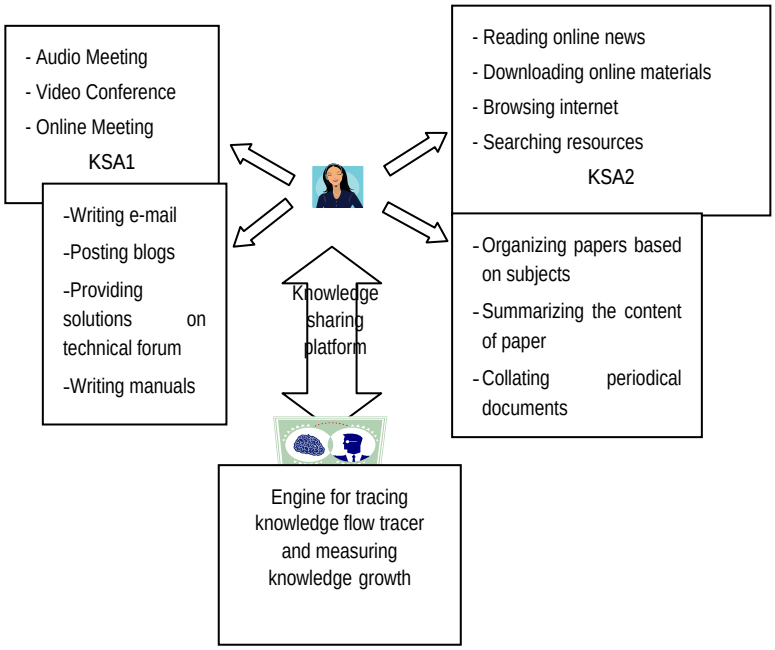
Introduction

In the past decade, there is a continuous growth of interest in knowledge management system as reflected through the research activities reported by the industrialists and academicians. The research in this field is driven by the compelling factors of building knowledge worker, generating knowledge economy and developing knowledge-based society. Organizations are enthusiastic to observe these changes and the impacts to the entire organizational settings, cultures and behaviours. Recently, the hopes for the knowledge management successes have been argued and presumed as hypes or failures. The analyses for these failures have spawned many hypothetical solutions, methodologies and approaches to overcome the shortfall. The implication of these studies still emphasizes the importance of knowledge as the essential element to the modern practices of knowledge-intensive organization. The drawbacks lie within the human factors more than the technology. As such, knowledge management approaches have gradually evolved from technology driven to technology-and-community driven. Our approach, so-called Intelligent Community Informatics (thereafter, ICI) focuses on the three knowledge management components which are the community, knowledge and learning in equal balances. This paper discusses our work where the three components are integrated and operated symbiotically to create the learning community environment by sensing the learning flow of the individual as well as the community and measures the growth of knowledge of different domain. This is demonstrated in our system called KFTGA (Knowledge Flow Tracer and Growth Analyzer) as discussed in section 2. Section 3 describes the application of KFTGA on SECI model environment; Section 4 gives the conclusion of the work.

KFTGA (Knowledge Flow Tracer and Growth Analyzer)

The engine for tracing the knowledge flow and measuring the knowledge growth resides as the backend system incorporated in the knowledge sharing platform. Figure 1 shows the four types of KSAs that an individual like Allen can do in any knowledge sharing applications (such as blogs, forums, e-mails, instant messenger, video calls). There are many knowledge sharing applications that are available in the market that are currently used in

the research laboratory which supports different modes of communication (synchronous or asynchronous, text, audio or video).



KFTGA role is to monitor the knowledge that transpired around the KSA application in terms of the topics and the subject domain that an individual or a community is engaging in. For example, KFTGA can sense whether Allen and her friends share their knowledge through various stages of KSA on a specific subjects or topics. Another way, the system is also able to identify what are the various subjects or topics that linger around the community during the knowledge sharing activities. In addition to that, the KFTGA system can determine which topics or subjects are more dominant than the others. To the best of our knowledge, there is no technology that has been built that integrates the content of the knowledge transpired among various knowledge sharing applications such that the content can be analysed for the purpose of tracing knowledge flow and measuring knowledge growth. The three aspects of content analysis are given below:

1. Relevancy – given the knowledge domain, the system will inspect whether the topics/subjects covered in the knowledge sharing activities are within the ambit of the knowledge domain.
2. Growth – the system periodically examines the size of the knowledge repository enlarging based on the two different point of time on a specific set of knowledge domain.
3. Variation – the system detects whether there are variations of side topics being discussed within the main topics/subjects of the knowledge sharing activities. For example, within the main topic of "knowledge management" there are variations such as "Analytic Hierarchy Process", "collaborative learning" or "ontology".

The content that is captured from various knowledge sharing applications comes in different format needs to be transcribed into text format to be processed and analysed with respect to the Relevancy, Growth and Variation. The challenging parts of the work are

i. to extract the conceptual meaning for each transcription generated from the informal discourse (discussion, blogs, emails or chat) as well as from the formal knowledge sources such as printed materials like manuals, scientific documents, reports. The challenge is that these sources are not predetermined in terms of the format, language style, document structure and presentation.

ii. to cluster the groups of the keywords which belong to the specific topics/subject such that these group can be the determinant factor to detect the relevancy of a given transcription. This approach is not similar to the traditional approach where the word dictionary such as WordNet or Ontology is used to classify the words.

iii. to measure the growth size of the transcriptions on specific subject between two progressive time points from the transcripts which are purely textual information. The measurement of the growth size needs to be translated into numerical value.

iv. to find the conceptual differences between a transcript and the domain knowledge where the difficulties would be to find the subtopics which are included in the main stream of the domain knowledge.

KFTGA technology

This paper addresses the methods of tracing the knowledge flow and measuring the growth of knowledge for community knowledge. The community knowledge is built on the formal and informal knowledge that are transcribed from several forms of communications and knowledge sharing activities such as online meeting, video conferencing, face-to-face, documents uploading and downloading, message posting etc. The reposition of these transcriptions is called community knowledge which evolves over a period of time. Therefore, it is essential to have a mechanism to be able to automatically analyze the content of the knowledge that flows within the repository and also measure the progress and development. By doing this, knowledge sharing attributes such as the learning pattern, the community interest and the knowledge building performance can be gauged. This research work provides three aspects of knowledge tracing and knowledge growth analyzing which are the relevancy, growth and variation. The interesting part is that these values are determined mainly from the textual transcripts generated from formal and informal knowledge sources.

Relevancy

Automated detection on relevancy is difficult since it requires one to understand the subject being discussed in the forum, e-mail, scientific document or report in which they may come in a short or long text. For a short text, there may not be enough words to be analyzed in order to determine the context of the discourse. Other difficulties are that the words usages may differ from one author to another even though they are used to refer to the same context. We avoid the usage of dictionary or word ontology as building them can be time consuming, laborious and costly. Our method allows new subject domain to be added into the community knowledge without the need to upgrade manually the new keywords definition as in the static ontology or dictionary. The KFTGA system builds the keywords definition automatically in incremental manner. The relevancy is determined from the two processes described in Figure 2 and Figure 3.

The set of keywords for a subject domain is built by analyzing formal sources that are published officially such as journals, proceedings, operation manuals, technical reports and others. The characteristics of these sources should be that the contents are specific to a certain topic, there exists finite set of discriminatory keywords that populates the document to represent the subject domain of the document and the representative discriminatory keywords are common to other sources of similar topics.

For example, the keywords such as “socialization”, “externalization”, “internalization” and “combination” are the discriminatory keywords for the articles on “SECI model” topic. The process flow in Fig. 2 produces the discriminatory words extracted from each document and these words are considered as the well-representatives for the chosen topic/subjects. That means if the collections of documents are about “Knowledge management”, then those DW words are the acceptable words being commonly used by the professionals, practitioners and researchers in that domain. The collection of words is now being treated as the pools of words under the category of a specific domain. This approach has an advantage that it can build its own recognizable word list without the need to build a dictionary or ontology which is usually built manually. Once the DW words are compiled, they can be used in the Words Mapper in the form of weight to identify whether a given document is Relevant to the domain as shown in Figure 3.

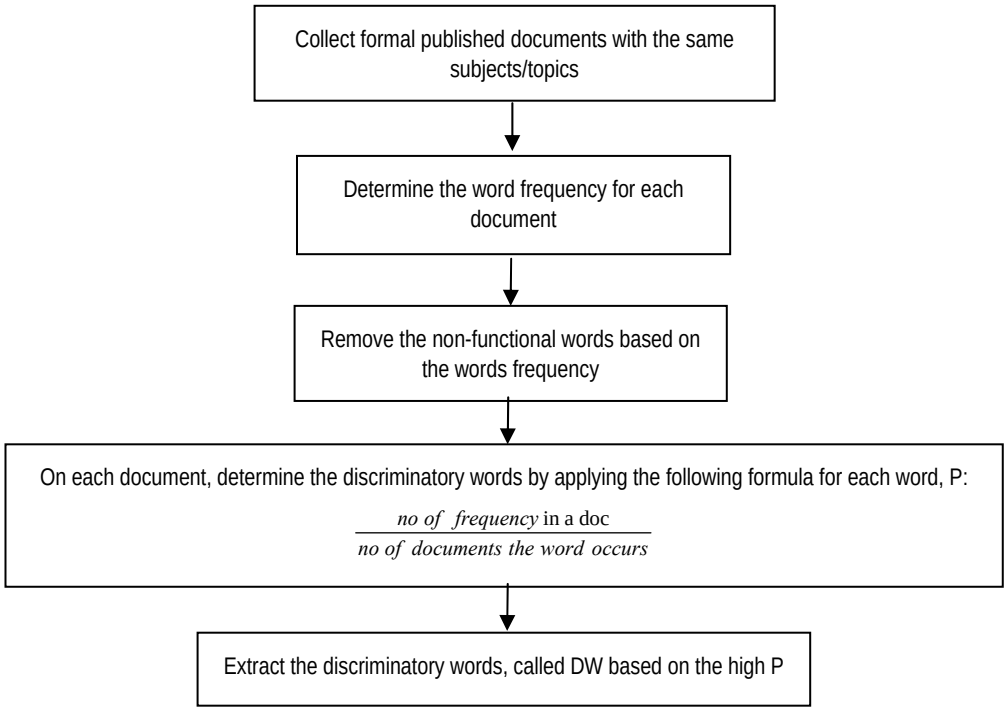
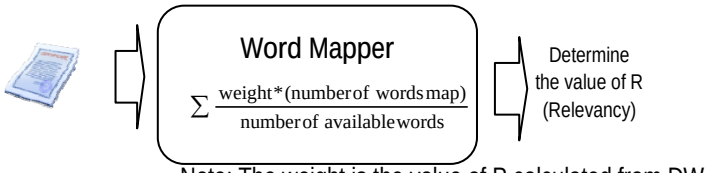


Figure 2. Determine the Discriminatory Words (DW)



Note: The weight is the value of P calculated from DW.

Figure 3. Word Mapper to determine Relevancy

Growth

The engine runs over a period of time that it is possible to measure the development aspect of the resources, so-called growth. There are four dimensions of growth that are The growth is defined as the increment of i) the number of resources that are made available in the knowledge repository (RG) ii) the number of terms that are relevant to the topic/subject (TG) iii) the number of references/usages (RUG) iv) the number of resource maintenance activities (MG).

Resource Growth

$$T_1 = \frac{(R)_{T1} - (\overline{R})_{T1}}{total\ number\ of\ resources}$$

$$T_2 = \frac{(R)_{T2} - (\overline{R})_{T2}}{total\ number\ of\ resources}$$

$$RG = T_2 - T_1$$

R – number of relevant resource

$\overline{R}$ - number of irrelevant resource

where T_1 and T_2 are two different time points and $T_1 < T_2$

Term Growth

$t_i \in (t_1, t_2, ...t_n)$ in T where T is the term in a document D

$$t_{i\delta} = \frac{\left[\left(\sum_1^n 1 \text{ for each } t_i \text{ in D}\right)^{T_1} - \left(\sum_1^n 1 \text{ for each } t_i \text{ in D}\right)^{T_2}\right]^2}{(\text{number of terms in D})^2}$$

$$\text{TG} = \sum_i^n t_{i\delta} \text{ for } \forall \text{i} \in \text{D}$$

where T_1 and T_2 are two different time points and $T_1 < T_2$

Reference Growth

r_{D_j} – the number of references for document D_j

$\bigcup D_j$ – the total number of references for $\forall D$

$$RUG = \left[\sum_{j=1}^{N} \frac{r_{D_j}}{\bigcup D_j}\right]^{T_2} - \left[\frac{r_{D_j}}{\bigcup D_j}\right]^{T_1}$$

where N is number of document and T_1 and T_2 are two different time points and $T_1 < T_2$

Maintenance Growth

$A = \{a_1, a_2 ... a_n\}$ where a is the maintenance activity

$a_i = \bigcup_1^n f(a_i)$ where f is the frequency for activity a

$$\text{MG} = \left[\frac{\sum_1^n a_i}{N_A}\right]^{T_2} - \left[\frac{\sum_1^n a_i}{N_A}\right]^{T_1}$$

where N is the total number of maintenance activity

Variations

For every main concept, there are always sub concepts that are associated to it. For example, if the main concept is VOIP (voice over IP), the possible sub concepts that are associated to it could be “skype”, “economics of VOIP”, “phone adapter”, “Cisco”. The difference between the main concept and the sub-concepts are that the main concept appears as the frequent words in all documents while the sub-concepts appear as the frequent words in some specific documents. This can be determined from the following algorithm.

1. For all $d_i \subset D$ where $i = \{1,2,...N\}$ and D is a set of documents
2. Calculate f_{d_i} *for* all i on each $t_j \in d_i$, called $f_{d_i}^{t_j}$
3. For each $t_j \in d_i$, *calculate* N where
   N is the number of documents that t_j occur, called N_{t_j}
4. Given P as the total number of documents, then if $\frac{N_{t_j}}{P} > \varepsilon$
   where ε is the threshold, add t_j to list T'
5. Given P as the total number of documents, then if $\frac{N_{t_j}}{P} < \mu$
   where μ is the threshold, add t_j to list T"
6. T' = main concepts and T" = sub concepts

KFTGA application on SECI model

KFTGA uses SECI as the knowledge creation model for its community of learning. SECI model, introduced by Nonaka and Takeuchi [Nonaka and Takeuchi, 1995], has received world-wide attention by many researchers. It proposes four types of knowledge conversion based on tacit knowledge and explicit knowledge as shown in Error! Reference source not found.. Each quadrant represents the conversion type, tacit-to-tacit (socialisation), tacit-to-explicit (externalisation), explicit-to-explicit (combination) and explicit-to-tacit (internalisation). The common understanding about the model is that the knowledge creation will take place as the result of several knowledge conversion spirals.

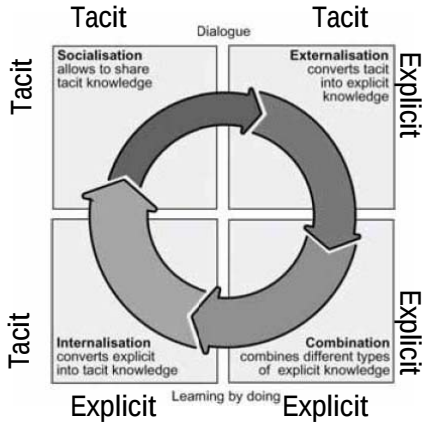


Figure 4 SECI model (adapted from [1]Error! Bookmark not defined.)

The SECI model has received an equal acceptance, rejection and modification from various researchers. Chatti [Chatti, 2007] issues a significant relevancy of SECI model to Web 2.0 by describing the emerging technologies that can be used for different quadrant of SECI model; social media (e.g. Wikipedia (reference), MySpace (social networking), gather.com (social networking), YouTube (video sharing), Second Life (virtual reality), Digg (news sharing), Flickr (photo sharing) and Miniclip (photo sharing)) for *Socialisation*; Discussion channel such as Blogs, Chats, E-mails, IM and Video-conference and any community publishing portals to express opinions such as ratings/voting/feedbacks support *Externalisation*; managing knowledge through a systematic source integration techniques such as RSS/Atom, Pod/vodcasting or any Mashup applications can be considered as *Combination*; and Multi-player gaming and simulations are examples of Web 2.0 applications that fall under *Internalisation*. Hämäläinen [Hämäläinen, 2003] reports the SECI models fits well in the software and research development work where all quadrants are found to be significantly exist in all of the three tested case in the R&D work even though the degree of its applicability may differ. Rice and Rice advocate the possible implementation of SECI model for project in a multi-organisational environment where the employees are not necessarily situated locally [Rice et al, 2005].

KFTGA captures the learning activities that can be characterised by the SECI model in order to trace the knowledge flow evolution of for an individual as well as the group.

Conclusion

KFTGA emphasizes on building the technology where the learning flow of the community can be gauged. The SECI model is used as the reference model for categorizing the four types of knowledge flow phases. The model covers well on every aspects of knowledge sharing activities (KSAs) that could take place on the web environment. It is hypothesized that the KFTGA system that we build could reside behind the knowledge sharing

platform and could perform two essential functions. Firstly, is to trace the knowledge of certain subject or topic that an individual is engaging with during the knowledge sharing session; and secondly, is to measure the growth of knowledge pertaining to the subject or topic. In other words, KFTGA will be able to determine whether an individual has delved a lot on certain topic and the quantity of knowledge the person has invested in.

Bibliography

[Nonaka and Takeuchi, 1995] Nonaka, I and Takeuchi, H (1995) The Knowledge-Creating Company: How Japanese Companies Create the Dynamics of Innovation. New York: Oxford University. http://commerce.concordia.ca/el-sayed/KCreation.pdf

[Chatti, 2007] Chatti, M.A. The Web 2.0 Driven SECI Model Based Learning Process, IEEE ICALT 2007 (to be published) http://en.wikipedia.org/wiki/Social_media

[Hämäläinen, 2003] Hämäläinen, L. Knowledge creation, communication and role of IT in traditional software R&D teams, e-BRF 2003. Abstract: http://www.ebrc.info/kuvat/2020.pdf

[Rice et al, 2005] Rice, J. L. & Rice, B. (2005). The Applicability of the SECI Model to Multi-Organisational Endeavours: An Integrative Review. *International Journal of Organisational Behaviour*, 9, 671-682. http://www.usq.edu.au/resources/vol9no8rice.pdf

Authors' Information

S.M.F.D Syed Mustapha – Associate Professor in the Faculty of Information Technology, University Tun Abdul Razak, 16-5 Jalan SS6/12 Kelana Jaya, Selangor Darul Ehsan, MALAYSIA; E-mail: syedmalek@unitar.edu.my

Charles A.H. Chong – Researcher; BT Multimedia Sdn Bhd, 1B-17, Block 1B, Plaza Sentral Jalan Stesen Sentral 5 Kuala Lumpur Sentral 50470 Kuala Lumpur Malaysia; e-mail: charles.chong@bt.com