KFTGA: A tool for tracing knowledge flow and knowledge growth in knowledge sharing environment

S.M.F.D. Syed Mustapha
School of Information and Communication Technology, Asia e University, Main Block, No 4, Jalan Sultan Sulaiman, 50000, Kuala Lumpur, Malaysia
Tel.: +60 192663064; E-mail: syed.malek@aeu.edu.my/smfdsm@yahoo.com

Abstract: Knowledge management is an interdisciplinary research area and it has been perceived in different perspectives by researchers and practitioners who come from various backgrounds. Nevertheless, the ultimate goal in knowledge management is to create a learning organization in a knowledge sharing environment. In a knowledge sharing environment, the learning communities are actively involved in building and developing the community knowledge through message postings, resource management and collaborative work besides performing the individual learning. A system, called KFTGA, has been built in this project that is capable to trace the knowledge flow of a pre-defined knowledge domain developed by the community members. Knowledge in this context is in the form of opinion, suggestions, and tangible objects from knowledge sharing activities which are traceable from the textual descriptions. Two salient aspects in KFTGA are having the ability to determine whether the knowledge that flows among the members is relevant to the knowledge context in the community and to trace whether there is a growth in knowledge building in relation to predefined topics. In order to achieve that, this paper describes three mathematical models for computing the relevancy of the knowledge flow, measuring the knowledge growth and examining the variations between documents. Finally, it elaborates how KFTGA can be applied in various knowledge sharing and knowledge flow model predominantly known as SECI model and its extended variations applied in virtual organizations, peer-to-peer team and multinational corporations.

Keywords: Knowledge tracing, knowledge measuring, SECI model, Web 2.0

1. Introduction

Knowledge management is a long standing research area that has attracted the academics and industrialists in formulating frameworks, solutions and methodologies from various perspectives ranging from business, management, and economics to information technologies. Understanding the employees' competencies is one of the knowledge management strategies for maintaining organizational productivity through knowledge retention [5,11]. Personal knowledge management provides personal knowledge resources which are built in collaboration and contribution of every knowledge worker [12]. Organizations perceive the knowledge management as one of the tool in the business strategy through the development of knowledge-based communities. However, industrial reports have also shown that many organizations have failed in the knowledge management implementation [1,13]. While documents may be reposited in a systematic manner using Document Management System and searching and retrieving techniques are improving, the issues of developing knowledge-based workers from DMS are not properly addressed.

The KM technology that is emphasized in this paper is to look at how to trace the learning activities among the knowledge workers and how to measure the knowledge flow and growth.

A Web 2.0 phenomenon has come at the right time with the new emphasis on knowledge management approaches. Web 2.0 stresses the collaborative effort of the community as the rapid approach in knowledge building and maintenance [6]. It is acceptable to claim that the ultimate objective in Knowledge Management strategy is not only about building IT infrastructure to support the management of knowledge resources but also to ensure the knowledge transfer and knowledge building among the communities do take place. We have developed Knowledge Flow Tracer and Growth Analyzer (thereafter, KFTGA) that is capable to trace the knowledge flow based on certain subjects or topics that are of interest among the communities and also to measure the growth in the amount of knowledge that is built as the result of resource sharing, resource utilization or knowledge regeneration within an individual or community.

Our approach aims towards building learning support with the capability to trace individual learning for different classes of knowledge flows and knowledge sharing models. There are many related works that have described the concept of knowledge flow and how knowledge sharing models have been applied. Those can be categorized as follows: i) analyzing the knowledge flow based on the workflow in an organization ii) technological approach in the detecting knowledge flow and iii) the framework for modeling knowledge flow.

1.1. Knowledge flow in an organization

Newman et al. [14] established a framework to analyze the knowledge flow based on the knowledge management activity areas and how knowledge artifacts are created, utilized, retained and transferred in an organization's work process. Organizational structure is also another factor being studied in analyzing how knowledge flow is affected and how it is vulnerable to the changes of the workflow [8]. Another interest shown in modeling the dynamic knowledge flow for virtual organization is on the definitive Knowledge Conversion Process (KCP) which describes how knowledge transforms in a complex organizational environment and how knowledge is stored, shared and managed using the Knowledge Sharing Space (KSS) conceptual model [19].

1.2. Technology approach in a workflow

Anjewierden et al. [2] described the work on extracting the terms from the weblogs to determine the conceptual structures in order to detect the knowledge flow in the community blogging. Nissen and Levitt demonstrated the computational simulation of knowledge flow for the Virtual Design Team for software development activities [15]. It demonstrated the system's capability to track the learning flow among the members through the formal training and how the learning has benefitted in terms of project cost and human skills. Another effort is to look at using the technology in identifying techniques for knowledge diffusion to relevant users based on user profiling and the organization's workflow. The technology can be used to build a network of knowledge resources that are relevant by the interested subjects and subsequently interest matrix for the users [3].

1.3. Framework for modeling knowledge flow

Zhuge [25] has given an extensive overview of various types of generic knowledge flow models that prescribe three attributes, content, direction and carrier that are commonly applicable to team work in organizations. The knowledge flow models need a vivid definition of how knowledge can be shared
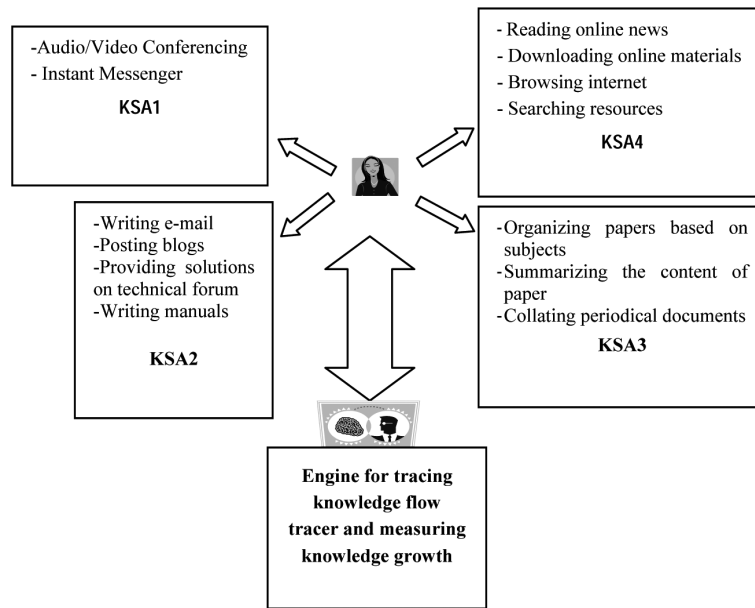
Fig. 1. Four types of knowledge sharing activities (original source [22]).

among team members to ensure effectiveness in the planning and knowledge distribution. Zhuge provides the descriptive model for planning the knowledge flow network and describes the simulation based on the knowledge spiral model [26].

Based on the literature, the work described in this paper falls under the category of building technology in supporting knowledge flow process. The differential factor in the current work emphasizes on the automation of building knowledge domain from the relevant training set and the methods in tracing the knowledge. The training set is a collection of text-based documents that are processed syntactically to extract a group of keywords that represent the knowledge domain. The knowledge domain has a collection of keywords that are categorized as main concepts and subconcepts using the *Relevancy* model and *Variations* model. The knowledge domain may reside on individual's repository of reading collection in which the development of the repository can be monitored using the *Growth Analyzer*.

Section 2 describes how KFTGA (Knowledge Flow Tracer and Growth Analyzer) works in knowledge sharing environment and what are the challenges that it needs to overcome. The KFTGA technology is explained in Section 3 which describes how knowledge is traced and measured. Section 4 demonstrates the application of KFTGA in terms of knowledge domain building and query processing and Section 5 describes several types of knowledge flow and knowledge sharing model such as SECI model and how KFTGA can be integrated and functioned within the system. Section 6 concludes the contribution of the paper and future work.

2. KFTGA (knowledge flow tracer and growth analyzer)

KFTGA is designed as the backend engine that traces i) the knowledge that flows within the community members and ii) the utilization of documents during the knowledge sharing process. KFTGA engine is embedded into the knowledge sharing platform that is used by the community or organization. Figure 1 shows the four types of knowledge sharing activities (noted as KSA) that take place in the knowledge

sharing platform. In KSA1, knowledge is shared in a synchronous manner using various available technologies such as instant messaging, text chatting, audio/video conferencing to get immediate responses in a highly interactive session. The outcome of the discussion can be transcribed or minuted in a written format. KSA2 indicates that knowledge is shared mostly in a community communication channel that supports asynchronous mode such as emails, blogs, technical forums, Questions and Answers, FAQs and others in which most of them can be captured in their original forms of writing. Both KSA1 and KSA2 require participation of the members that regenerate the knowledge products in the form of personal arguments, opinions and debates. KSA3 allows an individual or the community of knowledge management to reorganize the knowledge sources such as documents, videos, learning materials, blogs, emails, scientific data, official references and others based on subjects to effectuate better document management process. Some actions in KSA3 involve in ameliorating, altering or revising the content of the knowledge sources in which knowledge sources are regenerated. For example, summarizing the content of scientific substance from few articles into a single report or revising the meeting minutes requires one to scrutinize knowledge resources. KSA4 is an individual effort in the knowledge sharing activities as the prelude to other knowledge sharing activities. For example, one may need to read the previous meeting minutes prior to the next meeting, reading and searching for materials on the internet to prepare some business proposal.

KFTGA is built to monitor the knowledge flow and knowledge growth in the knowledge sharing platform by identifying the topics covered in the knowledge sources and by tracing the individuals that have dealt with them throughout the knowledge sharing activities. This is done by performing the content analysis to determine the topics that are used by the individuals based on the following parameters mentioned below [22]:

1. Relevancy – performs the words matching between the knowledge resources and the knowledge domain to find the degree of similarities of the words that appear in both sources. Knowledge domain contains the words representing the context of the knowledge area that is built automatically using a set of articles related to subject.
2. Growth – knowledge resources that are accessed by an individual will be reckoned that the individual knowledge has grown and the "growth" will be reflected whenever they are continuously being accessed.
3. Variation – knowledge is traced by examining the selected words that are used to represent the main context of the knowledge resource as these collection of words appear mostly in every knowledge resource that are relevant to a specific topic. For example, if "knowledge management" is the context, so the selected keywords to represent this context would be "learning", "knowledge sharing", "knowledge", "learning organization" which co-occur in most of its appearance in the knowledge resource. The words that appear irregularly are considered as "variations" as they do appear together with the selected words in some selected knowledge resource. Words such as "Community of Practice" or "CoP" are related to "knowledge management" but do not appear together at all times.

The content that is captured from various knowledge sharing applications comes in different format and that needs to be transcribed into standard text format in order to compute the values on Relevancy, Growth and Variation. The challenging parts of the work are:

1. 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 since the input to

KFTGA system comes in various sources, therefore, the layout format, language style, document structure and presentation are not predetermined.

2. 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. In other words, our approach will not rely on any references such as ontology or dictionary.

3. to measure the growth size of the transcriptions on specific topics/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.

4. to determine the concept and sub-concept of a given transcription based on the several sets of knowledge domain that can be built automatically.

2.1. Learning process

KFTGA is a tool that resides on the knowledge sharing platform where knowledge sharing activities take place. Even though the learning process is not the focus in this paper, but it is assumed to take place in the knowledge sharing process. The theory of learning process was discussed by Säljö who fundamentally categorized that the learning took place based on the following process [21]:

1. Learning is a process of acquiring knowledge and quantitatively increased.
2. Learning is when facts and information are stored and regenerated.
3. Learning is when facts, skills and methods can be retained and used when necessary.
4. Learning is when the meaning of the information is abstracted and some relation can be made with other subject matter.
5. Learning is when one can make an interpretation and show understanding in different way.

The knowledge sharing process in KFTGA supports those categories of learning processes. One of the KFTGA capabilities is measuring the knowledge growth which determines whether a learner has increased the knowledge through the learner's involvement in the knowledge sharing activities that may generate and regenerate new knowledge as the result of learning process. It is also capable of determining the topics of discussion and the content of learning objects that are shared by the learners. The details of the KFTGA technology are described in the following section.

3. KFTGA technology

This paper addresses the methods of tracing the knowledge flow and measuring the growth of knowledge for the learning community. 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.
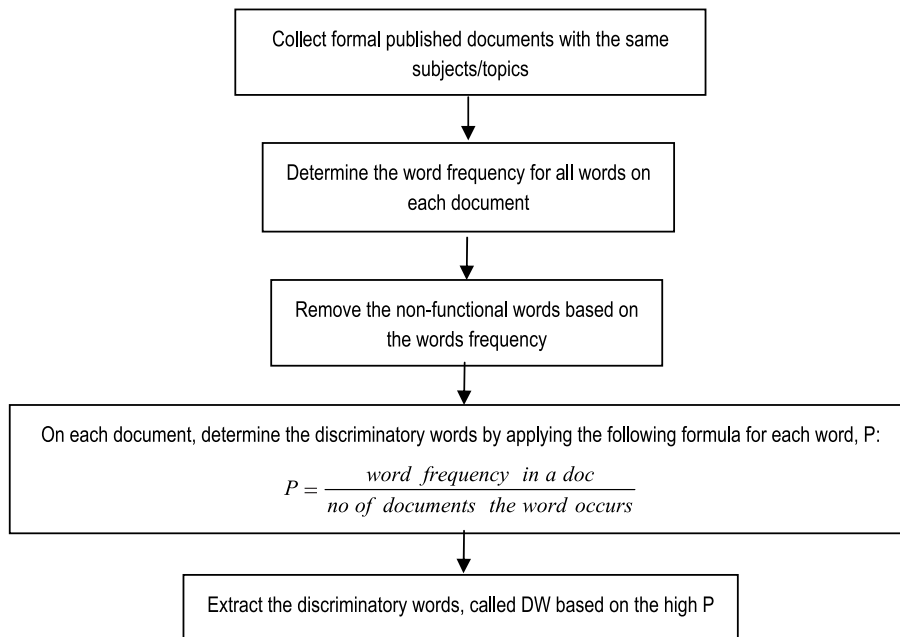
```
┌─────────────────────────────────────┐
│  Collect formal published documents  │
│        with the same subjects/topics │
└─────────────────────────────────────┘
                    │
                    ▼
┌─────────────────────────────────────┐
│  Determine the word frequency for    │
│       all words on each document     │
└─────────────────────────────────────┘
                    │
                    ▼
┌─────────────────────────────────────┐
│  Remove the non-functional words     │
│     based on the words frequency     │
└─────────────────────────────────────┘
                    │
                    ▼
┌───────────────────────────────────────────────────────────────┐
│ On each document, determine the discriminatory words by        │
│ applying the following formula for each word, P:               │
│                                                                 │
│         word  frequency  in a doc                               │
│   P = ─────────────────────────────                             │
│        no of  documents  the word occurs                        │
└───────────────────────────────────────────────────────────────┘
                    │
                    ▼
┌─────────────────────────────────────────────────────┐
│ Extract the discriminatory words, called DW based on │
│                   the high P                         │
└─────────────────────────────────────────────────────┘
```

$$P = \frac{word\ frequency\ in\ a\ doc}{no\ of\ documents\ the\ word\ occurs}$$

Fig. 2. Determine the Discriminatory Words (DW). *Note*: *The value P for a word is small (i.e. less discriminatory value) if it has high occurrences in many or all documents. The value P is used as the weightage to the calculation of relevant, R value.*

3.1. 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 which are the Discriminatory Words identification process and the Relevancy computational process. The set of keywords for a subject domain is built by analyzing formal sources which are published documents such as journals, proceedings, magazines, newsletter, technical manuals, operation manuals, technical reports and others. The characteristics of these sources should be that the contents are specific to certain topic and there exists a finite set of discriminatory keywords (DW) 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 which are extracted from each document and these words are considered to be unique representatives for the chosen topics/subjects.

If the collections of documents are about "Knowledge management", then the DWs are the acceptable words being commonly used by the professionals, practitioners and researchers in that domain. The
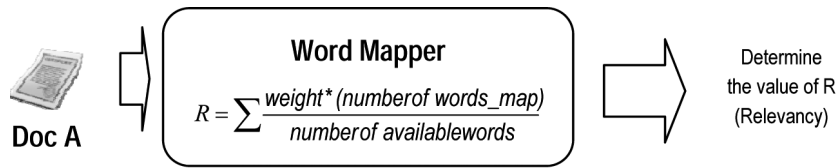
Fig. 3. Word mapper to determine relevancy.

collection of words is treated as the pool of words under the category of a specific domain. The advantage to this approach is that it can build its own recognizable word list without the need to manually build a dictionary or ontology. 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 Fig. 3. The *weight* is the value of P calculated from DW (process shown in Fig. 2). The value for the *words_map* is the word frequency of DW that appears in Doc A. The *number of available words* is the remaining words in **Doc A**. Word Mapper multiplies the strength of the DWs that appear in **Doc A** through weight value to increase the relevancy, R. However, the relevancy R will decrease its value when number of non-DWs overshadows the DWs.

3.2. 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 measured. The growth is measured by 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 a specific topic/subject (TG) iii) the number of references/usages (RUG) made by the knowledge sharing community iv) the number of resource maintenance activities (MG) performed by the knowledge sharing community such as uploading and downloading documents, posting messages etc.

3.2.1. Resource Growth (RG)

Resource Growth uses the Relevancy process (refer to Fig. 3) to determine the resources that are relevant to the specific topic that is measured at that time point. Only documents that are relevant to the identified topics are considered in the given Eq. (1). It calculates the number of relevant resources at time point T_1 and T_2 to determine the growth between the two periods.

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

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

$$RG = T_2 - T_1 \tag{1}$$

$$R - \textit{number of relevant resource}$$

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

where T_1 and T_2 are two different time points and $T_1 < T_2$. T_1 and T_2 are two different time points where the measurement of the difference between the two relevancy values of the resources at these two different time points is taken. Hence *RG* (*measurement of resource growth*) value is expected to be positive value if there is a resource growth.

3.2.2. Term Growth (TG)

Term Growth is another dimension of growth that is measured. While, Resource Growth (*RG*) measures the number of documents, Term Growth measures the number of terms in each document. The documents selected from the Resource Growth are used to measure the term frequencies that contain in the documents.

$$\lambda_i \in (\lambda_1, \lambda_2, \ldots \lambda_n) \; in \; \lambda \; where \lambda \; is \; the \; collection \; of \; terms \; in \; a \; document \; D$$

$$\lambda_{i\delta} = \frac{\left[\left(\sum_1^n 1 \, for \; each \; \lambda_i \; in \; D \right)_{T_1} - \left(\sum_1^n 1 \, for \; each \; \lambda_i \; in \; D \right)_{T_2} \right]^2}{(number \; of \; terms \; in \; D)^2} \tag{2}$$

$$TG = \sum_i^n \lambda_{i\delta} \, for \; \forall i \in D$$

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

$\lambda_{i\delta}$ calculates the ratio between the two time points T_1 and T in terms of number of terms at both time points. The ratio is calculated to obtain the normalized value since the number of terms differs for each document. The $\lambda_{i\delta}$ value is summed for every term in document D.

3.2.3. Reference Growth (RUG)

Reference Growth counts the number of times a document is referred by the members in the knowledge sharing activity.

$$r_{D_j} - the \; number \; of \; references \; for \; document \; D_j$$

$$\cup D_j - the \; total \; number \; of \; references \; for \; \forall D \tag{3}$$

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

where N is number of document and T_1 and T_2 are two different time points such that $T_1 < T_2$. *RUG* measures the ratio between the number of references of a particular document D and the total number of references that have been made to all documents. For example, in the knowledge sharing activities on online forum, each posted comment will have replies which can be counted as references.

3.2.4. Maintenance growth

$A = \{a_1, a_2 \ldots a_n\}$ where a is the type of maintenanc e activity and n is the type of activity
if N is the number of activities in A,

$f(a_i) = \sum_1^N 1$ is the sum of all activity type a_i

Given $F = \sum_i^n f(a_i)$ which is the total frequency for

all types of activities a_i, then

$$MG = \left[\left(\sqrt{\frac{(F)^2}{N}} \right)_{T_2} - \left(\sqrt{\frac{(F)^2}{N}} \right)_{T_1} \right] \tag{4}$$
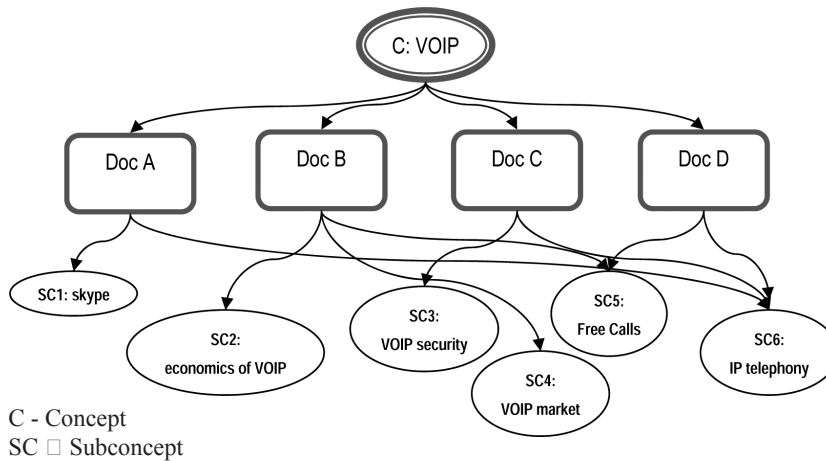
C - Concept
SC   Subconcept

Fig. 4. Concept and subconcepts of documents.

Maintenance growth counts the frequency for each type of activity during the knowledge sharing activity between two time points, T_1 and T_2. The activity type includes message posting, downloading and uploading of resources, viewing articles etc. F is the total frequency for the knowledge sharing activities. *MG* measures the number of activities that grow from time point T_1 to T_2.

3.3. 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", "VOIP security", "IP telephony". 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 as shown in Fig. 4. In order to determine the concept, a set of documents of a particular topic is identified and selected through subjective assessment. It is conjectured that the documents are identified to be categorized according to certain topic based on common keywords that are frequently used by all documents describing that topic. The keywords that appear repetitively in all documents under that topic are the concept to describe the domain of a paper. The subconcept is specific keyword to certain issue in the document and usually it is a collection of more than one keyword. In the collection, the set of keywords appears in some documents with certain level of frequency. For example, SECI model, internalization, socialization, combination and externalization are the subconcepts under the main topic of knowledge sharing or knowledge management but not all knowledge management papers contain these subconcepts.

The following variation algorithm describes the steps in determining the concept and subconcept keywords.

1. Given $d_i \in D$ where $i = \{1, 2 \dots N\}$ and N is the total number of documents in the set of document D
2. For $\forall d_i \in D, \forall t_j \in d_i$ where $j = \{1, 2, \dots M\}$ and M is the number of terms in document d_i, calculate $f(t_j)_{d_i}$
3. t_j is a concept if $\forall d_i \in D, f(t_j)_{d_i} > \varepsilon$ where ε is a threshold (term t_j appears in document d_i at a frequency $f > \varepsilon$) where ε is a threshold
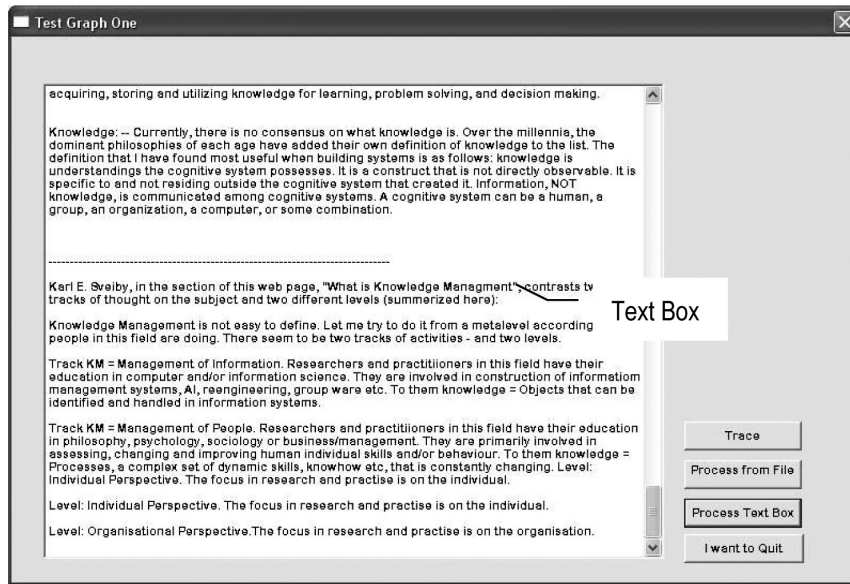
Fig. 5. Query processing. Trace: to display the knowledge domain mini window; Process from file: to process query; Process text box: to process query from the text box.
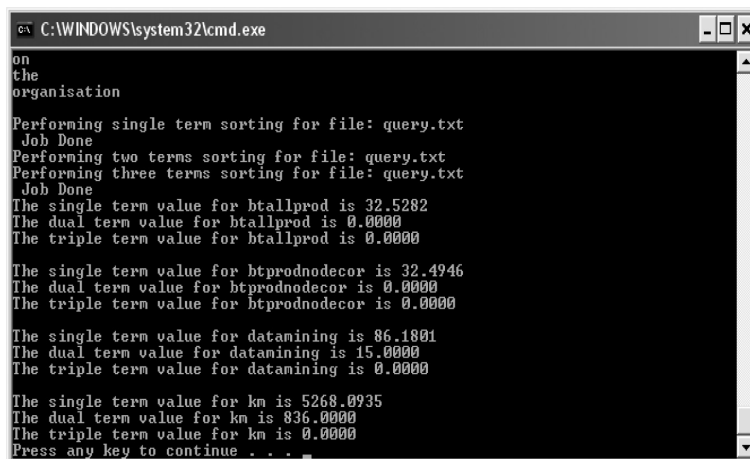
4. t_j is a subconcep t if $\exists d_i \in D, f(t_j)_{d_i} > \varepsilon$ where ε is a threshold (term t_j appears in document d_i at a frequency $f > \varepsilon$) where ε is a threshold

4. KFTGA demonstration

The two primary factors in demonstrating KFTGA are to show that the system is capable i) to sense the relevancy of documents to a given knowledge domain ii) to sense the knowledge growth of a specific knowledge domain based on cycles of knowledge sharing activities on the particular topic/subject. There are two types of operation in KFTGA which are the Query Processing and Knowledge Domain Building. The Query Processing processes the knowledge source (blogs, interview transcription, community channel, articles, manuals, e-mails etc) that a user is assumed to be operating within the knowledge sharing activity and exhibits the growth of the knowledge in the respective domain. The Knowledge Domain Building builds several knowledge sets for different domains. The knowledge domain is built from the collections of sources (i.e. pdf files) of a specific domain. The aim of the KFTGA system is to demonstrate the engine which is capable to trace the type of knowledge domain and to display the growth of knowledge.

4.1. Query processing

The query processing interface processes the knowledge source either from the text box or the text file. The knowledge source is an extraction from published materials or informal community's knowledge sharing platform such as blogs, forums, Q&As and community channel. The raw text from these sources is copied into the text box for submission to KFTGA engine or saved in an ASCII text file which can be called for processing. The difference is that the text box allows a limited number of characters compared to the text file. Figure 5 shows the text that is copied from a portal that describes about "knowledge

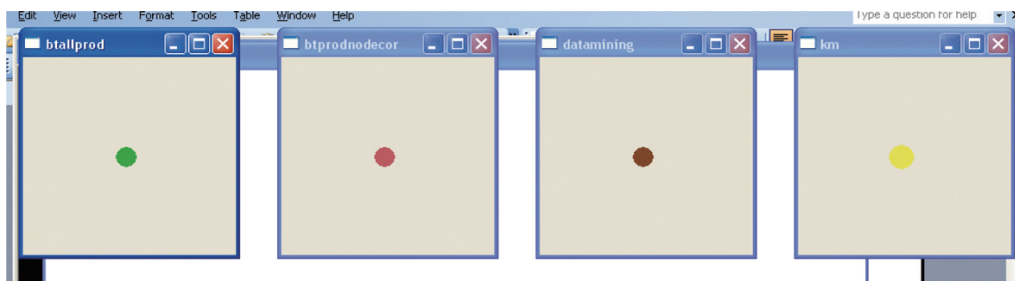Fig. 6. Windows showing matching value for different terms.

Fig. 7. Each mini window represents a knowledge domain and the size of knowledge growth. (Colours are visible in the online version of the article; http://dx.doi.org/10.3233/IKS-2012-0208)

management" with the assumption that the user is reading the topic from the portal. The "Process Text Box" button analyzes the contents of the text box against the pre-defined knowledge domain as shown in Fig. 6.

There are four pre-defined knowledge domain which are named as *btallprod*, *btprodnodecor*, *datamining* and *km*. For each of the knowledge domain, three types of terms which are the single term, dual terms and triple terms are used to compute their relevancy values against the knowledge domain. Figure 6 shows that the knowledge domain *km* has the highest values for its single term and dual terms. The knowledge growth for each knowledge domain is shown graphically when the "Trace" button (refer to Fig. 5) is clicked as shown in Fig. 7. The circle size for knowledge domain *km* enlarges each time the knowledge source on knowledge management subject is processed. The knowledge growth is shown on other knowledge domain depending on the knowledge domain of the knowledge source. The knowledge domain is built based on the collection of articles from related domain.

It is also important to demonstrate the capability of the Query Processor in handling the texts which are open to public for postings and the quality of the content is not moderated which may lead to uncontrolled used of colloquial language, abbreviation and style of writing. This is common in the public forum such as blogs, e-discussion etc as shown in Fig. 8. The Query Processor executed the basic information retrieval processes which include tokenization, stemming, indexing and application of recall and precision function. Even though KFTGA can manage un-moderated text as shown in
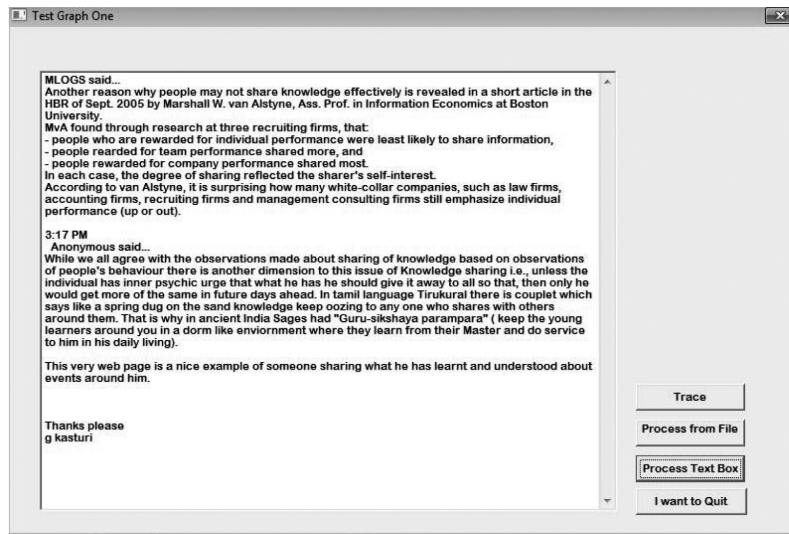
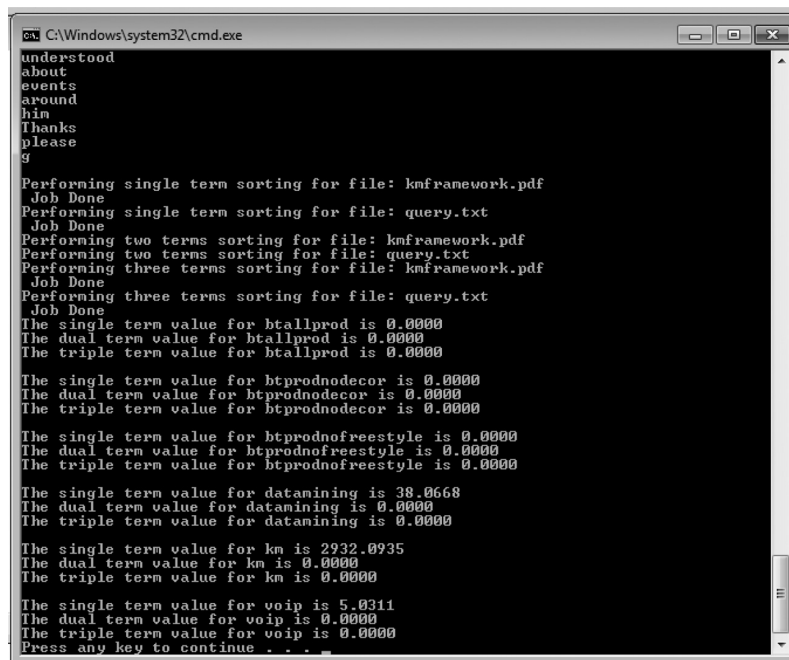Fig. 8. Query processing on public forum.

Fig. 9. Detection on KM Domain.

Fig. 8, there is certain level of tolerance in terms of its performances in handling which is not the scope of the research work presented in this paper. The outcome of applying the arbitrary text that is extracted from the website (http://knowledgemanagementforum.blogspot.com/2004/06/5-reasons-why-people-dont-share.html) related to Knowledge Management is shown in Fig. 9. The KM domain has shown the highest score for the *single term value* which is 2932 in comparison to other domains.
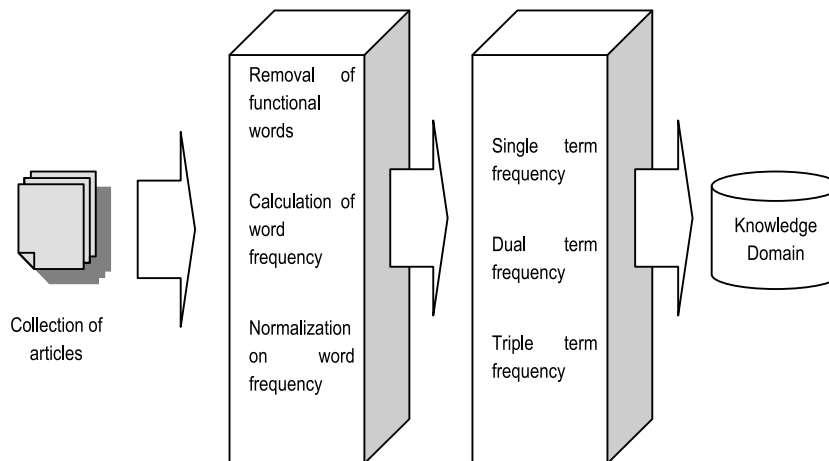
Fig. 10. Knowledge domain building.

4.2. Knowledge domain building

For each article, KFTGA determines the terms that uniquely represent the article based on single, dual and triple terms. These terms are compared against all articles that are categorized in the related domain. The common terms among all articles in the related domain will be generated to represent the overall knowledge domain. The common terms are determined based on the variation algorithm (cf. Variations, Section 3.0 KFTGA technology) where concepts and sub-concepts are identified. The concepts and sub-concepts that have high frequencies will have stronger representative values for the knowledge domain. For example, in the knowledge management domain, among the most popular dual terms are *knowledge management*, *knowledge flows*, *knowledge sharing*, *knowledge transfer*, *intellectual capital*, *organizational learning* and *explicit knowledge* and for triple terms are *knowledge management systems*, *knowledge management technology*. The process for knowledge domain building is simple since just user needs to submit a collection of articles in PDF format to KFTGA for the processing as shown in Fig. 10.

5. KFTGA applications

5.1. SECI model

KFTGA can be used in SECI modelled knowledge sharing environment. SECI model, introduced by Nonaka and Takeuchi, is the knowledge creation model for community of learning. It has received world-wide attention by many researchers [17]. It proposes four types of knowledge conversion of tacit knowledge and explicit knowledge as shown in Fig. 11. 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. The SECI model has received an equal acceptance, rejection and modification from various researchers. However, other alternative models shall also work with KFTGA technology. Chatti [7] argues that there is a significant relevancy of SECI model to Web 2.0 and describes how the emerging technologies can be used for
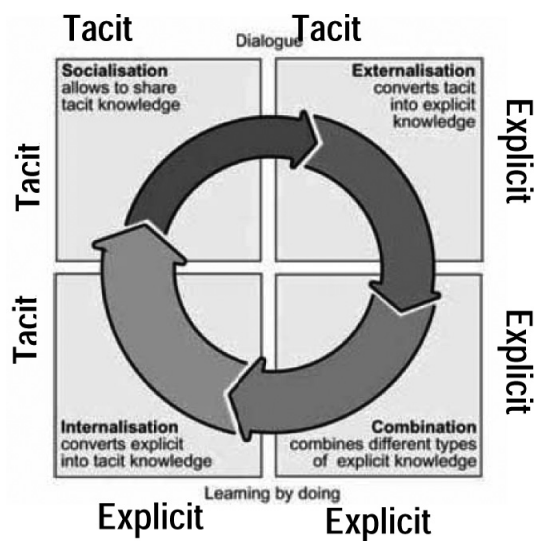
Fig. 11. SECI model (adapted from Nonaka [17]).

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 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 [9].

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 [20]. AcadeMix is an academic support system that encourages Japanese students and university communities to break the hierarchical barrier in communication, to encourage online collaboration, to stimulate academic material exchange and subsequently to promote Web 2.0 activities such as social publication, self-forming special interest group, community-based communication and semantic annotation and other personal support for Web 2.0 members [23]. Ouyang et al. [18] builds online learning system that adopts Web 2.0 in its learning environment where students are involved actively in the knowledge learning platform [18]. Web 2.0 applications such as blogs, Wiki Cyclopedia and RSS feeds are built as part of the learning platform. The underlying functions such as interest mining and recommender system use the information gathered from the participants' activities during the learning activities and learning processes.

KFTGA captures the learning activities that can be characterised by the SECI model in order to trace the knowledge flow evolution for an individual as well as the group. Based on SECI model, the following learning activities can be characterized by the four quadrant of SECI model: posting comments (externalization), summarizing articles (internalization), demonstrating skills and experiences (socialization), projecting arguments (externalization), managing learning materials and objects according to the context of discussion (combination) and reproducing new knowledge artefacts [14] (internalization
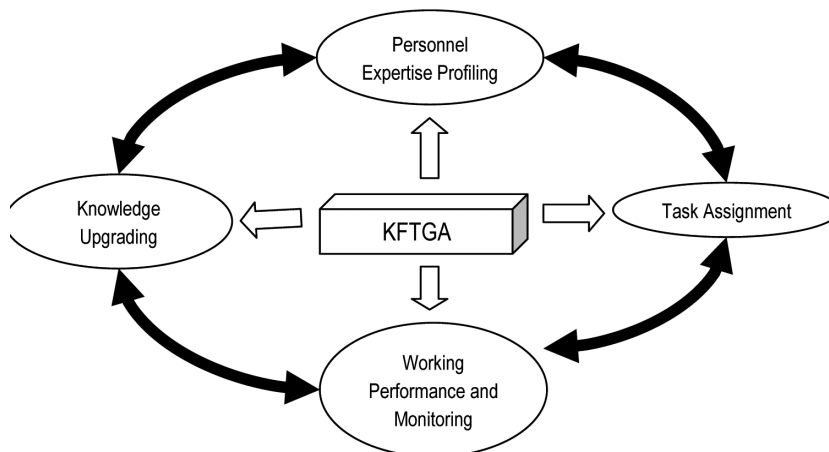
Fig. 12. Dynamic knowledge model in VDT system.

and externalization). KFTGA is developed to process text based knowledge sharing artifacts which are available in the form of forum, emails, postings, annotated video or images and others [14].

5.2. Organizational knowledge flow

There are different models that have been used to represent knowledge flow and knowledge sharing in the organizations and communities. The study on these models is essential to understand which part of the models that tracing the knowledge flow and knowledge growth can be used significantly. The followings are some of the examples of knowledge flow and knowledge sharing issues and their implementations.

5.2.1. Dynamic model of knowledge flow

Nissen and Levitt reported that the modelling of dynamic knowledge flow processes is built on the two synthesized components which are the ontology and the computational model [15]. The ontology models the organizational working processes, the personnel acting on the tasks and the working relationship between different roles. A more complex modelling at micro-organization level requires high consideration on the quantifiable and non-quantifiable elements of the organizational behaviour that appear in the form of culture, social and technicalities. For example, the management style, power delegation and authority, skills and knowledge of key personnel, task structures, and organizational structures are among the highly challenging elements to be captured and modelled. Nissen developed VDT (Virtual Design Team) system to emulate the design of physical system into the design of software development [16]. The system supports the dynamisms of the software development processes, evolutions and progressions throughout the complete life cycle. It defines the personnel attributes of the knowledge workers in terms of skills and roles, the connection between the personnel during the communication on assignment and supervision. The attributes show the dynamic relationships during the software development phases. At the beginning of software development phases, the task structures define the work assignment as software architect or software engineer. The execution of the tasks is captured in the VDT system that reflects the working relationship and working performance. The knowledge and skills that are associated with each personnel determines the task assignment. The system updates the task completion based on the projected timeline for each individual. The system captures the personnel's knowledge upgrading
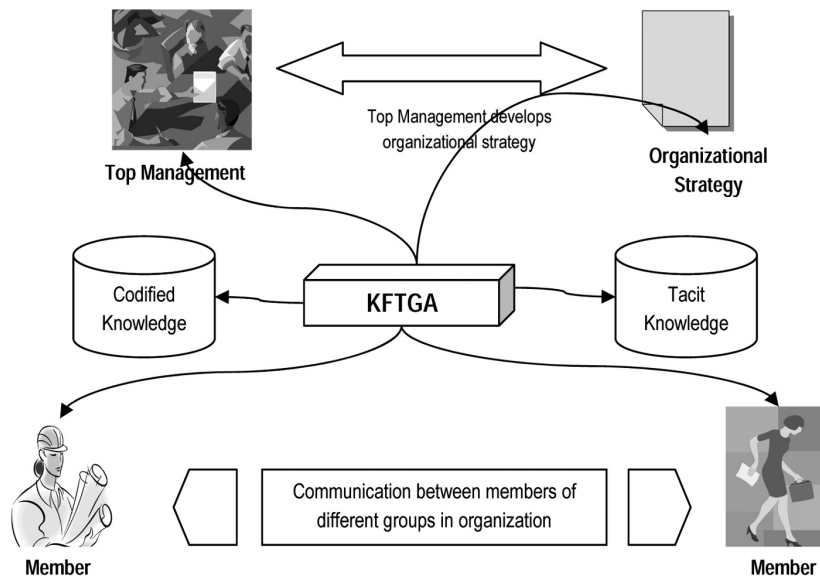
Fig. 13. Organizational knowledge transfer. (Colours are visible in the online version of the article; http://dx.doi.org/10.3233/IKS-2012-0208)

which is earned from the formal training or peer-to-peer coaching. It can be argued that the VDT system requires the knowledge flow tracing system as software development phases require an extensive knowledge searching, knowledge managing and knowledge eliciting for many tasks in the knowledge flow model. Figure 12 shows the round robin movement, as an illustration of how knowledge flow tracer could be used within VDT system. The personnel expertise profile is analyzed using the domain knowledge to determine the right talent for a certain expert area. The task has descriptive job requirement that states the knowledge and skills needed to be matched against the personnel expertise profile. The working performance and monitoring ties with each task assignment with indicators to show which skills or area of expertise are lacking that cause the detrimental effect to the expected performance. The work performance gaps have to be filled in with further staff training or peer-to-peer coaching for knowledge upgrading. KFTGA detects the appropriate knowledge flow from expertise profiling level to knowledge upgrading based on the knowledge domain that is built on the KFTGA system.

5.2.2. Organizational knowledge transfer model

The knowledge transfer model proposed by Ilovici and Han uses the concept of flow rate equation to emulate the organizational knowledge transfer [10].

The knowledge transfer model is claimed to exist if the following factors are fulfilled.

1. High degree collaboration and communication between members are the core component in the organizational working culture. Workers at the assembly line in the manufacturing company will have lower degree in comparison to software development team [4].
2. Knowledge transfer is the core competency required as the success factor in the organization strategy.
3. Top management initiative in pulling the information out from the organization as the routine steps in decision making.
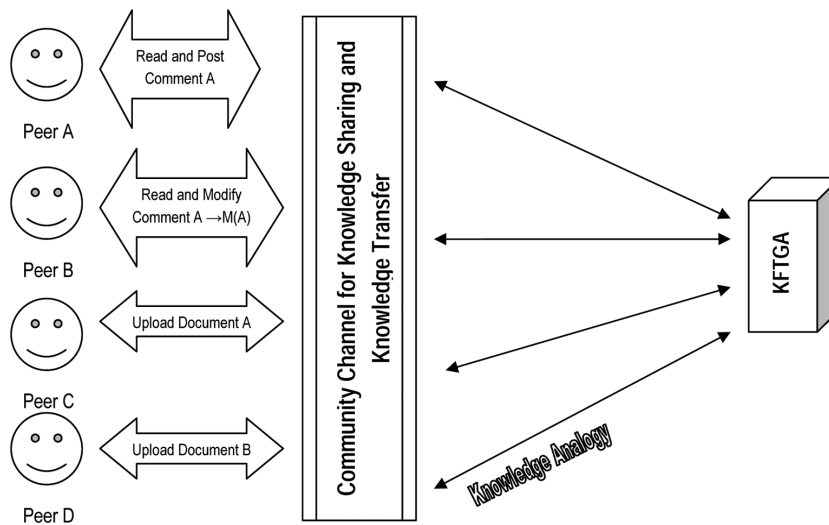
Fig. 14. KFTGA in the P2P knowledge sharing community channel.

4. Organizational structure layer that defines clearly the knowledge transfer flow between different levels of organizational community.
5. Knowledge is differentiated into codified knowledge and tacit knowledge in which the latter will be more challenging in knowledge transfer than the former.

Figure 13 illustrates the five factors above in a schematic form to demonstrate how KFTGA can be incorporated as the integral component in the organizational knowledge transfer model. The members are the employees of an organization or committee of the learning group. They share and exchange knowledge through various communication tools (such as email, community channels, blogs) mainly through text-based interface that can be analyzed by KFTGA to determine the subject area and topic of discussion. Top Management sets organizational strategy and communicates to different layers of people through middle and operational managers. The organizational strategy may contain the work instruction, company's regulation and policy, company's vision and mission and contemporary issues related to the core business of the organization that are recognized as learning materials of the employees. KFTGA monitors the knowledge flow and knowledge movement within these layers and determines the constant flow rate of knowledge. KFTGA is capable to measure the volumes of knowledge that flow around the organization and subsequently the flow rate. Therefore, KFTGA traces not only the subject or topic that flows among the employees or communities but also measure the flow rate through the calculation of knowledge growth. KFTGA supports the formal knowledge that can be codified from the official documents such as meeting minutes, corporate report, departmental report, scientific document and informal knowledge (tacit knowledge) which is less structured as those appear in emails, blogs, community channels, e-forums etc.

5.2.3. Peer-to-peer knowledge flow model

Zhuge has shown a formal presentation for the peer-to-peer knowledge flow model [24]. The formalism provides a detail and comprehensive knowledge flow definition that is scalable to describe a large knowledge flow network of an organization [24]. The knowledge flow describes the direction of the knowledge source to destination, the types of knowledge content and the carrier that supports the

transmission. The knowledge flow definition is extended to five important properties that include the aspects on knowledge accumulation, knowledge classification, knowledge abstraction, knowledge analogy and version management. The knowledge accumulation detects the knowledge content that is built over time, knowledge classification differentiates the subject matter of the knowledge content, knowledge abstraction provides the high level description of the knowledge content, knowledge analogy associates different sets of knowledge content that deemed to be related in certain context and version management tracks the changes of knowledge content made by the community. Zhuge [24] also describes possible connections on the knowledge flows, which are the sequential connection (SC), join-connection (JC), split (S) and broadcast (BC). Figure 14 demonstrates the role of KFTGA in the peer-to-peer knowledge flow model. SC occurs when two body of knowledge from the same domain are formed together in a sequential order. For example, in a threaded blog's text posting, one or more replies to an earlier posting will create a sequential connection of similar topics or subjects. In the case of blogs, there will be several sequential connections in the entire knowledge block. JC occurs when one or more body of knowledge from different topics or subjects are composed together in a single knowledge flow. The nature of a blog allows multifaceted topics of discussion that are contributed by various members who participated in various segments or sessions. Split states that a body of knowledge from a domain can be disseminated into several bodies of knowledge of several domains. While broadcast, disseminates similar piece of knowledge from a domain to different destinations. For example, the act of forwarding emails that were received from different source can be considered as split while forwarding the same email to different receivers is an example of broadcast.

6. Conclusion and future works

KFTGA emphasizes on building the technology where the learning flow of the community can be traced in terms of recognizing the type of knowledge domain. The learning flow as defined by the SECI model involves the four types of knowledge activities which can be supported by the KFTGA. This is possible since the recognition of the learning flow is based on the processing the learning artifacts which are text-based (blogs, textual documents, web pages etc). The model covers well on every aspects of knowledge sharing activities (KSAs) that could take place on the web environment.

It is described that the KFTGA system that we built could reside behind the knowledge sharing platform to 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 learning activities that the person has been involved in. Recent literature shows a growing interest in building learning system that supports the Web 2.0 learning environment and subsequently allows the use of KFTGA as the underlying function to trace the knowledge flow and the knowledge growth.

This paper discussed the methods and approach towards building the technology that is to identify the topics of the knowledge resource that is used by the learning communities. The results have shown that the KFTGA is able to trace the knowledge based on the knowledge domains that have been built earlier. Subsequently, there are some concerns that leave some spaces for future works such as to look at the effectiveness of the technology in implementing KFTGA in an organization, how the KFTGA system can be embedded into the knowledge sharing platform of an organization and how to integrate this technology.

References

[1] P. Akhavan, M. Jafari and M. Fathian, (2005). Exploring Failure-Factors Of Implementing Knowledge Management Systems In Organizations, *Journal of Knowledge Management Practice*, May. Online Access: http://www.tlainc.com/articl85.htm (Available on: 25 February 2012). Online Access: http://www.tlainc.com/articl85.htm (Available on: 25 February 2012).

[2] A. Anjewierden, R. Hoog, de, R. Brussee and L. Efimova, Detecting knowledge flows in weblogs. *Common Semantics for Sharing Knowledge: Contributions to ICCS 2005 13th International Conference on Conceptual Structures*, Frithjof Dau, Marie-Laure Mugnier and Gerd Stumme (eds.), Kassel University Press, Kassel, Germany, July, 2005, pp. 1–12. Online Access: http://staff.science.uva.nl/∼anjo/ (Available on: 25 February 2012).

[3] A. Arasteh and A. Afrazeh, Knowledge Flows Automation and Designing a Knowledge Management Framework for Educational Organizations International, *Journal of Industrial Engineering and Production Research* **21**(3) (September 2010), 155–162.

[4] O. Balogun, H. Hawisa and J. Tannock, Knowledge Management for Manufacturing: the Product and Process Database, *Journal of Manufacturing Technology Management* **15**(7) (2004), 575–584.

[5] K. Bisalyaputra, Knowledge Sharing as Sustainable Competitive Advantage, *International Engineering Management Conference* **3** (18–21 October 2004), 536–540.

[6] R. Boselli, M. Cesarini and M. Mezzanzanica, Customer Knowledge and Service Development, the Web 2.0 Role in Co-production, *Proceedings Of World Academy Of Science, Engineering And Technology* **30** (July 2008), ISSN 1307-6884.

[7] M.A. Chatti, The Web 2.0 Driven SECI Model Based Learning Process. *IEEE International Conference on Advanced Learning Technology, ICALT 2007*, 2007, pp. 780–782. Online Access: http://en.wikipedia.org/wiki/Social_media (Available on: 25 February 2012),

[8] A. Gachet and P. Brézillon, A context-based representation of knowledge flows in dynamic organizations. Revue d'Intelligence Artificielle, *Special Issue on Applying Context Management*, 19(3), S. Schultz, T. Roth-Berghofer and P. Brézillon, eds, 2005, pp. 557–574.

[9] L. Hämäläinen, Knowledge Creation, Communication and Role of IT in Software Design and Research Work, *Frontiers of E-Business Research 2003*, 2003, pp. 383–392. Online Access: http://www.ebrc.fi/kuvat/383-392.pdf (Available on: 25 February 2012).

[10] I. Ilovici and J. Han, Optimization of organizational Knowledge Transfer Model, *Proceedings of the 16th IEEE Symposium on Computer-Based medical systems* (*CBMS'03*), 2003, pp. 391–396.

[11] M.A. Jabar, F. Sidi and M.H. Selamat, Capturing Tacit Knowledge for Assessing Employees' Competency and Productivity, *American Journal of Economics and Business Administration* **3**(2) (2011), 358–362.

[12] H. Kim, J.G. Breslin, S. Decker, J. Choi and H. Kim, Personal Knowledge Management for Knowledge Workers using Social Semantic Technologies, *International Journal Intelligent Information and Database Systems* **3**(19) (2009).

[13] Y. Malhotra, Why Knowledge Management Systems Fail? Enablers and Constraints of Knowledge Management in Human Enterprises, In Michael E.D. Koenig and T. Kanti Srikantaiah, eds, Knowledge Management Lessons Learned: What Works and What Doesn't, Information Today Inc. (American Society for Information Science and Technology Monograph Series), 2004, 87-112. Online Access: http://www.brint.org/WhyKMSFail.pdf (Available on: 25 February 2012).

[14] B.D. Newman and K.W. Conrad, A Framework for Characterizing Knowledge Management Methods, Practices and Technologies, *Proc. of the Third International Conference on Practical Aspects of Knowledge Management* (*PAKM2000*). Basel Switzerland, 2000, 30–31st Oct.

[15] M. Nissen and R. Levitt, Dynamic Models of Knowledge-Flow Dynamics. *CIFE Working paper #76*, November, Center for Integrated Facility Engineering, Stanford University, 2002.

[16] M.E. Nissen, An Extended Model of Knowledge Flow Dynamics, *Communications of the Association for Information Systems* **8** (2002), 251–266.

[17] I. Nonaka and H. Takeuchi, *The Knowledge-Creating Company: How Japanese Companies Create the Dynamics of Innovation*. New York: Oxford University, 1995. Online Access: http://commerce.concordia.ca/el-sayed/KCreation.pdf (Available on: 25 February 2012).

[18] Y. Ouyang, C. Li, H. Li, P. Zhang and Z. Xiong, A Web 2.0 Based Computer Knowledge Learning Platform, *International Conference on Computer Science and Software Engineering* (*CSSE*), 2008, pp. 859–862.

[19] Y. Ren and X. Luo, Dynamic Models of Knowledge in Virtual Organizations, In Hai Zhuge, Geoffrey Fox (eds): Grid and Cooperative Computing – GCC 2005, 4th International Conference Proceedings. Lecture Notes in Computer Science 3795 Springer, Beijing, China, November 30–December 3, 2005, pp. 535–540.

[20] J.L. Rice and B. Rice, The Applicability of the SECI Model to Multi-Organisational Endeavours: An Integrative Review, *International Journal of Organisational Behaviour* **9** (2005), 671–682. Online Access: http://www.usq.edu.au/resources/vol9no8rice.pdf (Available on: 25 February 2012).

[21] R. Säljö, Learning in the learner's perspective. I. Some common-sense conceptions, *Reports from the Institute of Education, University of Gothenburg*, 76. Online Access: http://www.infed.org/biblio/b-learn.htm (Available on: 25 February 2012), 1979.

[22] S.M.F.D. Syed Mustapha, KFTGA: Knowledge Flow Tracer and Growth Analyzer for Community Learning. *International Journal Information Technologies and Knowledge* **2** (2008), Information Science and Computing – International Book Series, Number 5, pp. 85–91.

[23] Y. Tijerino, H. Masaki and N. Igaki, AcadeMix Juice – a Hybrid Web 2.0/Semantic Web Platform for Exchange of Academic Knowledge, *Proceedings of the 2006 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology* (*WI-IAT 2006 Workshops*), 2006, pp. 212–215.

[24] H. Zhuge, A Knowledge Flow Model for Peer-to-Peer Team Knowledge Sharing and Management, *Expert Systems with Applications* **23** (2002), 23–30.

[25] H. Zhuge, *The Knowledge Grid*. World Scientific, 2004.

[26] H. Zhuge, Knowledge flow network planning and simulation, *Decision Support Systems* **42** (2006), 571–592.

[27] H. Zhuge, J. Chen, Y. Feng and X. Shi, An Agent-Workflow-Federation Approach for Virtual Organization Development, *Information and Management* **39**(4) (2002), 325–336.