Using Key Terms to Measure and Visually Represent Community Knowledge

Abstract. This study explores possible ways of using key-terms to measure and visually represent community knowledge. The concept of community knowledge was first defined and then a set of key-term related measures were proposed based on this definition. A series of comparisons were employed to test if the proposed key-term measures added any values in complementing the conventional measures in assessing community knowledge. Findings suggested that number of key terms and frequency of key-term use were useful indicators of community knowledge. New ways of visually representing community knowledge by using key-term clouds was also discussed.

OVERVIEW

Knowledge building is a social process focused on the production and continual improvement of ideas of value to a community (Scardamalia & Bereiter, 2003), and defined by a set of knowledge-building principles which represent design challenges, ideals, and improvable objects in their own right (Scardamalia & Bereiter, in press). The principle of "community knowledge, collective responsibility" especially emphasizes the creation of community knowledge as the fundamental aim of knowledge building. As noted by Scardamalia (2002), contributions to shared, top-level goals of the community should be prized and rewarded as much as individual achievements; and team members should produce ideas of value to others and share responsibility for the overall advancement of knowledge in the community. Unfortunately, while the importance of advancing community knowledge has been greatly emphasized, the questions of "What exactly is community knowledge?" and "How is community knowledge measured?" have remained largely unexplored (Lee, Chan, & van Aalst, 2006). From an assessment point of view, it would be difficult to tell whether a community is truly advancing any knowledge if its community knowledge is not assessed. Indeed, it would be difficult to assess community knowledge without a clear working definition of it.

What is community knowledge?

Personal vs. public knowledge. To understand the nature of community knowledge, it is important to distinguish the differences between personal and public knowledge. The former emphasizes a psychological concept of knowledge and sees knowledge as possessed within an individual's mind (Hyman, 1999; Popper, 1972). In contrast, the latter highlights a social concept of knowledge and sees knowledge as a conceptual object that has a public life (Bereiter & Scardamalia, 1996; Hyman, 1999; Popper, 1972). Accordingly, community knowledge must be public knowledge, and knowledge can never become communal if it is possessed only within individuals. Knowledge must be made public (e.g., via intellectual conversation or note-posting in a forum) in order for it to be considered a candidate to become community knowledge. However, not all public knowledge can be treated as community knowledge. Knowledge must also be publicized in a place to which all community members can have access (e.g., on the class bulletin board or an online discussion forum) in order to increase the possibility of turning that knowledge into community knowledge.

Non-shared vs. shared knowledge. One way to define community knowledge is to see it as the sum of all community members' personal knowledge. For example, in some project-based learning, individual students may work independently on different science projects with some of their knowledge shared with others. Based on this definition, community knowledge will be regarded as the total sum of all students' knowledge derived from working in various projects, regardless of how much knowledge is shared (or overlapped) between community members. Alternatively, community knowledge may be defined as the core knowledge that is shared by all (not just some) community members. For example, in the kind of collaborative learning supported by the jigsaw method (Aronson & Patnoe, 1997), the goal is to have every community member master the same pre-specified knowledge. As such, only the part of knowledge shared by all members can be considered as community knowledge.

Clearly, the conceptual definition of the former seems too loose as it indiscriminately includes all non-shared knowledge. On the other hand, the latter concept of community knowledge seems too stringent to be practical, as it takes into account only the most densely shared knowledge. A more balanced approach may be to define community knowledge in between the two extremes. That is, as long as knowledge is shared between some (but not necessarily all) community members, it can be regarded as community knowledge.

Ephemeral vs. persistent knowledge. Specifically related to learning environments that involve certain online activities, there is an issue with regard to whether all public knowledge both occurred offline and online should be regarded as community knowledge. In a broad sense, the answer is yes, as typically in any class, much of public knowledge can take place in any form of knowledge discourse between students (e.g., in a hallway). The issue, however, is that such knowledge usually enjoys a rather short public life; consequentially, not all community members will have access to that specific piece of knowledge shared only between some members within a certain period of time. Another alternative is to define community knowledge in a more critical but constructive sense. That is, only when ideas, thoughts or knowledge are recorded in a public community space (e.g., an online forum) and can thus be continually accessed and improved (very much like published journals), they can be considered as community knowledge. Arguably, a conceptual definition as such would ensure any personal ideas to eventually have a true public, sharable and sustained social life, moving towards a trajectory of continual knowledge advancement in a community.

Breadth vs. depth of community knowledge. Arguably, regardless of what subject area is concerned, community knowledge can be measured in terms of the following two dimensions: breadth and depth. The former refers to the scope of all knowledge works collectively advanced in a community. The latter refers to the depth of collective understanding of these knowledge works in a community. While they are both important knowledge goals to pursue, there is also a tension between the two dimensions. All things being equal, if a community's aim is to cover a broad range of topics of inquiry, it would be unlikely for depth of community knowledge to take place in this community. As well, if a community's aim is to explore a specific topic in depth, it would be difficult for breadth of community knowledge to be achieved in this community. Figure 1 illustrates the conceptual definition of community knowledge employed in this study.

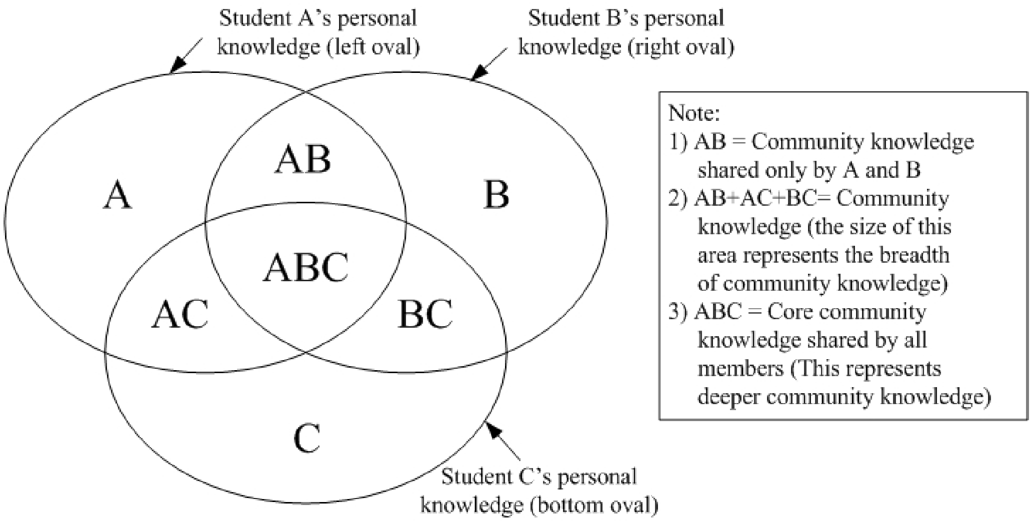


Figure 1. A conceptual definition of community knowledge

Measuring community knowledge

As a complex construct, the difficulty of assessing community knowledge lies in its involving not only each individual member's knowledge, but also the complex interactions between all members' knowledge. Typically in most learning environments where certain online activities are involved (e.g., WebCT), commonly used online measures are such as average number of notes created, average percentage of notes read, average number of notes revised, average number of words per note and the like. The strength of these behavioral measures is that they can be extremely useful in capturing the dynamics of online community activities, but they are poor in describing what content knowledge, whether it is personal and communal, is exactly being pursued online, not to mention assessing its depth or breadth.

To address this issue, content analysis has been commonly used as an alternative for capturing knowledge. While this approach is believed useful, there are issues as well. First, typically the kind of knowledge that content analysis is employed to capture is individual members' knowledge, not necessarily community knowledge. Indeed, what represents a valid approach to capturing community knowledge by using content analysis still remains unclear. Second, although content analysis is commonly used among researchers, it is not an easy-to-sue assessment approach for ordinary users (e.g., teachers and students) in an online environment. In other words, its feasibility for practical use in class is low. Third, it is a very time-consuming approach. Although efforts have been gathered to create tools for automatic content analysis, progress has been limited. Fourth, studies in the field of computer supported collaborative learning using content analysis as method for analyzing online knowledge works have also revealed that a lack of coherence in terms of theoretical base and the choice for the unit of analysis (see de Wever, Schellens, Valcke, van Keer, in press, for a review).

In between the above two assessment approaches, an alternative to measure community knowledge may be to use key terms. First, unlike conventional measures, it may be useful for representing content knowledge embedded in a database. As characterized by their power to represent important concepts or ideas, key terms have been widely used for subject-indexing (e.g., in books), for idea search (e.g., in academic papers or on Internet), and for knowledge representation (e.g., tag clouds, see Hassan-Montero & Herrero-Solana, 2006; semantic or propositional network, see Anderson, 2000; and knowledge or concept map, see O'Donnell, Dansereau, & Hall, 2002). Also, unlike most qualitative content analysis methods, key terms can be easily quantified in terms of their number and frequency of use; therefore, their potential usability to be developed as an automatic online assessment tool is high. Arguably, key terms as indicators can be useful to complement conventional measures (e.g., average number of notes) in better representing the actual knowledge advances in a community.

The purpose of this study is to explore possible ways of using key terms to measure and visually represent community knowledge. To this end, a comparison between traditional online measures and proposed key-term measures was employed to see if the latter can add any value to the former in assessing community knowledge.

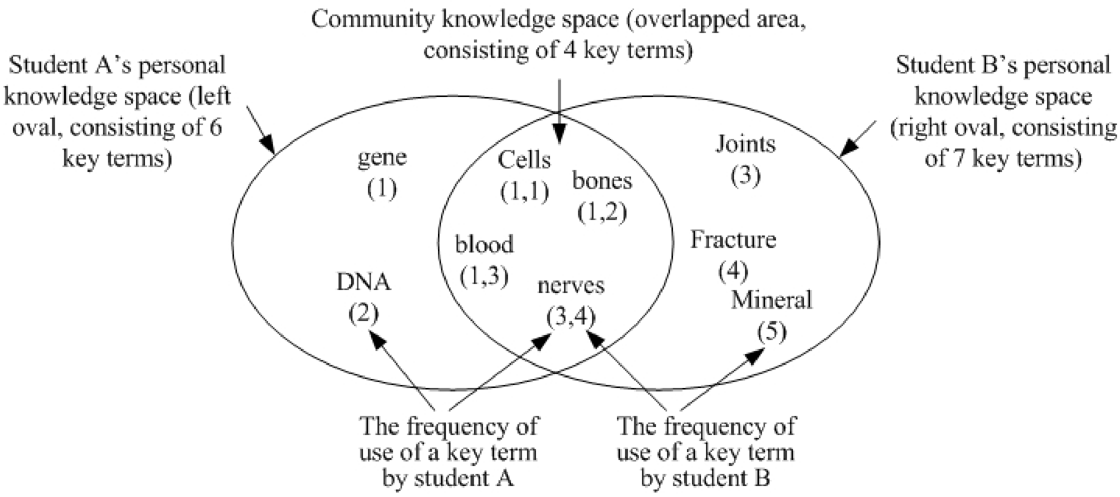
METHOD

Data source and context. Data mainly came from knowledge-building discourse among students in a Knowledge Forum database. There were two datasets in this database that was involved in this study. Both datasets were generated by the same students from a class at a school in downtown Toronto, Canada. The students were engaged in a course titled "integrated studies" and the theme to be explored was decided by the class as a whole to be "human body system." The aim of this course was to gain a deeper understanding, from an inter-disciplinary perspective, of the human body as a holistic system composed of many inter-related sub-systems. The course was divided into two major phases (thus two datasets): inquiry into the internal body as a system, e.g., studying how the brain, nerves, blood and cells work together (this phase covered the first nine weeks); and the physical body as a system, e.g. studying how the head, hands, legs, knees and feet coordinate to perform certain exercises such as a long jump (this phase extended from weeks 10-18). The two phases allow students to study human body system from both a biological perspective and a physics perspective.

Both phases had fundamentally the same class structure. All inquiry activities conducted in class, be it online or offline, were decided by students, not the teacher. More importantly, the implemented activities were guided by the knowledge-building principles (Scardamalia, 2002), not by any pre-determined procedures. Guided by principles, students were fully responsible for what and how they were to inquire, and were encouraged to improvise their inquiry activities so as to innovate knowledge (Messina, 2001; Caswell & Bielaczyc, 2002). These activities include, but not limited to, knowledge building talk[1] within small-group or whole-class discussion, watching videos, computer use, library search, reading, writing, conducting experiments, and working in a Knowledge Forum database etc. As a part of knowledge building culture, students were accustomed to recording all their ideas in their community database, be they from online or offline activities.

The students in this class consisted of 22 fifth and sixth graders (10 girls and 12 boys). They were all experienced knowledge builders as they had been engaged in knowledge building for five or six years depending on their grade level. The teacher was also an experienced knowledge-building practitioner given his seven-year experience of knowledge-building practice in teaching.

Proposed New Measures. Figure 1 exemplified how the proposed measures were used to assess community knowledge. In the left oval, the number of key terms (which is six) represents the breadth of key-term use in a student's (student A) personal knowledge space; and the frequency of appearance of these six key terms (which is nine times in total) represents the depth of key-term use by student A in his personal knowledge space. In the middle overlapped area, the number of key terms (which is four) represents the breadth of key-term use in the community knowledge space; and the frequency of appearance of these four key terms (which is 16 times in total) represents the depth of key-term use in the community knowledge space. In addition, in terms of how knowledge interacts between the personal and community knowledge spaces, student A' "knowledge contribution to the community" is represented by the percentage of his key terms contributed to the community knowledge space (four out of six, which is 66.7%). In contrast, "community contribution to individual member's personal knowledge" for student A is represented by the ratio of the frequency of key-term use in the community, over the frequency of key-term use by student A (16 times over 9 times, which is 1.78).



[1] Knowledge-building talk, in and of itself, is a guided intellectual discourse, whose goal is to facilitate students in shaping and improving their ideas through reflective social discourse.

Figure 1. An example of using number of key terms
and frequency of key-term use for knowledge representation

Method of Key-Term Extraction. Key terms were extracted in the following steps. First, all key terms within each student's notes in the database were extracted by using a Semantic Overlap Tool embedded in the Knowledge Forum software. At the same time, two researchers also manually went through each student's notes and extracted the remaining key terms missed by the tool (e.g., misspelled words). All automatically and manually extracted key terms were then merged into one complete set. Two researchers, both with science teaching background and extensive knowledge of the database content, then independently went through the whole set of key terms and removed those that were unrelated to the topics of inquiry under investigation in both inquiry phases (i.e. biological concepts in phase 1 and physical concepts in phase 2). Thirty irrelevant key terms were then removed from further analysis (see Appendix A for a complete list of all key terms extracted). The inter-rater agreement was 0.95, with differences resolved later by discussion.

Design and data analysis. Figure 3 shows the research design of this study. First, both the conventional measures (baseline) and the proposed key-term measures were respectively employed to compare the differences between phase 1 and phase 2, in terms of community knowledge that was recorded respectively in the two datasets. Then, a meta-comparison was further conducted to compare differences captured by using proposed key-term measures with differences captured by using the conventional measures. The purpose of this meta-comparison is to see if the proposed measures can help better capture some aspects of community knowledge that were not able to be captured by using the conventional measures. Table 1 summarizes all key measures employed in this study and the statistics used to compare the differences between the two datasets by the two sets of measures.

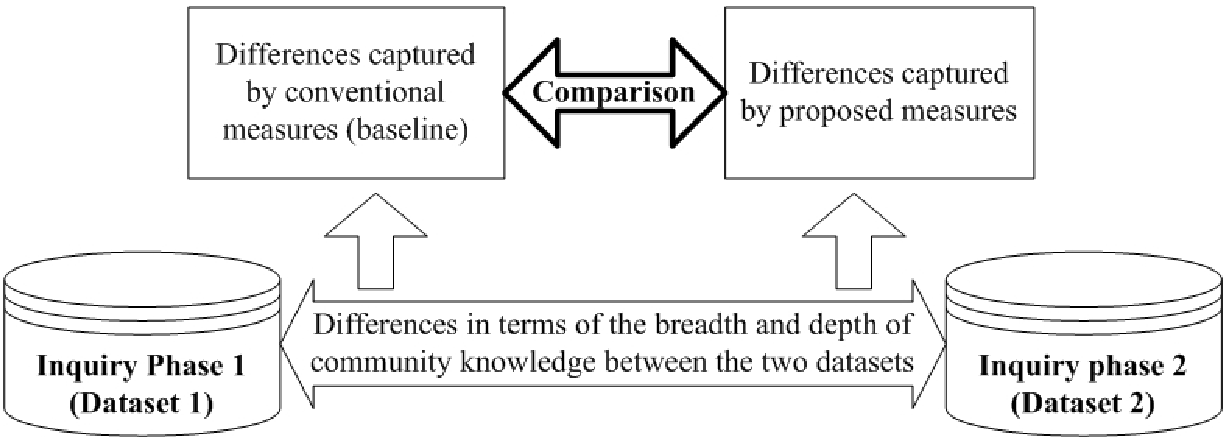


Figure 3. Research Design

Table 1. Key measures used in this study and related statistics

	Specific Indicators	Test Methods
1. Conventional measures	(1) Number of notes created, (2) percentage of notes read, (3) number of notes edited, (4) percentage of notes linked, (5) percentage of notes with keywords, (6) number of problems worked on, (7) number of non build-on notes, (8) number of build-on notes, (9) number of total words, and (10) number of words per note.	t-tests between phases 1 and 2
2. Proposed new	(1) Breadth of key-term use in a personal knowledge space,	(1) – (7) t-tests

key-term measures	(2) Depth of key-term use in a personal knowledge space, (3) Breadth of key-term use in the community knowledge space, (4) Depth of key-term use in the community knowledge space, (5) Personal contribution to community knowledge, and (6) Community contribution to personal knowledge, (7) Knowledge interaction between personal and community knowledge spaces.	between phases 1 and 2 (6) Pearson correlations

RESULTS AND DISCUSSION

Differences captured by conventional measures

First, as a base line comparison, a t-test found that there was no significant difference between the two phases in terms of the total number of notes contributed (t=1.89, df=21, p>.05, with N=254, M=11.55, SD=4.39 in phase 1, and N=212, M=9.64, SD=3.92 in phase 2). In a more detailed manner (see Table 4), t-tests indicated that there were significantly more problems (which were entered by students when contributing a note) being generated and worked on in the first phase. Also, there were significantly more notes being linked and built-on in the first phase. On the contrary, there is more words (per note) being generated by students in phase 2. Table 2 summarizes the results.

Table 2. Differences between the two phases as captured by the conventional measures (N=22)

Basic Indicators	Phase 1		Phase 2		t values
	Mean	SD	Mean	SD	
Number of notes created	11.55	4.39	9.64	3.92	1.89
Number of notes read	12%	7%	8%	4%	1.96
Number of notes edited	5.32	2.15	4.05	2.66	1.89
Percent of notes linked	25%	20%	10%	7%	3.33**
Percent of notes with keywords	13%	18%	6%	16%	1.98
Number of problems worked on	9.41	4.35	2.00	2.16	8.35***
Number of non build-on notes	7.32	2.50	6.64	3.30	0.82
Number of build-on notes	3.00	3.13	1.00	0.76	3.24**
Number of total words	574.77	227.20	699.82	300.26	-1.93
Number of words per note	51.03	17.73	72.61	15.83	-5.66***

** P<.01 *** P<.001

Differences captured by the proposed measures

In terms of the mean number of key terms each student contributed, it was found that there was no significant difference between the two phases (t=2.065, df=21, P>.05). On average, each student generated 40.10 key terms (SD=17.74) in phase 1 and 33.27 key terms (SD=11.35) in phase 2 (see Table 3). However, in terms of the mean frequency of key-term use by each student, there was significant difference between the two phases (t=-2.483, df=21, P<.05) in that the mean frequency of key-term use is 78.41 (SD=38.74) in phase 1 and 100.14 (SD=40.11) in phase 2.

In terms of the mean number of key terms overlapped in the community, there was no significant difference between the two phases (t=-0.617, df=21, P>.05; with M=29.10, SD=12.55, in phase 1 and M=20.64, SD=9.35, in phase 2). However, in terms of the mean frequency of use of overlapped key terms in the community, there was a significant difference between the two phases (t=-7.686, df=21, P<.001), with the second phase (M=324.36, SD=68.0) having a higher value than the first phase (M=203.77, SD=63.77).

Further, how knowledge is interacted between the personal and communal knowledge spaces, i.e., how individuals contribute their knowledge to the community and how community in return helps individuals deepen their knowledge (as defined earlier), were also investigated. In terms of the former, the results showed a significant difference between the two phases ($t=-6.207$, $df=21$, $P<.001$) in that the mean percentage in phase 2 ($M=93.3\%$, $SD=6.0\%$) is higher than in phase 1 ($M=73.2\%$, $SD=12.5\%$), suggesting that students contributed more key terms to the community knowledge space in phase 2. In terms of the latter, the results also showed a marginally significant difference between the two phases ($t=-1.976$, $df=21$, $P<.10$) in that phase 2 has a higher ratio ($M=3.68$, $SD=1.41$) than phase 1 ($M=2.97$, $SD=1.05$), indicating that key terms are more likely to be repeatedly elaborated or referred in the community knowledge space in phase 2. Table 3 summarizes the results.

Table 3. Changes between the two phases as captured by the proposed measures (N=22)

	Phase 1		Phase 2		
	M	SD	M	SD	t values
Personal Knowledge Space					
Breadth of key-term use	40.10	16.74	33.27	11.35	2.065
Depth of key-term use	78.41	38.74	100.14	40.11	-2.483**
Community Knowledge Space					
Breadth of key-term use	29.1	12.55	30.64	9.35	-0.617
Depth of key-term use	203.77	63.77	324.36	68.02	-7.686***
Knowledge Interaction between the two spaces					
Personal contribution to the community knowledge space	73.2%	12.5%	93.3%	6.0%	-6.207***
Community contribution to the personal knowledge space	2.97	1.05	3.68	1.41	-1.976*

* p<.10 **p<.05 ***p<.001

Table 4 also summarizes all possible correlations as resulted from knowledge interaction between the two knowledge spaces. The high correlation coefficients (ranging from 0.79 to 0.99) suggested there were frequent knowledge interactions between the personal and community knowledge spaces in both phases.

Table 4. Pearson's correlation coefficients between personal and community knowledge

	Personal Knowledge Space			
	Phase 1		Phase 2	
	Breadth of key-term use	Depth of key-term use	Breadth of key-term use	Depth of key-term use
Community knowledge Space				
Phase 1				
Breadth of key-term use	0.89***	0.88***	0.44*	0.42
Depth of key-term use	0.90***	0.79***	0.38	0.32
Phase 2				
Breadth of key-term use	0.48*	0.56**	0.99***	0.84***
Depth of key-term use	0.43*	0.53*	0.88***	0.81***

*p<.05 **P<.01 ***P<.001

To further explore how community knowledge might differ between the two phases, an integral use of both the number of key terms and the frequency of key-term use was employed. Figures

4 and 5 illustrate the community knowledge of phases 1 and 2 as represented by two key-term clouds. Key terms displayed in both clouds (N=231 in phase 1 and N=140 in phase 2) represents the terms generated in each respective community knowledge space during each phase. They were alphabetically ordered and visually weighted by font size. The frequency of key-term use ranges from 2 to 75 in phase 1 and ranges from 2 to 352 in phase 2. Key terms with frequency of once were excluded from display in both key-term clouds (184 key terms in phase 1 and 38 key terms in phase 2) as they belong only to the personal knowledge space, not the community knowledge space.

Figure 4. Community Knowledge in Phase 1 as Represented by Key-Term Cloud

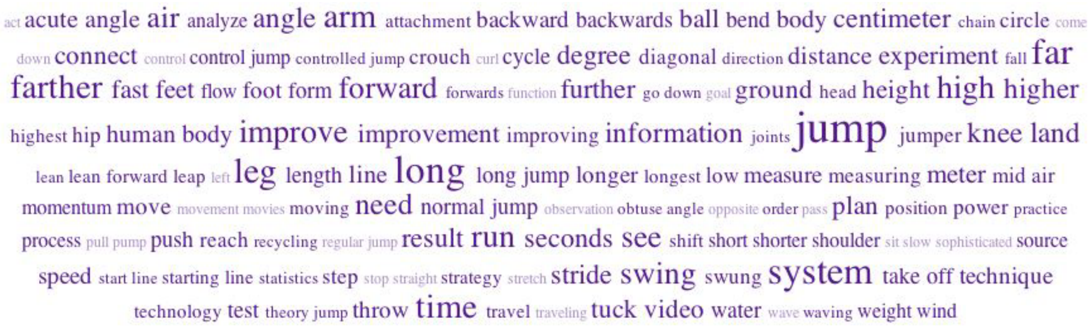


Figure 5. Community Knowledge in Phase 2 as Represented by Key-Term Cloud

Figure 6 also shows the relationships between number of key terms and frequency of key-term use in each phase. As indicated by its higher number (N=184) of key terms that appeared only once in phase 1 (as compared with phase 2; N=38), it seems that individual members were more active in their own personal knowledge space than in their community knowledge space in phase 1. And as also indicated by the higher total number of key terms (N=231) with lower total frequency of their collective use (f=1568) in the community knowledge space in phase 1 (as compared with phase 2; N=140 and f=2217), it seems the focus of community knowledge works was more breadth-oriented in phase 1, whereas the focus of community knowledge works was more depth-oriented in phase 2.

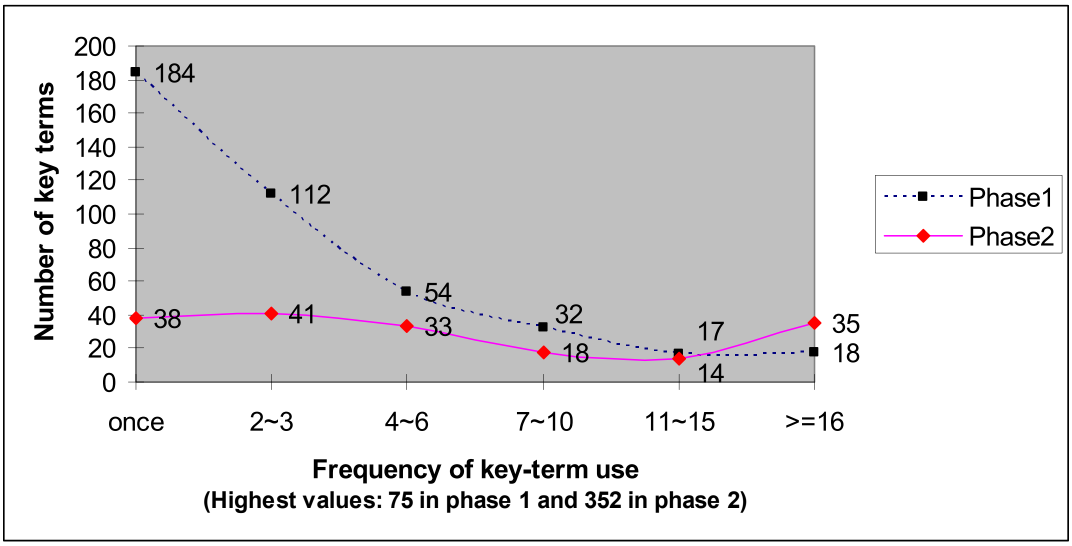


Figure 6. Relationships between the number of key terms
and frequency of key-term use in both phases

Comparisons between the conventional and proposed key-term measures

First, in terms of conventional measures, clearly, the results from comparison showed that there were differences in the community's knowledge building dynamics between phase 1 and phase 2. It seems that (1) less time was spent on note-linking, building-on, and problem generation, but (2) more time was spent on writing longer notes in phase 2. But what can these results tell about the actual breadth and depth of community knowledge? Not much. For example, the measure of "number of problems worked on" (in which phase 1 has a higher number than phase 2) may appear to be a good indicator of the breath of community knowledge, because given the same amount of time, the more problems a community is working on, the more likely the effort will be dispersed to deal with different issues and problems, and the less likely the depth of understanding will be achieved. But depending on the nature of problems, if the problems are all related to the same conceptual issue, then the high number of problems may indicate a course of progressive problem-solving in phase 1, which may eventually leads community members to achieve a deeper collective understanding. As another example, the findings also showed that the number of words per note in phase 2 was higher than phase 1. Intuitively, this measure may be used as an indicator of the depth of community knowledge, as the higher number of words per note seemed to suggest that students were more reflective and elaborative in phase 2. But an alternative explanation is that students were simply adding new ideas in their notes, not necessarily elaborating in-depth on the same issues. This is why conventional measures are not very reliable for assessing community knowledge.

On the other hand, findings from the proposed key-term measures seemed to clearly indicate that the approaches to knowledge advancement tend to be more breadth-oriented in phase 1 and to be more depth-oriented in phase 2. For example, the higher total number of key terms (N=231) and the higher number of the key terms that appeared only once (N=184) both occurred in phase 1 may be used as facts to complement the findings from conventional measures that the higher number of problems worked on" in phase 1 was associated more with the pursuit of a breadth-oriented community knowledge. As another example, the higher frequency of key-term use in the community space in phase 2 may be used to complement the findings from conventional measures that the higher number of words per note is associated more with the fact that students were actually more reflective and elaborative in deepening their community knowledge in phase 2. This is why the proposed key-term measures may be useful as they can be used to complement the conventional measures in better capturing the actual content knowledge (depth or breadth) in a community.

Expanding Possibilities

Admittedly, key-term use has its limitations, too. For example, one potential limitation may be related to its deficiency in capturing process-related community knowledge. Ryle (1949) argues that “know-that” and “know-how” are two essential kinds of knowledge. In various form, they are also referred to as declarative and procedural knowledge (Anderson, 2000). Corresponding to the concept of community knowledge, know-that may refer to the key ideas, concepts or problems collectively inquired in a community. Know-how may refer to the process knowledge that helps fulfill a deeper understanding of these ideas, concepts or problems. Arguably, as long as ideas or concepts are recorded in a community database, key terms are a great tool in capturing them and using them to assess the scope and depth of collective inquiry in a community via total number of key terms and frequency of their use. However, they may not be useful in assessing process-related community knowledge, e.g., how a certain key concept is actually pursued in depth over time. To supplement this deficiency may require other measures such as in-depth content analysis with a larger grain-sized unit of analysis (e.g., sentence).

The point in this study, however, is not to focus so much on the deficiency of key-term use but to focus on how to further develop its potential as a better tool for measuring and representing community knowledge. As noted earlier, key terms may be constructed as key-term clouds and used as a visual tool for knowledge representation. A tool that has been developed within Knowledge Forum is called a Semantic Overlap Tool (Figure 7). In its current version, it is capable of automatically extracting key terms from all notes contributed in a Knowledge Forum database, and of comparing key terms from different notes and then distinguishing shared from non-shared terms. Building on the findings from this study, several potential further developments may be added as new functions in the current Semantic Overlap Tool.

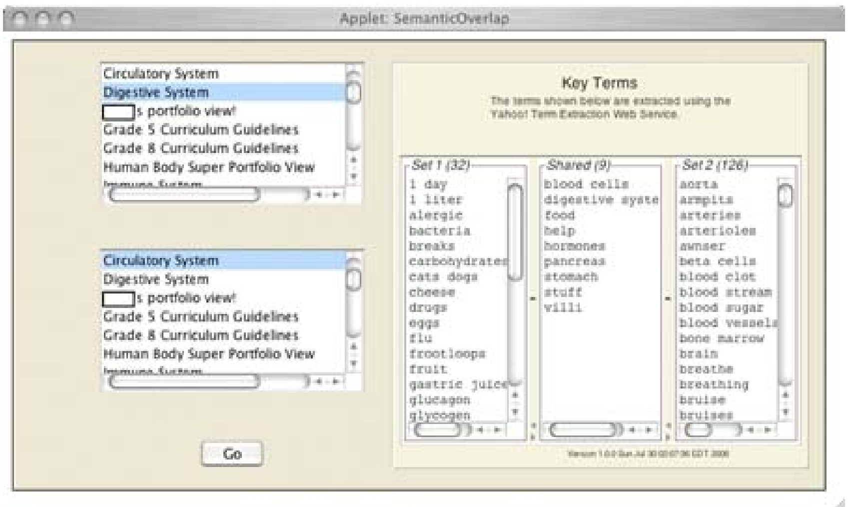


Figure 7. Semantic Tool: Illustrating shared ideas/key terms between two sets of *notes*

For example, the use of key-term clouds may has a great potential to be developed as a visual knowledge representation tool for better capturing and illustrating an overall content knowledge in a community. Key terms contained in a key-term cloud may be also used as subject index of a database, and if hyperlinks are provided, key terms can also serve as access points for members to further search for relevant notes or ideas within these notes for personal idea improvement or for opportunistic collaboration around similar ideas between members. So, it is more than an assessment tool. It is a knowledge-building tool. Moreover, as key terms can be used to represent both personal knowledge and community knowledge, the comparison between the two types of knowledge via various arrangement of key-term clouds (e.g., different coloring to distinguish personal knowledge from community knowledge) can also help individual members better understand how, and to what extent, they share their collective cognitive responsibility (Scardamalia, 2002) in a community.

Arguably, technologies designed as such would help extend members' social metacognitive capacity (e.g. knowledge of others' knowledge, Hong, 2005; Hong & Lin, 2005) to support epistemic agency (Russell, 2002; Scardamalia, 2002) for more effective knowledge building initiated by the members themselves. These new technologies should allow members to monitor and reflect more often on who has worked on which ideas (or sets of ideas), so members share a meta-perspective on their work. More effectively distributed knowledge building processes should result (Hewitt & Scardamalia, 1998).

References

Anderson, J. R. (2000). *Cognitive psychology and its implications* (5th ed.). New York: Worth.

Aronson, E., & Patnoe, S. (1997). *The jigsaw classroom: Building cooperation in the classroom*. New York: Longman.

Bereiter, C., & Scardamalia, M. (1996). Rethinking learning. In D. R. Olson & N. Torrance (Eds.), *The handbook of education and human development: New models of learning, teaching and schooling* (pp. 485-513). Cambridge, MA: Basil Blackwell.

Caswell, B., & Bielaczyc, K. (2001). Knowledge Forum: altering the relationship between students and scientific knowledge. *Education, Communication and Information, 1*(3), 281-305.

de Wever, B., Schellens, T., Valcke, M., & van Keer, H. (in press). Content analysis schemes to analyze transcripts of online asynchronous discussion groups: A review. *Computer & Education*.

Hassan-Montero, Y., & Herrero-Solana, V. (2006). *Improving tag-clouds as visual information retrieval interfaces.* Paper presented at the International Conference on Multidisciplinary Informatin Sciences and Technologies (InSciT), Merida, Spain.

Hewitt, J., & Scardamalia, M. (1998). Design principles for distributed knowledge building processes. *Educational Psychology Review, 10*(1), 75-96.

Hong, H. Y. (2005). *Effects of people knowledge on science learning in a computer-based learning environment.* Unpublished Dissertation, Teachers College, Columbia University, New York.

Hong, H. Y., & Lin, X. D. (2005). *Effects of people knowledge on science learning.* Paper presented at the annual conference of American Educational Research Association Montreal, Canada.

Hyman, J. (1999). How knowledge works. *The philosophical quarterly, 49*(197), 433-451.

Lee, E. Y. C., Chan, C. K. K., & van Aalst, J. (2006). Students assessing their own knowledge building. *International Journal of Computer-Supported Collaborative Learning, 1*, 277-307.

Messina, R. (2001). *Interactive learners, cooperative knowledge building, and classroom inventions.* Paper presented at the American Educational Research Association, Seattle WA.

O'Donnell, A. M., Dansereau, D. F., & Hall, R. H. (2002). Knowledge maps as scaffolds for cognitive processing. *Educational Psychology Review, 14*(1), 71-86.

Popper, K. R. (1972). *Objective knowledge: An evolutionary approach*. London: Oxford Univ. Press.

Russell, A. (2002). *The role of epistemic agency and knowledge building discourse to foster interprofessional practice in a Canadian hospital.* Paper presented at the the American Educational Research Association, New Orleans.

Ryle, G. (1949). *The Concept of Mind* London: Hutchinson.

Scardamalia, M. (2002). Collective cognitive responsibility for the advancement of knowledge. In B. Smith (Ed.), *Liberal education in a knowledge society* (pp. 67-98). Chicago: Open Court.

Scardamalia, M., & Bereiter, C. (2003). Knowledge building. In *Encyclopedia of Education* (pp. 1370-1373). New York: Macmillan Reference, USA.

Scardamalia, M., & Bereiter, C. (in press). Knowledge building: Theory, pedagogy, and technology. In K. Sawyer (Ed.), *Cambridge handbook of the learning sciences*.

Appendix A. Key terms contributed in each phase and their frequency of use

Key terms and their frequency of use
Phase 1: adult (1), alcohol (5), allergic (7), allergy (3), animal (4), ankle (3), antibodies (14), aorta (2), arm (8), armpit (1), arteries (4), arthritis (6), atrium (1), Avian flu (1), axis (1), babies (2), baby (5), bacteria (3), ball (2), ball and socket joint (2), B-cell (5), bend (1), benign tumor (1), beta cell (1), bird (2), birth (2), bleed (1), blood (54), blood cell (25), blood circulation (1), blood stream (1), blood sugar (11), blood transfusion (2), blood vessel (6), bodies (2), body (50), bone (74), booger (2), brain (49), brain imaging (3), brain tissue (1), brain tumor (3), break (6), break down (3), breast bone (1), breath (10), breathe (5), brittle (2), Broca's area (1), broke (3), brother (1), bruise (7), bug (1), calcium (4), cancer (2), canine (1), carbohydrate (2), carbon dioxide (12), cartilage (6), cast (2), cat (1), cell (54), central control (1), chalk dust (1), chamber (3), chimpanzee (1), chromosome (4), circulate (1), clog (2), cloning (2), cold (6), collar (1), collar bone (9), coma (1), compact bone (1), conjoined twins (1), connection (2), contagious (1), contagious disease (2), cool down (1), cord (2), core of red marrow (5), cough (2), crack (2), crash (1), crush (2), current (3), cut (11), DNA (6), damage (2), damaged cell (1), dangerous (1), death (1), defense (3), diabetes (8), dial (1), diaphragm (2), die (9), digest (7), digestive system (1), disabilities (1), disease (23), dislocate (1), disorders (3), division (1), dizziness (1), dizzy (1), doctor (3), dog (1), down syndrome (9), dream (3), drink (9), dust (1), ear (7), ear bone (1), ear wax (1), eat (5), Egypt (1), elbow (1), electric (1), electrical signal (1), ellipsoidal (2), ellipsoidal joint (1), embryo (4), embryonic stem cell (3), enemy (1), energy (2), enzymes (1), erythrocytes (2), evolve (2), exercise (2), facial (3), faint (1), fat (4), fatigue (1), fats (1), feather (1), feel (21), fetus (1), fever (2), fibula (1), fight (2), fight disease (2), filter (1), flat bone (2), floppy (1), flow (3), flow of blood (1), fluid (4), food (15), foot (5), foreign particle (1), formulation (1), fracture (11), frontal lobe (1), gas (1), gastric juice (1), gene (5), genetic (2), genetic code (1), germ (17), gland (1), glucagon (2), glucose (1), glycogen (1), gold (1), gorilla (1), grandparent (1), gravitation (1), grow (11), growl (2), hair (3), hammer (1), hand (3), head (11), headache (3), heal (7), hear (24), hearing loss (1), heart (19), heat (1), hemoglobin (4), hereditary (3), hinge (1), hinge joint (3), homing device (1), honeycomb (1), hormone (8), hot (3), human (15), humorous (1), hungry (5), ice (2), illness (5), immune system (11), immunity (1), infection (1), inject (2), injure (1), injury (2), insulin (10), intestine (6), invader (1), irregular bone (1), itch (1), jaw (1), joint (19), juice (1), jump (1), kidney bean (1), kill (2), kiss (4), knee (8), knee cap (1), knuckle (4), language area (1), learn (4), leg (10), life threatening (1), ligament (1), liquid (4), lisp (3), live (18), liver (13), liver disease (6), living tissue (1), long bone (4), lower neck (1), lump (2), lung (6), lymph (23), lymph gland (2), lymph node (8), lymphatic system (2), lymphocyte (4), lysosome (10), macrophage (15), malignant tumor (1), marrow (5), medicine (3), membrane (2), memories (2), memory task (2), men (5), message (10), milk (2), mind (2), mineral (2), mitochondria (2), molar (1), monkey (4), mononucleosis (3), mosquito bite (1), mother (1), motor strip (1), mouth (7), mucous (1), mucus (1), multiply (1), muscle (8), neck (3), needle (1), nerve (37), neuron (1), neutron (2), Nile (1), non-disjunction (1), nose (4), nucleus (10), numb (3), number (2), nut (3), nutrient (2), old age (1), orangutan (1), organ (5), oval cup (1), overweight (1), oxygen (12), pain (7), pancreas (6), paralysis (1), paralyze (2), parent (6), passage (1), patella (1), pathogen (5), people (14), periosteum (2), pesticide (1), petechiae (1), phagocyte (3), pine (1), pivot (2), pivot join (1), plane joint (2), plant (1), platelet (1), pollen (1), power (3), predator (1), pregnancy (1), pulmonary (1), pump (9), purpura (1), radius (1), rat (4), reaction time (1), recognition (1), red marrow (8), refill (1), reflex (7), regional blood flow (1), remember (5), reproduce

(2), reproducing (1), rib (5), root (1), rub (2), saddle joint (1), saliva (2), salt (2), scientist (2), scull (1), see (8), sense (8), separate (1), separation (1), shinbone (1), ship (1), short bone (1), short term (1), shorten (1), shoulder (5), Siamese (3), sick (7), side affect (1), signal (14), sister (1), skeletal system (1), skeleton (1), skin (13), skull (3), sleep (2), sliding (1), small intestine (3), smell (1), sneeze (3), sneezing (1), socket joint (1), sooth (1), speed (3), spinal cord (1), spine (5), split (1), spongy bone (3), spread (5), steel (2), sternum (1), stomach (16), straight (1), strength (2), stress (11), sugar (1), sun flower (1), sunscreen (1), sweat (7), swollen (9), Sylvian fissure (1), taste (1), T-cell (3), teeth (5), temperature (1), tendon (1), thigh bone (1), throw up (1), thumb (1), thymus (1), tickle (1), tired (6), tissue (1), toe (2), tongue (2), touch (4), transmit (1), triplet (2), trunk (1), tube (1), tumor (8), twin (5), ulna (1), uncontrolled (1), universal joint (1), upright (3), vein (6), ventricle (1), villi (4), viral (1), virus (22), voluntary action (1), vomit (1), walk (5), wall (1), waste (1), water (17), weak (3), weight (4), Wernicke's area (1), white cell (2), wind (1), woman (1), womb (2), women (5), working memory (1).

Phase 2: act (2), acute angle (13), air (31), analyze (5), angle (25), arm (61), athlete (1), attachment (4), average (1), backward (10), backwards (9), ball (14), bend (9), body (13), centimeter (20), chain (3), circle (7), cold (1), come down (2), connect (16), control (2), control jump (5), controlled jump (3), crouch (6), curl (2), cycle (9), degree (20), diagonal (6), direction (3), distance (12), experiment (17), fall (3), far (87), farther (61), fast (17), fault (1), feet (17), flow (5), foot (11), form (9), forward (36), forwards (3), friction (1), function (2), further (21), go down (3), goal (2), ground (14), gymnastics (1), head (4), height (16), high (42), higher (23), highest (4), hip (6), hot (1), human body (13), improve (46), improvement (16), improving (6), information (23), invention (1), joints (3), jump (352), jumper (6), kinetic energy (1), knee (23), land (24), lean (3), lean forward (4), leap (4), left (2), leg (77), length (9), life (1), line (13), linear (1), live (1), long (141), long jump (8), longer (10), longest (4), low (7), math (1), maximum (1), maximum height (1), measure (11), measuring (6), meter (17), mid air (7), momentum (5), move (11), movement (2), movies (2), moving (4), muscle (1), natural (1), need (29), normal jump (9), observation (2), obtuse angle (3), Olympics (1), opposite (2), order (3), pass (2), pattern (1), plan (15), position (4), power (6), practice (3), pressure (1), process (5), professional (1), progress (1), pull (2), pump (2), push (9), ramp (1), reach (6), record (1), recycle (1), recycling (3), regular jump (2), relay (1), result (18), run (43), seconds (21), see (39), shift (4), short (5), shorter (5), shoulder (5), sit (2), slow (2), sophisticated (2), source (5), speed (11), spin (1), split (1), sport (1), sprawl (1), start line (3), starting (1), starting line (4), statistics (3), step (7), stop (2), straight (2), strategy (4), stretch (2), stride (20), swing (37), swung (6), system (115), take off (9), tall (1), technique (11), technology (4), test (9), theory jump (3), throw (10), thrust (1), time (57), travel (4), traveling (2), trunk (1), try out (1), tuck (16), vary (1), video (15), water (9), wave (2), waved (1), waving (3), weight (5), wind (5).

Note: (1) Terms are listed by alphabetical order; the number inside the parentheses ( ) is the frequency of use of each key term; (2) The focus of analysis in phase 2 was on key terms related to physics concept. Therefore, many other key terms found related to the previously inquired topics such as internal human body system (as resulted from discourse continued from phase 1) and government system and electricity system (as resulted from discourse continued from previous semesters) were not included here.