WordWhispers: A Tool for Bootstrapping Vocabulary Development in Knowledge-Building Communities

Bodong Chen, Xing Liu, University of Minnesota
Email: chenbd@umn.edu, liux3099@umn.edu

Abstract: In this short paper, we introduce a work-in-progress feature named WordWhispers in Knowledge Forum 6. WordWhispers is designed to bootstrap vocabulary development and hereafter to facilitate collective conceptual engagement in knowledge-building communities. The major functionality implemented in WordWhispers turns current conceptual artifacts (in the form of Knowledge Forum notes) into resources for ongoing written discourse, by feeding relevant terms mined from current artifacts to writers in real-time. Preliminary user studies suggested areas WordWhispers could be further refined before its first pilot this fall. We conclude this paper by discussing its current limitations and future directions.

Keywords: knowledge building, vocabulary acquisition, literacies, learning analytics, emergence

Introduction

Knowledge building (KB) is defined as the production and continual improvement of ideas of value to a community (Scardamalia & Bereiter, 2003). It involves high levels of literacy in various disciplinary areas where it is applied (Chen & Hong, in press), as knowledge building in a community is essentially an intensive literacy practice where literacies of all kinds are geared to solving problems and developing new knowledge. High-level literacy in KB covers disciplinary core constructs identified in Goldman et al. (2016), which include (a) epistemology; (b) inquiry practices/strategies of reasoning; (c) overarching concepts, themes, and frameworks; (d) forms of information representation/types of texts; and (e) discourse and language structures. Efforts are made in the KB community to advance literacy development in these areas (see Chen & Hong [in press] for a review).

Among all types of literacies, basic or traditional literacies (i.e., reading and writing) remain an important concern. To foster the development of basic literacies, KB immerses students in complex literate worlds from early years. Students engaged in day-to-day KB practices have rich opportunities for literacy development in meaningful contexts, with consequent advances in it (Bereiter & Scardamalia, 1987). Prior studies have explored vocabulary development in KB, by tracing lexical frequency profiles of students across years of KB practices (Chen, Ma, Matsuzawa, & Scardamalia, 2015; Sun, Zhang, & Scardamalia, 2010; Zhang & Sun, 2011). In these studies, the development of basic literacies, or more specifically vocabulary development, is treated as a by-product. What seems a fruitful direction is to develop effective strategies that directly tackle literacy development in KB and may promote knowledge advancement in the meantime.

This short paper reports a work in progress that attempts to fill this gap of devising new supports for vocabulary development in KB. In the following sections, we review relevant KB and language acquisition literature to ground this work. Then we introduce WordWhispers, a new feature in Knowledge Forum 6 designed to support vocabulary development in KB. We then report results from preliminary user studies with WordWhispers and then conclude this paper by discussing next steps.

Vocabulary Development in Knowledge Building

KB promotes reading with a purpose and encourages a departure from traditional monologic writing to communicative and dialogic writing (Haneda & Wells, 2000). A norm for a KB community is that all members contributing ideas to the collective enterprise and working together to improve them (Scardamalia, 2002). In KB, students read each other's entries, search for information to address knowledge problems, design and report experiments, read authoritative sources, and synthesize across texts and information modalities. In this situation, gains in traditional literacies (i.e., reading and writing) become natural, often in absence of explicit literacy instruction. Prior research characterizes reading practice in KB into four themes: reading for the purpose of advancing community knowledge, reading as progressive problem solving (not only comprehending but also problematizing understanding), reading embedded in KB discourse, and reading as dialogues between local understanding and knowledge out in the world (Zhang & Sun, 2011). Likewise, writing in KB is treated as a tool for building knowledge—with text serving knowledge purposes in the community, and ideas rather than structure and lexico-grammar at the center (Haneda & Wells, 2000).

More specifically for vocabulary development, earlier studies have confirmed the efficacy of KB's approach in this area. In one study, an analysis of lexical frequency profiles of 22 students' Knowledge Forum entries in Grade 3 and 4 revealed growth in both domain-specific and academic words; over two years, each student on average produced 3231 words and 715 unique words, with the composition shifting towards more sophisticated words (Sun, Zhang, & Scardamalia, 2008). Occurrences of domain-specific and academic words correlated with the quality of ideas (Pearson rs > .50), indicating an underlying connection between vocabulary development and conceptual understanding (Sun et al., 2008). Similarly, a longitudinal study mining a student cohort's six years of KF writing activities also uncovered significant growth in productive writing vocabulary; in particular, note revision—a strong indicator of knowledge transforming and idea improvement less commonly found among novice writers (Scardamalia, 1981; Scardamalia & Bereiter, 1987)—was identified to be the strongest predictor of vocabulary growth rate (Chen et al., 2015). Co-elaboration of word meanings was evident in student discourse; so was sustained collective engagement with a same term across multiple years. For example, students were engaged with the term gravity across six years in various discourse contexts (e.g., explaining rains, building planes, and studying astronomy) and with different levels of understanding that improved in sophistication over time.

It is evident from earlier studies that KB helps "bootstrapping" vocabulary development in a community. Bootstrapping as a term in linguistics refers to the idea that a child is innately ready to acquire language (Höhle, 2009). By acquiring some linguistic knowledge early, the child becomes more prepared for acquiring further knowledge. When we discuss bootstrapping language development in KB in this paper, the general sense of this term applies. That is, "bootstrapping is the leveraging of a small initial effort into something larger and more significant" (Höhle, 2009, p. 360). When students initiate knowledge-building efforts, by operating as a community, they start from something small, then exchange ideas through discourse, and progressively improve their understanding (Chen et al., 2016; Scardamalia & Bereiter, 2003). In other words, socio-cultural practices in KB creates a scenario where students become resources for each other's vocabulary development and thus expand their proximal zones of development. Earlier analysis of student writings in KB has uncovered collaborative bootstrapping of vocabulary, concepts, and ideas (Chen, Ma, Matsuzawa, & Scardamalia, 2015), meaning that earlier writing contributions become resources for unlocking deeper ones provided that students share cognitive responsibility as a community. If KB technology is indeed becoming smarter, supports for self-organization around writing contributions to encourage emergent conceptual engagement are needed. What is missing in current knowledge-building environments is a mechanism to explicitly support such collaborate bootstrapping of vocabulary and conceptual engagement using data-intensive analytics. To this end, we developed a prototypic feature named WordWhispers in Knowledge Forum 6 to explore this possibility.

WordWhispers

The main design goal of WordWhispers is to use a community's current knowledge, represented as conceptual artifacts in Knowledge Forum (KF), as resources for ongoing knowledge building. To this end, WordWhispers tapps into existing notes in a KF database for potentially useful words for an ongoing attempt to create or revise a note. By doing so, WordWhispers attempts to feed recommended words in an unabstrusive manner—just like whispers—for the student's note-composing processes. Below, we briefly explain the current technical implementation of WordWhispers.

To begin with, we define the terminology of WordWhispers as follows:

- A *word* is the basic unit of composition in KF, denoted by **w**.
- A *document* (aka. a KF *note*) is represented by a sequence of N words, denoted by a numeric vector $\mathbf{v} = (v_1, v_2, \ldots, v_N)$, which has names $(w_1, w_2, \ldots, w_N)$, where w_n is the nth most used word in document and v_n represents the count of occurrences of w_n.
- A *corpus* is a collection of M documents denoted by $\mathbf{D} = \{d_1, d_2, \ldots, d_M\}$.

The algorithmic computation that powers WordWhispers is illustrated in Figure 1. The computation includes three phases. First, when a note is opened by the current KF user, either when creating a new note or revising an existing note, WordWhispers extracts words from the note, apply a part-of-speech (POS) tagging algorithm to filter words that are not nouns, and then generate a vector representation of the current document, i.e., $\mathbf{v} = (v_1, v_2, \ldots, v_N)$. Then the KF client/browser side will pass $\mathbf{v}$, together with *viewId* (id of the current view) and *sessionId* (id of the current user session), to the server side to look for "adjacent" words for this KF note.
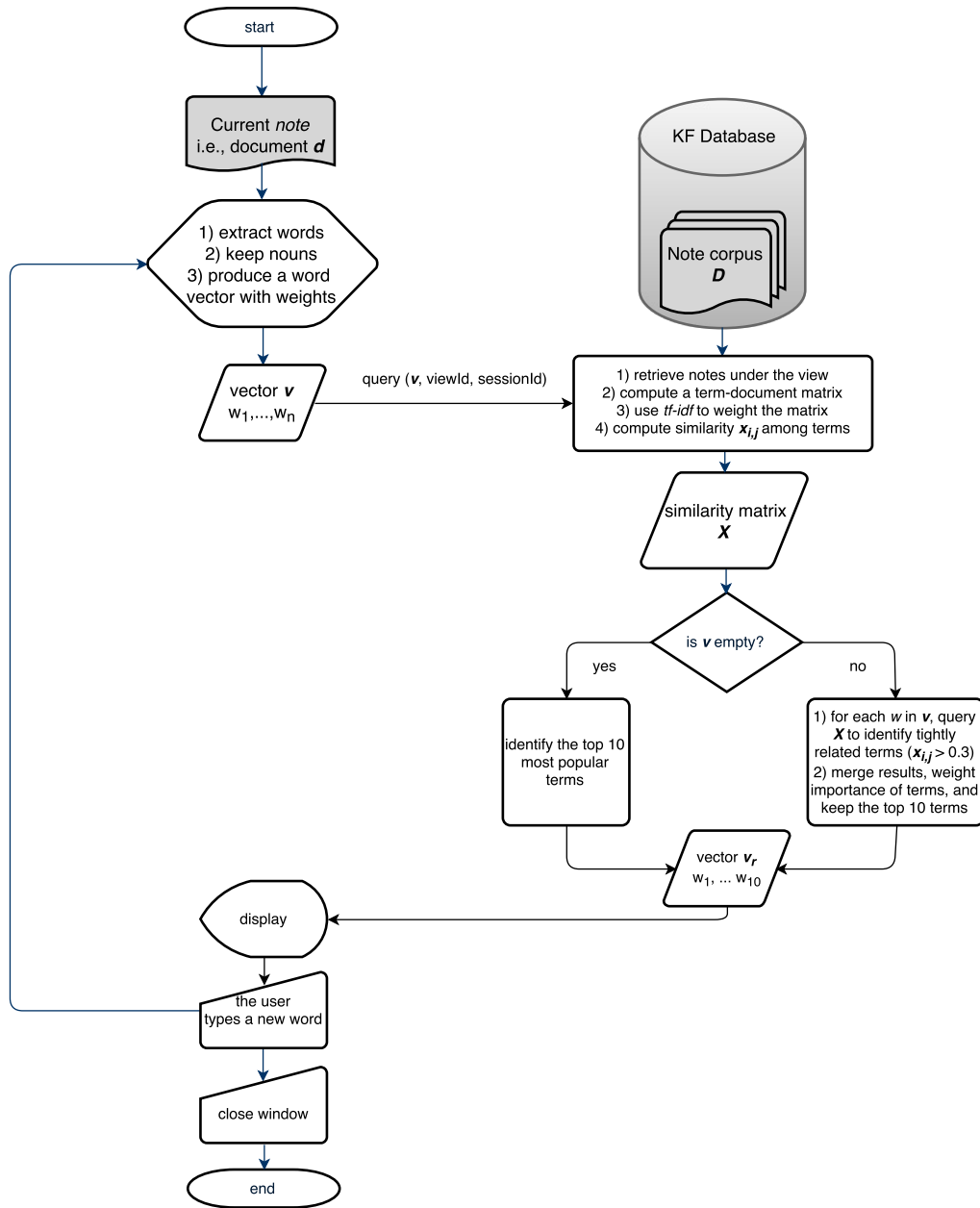
Figure 1. A flow chart of WordWhispers.

Second, after the query is triggered, WordWhispers first retrieves all notes (documents) under the current *view* and generates a term–document matrix based on the occurrences of all terms in each document. The popular *tf-idf* approach (Salton & McGill, 1986) is applied on the raw term–document matrix, resulting in a *tf-idf* term–document matrix, **X**, which contains adjusted values reflecting how important a word is to a document in the corpus. With **X**, for each term in the query vector **v**, WordWhispers identifies "tightly" related terms that have a similarity score higher than 0.3. Then WordWhispers merges results from all query terms, with the weights of query terms in **v** considered. This computation results in a list of ten terms that are the most related to the current note. If **v** is empty (which happens when creating a new note), WordWhispers simply returns the most dominant terms in the current view.

Finally, returned terms (i.e., word "whispers") are displayed in the note editing window (see Figure 2). The size of a word corresponds to its relative weight returned by the computation. In this display, the user can simply click on a term to insert it into the current note being edited. Meanwhile, WordWhispers will constantly update the whispers—every time when a new word is added—while the user is typing in this note.
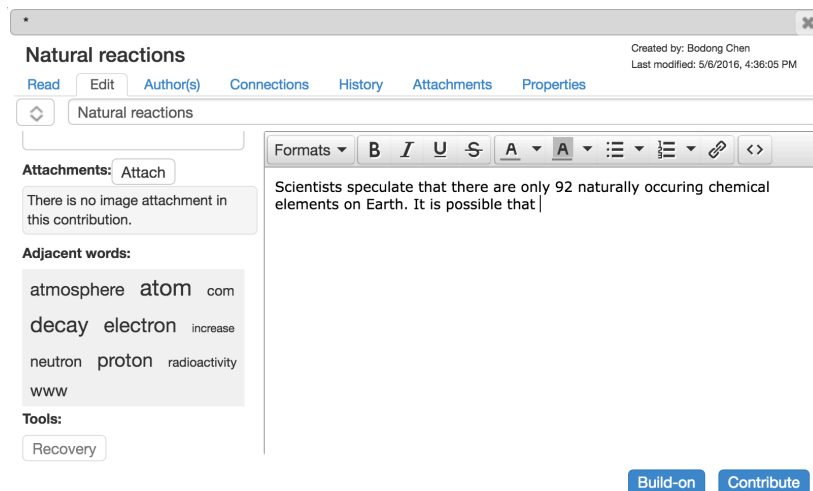
Figure 2. Word whispers are displayed on the left-side panel of the KF note-editing window, with the size reflecting a word's relevance to the current note. By clicking on a word it will be directly inserted in the note.

Preliminary User Studies

Preliminary user studies are underway, focusing on the usability and usefulness of WordWhispers before it is piloted in real-world classrooms. To this end, three participants have been recruited to use WordWhispers and report their evaluation of the tool using the System Usability Scale (SUS). SUS is an empirically validated questionnaire for usability testing (Bangor, Kortum, & Miller, 2008). It contains 10 items designed to measure users' perceived usability of a system. In addition, open-ended interviews were conducted to receive feedback from testers. Three participants included a high-school teacher who is currently using KF, a graduate student who used KF earlier, and an undergraduate student who has not used KF. Their varied backgrounds would help elicit a diverse range of perceptions of WordWhispers.

User studies with recruited participants are ongoing, and are expected to be finished by the Summer Institute. Below we report some initial qualitative feedback collected from one participant:

- "About the design, I'm trying to use WW like might a student and … I think the tools look is too complex–requires too much processing that gets in the way of the searching (by searching I mean like fishing for a word that fits my cognitive need). Idea to make it more simple: make it a single column of words with the most common on top and the least on the bottom. Instead of a left justify I'm feeling like a center-justify of the words would be more in tune of what my brain is feeling during a search."
- "Now I'm considering its validity as a suggestor of useful words. Mostly it seems good–there is one point that still provides dissonance for me, but it might well be a lack of understanding."
- "The other thought is, without spell-check there are a lot of 'distracter', space-wasting words in the 'cloud' such as 'dont'. As a great work-around it would be cool if the teacher could edit this list. Probably necessary at this point since we can't have students go back and fix all misspelled words. … there needs to be some way to edit the back-end list of vocab words, both to allow e.g. dont=don't and reactin=reaction, but also to allow kool aid to equal: 'kool aid' and not: 'kool' and 'aid'. The WW's need to not include obvious 'ugly' suggestions or when students are searching for a word, their eyes won't so willingly wander to the WW's…"

As demonstrated in these preliminary feedback from the participant, there are concerns with several aspects of the tool including (a) its user interface design, and (b) the validity/quality of suggestions. We plan to incorporate results from the user studies to refine this tool before testing it in classrooms this fall.

Discussion and Conclusions

In this short paper we introduce WordWhispers, a work-in-progress feature of Knowledge Forum 6, developed to bootstrap vocabulary development in a knowledge-building community. The central affordance of WordWhispers is to turn writings in a community into resources for ongoing or 'future' writings, by feeding pertinent vocabularies in real-time to writers. The algorithm that powers WordWhispers is uncomplex and open to further refinement. But the

tool itself is of significance in facilitating collective conceptual engagement with key vocabularies in knowledge building. Benefits of such scaffolding mechanisms are demonstrated in earlier design research using simple word clouds (Resendes, Scardamalia, Bereiter, Chen, & Halewood, 2015). With WordWhispers, we attempt to make such scaffolding efforts even more dynamic, concurrent, and emergent. Planning of new design research initiatives is underway to develop pedagogical principles for incorporating WordWhispers in knowledge-building classrooms.

In addition to user feedback reported earlier, there are several challenges we need to tackle. First, we need to improve efficiency of WordWhispers to make it more responsive. Computation involved in WordWhispers is relatively expensive and may undermine responsiveness of KF when it tries to process a larger volume of notes. We need to find innovative solutions to tackle this technical challenge. Second, we are tempted to test more powerful (and also more computationally expensive) algorithms such as Latent Dirichlet allocation (LDA; Blei, Ng, & Jordan, 2003), in order to improve accuracy/usefulness of "word whispers." These two key challenges compete with each other and a balance needs to be struck in future refinement of this tool. Finally, we also want to give voice to the users, both the teacher and student, so that they would have influence on word whispers. By doing so, future versions of WordWhispers would solicit input from and thus 'collaborate' with users instead of impose suggestions on them. Furthermore, students' intentional choices of terms through WordWhispers could potentially provide important proxies for high-order competencies in knowledge building (Chen & Zhang, in press).

References

Bangor, A., Kortum, P. T., & Miller, J. T. (2008). An empirical evaluation of the system usability scale. *International Journal of Human-Computer Interaction*, *24*(6), 574–594. doi:10.1080/10447310802205776

Bereiter, C., & Scardamalia, M. (1987). An attainable version of high literacy: Approaches to teaching higher-order skills in reading and writing. *Curriculum Inquiry*, *17*(1), 9–30.

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *The Journal of Machine Learning Research*, *3*, 993–1022. Retrieved from http://dl.acm.org/citation.cfm?id=944919.944937

Chen, B., Ma, L., Matsuzawa, Y., & Scardamalia, M. (2015). The development of productive vocabulary in knowledge building: A longitudinal study. In O. Lindwall, HäkkinenP., T. Koschmann, P. Tchounikine, & S. Ludvigsen (eds.), *Exploring the material conditions of learning: The computer supported collaborative learning conference 2015, volume 1* (pp. 443–450). Gothenburg, Sweden: ISLS.

Chen, B., & Hong, H.-Y. (2016). Schools as knowledge-building organizations: Thirty years of design research. *Educational Psychologist*, *51*(2), 266–288. doi:10.1080/00461520.2016.1175306

Chen, B., & Zhang, J. (in press). Towards agency and design thinking: Choice-based analytics in knowledge building. *Journal of Learning Analytics*.

Goldman, S. R., Britt, M. A., Brown, W., Cribb, G., George, M., Greenleaf, C., … Project READI. (2016). Disciplinary literacies and learning to read for understanding: A conceptual framework for disciplinary literacy. *Educational Psychologist*, 1–28. doi:10.1080/00461520.2016.1168741

Haneda, M., & Wells, G. (2000). Writing in knowledge-building communities. *Research in the Teaching of English*, *34*(3), 430–457. Retrieved from http://www.jstor.org/stable/40171501

Höhle, B. (2009). Bootstrapping mechanisms in first language acquisition. *Linguistics*, *47*(2), 359–382.

Resendes, M., Scardamalia, M., Bereiter, C., Chen, B., & Halewood, C. (2015). Group-level formative feedback and metadiscourse. *International Journal of Computer-Supported Collaborative Learning*, *10*(3), 309–336. doi:10.1007/s11412-015-9219-x

Salton, G., & McGill, M. J. (1986). *Introduction to modern information retrieval*. New York, NY: McGraw-Hill, Inc.

Scardamalia, M. (1981). How children cope with the cognitive demands of writing. In C. H. Frederiksen, M. F. Whiteman, & J. F. Dominic (eds.), *Writing: The nature, development, and teaching of written communication* (Vol. 2, pp. 81–103). Hillsdale, NJ: Lawrence Erlbaum.

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, IL: Open Court.

Scardamalia, M., & Bereiter, C. (1987). Knowledge telling and knowledge transforming in written composition. *Advances in applied psycholinguistics*, *2*, 142–175.

Scardamalia, M., & Bereiter, C. (2003). Knowledge building. In J. W. Guthrie (ed.), *Encyclopedia of education* (2nd ed., Vol. 17, pp. 1370–1373). New York, NY: Macmillan Reference.

Sun, Y., Zhang, J., & Scardamalia, M. (2008). Knowledge building and vocabulary growth over two years, Grades 3 and 4. *Instructional Science*, *38*(2), 147–171. doi:10.1007/s11251-008-9082-5

Sun, Y., Zhang, J., & Scardamalia, M. (2010). Developing deep understanding and literacy while addressing a gender-based literacy gap. *Canadian Journal of Learning and Technology*, *36*(1).

Zhang, J., & Sun, Y. (2011). Reading for idea advancement in a Grade 4 knowledge building community. *Instructional Science, 39*(4), 429–452.