

Breaking the Knowledge Acquisition Bottleneck Through Conversational Knowledge Management

Christian Wagner, City University of Hong Kong, Hong Kong,
and Claremont Graduate University, USA

ABSTRACT

Much of today's organizational knowledge still exists outside of formal information repositories and often only in people's heads. While organizations are eager to capture this knowledge, existing acquisition methods are not up to the task. Neither traditional artificial intelligence-based approaches nor more recent, less-structured knowledge management techniques have overcome the knowledge acquisition challenges. This article investigates knowledge acquisition bottlenecks and proposes the use of collaborative, conversational knowledge management to remove them. The article demonstrates the opportunity for more effective knowledge acquisition through the application of the principles of Bazaar style, open-source development. The article introduces wikis as software that enables this type of knowledge acquisition. It empirically analyzes the Wikipedia to produce evidence for the feasibility and effectiveness of the proposed approach.

Keywords: knowledge acquisition; knowledge artifacts; knowledge management; open source development; wiki

INTRODUCTION

Ever since the development of artificial intelligence (AI) and expert systems, there has been the promise of capturing an organization's knowledge on a large scale and making it available to the entire organization. Unfortunately, these promises did not materialize (Buchanan & Smith, 1988; Ullman, 1989). While there have been several early success stories, such as

American Express' Credit Advisor or Digital's Expert Configurer (XCON), attempts to acquire the broad knowledge of organizations have been less fruitful. More than a decade later, a decidedly optimistic survey by Frappaolo and Wilson (2003) found that no more than 32% of the knowledge was available in computerized form. Obviously, knowledge acquisition is a challenge. How can we extract more of the ex-

isting knowledge from organizational sources, especially from people? And how can we manage the maintenance so as to assure that the stored knowledge is accurate and up-to-date? Discovering answers to these questions is important for organizations as information work becomes knowledge work, thus requiring knowledge to support non-routine decision making (Drucker, 1993, 1999). It is similarly important for organizations whose corporate portals that were set up years ago increasingly are becoming dated and stale (Newcombe, 2000). Furthermore, it is important for organizations in the business of creating knowledge assets who are faced with increased costs of knowledge creation, shorter knowledge life cycles, and increased knowledge obsolescence.

Seeking a solution to the problems of organizational knowledge acquisition, the article makes the following argument. First, it introduces previous approaches to knowledge acquisition, identifies four limitations, and offers evidence for these limitations. The article then refers to Bazaar style (software) development (Raymond, 2001) as a potential direction for knowledge asset creation. It then explains the concept of conversational knowledge management and advocates wiki technology and the "wiki way" (Leuf & Cunningham, 2001) as a possible approach to using Bazaar-style methods in conversational knowledge management. An empirical analysis of the viability and effectiveness of the approach follows. The article ends with implications and conclusions about the future of conversational knowledge management.

KNOWLEDGE ACQUISITION

Approaches to Knowledge Acquisition

Organizations that try to acquire organizational knowledge formally (based on artificial intelligence methods) have relatively few available alternatives. For application areas with large amounts of transaction data, data mining can induce rules from that data. Data mining solu-

tions work well for high-volume applications such as credit approval. Even then, the knowledge creation effort is highly resource-intensive (Lee, 2001). When insufficient data volumes thwart data mining efforts, the acquisition activity has to elicit knowledge directly from experts as rules and facts or similar formal representations. This should be done under the guidance of knowledge engineers trained in knowledge elicitation, formalization, and representation. Yet a knowledge engineer's productivity is limited to hundreds of rules per year for development and maintenance (Sviokla, 1990; Turban & Aronson, 2000). This productivity level may be acceptable for high value-added projects but limits the broad applicability of the approach. Smaller projects have attempted to rely on capturing knowledge without knowledge engineers, relying on end-user development. The latter has not been very successful (Wagner, 2000, 2003). Wagner found end-user expert systems often to be poorly structured, incomplete, highly coupled, and thus, difficult to maintain. Artificial intelligence-based methods thus are facing considerable applicability constraints. Consequently, organizational knowledge management efforts have sought to capture knowledge in less formal ways; for instance, by extending document management and groupware systems into knowledge management systems (Davenport & Prusak, 1998; Holsapple & Joshi, 2002) in part through better indexing, search engines, and linking.

Yet challenges remain. When organizations try to make sense out of large volumes of documents in their document management systems, they usually need search engines, text mining, and automatic indexing tools, resulting in an expensive solution with limited success (Bygstad, 2003). Furthermore, this approach is well suited only for relatively stable and centralized knowledge bases. Users of such knowledge bases often encounter information overload, irrelevant responses, or no response to queries. Alternatively, organizations might use expert reports and harvest expert knowledge to capture the methods used by domain experts (Snyder & Wilson, 1998). Again, this method

often is limited to niche applications, requires considerable effort, and still faces knowledge maintenance difficulties (Malhotra, 2000). Other solutions, such as corporate controlled portals, can quickly suffer from outdated knowledge and lack of maintainability (Newcombe, 2000).

Knowledge Acquisition Bottleneck

In summary, we can describe the knowledge acquisition bottleneck as follows (Wagner, 2000; Waterman, 1986):

- **Narrow bandwidth.** The channels that exist to convert organizational knowledge from its source (either experts, documents, or transactions) are relatively narrow.
- **Acquisition latency.** The slow speed of acquisition frequently is accompanied by a delay between the time when knowledge (or the underlying data) is created and when the acquired knowledge becomes available to be shared.
- **Knowledge inaccuracy.** Experts make mistakes and so do data mining technologies (finding spurious relationships). Furthermore, maintenance can introduce inaccuracies or inconsistencies into previously correct knowledge bases.
- **Maintenance trap.** As the knowledge in the knowledge base grows, so does the requirement for maintenance. Furthermore, previous updates that were made with insufficient care and foresight (“hacks”) will accumulate and render future maintenance increasingly more difficult (Land, 2002).

Given these challenges, it appears that there are few opportunities for breaking the knowledge acquisition bottleneck. The next section will propose one possible remedy.

LEARNING FROM SOFTWARE DEVELOPMENT

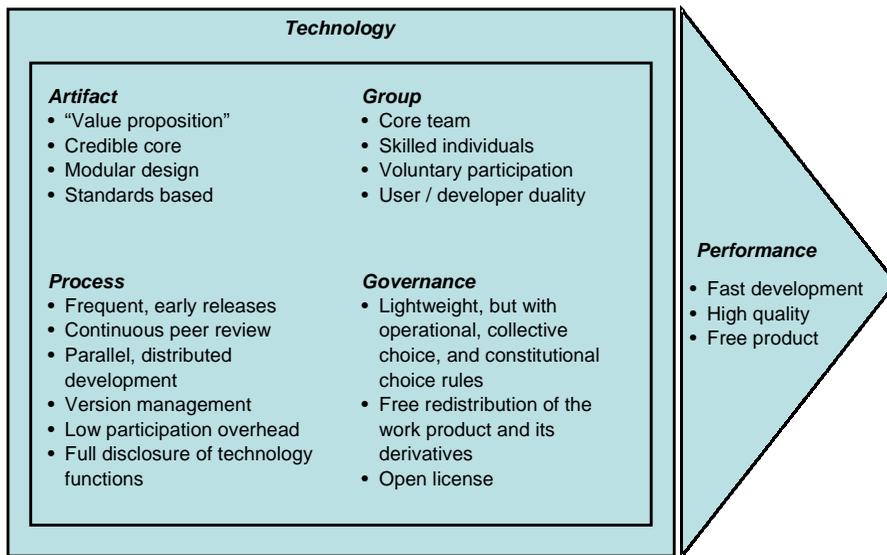
One area that has offered lessons for the successful creation of knowledge assets is software development, and specifically open source software development by distributed teams of

volunteers. Open source projects engage software developers, wherever they may reside, and have them collaboratively develop the knowledge asset (the software). Surprisingly, this activity takes place with little centralized management. Raymond (2001) characterized this approach to software development as the *Bazaar* style in contrast to the traditional *Cathedral* style of development. Cathedral is a metaphor for the development of a large monolithic artifact through a structured and lengthy development process. Fundamental to the cathedral style approach is that source code is only widely available at release dates with access restricted to a few developers between release dates. Bazaar style development, however, occurs over the Internet in constant public view. Raymond identified principles of this development style that challenge the assumption that large and complex software assets need to be built via an *a priori*, centralized approach. Overall, four themes guide this development approach, which can be characterized as follows: (1) design simplicity of the artifact, (2) team work, (3) frequent creation of a visible work product, and (4) development as an ongoing conversation. This section introduces a framework of open source (software) development, identifies its benefits, and derives lessons about the applicability for knowledge assets other than software.

Open Source Software Development

Open source software development, as described, for instance, by Raymond (2001), Benkler (2002), and Markus, et al. (2000), relies on several factors to achieve success (and thus, performance of the knowledge creation effort). Key success factors (see Figure 1) consist of a suitable artifact, a skilled and motivated team of volunteer users and developers, a lean and transparent development process, and lightweight but effective governance. Added to this is an enabling factor; namely, an appropriate technology infrastructure, which, for instance, permits frequent releases, accommodates voting mechanisms to govern the community, or enables fast and reliable version management,

Figure 1. Framework for open source software development



all with little overhead and few transaction costs. With these factors in place, open source software development promises faster development speed than proprietary approaches (including higher developer productivity) and a better quality product, which is also free.

Open source software development has had remarkable successes, creating software that appears to break long-standing rules of software evolution (Scacchi, 2004). For example, open source software size has been shown to grow super-linear (exponential) rather than linear or inverse-square (Mockus et al., 2002).

Bazaar-Style Knowledge Management

Can Bazaar-style development be applied successfully to the creation of knowledge assets other than software? Several leaders of the open source community have hypothesized this, including Torvalds (Hamm, 2004). Yet Torvalds also acknowledged that not all knowledge assets are equally suitable, as the creation process may be too personal or too linear. Hence, in order to extend the lessons and benefits of Bazaar-style development, we should target applications where the core themes can

be applied: (1) simplicity of design and frequent redesign (refactoring) to maintain simplicity, (2) teamwork (3) frequent creation of a small work product available for review and testing, and (4) development as conversation to facilitate back-up, clarity, and shared understanding. Applications of this kind exist within organizations, and among organizations and people. For example, companies could conceivably turn their traditional help desks into open help desks, where customers would openly share their problems with others, help each other, and free up company experts to tackle only the most difficult problems. Unfortunately, companies frequently do not want to relinquish control of their (closed) help desk. Open help desks exist on the Web, typically as discussion forums of questions and answers. While they embody teamwork and conversation, the resulting work product often is not simple and well-structured but lacks organization and is filled with repetition and inconsistencies.

Consequently, one necessary condition for this research was to find a knowledge asset that was highly amenable to the Bazaar-style development approach and that used a technology that facilitated this type of development.

The selected asset was an online encyclopedia — Wikipedia (wikipedia.org) — that employs wiki technology and the “wiki way” of knowledge asset creation (Leuf & Cunningham, 2001). The article will provide more detail on the Wikipedia application, following a briefing on knowledge management with wiki technology and the wiki way.

CONVERSATIONAL KNOWLEDGE MANAGEMENT WITH WIKIS

Knowledge management with wikis has recently drawn media attention (Brown, 2004; Hof, 2004; Ripley, 2003) as a new, end user developed approach founded on collaboration and conversation. Collaborative knowledge management means that many people work together to create or acquire knowledge instead of a few individual experts. In other words, a community (of practice) will jointly create and maintain the knowledge. Research elsewhere (Cheung et al., 2005) suggests that conversational knowledge management is well suited for this challenge, whereby conversations (i.e., questions and answers) become the source of relevant knowledge.

Conversational knowledge management has become popular in communities that form around discussion boards. Leading solutions such as ezboard or Yahoo groups are now used by millions of communities¹. Yet while discussion forums have been a simple and practical solution to share knowledge through conversation, they lack several useful knowledge representation and maintenance features. For example, discussion forum postings, even within a single thread, often do not build upon each other. As a result, the latest post may not be an incremental improvement of earlier ones. An alternative technology, which combines the most desirable features of other conversational technologies, is the wiki. This section discusses wiki technology and its suitability for knowledge management.

Wiki Structure and Principles

A wiki is a set of linked Web pages created through the incremental development by a group of collaborating users (Leuf & Cunningham, 2001) as well as the software used to manage the set of Web pages. Ward Cunningham developed the first wiki in 1995 as the PortlandPatternRepository in order to communicate specifications for software design within a large, heterogeneous community. The term *wiki* (from the Hawaiian *wikiwiki*, meaning *fast*) references the speed with which content can be created with a wiki. Wiki key characteristics are as follows:

- It enables Web documents to be authored collectively;
- It uses a simple markup scheme (usually a simplified version of HTML, although HTML frequently is permitted);
- Wiki content is not reviewed by any editor or coordinating body prior to its publication; and
- New Web pages are created when users navigate a hyperlink that points nowhere.

Underlying these characteristics are specific principles that have shaped wiki software as well as its use. They are intended to produce a development environment where multiple people easily can create and modify a set of jointly owned Web pages. Wiki pages are expected to be open, incrementally developed, and organic; require little markup; have consistent edit functions and clear naming, be heavily hyperlinked and easily observable (found). As a result, wiki pages are expected to change and improve incrementally.

Wikis in Use

Creating Wiki Pages

Creating and editing wiki pages is a simple activity. A wiki author will use a Web-enabled *formfield* to enter a comment he or she wishes to publish. Authors can use plain text or a simplified markup language. The system then auto-

matically generates and publishes a Web page with a unique URL that can be indexed and linked to. Hence, users with virtually no Web publishing knowledge can create Web content about as quickly as they can write a text document.

Linking Wiki Pages

A fundamental aspect of knowledge management with wikis is the use of simple hyperlinks. Hyperlinks link topics and create context. Wikis drastically simplify hyperlinking. To link pages within a wiki, users do not have to create and use URLs (although they can). Instead, they normally use CamelCase (multiple words capitalized and concatenated) or double parentheses around a term ([[term]]) in order to create a link. Links whose destination (page) does not exist are depicted as question marks (or similar) as if the author were asking a question. Another author (or the original creator) then can respond by clicking on the question mark, thus navigating the hyperlink to a new page and invoking an editor to write that page. Upon completion of the edit, the question mark automatically will be rendered as a regular hyperlink (now underlined text) pointing to the new page.

Versioning

As multi-user systems, wikis enable every user to modify any other user's Web pages (unless explicitly forbidden by access right settings). This creates challenges in version management. Wikis solve them by keeping prior versions of any Web page in memory, and enabling rollback, comparison, difference identification, and similar capabilities, if so desired. Wikis also track the history of prior changes with author, date, and related information.

Wikis and Open Source Principles

Knowledge management using wikis and the wiki way (see, for instance, "WhyWikiWorks" at <http://c2.com/cgi/wiki?WhyWikiWorks>) appear to bear considerable resemblance to open source software development, described in part by the following traits:

- Sense of responsibility in contributing to a common good;
- Openness to change and modification by anyone;
- Meritocracy (anyone can play, but only good players last);
- Self-governance of the developer team;
- Task decomposition and incremental development;
- Use of technology for communication and coordination, as well as norms for their use, including objectivity (neutral point of view); and
- Ease of use for knowledge creation and maintenance.

Thus, as an enabling technology, wikis establish an environment to develop the right artifact, to use a Bazaar-style process, to engage teams in voluntary collaboration, and to govern the effort with a lightweight structure (Figure 1), thus offering the potential for open source knowledge management. In open source software development, the corresponding results are ultimately lower error rates (compared to closed source); fast(er) development speed; and the ability to develop large(r) applications, accelerated development, and high(er) maintainability of the source code (Mockus et al., 2002). Whether these same benefits accrue in wiki-enabled open source knowledge management must be determined empirically.

ASSESSING CONVERSATIONAL KNOWLEDGE MANAGEMENT

Can principles of Bazaar-style development be applied to knowledge management, and if so, will they improve knowledge acquisition effectiveness? To begin to answer these questions, the research analyzed a single case of wiki-enabled knowledge asset creation — Wikipedia.

Knowledge Asset: Wiki-Based Encyclopedia

Encyclopedias reasonably can be characterized as knowledge assets. While one may

debate how much of their content is information instead of knowledge, encyclopedias contain insights (factual), rules (inferential), principles (inferential), and so forth. They also fit the definition of *information in context* (Davenport & Prusak, 1998), since they frequently link concepts to other concepts (cross-referencing). By design, encyclopedias also are relatively loosely coupled knowledge assets, whose components (articles) can exist independently. Encyclopedias frequently are compiled from the work of a group of authors who know little about each other or each other's work. Encyclopedia articles have common structural elements, since all articles are definitions. They typically also follow some standards for articles of a similar type (e.g., all biographies are structured similarly and different from city descriptions).

The majority of digital encyclopedias, such as Britannica, Encarta, Compton, or Grolier, is closed source. They are compiled by a relatively small group of commissioned writers and editors. The result of their work only becomes available to the readership once the entire edit process has been completed and the new encyclopedia version is released. Yet, because of their loosely structured nature, encyclopedias (and other, similar knowledge assets) also can be created in Bazaar style, given certain conditions. The work product cannot be an off-line product such as a book or a CD; the technology in general has to be amenable to Bazaar-style knowledge acquisition and representation, and the organization creating the encyclopedia has to formulate procedures and methods that enable this type of knowledge acquisition. Bazaar-style knowledge acquisition, therefore, becomes a possibility when the asset is created following the wiki way. Hence, Wikipedia, the online encyclopedia developed as a wiki, was used as the knowledge asset to be analyzed for this research. Wikipedia is one of several knowledge products developed over the last few years with wiki technology and the wiki way of development. Other applications include Wikitravel and Wikibooks. Development of Wikipedia began in 2001. As of May 2004, less than three and a half years

later, the (English) Wikipedia contains about 280,000 articles.

Wikipedia, applying wiki principles, appears to enable its developers to use a Bazaar-style approach. Specifically, writers can make incremental changes and then commit and publish them immediately. Also, articles can be written by numerous writers as joint authors, thus building on the work of others or correcting mistakes. Furthermore, Wikipedia rules stress an authoring etiquette that incorporates rules of article design and redesign targeted toward simple and clear articles. In other words, it is possible for Wikipedia authors to follow the main themes of Bazaar-style development. Whether authors do so and whether the outcome of their efforts is consistent in its effectiveness with Bazaar-style software development needs to be determined empirically.

Research Questions

The research sought to address two questions through empirical analysis.

1. Is conversational knowledge management, as demonstrated in Wikipedia, consistent with Bazaar-style knowledge asset creation?
2. Is conversational knowledge management, as illustrated by Wikipedia, able to achieve the benefits of Bazaar-style development?

The research thus needed to determine whether "Wikipedians" would follow Bazaar-style knowledge acquisition and whether the effect would be improved knowledge acquisition. Based on the criteria in Figure 1, numerous questions would have to be addressed. Yet, as Table 2 illustrates, compliance with the majority of criteria was confirmed from Wikipedia information (Wikipedia Web site and Wikimedia Meta-Wiki), leaving four core questions to be answered.

Thus, the research questions focused on the incremental nature of the knowledge acquisition effort, the multi-person effort, and the effect on the growth and quality of the work product, as described in the following subsections.

Table 2. *Wikipedia fit with Bazaar-style development criteria*

Dimension	Bazaar-Style Development	Wikipedia Adaptation
Artifact	“Value proposition”	Yes – create free and open encyclopedia
	Credible core	Yes / No – not a <i>content</i> core, but a developer group core
	Modular design	Yes – loosely coupled articles
	Standards based	Yes – article structures
Group	Core team	? – <i>Is Wikipedia development a team effort?</i>
	Skilled individuals	Yes/No – participants from the Nupedia initiative all had PhD degrees, but no control over new participants
	Voluntary participation	Yes – only one paid chief editor, Larry Sanger (until 2002)
	User / Developer duality	Yes – author and users
Process	Frequent, early releases	? – <i>Are Wikipedia articles developed through an incremental approach with continuous releases?</i>
	Continuous peer review	
	Parallel, distributed development	Yes – 7,000 authors worldwide
	Version management	Yes – through wiki technology
	Low participation overhead	Yes – through wiki technology
	Full disclosure of (technology) functions	Yes – through wiki technology
Governance	Lightweight, operational, collective choice	Yes – Wikimedia organization with meritocracy as governing structure
	Free redistribution of the work product	Yes – GPL license
	Open license	
Performance	Fast development	? – <i>Does conversational knowledge management result in linear or better growth of knowledge assets?</i>
	High quality	? – <i>Does conversational knowledge management result in improved knowledge asset quality?</i>
	Free product	Yes – GPL license

Incremental Development with Frequent Releases

Incremental development and frequent releases are fundamental to Bazaar-style development. Would Wikipedians follow this approach, or instead would they prefer to write an authoritative article in an effort burst with few revisions in the process and even fewer thereafter?

To answer this question, the research explored (1) the frequency of article edits and (2) the change in article size. If the effort were non-incremental, one would expect a relatively short development period of high activity (since an article is typically a few hundred to a few thousand words long) followed by little editing activity thereafter, possibly with some maintenance and some extensions. An incremental effort, in contrast, would result in a high level of activity with many edits during an extended development period followed by a much-extended maintenance period with lower yet still considerable update efforts. To operationalize the assumption, the research adopted the

Pareto rule, thus hypothesizing that if Wikipedia articles were written in a non-incremental effort, then 80% of their size growth and 80% of the edit efforts should occur during the first 20% of their existence:

H1: Wikipedia articles are the outcome of an incremental development, and therefore, their growth and edit pattern does not follow the 80-20 Pareto rule.

Multi-Person Effort

There is little doubt that Wikipedia is a multi-person effort with presently more than 7,000 people contributing to it and more than 500 people making more than 100 contributions each per month (see Wikistats at <http://www.wikipedia.org/wikistats/EN/TablesWikipediaEN.htm>). However, according to the principles of Bazaar-style development, one would expect Wikipedia development to be a team effort at a more detailed level with multiple authors working on each article in order

to extend it and possibly to correct mistakes. This would reflect one of the key themes of open source, also called *Linus' [Torvalds] Law*; namely, that "given enough eyeballs, all bugs become shallow." Hence, the research sought to determine whether enough eyeballs were scrutinizing each article, at least more than two. Hence, the analysis focused on whether article publication and maintenance was a multi-person effort.

H2: Knowledge acquisition and maintenance in individual Wikipedia articles is a multi-person effort.

Effectiveness

The research sought to determine whether encyclopedia development adopting the wiki way would be effective. In this exploratory study, effectiveness was measured through two variables; namely, (1) growth of the knowledge asset and (2) quality improvement efforts. Growth of the knowledge asset was determined, based on the increase in the number of articles in the Wikipedia over time. In line with other open source successes (Mockus et al., 2002), the expectation was that growth would be linear or better (super-linear).

H3: Wikipedia growth in terms of number of articles will be linear or super-linear.

Unable to assess the overall quality of the Wikipedia objectively vis-à-vis other encyclopedias, the research focused on process quality and specifically quality improvement efforts. These efforts were operationalized by the ratio of edit efforts vs. the growth of Wikipedia articles. In other words, the research tested whether editing efforts were devoted to increasing the size of articles or to refining existing articles. The assumption was that refinements (without significant increase in size) would improve overall quality, for instance, through an increase in presentation quality, content quality, or the inclusion of more viewpoints (diversity).

Hence, we computed a words-per-edit ratio based on the number of words (per article)

written and the number of edits it took to create the article version. This ratio was calculated for articles in their early stages (20% of development effort) and at their present state. Decreasing ratios would indicate more effort being spent over time on article refinement. To exclude insignificant edits, the research only considered non-minor changes (counted separately in Wikipedia). The expectation was that, over time, more effort would be devoted to increased article quality. It is a stated Wikipedia goal to increase quality as articles mature (see, for instance, the reply to objections concerning Wikipedia, which discusses quality and growth issues, at http://en.wikipedia.org/wiki/Wikipedia:Replies_to_common_objections).

The corresponding hypothesis concerning quality improvement was as follows:

H4: Edit effort targeted at quality improvements for individual Wikipedia articles will increase over time, demonstrated by reduced article growth per edit.

Data Source

Wikipedia is an open encyclopedia in many ways. In addition to articles being freely accessible, so is the history of their creation, including dates, content of each version, and author information. Hence, it was possible to trace changes, change frequencies, and author contributions. To address the first two questions, 80 articles were randomly selected with the one qualification that 40 of them had to be created originally in 2001 and 2002. More recent articles were ignored because of their short history. To determine knowledge asset growth, Wikipedia summary statistics were accessed, which logged the number of articles written each month from the start of Wikipedia.

Results

Incremental Development (Release Early and Often)

This analysis focused on two samples of 40 articles from 2001 and 2002. For both of these samples, the edit efforts for the first 20% of

Table 3. Wiki development activity

	20% Avg. Actual	Avg. Expected (80-20 Rule)	t (df = 39)	Significance p
2001 Articles, Size (Words)	793	1,855	6.468	0.0000
2001 Articles, Edits	17	230	8.841	0.0000
2002 Articles, Size (Words)	811	1,795	4.212	0.0000
2002 Articles, Edits	24	133	6.820	0.0000

Table 4. Wikipedia article author statistics

	Min. No. Authors	Max. No. Authors	Avg. No. Authors	t-Statistic	Significance p
2001 Articles	33	285	121.4	10.33	0.0000
2002 Articles	18	268	70.8	7.870	0.0000
All Articles	18	285	96.1	12.21	0.0000

each article's existence (up to the measurement point in March 2004) were compared against the entire development effort. The results do not support the notion of a short effort burst but one of incremental development, as shown in Table 3.

For articles started in 2001, the first 20% of an average article's existence accounted for about 34% of the article's size (793 words out of 2,319) and less than 6% of its edits (17 out of 288). For the 2002 articles, the first 20% accounted for about 36% of article size and 15% of article edits. Overall, this was considerably less than expected according to the Pareto rule. All results are highly significant ($p < 0.0001$). Hence, size grew relatively incrementally with a somewhat larger upfront effort (about 35% of size produced in 20% of the time). Wikipedia edits were even more incremental with a disproportionately small number during the early existence of an article (15% or less of the edits in 20% of the time).

Multi-Person Effort

Each of the 80 articles in the two samples also was evaluated according to the number of authors. None of the articles in the sample was co-authored by fewer than 18 people, and the maximum number of authors for any article was 285. On average, more than 96 authors worked on an article (Table 4).

Given these results, what is the likelihood that articles overall were predominantly single-authored? Virtually none. A t-test showed very significant differences between the actual author numbers and the possibility of single authorship. This is a strong result, yet the reader is reminded that the sample articles were old articles. More than half of the Wikipedia articles were less than 12 months old (as of June 2004) and will have been edited by fewer people. An additional sample of 40 randomly selected articles started in 2003, though, still corroborated the results (average of 48 authors, $t = 7.164$, $p = 0.0000$). In other words, as time progresses, Wikipedia articles are scrutinized by "many eyeballs".

Wikipedia Growth

Data points concerning the growth of Wikipedia illustrate dramatic growth. Although Wikipedia has existed since 2001, more than half of its approximately 280,000 articles (English articles as of May 2004) were written since June 2003. (See <http://www.wikipedia.org/wikistats/EN/TablesWikipediaEN.htm>).

To explore the growth pattern further, the analysis targeted the numbers of new articles created each month. Three different time series were compared: number of articles, log of number of articles, and square root of number of articles. For each series, the fit was computed

Table 5. Growth in Wikipedia articles (articles official count, March 2004)

Relationship	R ²	p
Linear	0.932	0.0000
Exponential (log)	0.819	0.0000
Quadratic (square root)	0.988	0.0000

Table 6. Words per edit by article age

	20% - Avg. Words / Edit	80% - Avg. Words / Edit	t (df = 39)	Significance p
2001 Articles	57.8	6.7	5.848	0.0000
2002 Articles	64.4	10.7	4.239	0.0001

to determine which one best predicted Wikipedia growth. As Table 5 illustrates, Wikipedia growth is best explained by a quadratic function ($R^2 = 0.988$, highest). In other words, Wikipedia article growth is most likely quadratic and, thus, super-linear, which is an aggressive growth pattern. Quadratic growth also best explained the increase in the number of Wikipedians and in the number of edits (changes) made to Wikipedia.

Quality Improvement

The second effectiveness measure, the allocation of effort to quality improvement, suggested a shift toward more quality as Wikipedia articles aged. Table 6 illustrates that during the first 20% of an article's existence, each edit resulted in about 60 additional words vs. 11 or fewer words for the remaining 80% of the article's life (up to the measurement date in March 2004).

The differences in the means of these ratios were highly significant, confirming that later effort is an investment in article quality rather than article length.

Discussion

Results of the exploratory study confirm what has been expected. Hypotheses H1, H2, H3, and H4 were all confirmed. Knowledge acquisition efforts apparently can successfully adopt Bazaar-style development with multi-user involvement, incremental changes, and quick

releases in an environment that enables conversational knowledge acquisition. In the case of Wikipedia, this was possible for several reasons. First, Wikipedia was able to draw a large and quadratically growing developer group (approximately 7,000 as of May 2004).

Second, Wikipedia pages are highly decoupled from each other so that new authors can write with little concern for the current content of other pages. When an author breaks a hyperlink or negatively affects content, it becomes quickly apparent, and other Wikipedians will fix the problem. Third, when authors make a contribution, whether writing a new page or changing an existing article, the result is immediately visible to the entire community, thus enabling quick releases with minimal latency and multi-user quality assurance. Therefore, the transaction cost of making a contribution is low, much lower than in any peer-reviewed or closed source authoring environment (Ciffolilli, 2003). Fourth, there is no individual ownership of Wikipedia pages, which are developed by volunteers; thus, everyone works to improve everyone's contributions. Quality is everyone's responsibility. Fifth, Wikipedia has strong editing guidelines that are motivated by the refactoring rules of software development and principles of objectivity. This ensures that articles, which might have suffered in readability from the disjointed work of multiple contributors and commentator, ultimately become very readable again.

As a result, in three and a half years of existence, Wikipedia has challenged the otherwise largest but closed authorship *Encyclopaedia Britannica* (Britannica Online) for leadership in content (*Britannica* has self-reportedly about 100,000 entries, although with a larger word count per article). Other wiki-supported knowledge assets, such as Wikitravel, for instance, may be able to achieve similar leadership roles in their knowledge domain. The open, multi-user model also appears to scale well by interesting an increasingly larger user population to contribute their efforts, thus keeping the article latency at about 10 days for old articles (initially created in 2001) and less than two days for newer articles (i.e., 2003). However, since wiki technology is relatively new and contrary to many organizations' cultures, we should not expect this approach to become predominant soon. In fact, the successes are few at present. However, one should expect an increasing number of wiki software products to emerge in the future and an increasing number of communities to replace their inferior conversational technologies with wikis.

CONCLUSION

The challenge of capturing and maintaining exponentially growing volumes of knowledge requires new ways of knowledge acquisition; namely, on approaches that rely on the contributions of many rather than the expertise of a few. Wiki technology and the wiki way of collaboration show a feasible model for knowledge acquisition and maintenance. Wikipedia offers an illustration of the effectiveness of this approach. The research demonstrates that users of a wiki-based knowledge asset (i.e., Wikipedia) apply Bazaar-style methods and techniques in their conversational knowledge asset creation. The research also suggests that knowledge acquisition through collaboration and conversation can lead to super-linear knowledge asset growth and continuous quality improvement.

Not surprisingly, there are several caveats. For instance, knowledge quality cannot be measured or managed easily. The quality of

Wikipedia articles, for instance, remains a source of arguments. Therefore, future research will need to investigate the quality of the resulting knowledge based on content. In addition, knowledge creation with wikis relies on a strong and positive social contract among its contributors and on subject matters that are not controversial. These conditions are not always present. Wikipedia does have guidelines in place to handle disorderly participants and to maintain a neutral point of view (NPOV) in articles. But Wikipedia clearly relies on the social capital within its community. Studies of less strong communities will have to be part of the future research in order to determine knowledge losses due to lack of social capital. Furthermore, Bazaar-style knowledge management relies on volunteers who are genuinely interested in the cause. This may not be a paradigm for organizations where knowledge assets are not free. Future research will need to explore the applicability of open source knowledge management when the intellectual property is at least partially proprietary. Finally, the discussed approach to knowledge management appears to work, partly because it can engage increasing numbers of participants to deal with a growing task domain. One has to wonder about the limits of growth of this scenario. Considering both the positive findings and the challenging questions, it appears that Bazaar-style knowledge acquisition using wikis will be a promising application for the practice of knowledge management as well as a rich source of interesting research questions.

ACKNOWLEDGMENTS

The research described in this article was supported in part by CityU grant SMA#9030992. The author completed a substantial portion of the research leading to this article during two sabbaticals at the School of Information Science, Claremont Graduate University. The research assistance of Karen Cheung is thankfully acknowledged.

REFERENCES

- Benkler, Y. (2002). Coase's Penguin, or, Linux and the nature of the firm. *The Yale Law Journal*, 112(3), 369-446.
- Brown, E. (2004). Veni, vidi ...Wiki? *Forbes*, 173(13), 062a.
- Buchanan, B., & Smith, R. (1988). Fundamentals of expert systems. *Annual Review of Computer Science*, 3(7), 23-58.
- Bygstad, B. (2003). The implementation puzzle of CRM systems in knowledge-based organizations. *Information Resources Management Journal*, 16(4), 33-45.
- Cheung, K. S. K., Lee, F. S. L., Ip, R. K. F., & Wagner, C. (2005). The development of successful on-line communities. *International Journal of the Computer, the Internet and Management*, 13(1).
- Ciffolilli, A. (2003). Phantom authority, self-selective recruitment and retention of members in virtual communities: The case of Wikipedia. *First Monday*, 8(12). Retrieved from http://firstmonday.org/issues/issue8_12/ciffolilli/index.html
- Davenport, T., & Prusak, L. (1998). *Working knowledge. How Organizations Manage What They Know*. Harvard Business School Press.
- Drucker, P. (1993). *Post-capitalist society*. New York: Harper Business.
- Drucker, P. (1999). *Management challenges for the 21st century*. New York: Harper Business.
- Frappaolo, C., & Wilson, L. T. (2003). After the gold rush: Harvesting corporate knowledge resources. *intelligentKM*. Retrieved from <http://www.intelligentkm.com/feature/feat1.shtml>
- Hamm, S. (2004, August 18). Linus Torvalds' benevolent dictatorship. *Business Week Online*. Retrieved from http://businessweek.com/technology/content/aug2004/tc20040818_1593.htm
- Hof, R. D. (2004, June 7). Something wiki this way comes: They're Web sites anyone can edit — And they could transform corporate America. *Business Week*, 128.
- Holsapple, C. W., & Joshi, K. D. (2002). Knowledge management: A threefold framework. *The Information Society*, 18, 47-64.
- Jones, C. (1996). *Patterns of software systems failure and success*. Boston, MA: International Thomson Computer Press.
- Land, R. (2002). Software deterioration and maintainability: A model proposal. In *Proceedings of the Second Conference on Software Engineering Research and Practice in Sweden (SERPS)*, Karlskrona, Sweden.
- Lee, W. A. (2001). Experian and fair, isaac tweak subprime scoring. *American Banker*, 167(247), 7.
- Leuf, B., & Cunningham, W. (2001). *The wiki way: Collaboration and sharing on the Internet*. Reading, MA: Addison-Wesley.
- Malhotra, Y. (2000). From information management to knowledge management: Beyond the "hi-tech hidebound" systems. In K. Srikantaiah, & M. E. D. Koenig (Eds.), *Knowledge management for the information professional* (pp. 37-61). Medford, NJ: Information Today Inc.
- Markus, M. L., Manville, B., & Agres, C. (2000). What makes a virtual organization work — Lessons from the open source world. *Sloan Management Review*, 42(1), 13-26.
- Mayfield, R. (2002, November 19). Occupational spam. *Ross Mayfield's Weblog: Markets, Technology and Musings*. Retrieved from <http://radio.weblogs.com/0114726/2002/11/19.html>
- Mockus, A., Fielding, R. T., & Herbsleb, J. D. (2002). Two case studies of open source software development: Apache and mozilla. *ACM Transactions on Software Engineering Methodology*, 11(8), 309-346.
- Newcombe, T. (2000). Opening the knowledge portal. *Government Technology*. Retrieved from <http://www.govtech.net/magazine/gt/2000/sept/knowledge/portal.phtml>
- Nonaka, I. (1994). A dynamic theory of organizational knowledge creation. *Organization Science*, 5(1), 14-37.
- Raymond, E. S. (2001). The cathedral and the bazaar. *First Monday*, 3(3). Retrieved from http://www.firstmonday.dk/issues/issue3_3/

- raymond/
Scacchi (2004). Understanding free/open source software evolution: Applying, breaking and rethinking the laws of software evolution. In N. H. Madhavji, M. M. Lehman, J. F. Ramil, & D. Perry. (Eds.), *Software evolution*. New York: John Wiley and Sons.
- Snyder, C. A., & Wilson, L. T. (1998). The process of knowledge harvesting: A key to knowledge management. In *Proceedings of the 8th Annual BIT Conference*, Manchester, UK.
- Sviokla, J. J. (1990). An examination of the impact of expert systems on the firm: The case of XCON. *MIS Quarterly*, 14(2), 127-140.
- Turban, E., & Aronson, J. (2000). *Decision support and intelligent systems* (6th ed.). Prentice Hall.
- Ullman, H. (1989, July 17). Machine dreams: Future shock for fun and profit. *The New Republic*.
- Wagner, C. (2000). End-users as expert system developers. *Journal of End User Computing*, 12(3), 3-13.
- Wagner, C. (2003). Knowledge management through end user developed expert systems: Potential and limitations. In M. A. Mahmood (Ed.), *Advanced topics in end user computing, Volume 2* (pp. 148-172). Hershey, PA: Idea Group Publishing.
- Waterman, D. A. (1986). *A guide to expert systems*. Reading, PA: Addison-Wesley.

ENDNOTES

- ¹ ezboard.com announced that it had hosted more than 1 million communities on March 1, 2002, and claims 14 million registered users as of June 2004.

Christian Wagner is an associate professor of information systems at the City University of Hong Kong. He has carried out research on knowledge acquisition and knowledge management since 1986 and has published several articles on the topic. Wagner's recent research activities are documented at <http://wagnernet.com>. He believes that embracing people's dual role as both knowledge providers and knowledge "consumers", paired with technology that enables knowledge contribution and aggregation at very low effort, leads to a paradigm shift in knowledge management. Wagner obtained his PhD at the University of British Columbia, and previously served on the faculty at the University of Southern California.