EMPIRICAL EVALUATION OF CYCLADES: A COOPERATIVE KNOWLEDGE MANAGEMENT ENVIRONMENT

Tom Gross[1], Dian Tan[2], Wido Wirsam[2],

Knowledge management is often viewed as a structured process of eliciting knowledge, storing knowledge, and later retrieving by individuals. In this paper we argue that knowledge management should be seen as a dynamic process—an interaction between experts. We start with a motivation for this cooperative perspective of supporting knowledge management through support for the creation and exchange of knowledge in communities. We introduce the CYCLADES environment—an open cooperative virtual archive environment based on open archives—and we present an empirical study of its use. A final discussion will reveal some important lessons for the design of cooperative knowledge management environments.

1 Introduction

Knowledge management is often viewed as a structured process of eliciting knowledge, storing knowledge, and later retrieving by individuals. In this paper we argue that knowledge management should be seen as a dynamic process—an interaction between experts. We would like to depart from some epistemological considerations. A thorough discussion of epistemology would go beyond the scope of this position paper; nevertheless, we need to clarify some basic notions: in the context of this paper data is seen as raw; information is data with meaning; and knowledge is verified information. Expertise is seen as the 'embodiment of knowledge and skills within individuals' [11]. Whereas data and information can be easily processed by and exchanged via computers, the verification that is inherent in knowledge is typically a human activity, which can only be delegated to computers and mediated through information systems and knowledge management system to some extent. Expertise is *per definitionem* human. It is important to also see the 'knowing' and not only the 'knowledge' [1]. Therefore, sharing and exchanging knowledge and expertise cannot be automated totally. In a study Twidale and associates [13] identified several tactics of information seekers. Information seekers can consult colleagues and ask them for information and references;

[1] Faculty of Media, Bauhaus-University Weimar, Germany
tom.gross(at)medien.uni-weimar.de
[2] Fraunhofer Institute for Applied Information Technology FIT, 53754 St. Augustin, Germany
{dian.tan, wido.wirsam}(at)fit.fraunhofer.de

they can wander around to search for information and hope to meet others coincidentally; they can do brainstorming to generate numerous ideas; or they can bibble—that is, they can look for existing bibliographies to the same or a similar topic (this strategy is often used on the WWW, when users use the bookmark list of other users with similar interests).

The results of the above studies have clear implications for the design of systems that aim at supporting the flexible sharing and exchanging of knowledge. First, technology has to extend the reach of the single knowledge worker; so, she can easily and flexibly contact other persons and share and exchange knowledge when needed. Flexible switching between solitary work and information seeking and spontaneous contacts and dialogues with others are a core requirement for flexible sharing and exchange of knowledge. Kuhlthau [8] developed the notion of zone of intervention as an area in which users are very efficient in gathering the information needed. She defines a zone of intervention as: 'that area in which an information user can do with advice and assistance what he or she could not do alone'. Secondly, technology needs to empower communities of knowledge workers to establish a common knowledge base over time—often called community memory. Marshall and associates [10] argue that 'when people work together—whether in designing a product, or creating training materials from video-based documentation, or writing a coherent analysis of a complex situation in the world—they require, and put effort into constructing and maintaining, shared understandings of what they are doing'. They define a community memory as an 'open-ended set of shared interpretations and understandings developed and maintained by the group'.

Subsequently, we introduce the CYCLADES environment, which supports this cooperative approach to sharing and exchanging knowledge. The CYCLADES environment is an open collaborative virtual archive environment that supports knowledge workers with functionality for searching and browsing in pre-print (e-print) archives as well as for sharing and exchanging their findings in virtual communities. Existing systems typically only provide a part of the functionality needed. CYCLADES combines functionality that is partly available in other, more or less isolated systems. For instance, some systems support the information retrieval of online literature such as the digital libraries of the ACM and of IEEE for computer science, Math-Net for mathematics [3], or arXiv for physics [15]. Other systems such as Wikis [16] or BSCW [2] allow users to share and exchange documents in groups or communities, but do not provide functionality for information retrieval of online literature. In general, most of these systems are designed and implemented as stand-alone, closed applications. The CYCLADES environment combines functionality for information retrieval and for sharing and exchanging information in groups and communities. It is compliant to the Open Archive Initiatives (OAi) standard—a technical and organisational framework facilitating the discovery of content stored in distributed e-print archives [12]—and can

connect to any e-print archive. The main focus of this paper is on an empirical evaluation of the CYCLADES environment.

2 CYCLADES

The CYCLADES system integrates functionalities that support a knowledge worker in searching a large number of different digital archives. The most important features are: retrieval of information from many large, distributed digital archives; feedback on the degree of relevance of the retrieved articles; regular information about new publications; automatic retrieval of users' long-term information needs (user profiling); automatic dissemination of relevant information to community members through recommendations; quick on-line annotations on search results; support in carrying out community services such as peer review. Those community functionalities enable the members to learn from, contribute to and collectively build upon the community's knowledge. More information on the CYCLADES functionality can be found in [4, 7].

The users' requirements were surveyed and accounted for in the design of the CYCLADES functionality: the system concept was presented to prospective users, and the users' feedback was captured via a Web-based questionnaire. Results were reported in [6].

CYCLADES is implemented as several interoperable services. The communication among these services is based on standard Internet technology using the XML-RPC protocol [14]. This specification also gives a great flexibility to integrate new functionality as well as the possibility to easily update existing components. The distributed architecture allows the different servers to achieve the best performance for the specific task:

- The Mediator Service (MS) manages the communication among all other services. Through this the system gains flexibility and transparency. The MS is also responsible for the registration and login of users via the Web front end.

- The Access Service (AS) is responsible for accessing all connected archives. It harvests, indexes and retrieves metadata records from the underlying archives.

- The Collaborative Work Service (CWS) provides a folder-based environment for managing metadata records, queries, collections, annotations as well as links to external documents. The CWS supports users in managing their individual and shared folders. It enables communication among community members including special awareness features reflecting the groups' activities in the shared information space.

- The Rating Management Service (RMS) handles ratings of items stored in the CWS and provides this information to users as well as to the other services.

- The Search and Browse Service (SBS) supports searching for records from the various collections, formulating search queries and browsing the results.

- The Filtering and Recommendations Service (FRS) provides personalised filtering of queries and query results, as well as recommendations of relevant records, collections, users and communities. The FRS analyses a user's behaviour, generates a user profile describing the user's interest, and provides information according to this profile.

- The Collection Service (CS) supports the creating and editing of collections. Collections partition the information space according to the users' interests and make the individual archives transparent to the user.

The breakdown of the CYCLADES functionalities into several distributed services was hidden from the user. From the users' perspective the CYCLADES user interface acted as one homogenous system. The only technical requirement for potential users is a Web-browser. Further details on the design and implementation of the CYCLADES environment can be found in [5].

3 System Evaluation

The user evaluation aimed at gathering information about the users' satisfaction concerning the usability and usefulness of CYCLADES. The CYCLADES prototype was available on the Internet. Anyone was allowed to register and test the system. We spread the information in Usenet newsgroups, on mailing lists, on Websites and in journals. We prepared a 'QuickStart to CYCLADES' as a fast introduction into the system. After reading this the user could easily explore the system. The data collection took place from May 5 to July 5, 2003.

3.1 Methods

We created a Web-based questionnaire. The questions were implemented using four point Likert scales [9]. These are apt to elicit subjective data about how the users perceive the system. But since the data are ordinal scaled, it is not permissible to perform further statistical calculation with them. Free text fields were offered for further explanations, suggestions for improvement or remarks.

A pre-test was conducted to gain information about the suitability and comprehensibility of this questionnaire. Considering the user comments the questionnaire was redesigned and released for the public.

The final questionnaire contained questions regarding the responding user group as well as some general information about the system use (e.g., how much time the user had spent with the system, the browser environment). The principal part consisted of six sections, each covering a set of services and features that users typically use together (Communities and Folder Handling; Searching Archives; Collections; Advanced Search Features; Community Management and Recommendations). At the beginning of each section the user stated if she had used this function. If a function had not been used, the user specified her reasons for not using it, and then proceeded to the next section. For those who used a function, some specific questions about the according function and a concluding appraisal followed. After assessing the single functions the user rated the usefulness and usability of CYCLADES as a whole.

Technically the questionnaire was implemented as Web pages using PHP scripts to fill the values entered by the user in a MySQL database. For the data acquisition SQL queries were performed to extract the specific information needed.

In addition to the questionnaire we recorded all actions the users performed during the system use, to get an impression about the frequency and area of use of the particular functions.

3.2 Results

Altogether 238 new users created an account that was needed to use the CYCLADES system. The high number of registrations shows the substantial interest in CYCLADES. About one third of all registered users started to fill in the questionnaire, but only 35 completed most of it and were included in the analysis. We assume that this was due to the complexity of CYCLADES; some experience with the system was mandatory to reasonably answer the questions. Still there were some blank fields, these are referred to as "not specified".

The age of the respondents spread from 21 to 60 years, with an average age of 29 years. Although the CYCLADES system and the QuickStart were accessed from countries all over the world the questionnaire was answered mainly from German residents. There were also respondents from Greece, France, Brazil and other countries. The majority of them worked in computer science, but there were also among others respondents from physics, business administration, and librarianship.

Firstly, the respondents gave a *self-assessment* how good they could handle the system (cf. Figure 1). Every respondent categorised herself as beginner, intermediate, advanced or expert in handling the according function. The function 'Recommendations' is an exception, since the user is passive in the actual process of creating the recommendations. She only chooses if she wants to get recommendations and which kind of recommendations when creating a new folder or community. The recommendations are then created by the CYCLADES system. Therefore the recommendation service is not accounted for in this analysis. The figure shows that only few users reached an advanced or expert level. This is supposedly due to the relatively short time of exploration.

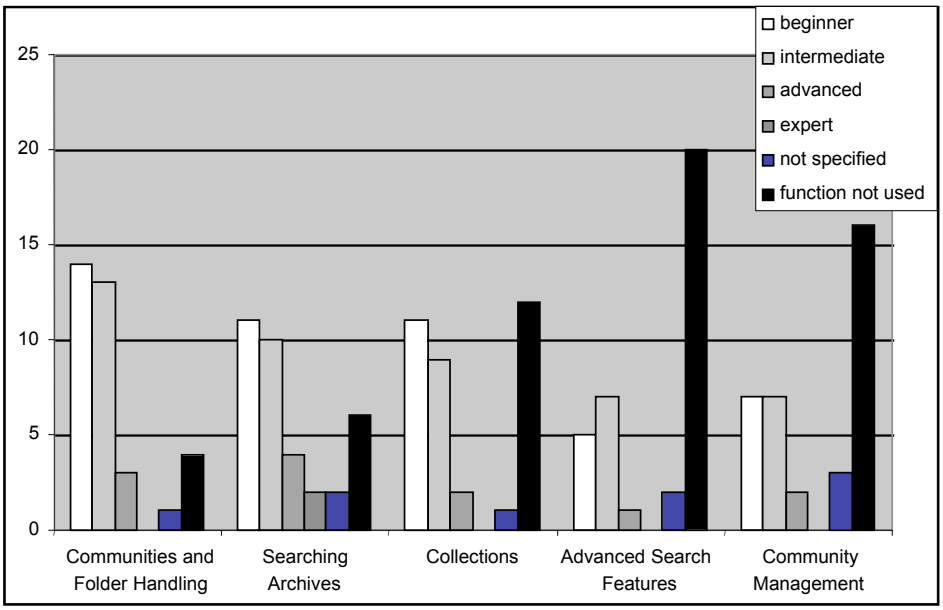

Figure 1. Self-assessment concerning system expertise.

Finally, the users reconceived each function under the aspects of usefulness and usability. For the overall appraisals we combined closely related function sets to larger categories. The resulting sets were Community Support, Search Functions, Collections and Recommendations. Additionally the CYCLADES system as a whole was rated. The evaluation of the *usefulness* per function set is mapped in Figure 2.

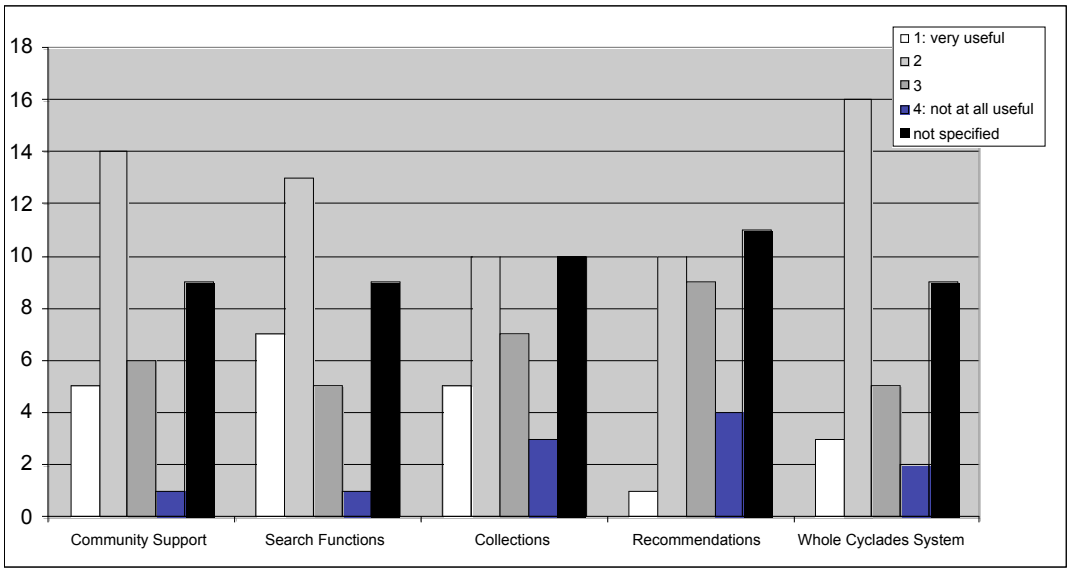

Figure 2. CYCLADES overall usefulness per function set.

The majority of the users assessed most function sets as very useful or useful. The search functions received the best rating of all functions; one third of the respondents chose the best benchmark 'very useful'. The community support was also appraised as useful, 54 percent of all respondents chose the best or second best benchmarks. CYCLADES as a whole, considering the combination of all functions, made a good impression: only five persons selected the benchmark '3' and two users the '4', the median is '2'. The most often chosen benchmark, the mode, is '2' for each function set as well as for the whole CYCLADES system.

A lot of users thought functions sets to be useful for their work, even though they did not use it in the exploration. Reasons for this low use were among others the lack of time, but also that the users did not to know how to use the function or they could not find the function. Some users stated that certain activities within CYCLADES caused trouble with their computer.

The *usability* of CYCLADES (cf. Figure 3) was assessed on a scale ranging from 'self-explaining' to 'difficult'. Since the usability can only be judged by those who know a function the extra option "not used" was offered. Regarding the usability the Community support received the best rating. The collections function was the most difficult to use. The usability of the whole CYCLADES system received a good to intermediate grate.

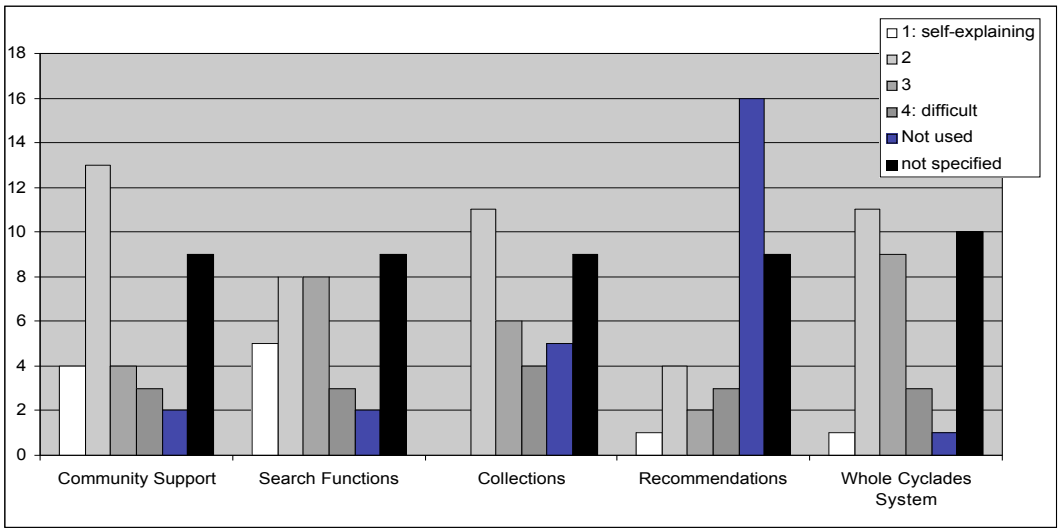

Figure 3. CYCLADES overall usability per function set.

4 Discussion

On a whole the evaluation clearly showed that users used the community support and the search functionality. The functions related to these two sets also received the best results concerning usefulness. We think that this is the case for two reasons: first, these functions are basic for any cooperative knowledge management system—that is, they are real prerequisites for successful sharing, exchanging, and learning on demand—and, secondly, they were easy to use. While some functions related to collections might be interesting for long-term and/or experienced users, these functions are either too difficult to use for non-regular users and the benefit of these functions is not clear if the system is used for a short- to medium-term duration only. With the recommendations it is similar: for novel users the system, the information, as well as the communities are new, and therefore novel and interesting. It is only after a longer time (around six months) that the users have a very good overview of the contents offered and the communities around. Only then, the recommendation of other users, communities, and so forth that are still not known is a positive surprise. Furthermore, recommendations often require a thorough understanding of users about their genesis; otherwise, users do not accept them.

All in all we are satisfied with the evaluation: on the one hand it gave us some positive feedback on designs and implementations that are accepted by users, and on the other hand we got valuable feedback concerning possible improvements of some functions.

References

1. Cook, S. and Brown, J.S. Bridging Epistemologies: The Generative Dance Between Organisational Knowledge and Organisational Knowing. *Organisation Science 10*, 4 (1999). pp. 381-400.

2. FIT. *BSCW Home Page*. Fraunhofer FIT and OrbiTeam Software GmbH, http://bscw.fit.fraunhofer.de, 2004. (Accessed 28/5/2004).

3. Groetschel, M. *Math-Net.* http://www.math-net.org/, 2004. (Accessed 28/5/2004).

4. Gross, T. CYCLADES: A Distributed System for Virtual Community Support Based on Open Archives. In *Proceedings of the Eleventh Euromicro Conference on Parallel, Distributed, and Network-Based Processing - PDP 2003* (Feb. 5-7, Genova, Italy). IEEE Computer Society Press, Los Alamitos, CA, 2003. pp. 484-491.

5. Gross, T. Design, Specification, and Implementation of a Distributed Virtual Community System. In *Proceedings of the Twelfth Euromicro Conference on Parallel, Distributed, and Network-Based Processing - PDP 2004* (IEEE Computer Society Press, (accepted).

6. Gross, T. and Kreifelts, T. Sharing and Exchanging Knowledge in Virtual Communities: The CYCLADES System. In *Proceedings of the Tenth International Information Management Talks - IDIMT 2002* (Sept. 11-13, Zadov, Czech Republic). Universitaetsverlag Rudolf Trauner, Linz, 2002. pp. 71-81.

7. Gross, T. and Wirsam, W. Distributed Development of a Distributed Virtual Community System. In *Proceedings of the Eleventh International Information Management Talks - IDIMT 2003* (Sept. 10-12, Budweis, Czech Republic). Universitaetsverlag Rudolf Trauner, Linz, 2003. pp. 103-113.

8. Kuhlthau, C.C. Inside the Search Process: Information Seeking from the User's Perspective. *Journal of the American Society for Information Science 42*, 5 (May 1991). pp. 361-371.

9. Likert, R. A Technique for the Measurement of Attitudes. *Archives of Psychology 140* (June 1932).

10. Marshall, C., Shipman, F.M. and McCall, R.J. Putting Digital Libraries to Work: Issues from Experience with Community Memories. In *Proceedings of the Digital Libraries Workshop - DL'94* (May 19-20, Newark, NJ). Springer-Verlag, Heidelberg, 1994.

11. McDonald, D.W. and Ackerman, M.S. Just Talk to Me: A Field Study of Expertise Location. In *Proceedings of the ACM 1998 Conference on Computer-Supported Cooperative Work - CSCW'98* (Nov. 14-18, Seattle, WA). ACM, N.Y., 1998. pp. 315-324.

12. OAi. *Open Archives Initiative.* http://www.openarchives.org/, 2004. (Accessed 28/5/2004).

13. Twidale, M.B., Nichols, D.M. and Paice, C.D. Browsing is a Collaborative Process. *Information Processing and Management 33*, 6 (1997). pp. 761-783.

14. UserLand Software Inc. *XML-RPC Home Page.* http://www.xmlrpc.com/, 2003. (Accessed 28/5/2004).

15. Webmaster. *arXiv.org.* http://arxiv.org, 2002. (Accessed 28/5/2004).

16. Wiki. *Wiki: Welcome Visitors.* http://wiki.org/, 2002. (Accessed 28/5/2004).