

Supporting Collaborative Information Spaces for Tourists

Marco Prestipino
University of Zurich

Abstract

We show that virtual communities are able to serve individual, highly specific information needs. Our analysis also indicates that existing technology for communities might not support collaborative creation and maintenance of knowledge well and thus improvements to the interaction and visualization mechanisms might be beneficial. We present and briefly discuss an architecture for a community system supporting communication and collaboration on shared material.

1 Introduction

Tourism is an information-intensive activity. Starting with destination choices and itinerary planning, information needs also arise spontaneously during travel. More so on self-organized tours, where itinerary and activities are completely free to be adjusted, e.g. a traveler could extend his stay at a particular place when being attracted or shortening his stay because of bad weather or other unpleasant but unforeseeable events. Information needs of tourists can range from open, complex questions to highly specific ones. Consider the following description of a tourist's information need: "I have three weeks to travel with a certain amount of money, and I would like to see a part of central Africa, but also visit a certain city a friend of mine is living in. What places should I visit?". This information need could be answered with dozens of different propositions, but it is obvious that not all the preferences have been expressed, so there is a need for communication and feedback allowing for refinement and clarification. Also, the more individualization is permitted, the more variables have to be taken into account and the more information needs to be available. The user can acquire information through communities, brokers or information products. Communities are networks of people which can exchange information. This might be personal networks like family, workmates, but also large, fuzzy groups centered around shared values or interests, like participants in Internet newsgroups. Information brokers in tourism are travel agencies, tourism offices, tour organizers and the like, where humans communicate with the consumer, gather information from other sources and provide tailored value-added information. Brokers may also use personal networks or information systems, and then condense and tailor the information to the specific information need. Communities

and brokers deliver information as a service, whereas information products are created without customer involvement and are also consumed autonomously by the consumers. Communities and information broker refine information through mental processes before delivery, whereas information systems provide information without human intervention based on selections made by the customer. The rest of the paper is structured as follows: chapter two discusses the deficiencies of information systems and the advantages of communities when dealing with complex information needs. In chapter three we analyze the performance of a community and compare it with an information product. Following, results of an explorative analysis of usage data are presented in chapter four. In chapter five we present an architecture for a community support system.

2 Information for Travelers

There are three problems when trying to support travelers with information products, e.g. databases or books:

- Information needs being too complex and multidimensional. As the example above showed, some questions must be either clarified in a communication process, or can only be answered by generic all-encompassing answers. The latter could produce so many answers as amounting to information overflow.
- Not only does the information system need to hold a large amount of data, there must also be a way of specifying preferences. It would be a daunting task to include all the relevant information. And so far no user interface or structure of a book achieves the expressiveness of human language. The fine distinctions possible in human language are necessary to assign the appropriate weight to each preference. Information needs of travelers are characterized by highly dynamic and individual factors, as preferences for sites, weather conditions, prices, transportation, bank holidays, political and economical changes, appointments and so on.
- It is difficult for humans to deal with information provided by a “black box” information system: how trustworthy is the information and is it really the best information for a given need? (cf. Kuhlen 1999).

This is aggravated by the fact that travelers face a “market for lemons”, as a market with uncertainty about product quality was coined by Akerlof (1970). Information goods need to be consumed before their quality can really be judged. Trip itineraries, hotels etc. might differ strongly from the description given beforehand because sellers try to advertise as much as possible. To reduce uncertainty, sellers give out samples, which is cost-effective for digital products. But this is best suited for information goods: more pictures will not help travelers but simply add to information already available. So another way, successfully demonstrated by eBay, is to generate trust by letting consumers write their opinion about products or sellers. Information by an independent third party is more likely to be accepted by customers. Tourism communities of this type have not been promoted from official or commercial tourism organizations. This is interesting because survey data (Freyer 2002) indicates that 34.4% of travelers use personal experience to choose their destination and 38.1% use

experience of friends. The data used for this conclusion is from 1986, 1990 and 1992, when the Internet was almost unknown to the public. We contend that today personal networks may now play an even more significant role, as electronic communication media extend and facilitate personal networks. Many information in tourism involves judgment, e.g. about the beauty of a city. Then, a consumer does not want a view based on a single producer or a group of producers unknown to him, but a more balanced view of many people explaining and arguing their position. The conversation also reveals traits of the contributing persons, which can be used to value their statements according to personal preferences.

The advent of the World Wide Web brought about easier access and navigation in information spaces, but soon the amount of Web documents could not be managed by manually created lists, and search engines boomed. Whereas professional database provider or guidebook publisher only include quality content into their products, the Web has no such restrictions, and therefore quality is very heterogeneous. Search engines are black boxes, as the user does not know why he is given a certain result. Because of their limited capabilities, quality of information obtained by search engines is extremely heterogeneous. Thus, there is a limit in using information products to automatically answer information needs. We claim that blending information technology with human communication and collaboration results in an information system with higher performance. In the context of the Internet, a virtual community is such a system, although traditional research on virtual communities focuses more on the social relationships, whereas here we look at the capability of creating, storing and distributing information. The benefit of this approach is that we can compare communities and information systems in terms of performance and determine beneficial changes of the underlying technology.

Figure 1 depicts the relationship between a person's information needs, what information need she perceives and consequently is able to express and what is available through an information system. Only the intersection of all three circles is returned. Note that the actual information need may be much larger and/or different than the perceived need. Also, a user might not be able to express her need adequately, because she does not know what the answer will be like and what terms are relevant to her problem. This is a well-known problem in information retrieval research, labeled anomalous state of knowledge (Belkin 1982).

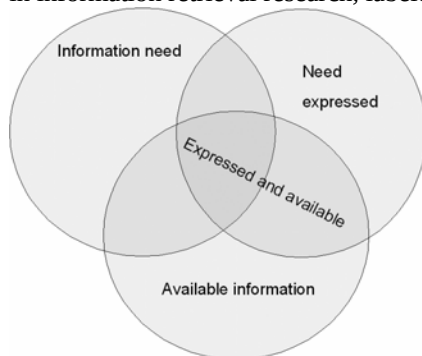


Figure 1 The user is unable to express his full Information need (cf. Belkin 1982)

Our hypothesis is that communities change the information situation as show in fig. 2. The community allows for feedback and is therefore able to clarify with a questioner if the expressed need describes indeed his information need. The natural language helps circumscribing missing domain specific terms and therefore allows expressing a larger information need. Human communication makes it much easier to judge trustfulness of information or persons. In a discussion, the questioner may discover his information need is much larger than he initially expressed. Finally, we contend that an active community of a certain size keeps more information available than e.g. a book, as an information product produced with limited resources.

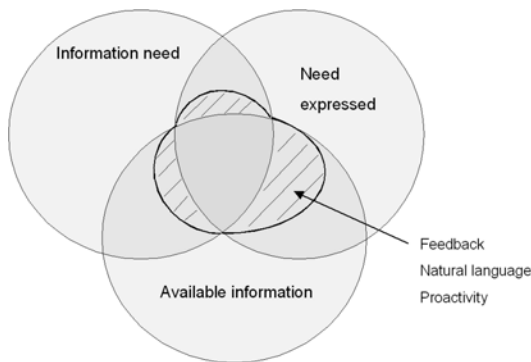


Figure 2 Effects of a community on a user's information situation

Aside from psychology or sociology, community research has focused on factors in the use of technology for communities (cf. Leimeister et al. 2004), but not about changes in the technology itself or evaluation of the performance, i.e. how well a community serves information needs. With regard to the technology, this seems rather odd in our fast-paced world: most communities use tools based on USENET¹, a technology invented in 1979 for system administrators (cf. Pfaffenberger 2003). Other research has been carried out on providing filtering mechanisms based on a user's previous actions or collaborative filtering (Lueg 2003).

3 Comparing a community with a guidebook

So far we have been theorizing about the value of communities as information systems. To justify this hypothesis, we conducted a formative evaluation of a tourism community and compared it with a guidebook. The community was started in 1998 as a simple web-based forum. It is mainly aimed at people interested in Brazil and Brazilians living in German-speaking Europe. Membership is free, and the community is not run for profit, i.e. there is almost no advertising. No registration is required to read or write in the forum, but it is possible to register in order to keep a permanent username and be able to write private messages to other users. The community discussion area is structured into 24 topic categories

¹ UNIX User Network

(“forums”). Besides, the community features picture galleries, chatrooms, a calendar for community-related events, a database with addresses and links to web resources. The discussions are moderated. Moderators can delete posts and enforce community rules. We chose two forums: one about insider tips about cities and another concerned with traveling to Brazil. We used forum usage data of a six-month period (27.09.2002 – 27.03.2003) and conducted several data analysis. General data about forum usage is shown in Table 1. We excluded two Threads (see number in brackets) because they were obviously related to user errors, i.e. threads posted verbatim two times in a row or posted as an answer to a thread in another forum.

	Number of Threads	Average Number of Threads / day	Average number of participants	Average number of replies
Insider tips	63	0.35	3.92	4.41
Travel to Brazil	143 (145)	0.79	3.97	4.76
Total	205	1.13		

Table 1 Analysis of two forums for a period of 181 days (27.9.2002 - 27.3.2003)

We analyzed the performance of the community in terms of efficiency and effectiveness. Efficiency relates to how fast questions are answered, effectiveness relates to quality, i.e. a measure of the number of questions answered satisfactorily. The number of participants in a discussion gives a hint on quality, as wrong or one-sided information is likely to be corrected if many persons contribute. This is a kind of peer review-mechanism (cf. Kuhlen 2002, p. 37). The data shows that usually more than two people engage in a discussion and a certain amount of peer review occurs. Of course much more people only read without contributing. It is a typical problem when observing communities not entirely closed that the actual number of participants, including lurkers, is difficult to establish.

To determine the performance we chose travel-related questions about Brazil posted in German out of the two forums and analyzed how many questions were answered satisfactorily and how fast. The results, shown in table 2, describe an efficient and effective information system, as more than 80 percent of questions asked were answered. As the time spans differed vastly, we also included the median. The data shows that this community reacts to questions quickly.

	Insider tips	Travel to Brazil
Relevant questions	32	62
Answered	24 (83.9%)	52 (80.9%)
Average time span to last relevant reply	108 hours	106 hours
Median time span to last relevant reply	66 hours	52 hours
Average time span until first reply	14 hours	39 hours
Median time span until first reply	5 hours	5 hours

Table 2 Efficiency and effectiveness of the two forums

We claimed that natural language and communication facilities constitute an advantage of communities compared to using information products. This claim seems justified: looking at

the threads with answered questions, in sixteen threads (21.1%) community members posed questions to the questioner to further clarify the information need. In twenty-five threads (32.9%) additional questions by the questioner emerged, i.e. the questioner's real information need was made more explicit by the communication process (cf. table 3).

	Insider tips	Travel to Brazil
Feedback on question	5 (20.8%)	11 (21.2%)
Additional questions	6 (25.0%)	19 (36.5%)

Table 3 Clarification of information needs

We compared the information given by the community in the answered threads with the information given in the leading travel book Lonely Planet Brazil (5.th edition 2002). We rated information on a scale ranging from zero (not answered) to six (excellent answer). The rating was done by one student, to ensure consistency of rating. The community scored an average of 5.03, whereas the book scored 3.31. The community often provided more accurate and comprehensive, up-to-date information, tailored for individual information needs. The book in turn delivered better answers for very broad questions, e.g. "what to see in Rio". We conclude that active communities are able to deliver content of a quality that can stand up against commercial information products. The result of this analysis is in accordance with our model, in that it shows that a community is capable of providing tailored information for individual information needs, whereas general information could be delivered by information products.²

4 Interaction Issues in a Forums-based Community

Considering the visualization typically used in forums and the methods of interaction, the following issues could be expected to arise:

- Threads are isolated: several Threads might deal with the same or related topics, so information might be fragmented. E.g. the same questions may repeatedly be posed.
- Threads mimic a discussion transcript, i.e. it is often required to read the whole thread to get information. The information is not structured for later use.
- As the information in newsgroups is not condensed in one information object, it is difficult to update them. Usually topics are visualized as a list sorted by date of the last contribution. Thus, topic threads are abandoned and forgotten. The community might mitigate this by pointing questioners to existing threads.

To determine if the human-computer interaction really causes problems and to identify areas for enhancements, we extended the analysis on the community. Several interviews with the administrator of the community and input from moderators and community members confirmed our hypothesis that in the existing structure, similar questions are repeated over and over again, which is perceived as annoying. Another hint is the frequent quotation of

² More details of the comparison are available online: <http://www.ifi.unizh.ch/im/imgt/evaluation/>

older threads. In USENET, this problem led to the creation of Frequently Asked Questions³ (FAQs), lists of common questions with answers that are available at a prominent place. But FAQs themselves need to be maintained and the mentioned problems hold for them too. We then used an explorative approach by clustering discussion topics. Initially we found questions which were rather broad in comparison to very specific questions. This led to the category ‘general question’. Though, as we categorized all postings, we found it difficult to differentiate between ‘general’ and ‘travel related’ questions. Our criterion for qualifying for the category ‘travel-related’ had been that a question be related to a concrete travel. This proved inadequate, because there are questions like ‘where can I find information about viewing soccer games in Brazil?’ or ‘how can I use a mobile phone in Brazil?’ which do not provide enough context to be categorized using this criterion. Also, as these examples show, location is not a criterion either, because the mentioned questions might be important questions for somebody traveling, but are only related to a broad context. Actually most postings are ‘travel-related’ to a different extent, aside from postings which are simply off-topic in these forum categories. Also, the percentage of questions in this category is low and does not indicate a problem or opportunity for better technological support. The categories and the distribution of topics in the two forums are shown in table 4. Note that the categorization was based on the initial posting, and sometimes following messages in a thread could be of a different category.

The most valuable finding is that only 52% of the topics are questions, whereas the remainder consists of postings which were not solicited in the first place, but nevertheless are of interest to the community. Many of these unsolicited postings are *merely informative* and *not meant to start a discussion*. Newsgroup technology, which was invented as a tool for discussion and announcements, does not support this kind of information very well. Because of the different granularity and heterogeneity of information posted and the lack of structure in the forum, it is difficult to find information. Unsolicited information adds to the problem of information being hidden, as it increases the speed at which postings are scrolling out of the visible area. This information and the information that is the outcome of discussions should be kept in a structured way as shared material, such that it can be corrected, updated and synthesized.

	Insider tips		Travel to Brazil	
Thread category	63 Threads		143 Threads	
Travel-related question	33	52.38%	72	50.35%
General question	1	1.59%	10	6.99%
Meeting / Contact	3	4.76%	7	4.90%
Travel diaries	4	6.35%	9	6.29%
Unsolicited information	22	34.92%	41	28.67%
Unsolicited Opinion	0	0.00%	2	1.40%
Provocation	0	0.00%	1	0.70%
Amusement	0	0.00%	1	0.70%

Table 4 Results of categorizing two forums

³ Cf. <http://www.faqs.org/faqs/about-faqs>

5 An Architecture for Community Support Systems

Based on our findings, we propose an architecture more suitable for community based creation, storage and diffusion of information, shown in figure 3. We suggest a combination of two distinct, interconnected functional blocks (denoted by the grey box). First, the system must support communication, e.g. through a forum. Second, the system must provide facilities to collaboratively build and maintain the community knowledge (“shared material”). This information can be viewed but also modified by community members. Information can directly be entered in this repository, or be an outcome of communication supported by the system’s communication tools. The shared material may in turn be the topic of discussions. The system must provide visualization and interaction facilities to enable the community to effectively enact both processes. We believe this architecture circumvents the disadvantages we theorized and confirmed about traditional community systems. A user may search or contribute to the repository, or communicate with the community. In this latter case, the community may act as a gateway to the repository, when a user is not able to find the right information himself or does not know about his information need. The dotted line shows this information flow, which is a distinctive feature of a community information system when compared to traditional information systems. The shared material constitutes an information product, which could be used without interfering with the community. The difference is that

- the community provides an additional, natural-language interface to the information product (cf. chapter one)
- the fluidity of this information product is very high (it can be modified easily)
- the community provides additional highly specific information (cf. chapter three)
- the information product is transparently created by a group of people and peer-reviewed, as opposed to classical information products created by a small group of producers.

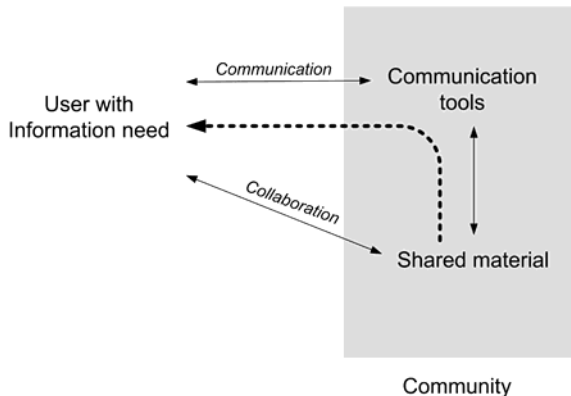


Figure 3 Architecture of a community support system

Our approach is similar to Ackermanns (1996) redesign of the Answer Garden system in that both support and rely on a process to build a repository of knowledge out of raw information

by intellectual work. The process in AG 2 consists of four activities, namely collecting, culling, organizing and distilling. We are evaluating the use of such activities for creating personal information spaces. AG 2, as its predecessor, is designed towards an organizational setting. It allows escalating questions not answered by a database to collaborative systems, and further to human experts, which are located by the system. So far, no evaluation of AG 2 has been published. Also, Pipek and Wulf (2003) show that AG seems to be based on the assumption that “information seekers know exactly what their problem is”, whereas our system supports collaborative and iterative identification of information needs.

Wikis (Leuf & Cunningham 2001) are an example of a suitable tool for collaborative use of shared material. Wikis are web-sites whose content can easily be changed by users using their web-clients. While a Wiki is not a good communication tool for a community (e.g. Wikis lack visualization features for thread-like discussion structures), it fits nicely into our architecture as a tool for the shared material.

When designing according to the presented architecture, core questions need to be answered:

1. How is communication transformed into shared material?
2. How do users find information in the shared material?
3. How do community member realize that a question is posed on a topic of the shared material?
4. How does the system help the user to appraise the trustfulness of information

As an example, consider just running forums and a separated Wiki. The transformation of the information into the shared material would require a user to find the appropriate place in the Wiki to copy the material to. It is obvious that only very limited amount of transfer would occur. Also, users searching information would have to find the appropriate Wiki pages, and then the appropriate forum to pose a question. An example of slightly tighter integration of communication tools and shared material is to map the discussion categories of a forum to thematic starting points in a Wiki. Other examples are interaction mechanism for easily creating shared content out of discussion content and the visualization of open question and topics inside the shared material, e.g. by showing recent question at the bottom of related Wiki pages. We are currently building a new system according to this architecture. Enhancing the performance of virtual communities would significantly ease the overload problem we face in the information age.

6 Literature

- Ackermann, M. (1996): Answer Garden 2: Merging Organizational Memory with Collaborative Help. Proceedings of the ACM Conference on Computer Supported Cooperative Work (CSCW96), ACM
- Akerlof, G. A. (1970): The Market for "Lemons": Quality Uncertainty and the Market Mechanism. The Quarterly Journal of Economics 84 (3), pp. 488–500.
- Belkin, N.J., Oddy, R.N, Brooks, H.M. (1982): ASK for Information Retrieval: Parts I&2, Journal of Documentation, 38(2,3), pp. 61-71; 145-164.

- Freyer, W. (2001): *Tourismus. Einführung in die Fremdenverkehrswirtschaft*. 7. ed., München: Oldenbourg.
- Kuhlen, R. (1999): *Die Konsequenzen der Informationsassistenten. Was bedeutet informationelle Autonomie oder wie kann Vertrauen in elektronische Dienste in offenen Informationsmärkten gesichert werden?* Frankfurt: Suhrkamp.
- Kuhlen, R. (2002): *Wissensmanagement über Elektronische Kommunikationsforen*. In: *Proceedings SEL-ALCATEL summer conference*, 12 .July 2002, Berlin. URL: <http://www.inf-wiss.uni-konstanz.de/People/RK/Publikationen2002/sel-alcotel-rk-wissensmanagement.pdf>, downloaded 5.10.2003.
- Leimeister, J.M., Sidiras, P., Krcmar, H. (2004): *Success Factors of Virtual Communities from the Perspective of Members and Operators - an Empirical Study*. In: *Proceedings of the Hawaii International Conference on System Sciences*, January 5 – 8, 2004, Big Island, Hawaii.
- Leuf, B., Cunningham, W. (2001). *The Wiki Way: Quick Collaboration on the Web*. Boston: Addison-Wesley.
- Lueg, C. (2003): *Exploring Interaction and Participation to Support Information Seeking in a Social Information Space*. In: Lueg, C.; Fisher, D. (Ed.): *From Usenet to CoWebs*. London; New York : Springer, p. 232ff.
- Pfaffenberger, B. (2003). *A Standing Wave in the Web of our Communications: Usenet and the Socio-Technical Construction of Cyberspace Values*. In: Lueg, C.; Fisher, D. (Ed.): *From Usenet to CoWebs*. London; New York: Springer, p. 20ff.
- Pipek, V. and V. Wulf (2003). *Pruning the Answer Garden: Knowledge Sharing in Maintenance Engineering*. in: *Proceedings of the Eighth European Conference on Computer Supported Cooperative Work (ECSCW 2003)*, Kluwer, Dordrecht 2003