2009 3rd IEEE International Conference on Digital Ecosystems and Technologies
Lek, H.H., Poo, D.C.C. & Agarwal, N.K. (2009). Knowledge community (K-Comm): Towards a digital ecosystem with collective intelligence. *Third IEEE International Conference on Digital EcoSystems and Technologies, Istanbul, Turkey, Jun 1-3* (pp. 211-216). https://doi.org/10.1109/DEST.2009.5276690

# Knowledge Community (K-Comm): Towards a Digital Ecosystem with Collective Intelligence

Hsiang Hui Lek[1], Danny C.C. Poo[2] and Naresh Kumar Agarwal[3]

[1, 2, 3] Department of Information Systems, School of Computing, National University of Singapore
E-mail: {lekhsian, dpoo, naresh}@comp.nus.edu.sg

***Abstract*—In this paper, we present a platform called Knowledge Community (K-Comm) that embodies the principles of a digital ecosystem with collective intelligence. This platform is a knowledge-based social network which allows users to contribute and seek information. Users can ask questions or answer questions asked by other members. K-Comm also captures the user profile, and based on a user's participation, identifies his/her area of interest and expertise. The platform is built around the philosophy that each individual has knowledge in a number of areas and that everybody can be an expert in one area or the other. The whole platform is driven by community effort. From a researcher's point of view, this platform provides the researcher with a rich resource in multiple domains.**

***Index Terms*—digital ecosystem, collective intelligence, expert system, knowledge community, knowledge capturing**

## I. Introduction

A natural life ecosystem is defined as a biological community of interacting organisms plus their physical environment [1]. A digital ecosystem (also called Digital Business Ecosystem or DBE) is a self-organizing digital infrastructure aimed at creating a digital environment for networked organizations that supports the cooperation, the knowledge sharing, the development of open and adaptive technologies and evolutionary business models [1]. The organizers of this conference highlight [2] the role of digital ecosystems in "transcend[ing] the traditional, rigorously defined collaborative environments, such as centralised (client-server) or distributed (such as peer-to-peer) models or hybrid models (web services) into a self-organised, interactive environment which offers cost-effective digital services and value-creating activities that attract human, organisation and software agents that participate in it and benefit from it.

Such a digital ecosystem depends on collective intelligence for its success. Collective intelligence (CI) is a shared or group intelligence that emerges from the collaboration of many individuals [3]. There have also been many different interpretations of collective intelligence like crowd sourcing [4], smart mobs [5], and wisdom of the crowds [6] – all of which boil down to the fact that the final output is achieved by the collective efforts of individuals or groups of individuals. The MIT Center for Collective Intelligence (cci.mit.edu) puts forth its basic research question as 'How can people and computers be connected so that—collectively—they act more intelligently than any individuals, groups, or computers have ever done before?', which sums up the idea of collective intelligence [7].

With the increasing connectivity and growing popularity of the web, we see that the Internet has provided an environment for CI. Especially since the Web 2.0 [8] generation, we see that collaboration and community efforts have become a norm, where a digital ecosystem is living and thriving. Unlike web 1.0 which could be considered as mainly "read-only", web 2.0 is based on the idea of a read-write ideology where the users define the content. This can be seen from popular sites like Wikipedia (www.wikipedia.org), YouTube (www.youtube.com), Digg (www.digg.com) and Facebook (www.facebook.com), where there is no one central party providing content. Instead, users contribute and collaborate with one another to define both the content of the sites and how they want the sites to be like. This is what makes a digital ecosystem.

In other areas like open source, we can also see CI in play. These exploit the efforts of the community consisting of end-users and developers. Taking as an example the Mozilla Firefox browser, an open source web browser, we see that there is community involvement from different groups of users. Since the source code is freely available, developers all over the world can participate in the development work. The participation can range from developing browser add-ons to checking correctness of the web browser. Furthermore, end users themselves knowingly and unknowingly contribute in the form of bug reports or suggestions for new features. This demonstrates that different groups of people are involved, leading to a scenario where each of them ultimately benefits from each other. Similarly, major companies like IBM too have been shifting their business model from total proprietary software to partially open source and earning by rendering services [9].

The community is willing to share, contribute and participate in providing better solutions because it ultimately benefits and affects them. E.g. Digg (www.digg.com) relies on the collaborative effort of the community to submit stories and moderate stories [10]. Interesting stories will get more votes and thus be promoted to the front page.

In this paper, we propose a platform that embodies the principles of a digital ecosystem with CI. Our platform, which we call Knowledge Community (K-Comm), is a knowledge-based social network that allows users to contribute and seek knowledge in a social networking environment with the purpose of sharing knowledge. Not only does this benefit the end-users, the platform also allows researchers to perform different kinds of research like information retrieval and information extraction.

The rest of the paper is organized as follows. In the next section, we describe K-Comm. We then look at the components of K-Comm, which include profile gathering, blogs, groups, question and answer (Q&A) and the incentive mechanism for user participation. This is followed by discussion, conclusion and future work. Let us now look at our platform – K-Comm.

## II. K-COMM

K-Comm is a knowledge-based social network platform where people can seek and contribute knowledge and link with like-minded people.

While traditional Knowledge Management Systems (KMS) are based in a company and focus on specific domain or areas the company specializes in, K-Comm has been designed to help every individual contribute knowledge gathered through experiences across a variety of areas – school, culture, countries visited, levels of expertise, interest spans, etc.

K-Comm is also different from conventional social networking sites in that the main focus here is knowledge sharing, as opposed to general socializing and keeping in touch as facilitated by other social networking websites. Currently, knowledge sharing is done mainly using a question and answer (Q&A) approach. Users can ask questions and answer questions. They can also leave comments and rate the questions or answers.

Furthermore, users can also form groups where they can connect with like-minded people who have the same interest. These groups allow them to participate in discussion and share information with one another.

The idea of K-Comm is based on the realization that every individual has something to share and contribute – from knowledge that might seem mundane or profound to oneself, but which helps another individual get an answer to a question. The answer could help the information seeker resolve a task at hand or satisfy a curiosity. It will provide for a feeling of self-worth in every individual, a feeling of usefulness to others/community/society, and at the same time, a ready place (and access to experts) to seek answers to one's questions. The end result is the utilization of collective intelligence towards a thriving digital ecosystem. Community effort is central to the success of such a digital ecosystem, where users help one another and define the content. Based on the users' profiling information and their involvement in K-Comm, we are able to identify their area of expertise and area of interest. Using this information, the platform is able to build an expert directory where every individual is an expert in a particular field. New questions relating to a field can be directed to the experts in those fields. Users do not need to be professionals in order to be an expert in a field. Thus, we are able to tap into the knowledge residing within individuals and unleash their knowledge potential.

We had previously developed a prototype of K-Comm with no social networking aspect [11] in an attempt to capture an individual's tacit knowledge and to increase the span of knowledge areas reachable by traditional KMS. Since KMS is limited to organizations with specific knowledge area of interest, the full potential of individuals is often left untapped. Based on our experience with the prototype developed, as well as a study of various social networking websites, we have identified that social networking is able to break barriers that prohibit an individual from contributing. See [12] for a detailed discussion on barriers to knowledge sharing. Thus, we have incorporated the social networking aspect into the concept of K-Comm and redesigned and rebuilt a totally new system with a social networking core. This system, which is knowledge centric in nature, allows us to tap into the knowledge potential of every individual across different domains. Individuals can contribute based on their strengths, experiences and passions. The contribution could range from trivial questions that could be answered easily by a fellow member to complex questions that involve substantial research. Complex questions have the potential to stir up interest in other users who might be interested in the same domain, and each can contribute with the limited knowledge they have.

## III. COMPONENTS OF K-COMM

### *A. Profile Gathering*

K-Comm captures the user profile which allows the system to personalize each user's page according to his/her needs and interest. Profile gathering also allows the system to identify a user's expertise. This could further allow us to analyze content and help us process the information on the system. Some of the information processing tasks include information organization (where we classify questions into the most probable category) or resolving contextual ambiguity of content. E.g. by capturing the list of categories in which a user is interested, and based on the searching history of that user, we could list out the interesting questions of potential interest to the user. We could also make use of the domain of interest of a user to automatically classify a question asked by a user into the most probable category. There are two ways of capturing the user profile – explicitly (Fig. 1) or implicitly.


Fig.1 Profile gathering during registration

User profile is captured explicitly during the registration process (See Fig.1) or when the user edits his/her profile information. From Fig.1, we see that a user can specify his/her interests during registration. The registration process also captures additional information. This installation of K-Comm has been deployed under a university setting, where the additional information captured is user type (student, staff or alumni member) or the faculty the user is associated with. However, the platform allows the administrator to customize the registration page by adding fields based on the needs of the organization where K-Comm is deployed. E.g. in a company setting, faculty could be replaced by department. The user type could be based on the different roles employees adopt in a particular organization. The platform is extensible and flexible. It provides an Application Programming Interface (API) for developers to program fields specific to the needs of a specific setting. These fields are then packed into a plug-in and added to the platform. Administrators can then add these fields to the registration page without touching any code. E.g. there could be a "school pack" that includes a dropdown list of the available courses offered by the school.

On the other hand, a user profile can also be captured implicitly. The system attempts to deduce some aspect of users like their interest and expertise based on their participation. This could be in the form of asking or answering questions, searching for answers, or writing blog entries or articles. E.g. if a user always answers a particular category of questions, the system will be able to deduce that s/he is an expert in that domain. K-Comm then comes up with a list of skill points for each category which varies over time. We will be incorporating a disclaimer where we seek the user's consent for implicit profile gathering when the user first registers into the system. This is to avoid any possible privacy issues.

### *B. Blogs*

K-Comm also offers blogging facilities. Users can keep their own personal blogs or make their blog entries publicly available to other users or their friends. Each blog entry can be tagged using keywords which defines the blog post. Research has shown that tags can help in recall and support searching [13]. Other than facilitating in the profile gathering process, they also provide rich resources for the platform to acquire new knowledge.

We are currently looking into automatic suggestion of tags for blog entries. This allows researchers in the area of text classification to test their classification techniques. Since users can modify or remove these suggested tags, this automatic suggestion provides relevance feedback to the system. E.g. when a user removes a tag suggested by our system, it implies that our system has recommended a wrong tag. Utilizing machine learning techniques, the system can then come up with more accurate classification based on past results. We recognize that suggesting tags is important because it unknowingly encourages users to tag their blog entries.

From the point of view of a researcher in information retrieval or information classification, this tagging implicitly provides us with human annotated training data. Conventionally, people are hired to annotate training data. Since involving people is always an expensive proposition, the platform provides a cheaper and less tedious solution to the problem of annotating data.

### *C. Groups*

Users can also create groups in K-Comm. Groups allow users to participate in discussion, post messages or share files. The system also allows members to connect to a group of members who are interested in a particular topic. Other than public groups that are available to everybody, groups can also be private, where a group owner has to invite other members to join. In private groups, only members of the group can access the content in the group.

In our current implementation in a university setting, we have identified that every batch of students taking a course usually has the same set of queries about the content of the course. Thus, lecturers have to answer similar questions again for each batch. To address this issue in the implementation based on our university, we have created a group for each of the courses. Fig. 2 shows the snapshot of a group for a particular course CS2103 on software engineering. During the registration process, we capture the list of courses taken by a student. The system ensures that students by default are added into the corresponding groups of the courses they have taken. This allows juniors to find seniors who have previously taken the same course(s), allowing them to connect with users who are knowledgeable in that topic.

### *D. Question and Answer (Q&A)*

The key component that K-Comm offers is a question and answer interface (Fig. 3). This component allows users to ask and answer questions. Questions are categorized under different categories or sub-categories. Currently, the category of a question has to be explicitly provided by the users.

Apart from category, questions can also be tagged in a similar manner to blogs to further provide any additional keywords. As seen from Fig. 3, our system also generates the

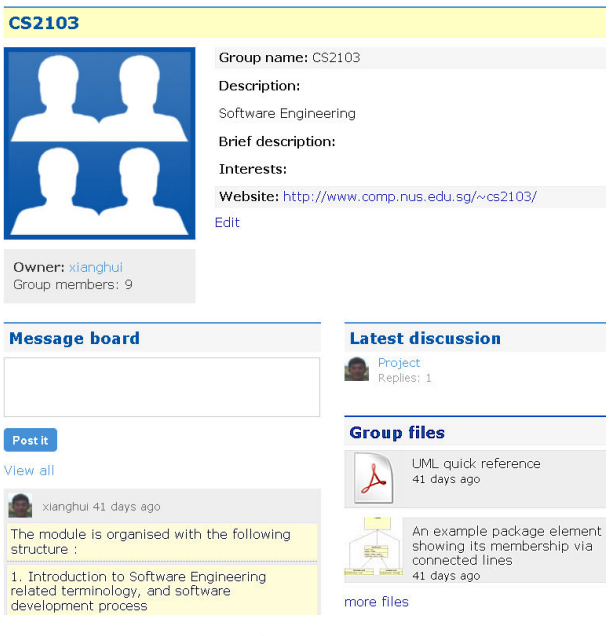


Fig.2 Groups

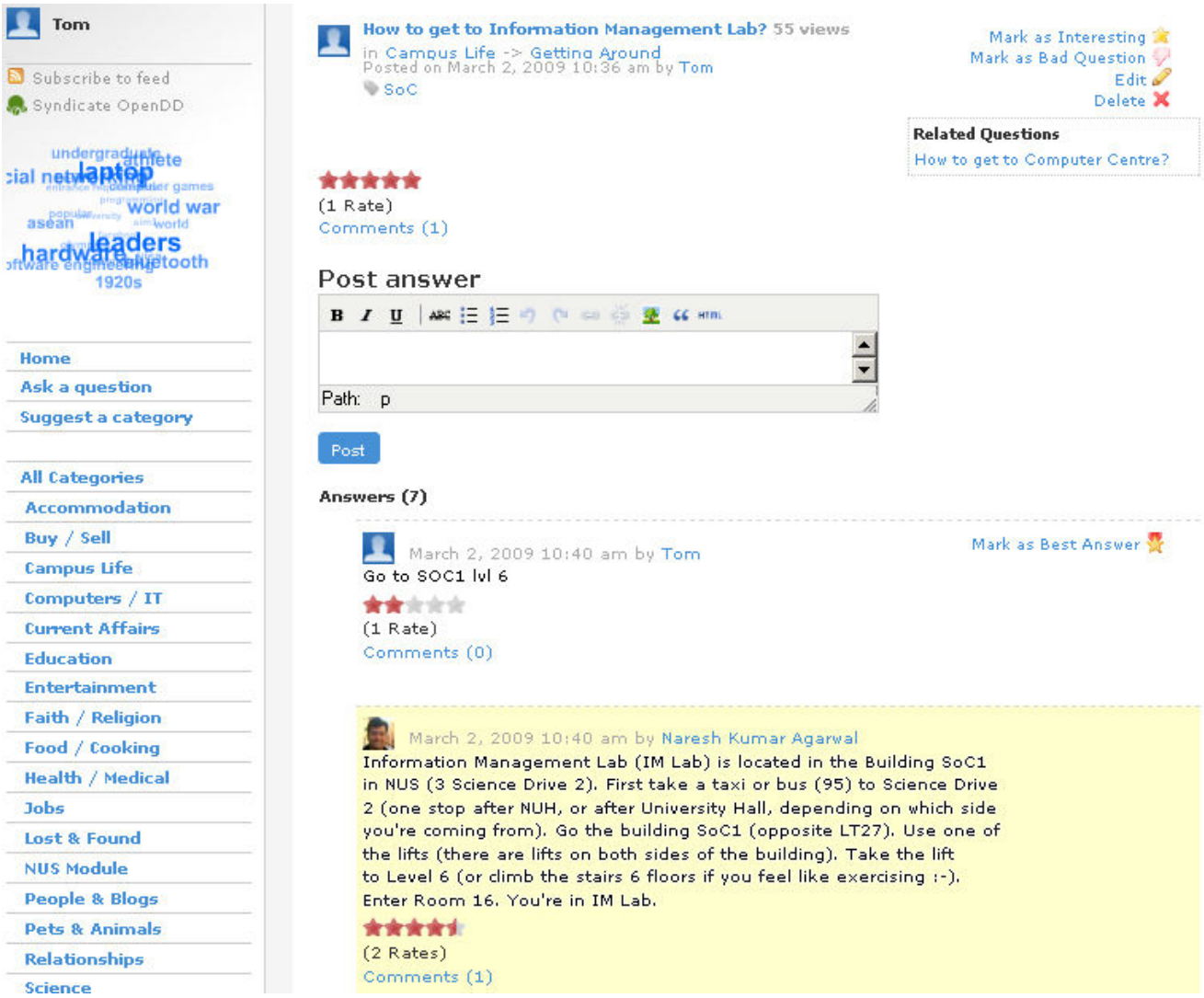


Fig.3 Question and Answer

list of tags in the form of a tag cloud on the top left column to show the list of interesting tags.

Apart from category, questions can also be tagged in a similar manner to blogs to further provide any additional keywords. As seen from Fig. 3, our system also generates the list of tags in the form of a tag cloud on the top left column to show the list of interesting tags.

Each question or answer can be rated by other members. This helps to verify the creditability and correctness of a question or answer. Users can also mark a question as a bad question if it contains inappropriate content. The person who asks a question can also choose a best answer which s/he feels is the most correct in his/her context. This further demonstrates that the content is not just audited by an administrator

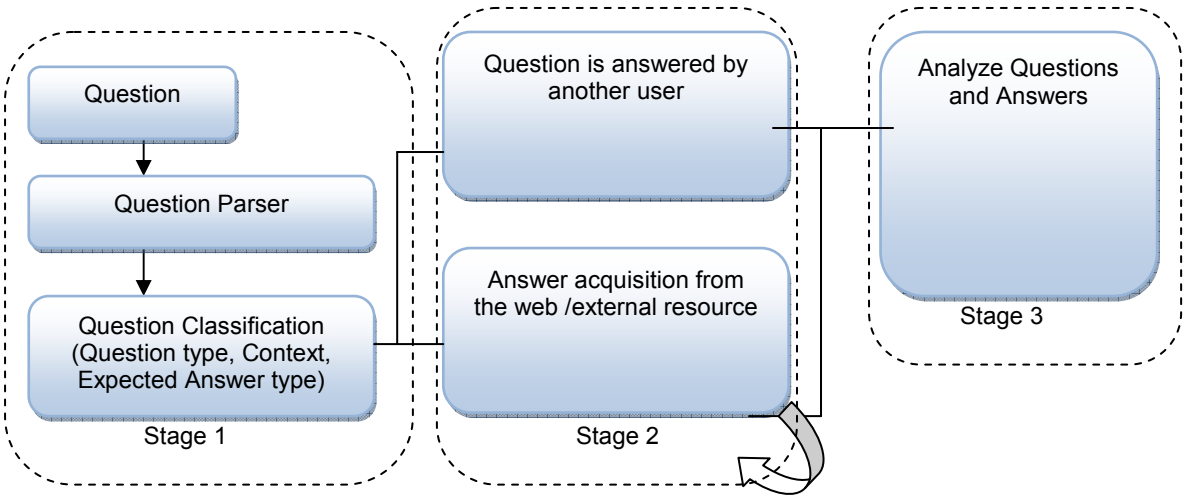


Fig.4 Stages of the Q&A process

or moderator. Instead, content moderation is a community effort involving every individual, as befits the principles of a digital ecosystem and collective intelligence.

Based on the tagging information or semantic meaning of the question, our system also attempts to recommend related questions that a user might be interested in. Since the system is able to discover the areas of expertise of each user and build a directory of experts, we are also able to direct a question to another expert in that domain.

We are currently working on analyzing the content and building an intelligent system that will:

1) Identify whether an identical or similar question has been asked before.
2) List questions of interests to the users based on the information in the user profile
3) Perform automatic categorization and tagging of questions
4) Represent answers in multiple formats e.g. thread-based, taxonomy or mind map
5) Perform automatic question answering

To tackle 5) i.e. automatic question answering, we have divided the Q&A process into 3 main stages (Fig. 4). In the first stage, when a user asks a question, the system will parse the question and based on other information like tags and categories as provided by the user, determine: 1) the question type, 2) the context of the question and 3) the expected answer type. In the second stage, the question could be answered by another user. Alternatively, the framework may also attempt to answer the question automatically by sourcing for the answer from the World Wide Web or other external sources. Users may provide relevance feedbacks to the answers generated. They can also rate the answers provided by other users. Using these relevance feedbacks and rating, the system is able to adapt over time providing better accuracy in future task which ensures sustainability. In the third stage, the question and the provided answers will be analyzed to produce useful data as part of the learning process. E.g. some pattern templates can be generated.

### *E. Incentive mechanism for user participation*

We have also built into the system a preliminary idea of an incentive system where points are given to users for participation like inviting a friend, posting of questions or answers and contributing answers that are rated as the best answers. In a university setting, these points could be used by lecturers to award students and to keep track of participation.

## IV. DISCUSSION

K-Comm utilizes the wisdom of the crowds [6]. It provides an environment where any individual can be an expert based on his/her experience, interest or passion in different domains, and can contribute to help one another. Table 1 gives a comparison of K-Comm with other technologies. The platform is focused on knowledge sharing and contribution. The whole process can be seen as a community effort not only from the user's perspective but also from a researcher's perspective. Researchers of information retrieval and extraction can mine rich resources from K-Comm. They can also experiment with different information processing techniques on K-Comm and obtain implicit relevance feedback from users. The existing content on K-Comm can be seen as training data for future information processing tasks. Thus, we can build a model and utilize machine learning techniques to better serve users' needs e.g. providing more accurate suggestions. This is what is suggested in [14], where sharing becomes mutually reinforcing, and helps in shaping the tools of the future. Researchers come up with better techniques to improve the system which also serves the needs of users. On the other hand, users also contribute, which in turn provides more data for researchers to work on, and they help one another mutually, which promised sustainability. This is what the idea of a digital ecosystem is all about. An advantage of

| | **Blogs** | **Forums** | **QA sites** | **SocialNWsites** | **K-Comm** |
|---|---|---|---|---|---|
| **Communication Architecture** | Users write about their thoughts | Communication generally discussion-based where multiple users come together and discuss about a topic rather than just commenting about what others say | Question and Answer approach where someone asks a question and people try to answer | Depending of different sites, can be interactive, message-based or posting made on friend's page | Communication can be Interactive where one talks to an expert or friend |
| | Even though there are comments/ratings, mode of communication is mostly single directional | | Can be interactive if allowed to connect to experts | | |
| **Organization** | Tags, Categories | Topics, Postings | Categories | Communities | Tags, Categories, Communities |
| **Focus** | Personal Opinion | Topic-Oriented | Question and answering | Human relationship | Knowledge Sharing and contribution |

Table 1 Comparison of K-Comm with other technologies

this platform over specialized software like ontology learning software is that any individual can participate. Participants do not need to be professional users who are very knowledgeable in one particular domain.

Different knowledge communities can also co-exist at the same time and interact with one another, aggregating their knowledge capability. There could be one K-Comm instantiation at school A, another instantiation at school B and another at a company C. The knowledge of each community could benefit members of other communities. Furthermore, each knowledge community can be a public community or a private community. Public community would mean that anyone on the World Wide Web can register to join the system. On the other hand, private community would be part of an enclosed scenario where accounts are pre-registered in advance and given to users.

## V. Conclusion and Future work

Digital ecosystems help transcend the traditional rigorously defined collaborative environments into a self-organised, interactive environment which offers cost-effective digital services and value-creating activities that attract human, organisational and software agents that participate in and benefit from it [2]. The basic challenge for collective intelligence is looking for ways and means whereby people and computers can be connected so that–collectively–they act more intelligently than any individuals, groups or computers have ever done before [7]. We have presented the design and development of a platform that provides for a digital ecosystem that helps meet the challenge of collective intelligence. This platform identifies the expertise of individuals and maximizes their knowledge potential, allowing them to contribute based on their experience, passion and interest. There is mutual involvement from both the end-users and the researchers. End-users can seek and contribute knowledge and help one another. Researchers, on the other hand, can mine useful information which can be helpful in research in the area of information retrieval and information extraction. The whole process is driven by a community effort.

In the current implementation of this platform, we have finished developing the basic skeleton of K-Comm. We are now researching into techniques to analyze the content and extract useful content from it. This involves research in Information Extraction and also requires learning about ontologies.

Future work will look into ways and means by which different communities can communicate with one another. This will involve having an interface layer that allows one K-Comm system to talk to another K-Comm system seamlessly – a scenario where the collective intelligence of one community benefits from the collective intelligence of another community.

## References


[1] F. Nachira, *The Information Resource about the European approach on Digital Ecosystems*, Directorate General Information Society and Media of the European Commission. Retrieved from http://www.digital-ecosystems.org/ on 25 March 2009.

[2] "Background and Objectives", *IEEE DEST 2009*. Retrieved from http://dest2009.debii.curtin.edu.au/ on 25 March 2009

[3] Collective Intelligence. Retrieved from http://en.wikipedia.org/wiki/Collective_intelligence on 16 Mar 2009.

[4] J. Howe, "The Rise of Crowdsourcing", *Wired*. Retrieved from http://www.wired.com/wired/archive/14.06/crowds.html on 26 Feb 2009.

[5] H. Rheingold, *Smart Mobs: The Next Social Revolution*, Perseus Books, 2002.

[6] J. Surowiecki, *The Wisdom of Crowds: Why the Many Are Smarter Than the Few and How Collective Wisdom Shapes Business Economies, Societies and Nations,* Little Brown, 2005.

[7] MIT Center for Collective Intelligence. Retrieved from http://cci.mit.edu/ on 25 March 2009.

[8] Web 2.0. Retrieved from http://en.wikipedia.org/wiki/Web_2.0 on 27 Feb 2009.

[9] P. Samuelson, "IBM's pragmatic embrace of open source," *Comm. ACM*, vol. 49, iss. 10, pp. 21-25, 2006.

[10] K. Lerman, "Dynamics of collaborative document rating systems," in *Proc. WebKDD/SNA-KDD '07: Proceedings of the 9th WebKDD and 1st SNA-KDD 2007 workshop on Web mining and social network analysis*, San Jose, California, 2007, pp. 46-55.

[11] N. K. Agarwal and D. C. C. Poo, "Capturing tacit knowledge across different domains: Knowledge Community (K-Comm)," *Int. J. Bus. Inf. Sys.,* vol. 3, iss. 6, pp. 668-685, 2008.

[12] N.K. Agarwal, D.C.C. Poo and K.H. Tan (2007), "Impediments to sharing knowledge outside the school: Lessons learnt from the development of a taxonomic e-learning portal" *Proceedings of the International Conference on Information Systems (ICIS 2007)* (Montreal, Canada, Dec 9-12, 2007).

[13] C. Marlow, M. Naaman, D. Boyd, and M. Davis, "HT06, tagging paper, taxonomy, Flickr, academic article, to read," in *Proc. HYPERTEXT '06: Proceedings of the 17th conference on Hypertext and hypermedia*, Odense, Denmark, 2006, pp. 31-40.

[14] A. Weiss, "The power of collective intelligence," *netWorker*, vol. 9, iss. 3, pp. 16-23, 2005.