Harnessing P2P Power in the Classroom

Julita Vassileva

Department of Computer Science, MADMUC Lab
University of Saskatchewan, 1C101 Engineering Bldg,
57 Campus Drive, Saskatoon, S7N 5A9 CANADA
jiv@cs.usask.ca
http://www.cs.usask.ca/faculty/julita

Abstract. We have implemented a novel peer-to-peer based environment called Comtella, which allows students to contribute and share class-related resources. This system has been implemented successfully in a fourth year undergraduate course on Ethics and IT. The intermediate results of the ongoing experiment show a significant increase in participation and contribution in the test version in comparison with a previous offering of the class where students contributed class-related resources via their own personal web-sites. Our ongoing and future work focuses on motivating higher levels of participation and getting students involved in the quality control of the contributions aiming at a self-organizing virtual learning community.

1 Introduction

The Computer Science Department at the University of Saskatchewan offers a fourth year undergraduate class called "Ethics and Information Technologies" which discusses broader social issues related to the development and deployment of IT in the real world, including privacy, freedom of speech, intellectual property, computer crime and security, workplace issues and professionalism. One week of classes and discussion is dedicated to each of these themes. Much of the content of the class involves legal cases, which are often ongoing and not yet resolved. The class has a strong communication component, involving class discussions and requiring students to summarize popular magazine articles related to the issues. For the class discussion, it is important to read and analyze the different viewpoints presented in the media and it is impossible to rely only on the textbook as a source of up-to-date cases and controversial viewpoints in discussion. Most newspapers and magazines, professional and public organizations have websites that represent their viewpoints on current controversial issues and cases. News about the development of lawsuits, new stories related to the use of IT appear constantly. Therefore, the Internet is an ideal resource of readings related to the class and can be used to create a repository for students' use. However, keeping current in the stream of news on such a wide variety of topics and locating appropriate resources for each week is an overwhelming task for the instructor.

Therefore, one of the class activities involves the students in the process of creating and maintaining of such a repository. Students are required to find on a weekly basis web-links to articles related to the issues discussed during the week and post them on their personal websites dedicated to the class. The instructor reviews these websites and selects from them several links to post on the class website. The students need to write a one-page summary - discussion for one of these selected articles.

The process described above is quite laborious both for the students and for the instructor. The students need to create and maintain personal class-websites on which to post the links they find. The instructor needs to review frequently differently organized student websites, to see which students have found links to new articles, to read and evaluate the articles and add selected good papers to the official class website where the students can pick an article to summarize. This process takes time and usually can be done only in the end of the week, therefore the students can only write summaries for the articles on the topic discussed during the previous week, which makes it impossible to focus all activities of the students to the currently discussed topic. Another disadvantage of this process is that the selected by the instructor articles posted on the class website reflect the instructor's subjective interests in the area; the students may prefer to summarize different articles than those selected by the instructor.

The process of sharing class related articles, selection of articles and summarization can be supported much better by using a peer-to-peer (P2P) file-sharing technology. Therefore we decided to deploy in the "Ethics and IT" class Comtella, a P2P system developed at the MADMUC lab of the Computer Science Department for sharing academic papers among researchers in a group, lab or department. The next section introduces briefly the area of P2P file-sharing. Section 3 describes the Comtella system. Section 4 explains how Comtella was applied to support the Ethics and IT class. Section 5 presents the first encouraging evaluation results.

2   P2P files-sharing systems

Peer to Peer (P2P) file-sharing systems have been around for 5 years and have enjoyed enormous popularity as free tools for downloading music (.mp3) files and movies. They have also gained a lot of public attention due to the controversial lawsuit that the RIAA launched against Napster and the ensuing on-going public debate about copyright protection. RIAA initially claimed that P2P technologies are used mainly to violate copyright and argued unsuccessfully for banning them. It succeeded in closing Napster, which used a centralized index of the files shared by all participants to facilitate the search. However, the widely publicized decision spurred a bunch of new entirely distributed and anonymous file-sharing applications relying on protocols such as Gnutella or FreeNet, which make it very hard to identify and prosecute file-sharers. Up to now, with the exceptions of P2P applications aimed at sharing CPU cycles (e.g. SETI@home which harnesses the CPU power of the participating peers' PCs to compute data from telescopes to search for signs of extraterrestrial intelligence and several projects like the Intel Philanthropic Peer-to-Peer project using P2P technology to

harness computer power for medical research), instant messaging applications like Jabber and AVAKI, and collaboration applications like Groove, the most widely used P2P applications are used for illegal file-sharing (e.g. KaZaA, BearShare, E-Donkey) of copyrighted music, films, or pornographic materials. Most recently, there have been initiatives to put P2P file-sharing for better use, e.g. MS Sharepoint or Nullsoft's Waste which serves a small private network of friends.

We see a huge potential for P2P file-sharing to tap the individual efforts of instructors, teaching assistants and learners in creating and sharing learning materials. These materials can be specially developed instructional modules or learning objects as in EDUTELLA (Neidl et al, 2002), or in the Ternier, Duval & Neven's (2001) proposal for a P2P based learning object repository. However, any kind of files can be shared in a P2P way, including PowerPoint files presenting lecture notes, web-based notes or references, research papers (used as teaching materials in graduate classes or during graduate student supervision). We propose a P2P system enabling learners to bring in and share course-related materials, called Comtella. The system is described in the next section. Section 4 presents results of an ongoing experiment with the system in the Ethics and IT course and compares the amount of student contributions using Comtella with the contributions of students taking the same class in the previous year, using their own websites to post links to the resources found.

3   The Comtella system

The Comtella system (Vassileva, 2002) was developed at the MADMUC lab at the Computer Science Department to support graduate students in the laboratory to share research papers found on-line. Comtella uses an extension of the Gnutella protocol, and is fully distributed. Each user needs to download a client application (called "servent") which allows sharing new papers with the community (typically, pdf files, but it can be easily extended for all kinds of files) and searching for papers shared by oneself and by the other users. The shared papers need to be annotated with respect to their content as belonging to a subject category (an adapted subset of the ACM subject index). The user searches by specifying a category and receives a list of all own papers and papers shared by others related to this category. From the list of results, the user can download the desired papers and view them in a browser.

While the research papers shared by users are not necessarily their own papers, but written by other authors, there is a copyright issue. However, these papers are typically found on the Web anyway (Comtella supports the user to seamlessly save and share pdf files that are viewed in the browser as that are typically found on the Web using search tools, such as Google or CiteSeer). Storing a local copy of the paper may be considered as a violation of copyright. However, users typically store local copies of viewed papers for personal use anyway, since they can not rely that they will find the file if they search again later (the average lifetime of a document on the web is approximately three months). Saving a copy for backup purpose is generally considered fair use. The sharing of papers happens on a small scale, among people interested in the same area within a research group or department, typically 5-10 people. Lending a book to a friend or colleague is normally considered fair use and in aca-

demic environment supervisors and graduate students typically share articles both electronically or printed on paper. Therefore, we believe that this type of sharing can not be considered copyright violation, since it has educational use and stimulates the flow of ideas, research information and assists the generation of new ideas.

In addition to facilitating the process of sharing papers, Comtella supports the development of a shared group repository of resources, by synergizing the efforts of all participating users. It allows users to rate the papers they share and add comments, which can yield a global ranking of the papers with respect to their quality and/or popularity within the group. Thus an additional source of information is generated automatically, that can be very useful for newcomers to the lab (e.g. new students) to get initial orientation in the assembled paper repository.

Comtella has been used on experimental basis with some interruptions and various successes for nearly one year in the MADMUC lab and for about three months across the Computer Science Department. We identified a number of technical issues related to instability of servents caused by Java-related memory leaks, and communicating across firewalls (so that the users could use the system from home), which have been mostly resolved. There were logistics issues, related to the fact that the system was fully distributed: a user who wanted to use the system both from home and from the office had to always leave his/her servents running on both machines, so that s/he can access from work his/her own papers shared by the servent at home and to access from home the papers shared at the work computer. In fact, Comtella considers the user's servents at home and at work as servents of two different users, with different ids, lists of shared papers etc.

In order to access the papers shared by another user, the other user has to be online. This proved to be a problem, because users typically switch off their home computers when they are at work. In addition, the users tend to start their servents only when they want to search for papers and to quit it afterwards. This leads to very few servents being on-line simultaneously, and therefore there are very few (if any) results to a query. It is very important to ensure a critical mass of on-line servents to maintain an infrastructure that guarantees successful searches and attracts more users to keep on-line their servents. Various solutions have been deployed in the popular file-sharing systems like KaZaA and LimeWire, for example, switching off the servent can be made particularly hard.

Finally, even when most of the users keep their servents running all the time, the system quickly reaches a "saturation" point, when all users download all the files in which they are interested from other users during their first successful searches. If there are no new resources injected into the system (by users bringing in and sharing new papers), very soon it makes no sense for a user to search in his/her main area of interest since there is nothing new. Ultimately, the system reaches equilibrium where everyone has all papers that everyone else has. In order to achieve a dynamic and useful system, the users have to share new papers regularly and thus contribute to the growing repository rather to behave as lurkers (Nonnecke & Preece, 2000). Motivating users to contribute is an important problem and we have researched a motivational strategy based on rewards in terms of quality of service and community visualization of contributions (Bretzke & Vassileva, 2003).

Since Comtella provides exactly the infrastructure allowing users to bring in and share resources with each other, we decided to deploy it in the "Ethics in IT" course to support students in sharing and rating on-line papers related to the topics of the course. We expected to have a higher participation and contribution rate than in the case where Comtella is used to share research papers within a lab, since within a class, the students are required to put a concerted effort scheduled by the class curriculum (weekly topics) to summarize papers and to contribute new papers to get a participation mark. We wanted to see also how the contribution level when students use Comtella will differ from the level in the previous offering of the class, when students had to add the links to their own class website. Finally, we wanted to experiment with some of our motivational strategies to see if they actually lead to increase in participation compared to a system with no motivational strategies.

4   Applying Comtella in the Ethics in IT Course

The first problem that had to be dealt with was ensuring a critical number of servents running at any given time, so that queries were guaranteed to yield some results. Our previous experience with peer-help systems shows that otherwise the users are unlikely to query again, and the usage of the system quickly winds down to zero. However, while in a research lab graduate students tend to leave their office computers running permanently, it is unrealistic to expect that undergraduate students will have a permanently running computer at home where they can keep running their Comtella client. Therefore, we had to make a compromise with the distributed P2P architecture, and move all servents to two server machines. We split the user interface (UI) part of the servent from the backend part (BE) that processes and forwards queries, pings and pongs to other peers and maintains the list of files shared by the user. The UI part is a jar file that students can download and run on their local machine to log into their own BE, which runs constantly on the server. Thus all queries by other users are served by the BE which is available all the time, even when the user is not on-line. Users log into their BE only when they need to search for papers or to share new papers. In this way, we also restricted the access to allow only class members and imposed unambiguous user identification through username and password. This is important since we need to keep track of users' contribution to reward it fairly with participation marks.

The next change to the Comtella architecture was necessitated by two factors:

- shared files by the servent of each user occupy a lot of disk space. Since now the BE on a central server, the disk space required may exceed the availability and become prohibitive.
- the files shared for the class are mostly from web-magazines and their format is varied. Some are html, xml, xhtml, some contain flash. Saving a copy of the file on disk is not trivial and depends on the browser and the settings. Sometimes a file can be associated with a directory of images, ad banners etc.

For these two reasons, we decided to modify the standard Gnutella servent functionality and instead of sharing the actual files, only their URLs are shared. To share a paper from the Web, the user needs to copy and paste the URL into the Comtella

"Share" window, to copy and paste the title of the article from the browser, and finally to select the category (topic) of the article, which is indicated by the week of the class it relates to (see Figure 1). The shared paper in this way consists of: title, URL, category, and optionally, rating and comment, if the user decides to provide them.
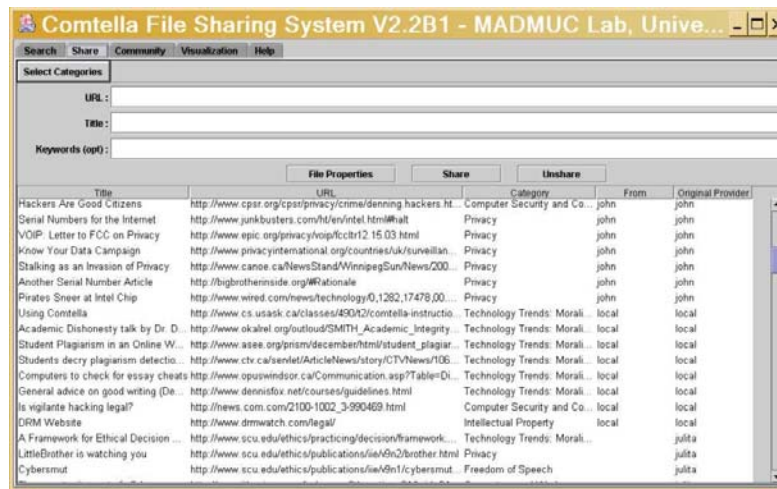

Fig. 1. Sharing new links in Comtella

Users who decide to search for papers related to the topic of a given week have to specify the topic from the category list in the "Search" window. The servent sends the query to its neighbour servents residing on the server and they forward the query to their neighbours. If any of the servents that receive the query share some papers about this topic, the results are sent back to the querying peer using the standard Gnutella protocol. In other words, the protocol for search is not changed; the only change is the physical location of the BE of the servents that reside now on two server machines.

Students can view and read the papers that were yielded as results of the search by clicking on the "Visit" button without actually downloading the paper (see Figure 2). Clicking "Visit" starts the default browser with the URL of the paper and the student can view the paper in the browser. The student can also view the comments of the user who shares the paper and his/her rating of the paper. If the student likes the paper and decides to share it him/herself, to comment on it or rate it, s/he can download it, by clicking on the "download" button. This initiates a download process between the servents (which follows again the standard Gnutella protocol). Rather than the actual paper, the title and URL are downloaded, while the comment and rating that the sharing user entered are not. In this way, each user who shares a paper has to provide his/her own comment and rating.
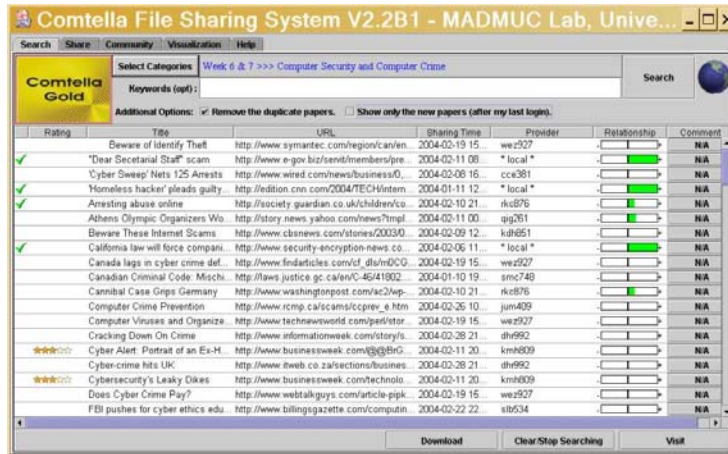
Fig. 2. Searching for papers about a given topic (category).

The ratings of the papers indicate the quality of the paper and the level of interest the students who downloaded the paper have in the topic of the paper. The students were instructed to select for their weekly summary a paper that was rated highly by two or more students of those who share the paper. The students could enter their weekly summary through Comtella too, by entering a summary for a selected paper from their shared papers.

If two students disagreed in their rating of a paper, their relationship strength decreases. The relationship between the student who does the search and each student who shares a paper is shown in the search results (see Figure 2). In this way, students can find other students who judge papers in a similar way, since the relationship value serves as a measure of trust that the student has in the papers provided by the other student.

Comtella became a focal point of all weekly course activities. The instructor did not need to find and share any new articles since the students provided an excessive number of materials, which were immediately accessible for anyone who wanted to search for papers with the same topic (category). It became also unnecessary for the instructor to review all contributed papers and select those appropriate to be summarized, since the ratings of the papers indicated the good papers and those in which the students were interested.

5. Evaluation

The deployment of Comtella in the Ethics and IT course is ongoing at the time of writing. The planned evaluation of the system includes data collected through system use (e.g. statistics on numbers of contributed papers, numbers of downloaded papers, ratings and summaries written, average time spent on line, frequency of logging in)

and data from student questionnaires. However, even though the experiment is only half-way through, comparing the average levels of student contributions during the same period of time (first 6 weeks) in the current offering and the last year's offering of the class shows evidence for the success of the system. In both cases the same instructor taught the class; the curriculum, scheduling of weekly themes and grading scheme were the same. We compare the first six weeks in the 2002/2003 offering of the class with the first six weeks of the 2003/2004 offering.

Table 1. Main characteristics of the two offerings.

| \ Class offering Parameters \	2002/03 - first 6 weeks (using personal websites)	2003/04 - first 6 weeks (using Comtella)
Number of students in class:	25	35
Total contributed links:	51	334
Average number of new links contributed per person (average from only those who contributed):	4.63	11.52
Number of students who didn't contribute at all:	14 or 56% of the class	6 or 17% of the class
Percentage of links contributed by the top five (5) students:	78% of all the contributed links	39% of all contributed links

Table 1 summarizes the student and participation data in each class. We can see that the average number of contributed new links per person in the 2003/2004 class where students used Comtella was nearly three times higher than in the 2002/2003 class. The bulk (nearly 80%) of contributions in the 2002/2003 class was done by five students, while in the 2003/2004 class the top five students contributed approximately 40% of the links and the contributions were spread more equally (see also Figure 3, which compares the distribution of contributions among the students in the two class offerings). We can see that 56% of the students in the 2002/2003 class did not contribute, versus only 17% in the 2003/04 class. Figure 4 shows how regularly students contributed over the course of the first six weeks of the experiment. As can be seen more students contributed regularly in the 2003/2004 class than in the 2002/2003 class.

One reason for these encouraging results is that it is much easier for the students to add new links in Comtella than to maintain a website and add links there. Another reason is that searching for relevant links with Comtella is much more convenient than visiting the websites of each student in the class, so the students tended to use the system much more often. They visited links shared by others in Comtella and when viewing these articles they found new relevant articles (since often Web-magazine articles have a sidebar field "Related Stories" or "Related Links") and shared them in Comtella "on the go".
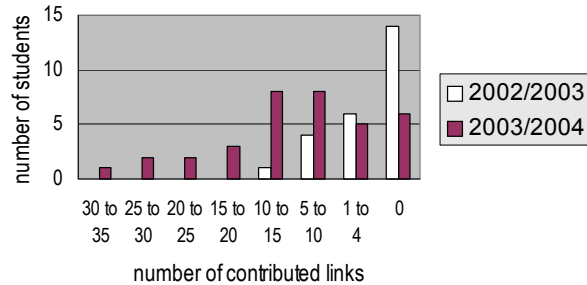
Fig. 3. Number of new contributions: comparing the first weeks of the two courses.
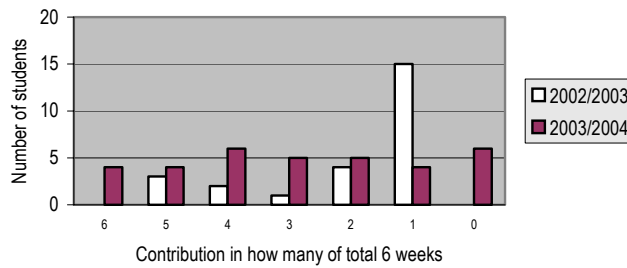
Fig. 4. Regularity of contributions.

While in the beginning the instructor had to evaluate the links added and to rate them, from the third week on the students started giving ratings to the articles themselves and the system became self-organized. Of course, monitoring is still necessary, since currently nothing prevents students from sharing links that are not related to the contents of the course or offensive materials. In our experiment, such an event has not happened to date, possibly, since the students are senior students in their last year before graduation. Yet, it would be good to incorporate tools that would allow the community of students to have a say on the postings of the colleagues and thus achieve an automatic quality control by the community of users, similar to Slashdot.

In the remaining six weeks of the course, we will experiment a three level "membership" in the Comtella community based on the level of contribution: bronze, silver and gold, that will give certain privileges for members who have contributed on regular basis papers that have been downloaded and rated highly by other students. This newer version contains also a visualization of the community showing the contribution level of each individual member, information about whether s/he is on-line at the moment in line with the motivation visualization described in (Bretzke & Vassileva,

2003). The goal is to create a feeling of community (Smith & Kollock, 1999, De Souza & Preece, 2004) and a competition among the students to find more and better links.

6. Conclusions

We have implemented a novel peer-to-peer based environment called Comtella, which allows students to contribute and share class-related resources. This system has been implemented successfully in a fourth year university course. The intermediate results of the ongoing experiment show a significant increase in participation and contribution in the test version in comparison with a previous offering of the class where students contributed class-related resources via their own personal web-sites. We believe that the system can be applied to support a wide range of courses requiring intensive reading of on-line resources, e.g. in the humanities, or even programming courses where code-examples can be shared. A similar functionality could have been implemented more efficiently using a centralized repository and search engine, without the natural duplication of resources implied by Gnutella. However, we did not want to sacrifice the advantages of logical decentralization offered by a P2P architecture, since they allows to move any of the servents freely around depending on the load on the servers. If the users were sharing actual papers, storing them on a server may have quickly become too costly. The shared papers remain in the control of the user, which may be important for their motivation to contribute. Our ongoing and future work focuses on motivating higher levels of participation and student involvement in quality control of the contributions aiming at a self-organizing virtual learning community.

References

1. Bretzke H., Vassileva J.: Motivating Cooperation in Peer to Peer Networks, Proceedings User Modeling UM03, Johnstown, PA, Lecture Notes in Computer Science, Vol. 2702. Springer-Verlag, Berlin Heidelberg New York, 218-227 (2003).
2. Comtella © 2002-2004 available from http://bistrica.usask.ca/madmuc/peer-motivation.htm
3. De Souza, C., Preece, J. A framework for analyzing and understanding online communities. Interacting with Computers. 16 (3), 579-610 (2004).
4. Nejdl, W., Wolf B. et al.: EDUTELLA : A P2P Networking Infrastructure Based on RDF, WWW2002, May 7 -11, Honolulu, Hawaii, USA (2002).
5. Nonnecke, B., Preece, J., Lurker Demographics: Counting the Silent. Proceedings of ACM CHI'2000, Hague, The Netherlands, 73–80 (2000).
6. Smith, M.A. and Kollock, P., 1999. Communities in Cyberspace. , Routledge, London.
7. Ternier, S., Duval, E., and Vandepitte P. LOMster: Peer-to-Peer Learning Object Metadata. Proceedings of EdMedia-2002, AACE: Blacksburg, 1942-1943 (2002).
8. Vassileva J.: Supporting Peer-to-Peer User Communities, in R. Meersman, Z. Tari et al. (Eds.) Proc. CoopIS, DOA, and ODBASE, LNCS 2519, Springer: Berlin, 230-247 (2002).