A Digital Repository Supporting Collaboration

Keith Irwin, Darina Dicheva, and Christo Dichev

Winston-Salem State University, Computer Science Department
601 Martin Luther King Jr. Drive, Winston-Salem, NC 27110, USA
{irwinke, dichevad, dichevc}@wssu.edu

Abstract. Today's digital learning repositories are focused on the process of sharing materials created elsewhere. We propose a new type of digital learning repository which is focused on on-line collaboration. This new type of repository will foster an active community which creates and shares resources. It will use a mix of existing and new technologies to encourage users to communicate, contribute, and collaborate. It will be focused on the needs of instructors and be tolerant of differences of opinion.

Keywords: collaborative authoring, learning repositories, web 2.0, semantic web, social networking

1 Introduction

Over the past few years, a number of digital repositories (e.g. MERLOT[1], LORNET[2], ARIADNE[3], CAREO[4]), have been developed to encourage the reuse of learning material. Their primary purpose is to allow instructors to share materials with one another. This is certainly a worthwhile goal, but it is also somewhat limited. Although the internet has become a powerful vehicle for sharing digital resources, it is not limited to just that. Recently the internet has increasingly been used for enabling collaborative development of digital resources. For example, developers collaborating on open source projects to create new software have become commonplace. Sites such as Wikipedia have allowed for the collaborative creation of information resources. We wish to enable people to use the internet to collaboratively develop courseware.

Users who wish to begin new collaborative software projects can use sites such as SourceForge[5], Google Code[6], and many others, which will enable their collaboration by providing them with source code hosting, bug tracking software, mailing lists, and other tools. Likewise, those who wish to begin a new project to create shared documents can turn to sites such as Wikia[7] and PBWorks[8], which will allow them to create their own wiki.

However, there are currently no sites aimed at enabling collaborative projects to develop course materials. Existing digital learning repositories are focused

[1] http://www.merlot.org
[2] http://www.lornet.org/
[3] http://www.ariadne.ac.uk/about/myth-new.html
[4] http://www.careo.org/
[5] http://sourceforge.net/projects/opencvlibrary/
[6] http://code.google.com/
[7] http://www.wikia.com/wiki/Wikia
[8] http://www.pbworks.com/

on sharing, not collaborating. Also, most of these repositories are not very active. Our examination shows that most of the contributions to existing digital repositories for computer science course materials correlate to times when stipends were offered in exchange for contributions. They have not successfully created self-sustaining collaborative communities, which create or share new materials.

There are several factors contributing to this limited success.

- Adaptation of learning resources (to fit a particular instructional goal) and resulting versions are typically not supported. This affects collaboration among content authors, who do not necessarily have to adhere to the same resource.
- Typically authoring tools are not integrated with the learning repositories where the content to be reused can be found (if even the intellectual property of the resources in the repository allows for adaptation or modification). This puts an additional burden on content authors to use several different tools in parallel while creating a new learning resource, which may reduce their creativity and productivity.
- Repositories are mainly used as centralized stores of learning resources and their metadata. However, today's Web 2.0 technologies demonstrate that a significant content sharing can be achieved by participating in a community and by leveraging social relations with peers (i.e., social networks of content authors).

The centralized model adopted by the most repositories contrasts the web's strongly decentralized nature. Repositories such as MERLOT, LORNET and CAREO are susceptible to the traditional centralized management problems that arise in distributed computing such as limited scalability and monolithic access rights. Also, technical problems such as managing the history of the learning resources have been largely ignored by the educational technology community, but are becoming more important as sharing of learning resources increases.

By contrast, there are several open-source software or wiki hosting sites which have successful collaborative communities. However, they are not well suited to the needs of the educational community. Sites which target collaborative software development are very specialized to their niche. And although wikis can be useful in some situations, they are not tailored to the needs of academia. They tend to hide authorship making it difficult to build reputations, a necessity in academia, and they are not structured to deal with situations where there is disagreement about either subject matter, or instructional approach, or presentation.

Our goal is to create a site which will encourage instructors to collaborate on-line to develop course materials. This site must focus on collaboration and not just sharing. It must be specifically tailored to the development of course materials and to the needs of academia. In specific, we are building a site which uses ideas drawn from successful community sites to make it easier to identify and build communities, which uses ideas from distributed source code management to enable simple collaborative resource development, and which includes a novel automated system, which add attribution of work and objective measures of contribution to allow for contributions to lead to academic reputation

building. In a long term, efficient support for creating new derivative works can reduce the amount of duplication of resources.

In the long run, this site and its techniques should allow for the collaborative development of many kinds of course resources. However, because we must start somewhere, we have chosen to begin with presentation slides. We make this decision because slides are a vital part of most modern courses and because they avoid some of the tricky issues surrounding who should be allowed to access materials such as homework problems or project descriptions.

2 Increasing Incentives

The biggest problem of existing digital learning repositories is a shortage of contributors. So our most fundamental goal is to find ways to attract contributors. The most immediate related question is what motivation people have to contribute. For most repositories, the answer is that they have very little motivation. In some cases, stipends have been offered, but this is only a short-term solution, which is not sustainable in the long-run and does not create communities. When stipends are not offered instructors have only altruism as a motivation for sharing their materials. For some instructors altruism may be a sufficient motivator, but the low rates of contribution suggest that for most it is not enough, especially given that many instructors are pressed for time. So, it seems logical that we need to create a site which will increase the rewards for contribution.

Our approach offers several benefits to contributors which are not offered by other sites. The first is that we allow users to make and share sets of changes to presentations which the initial uploader can choose to adopt. As such, one possibility for an instructor is that they will share their slides and in return they will get back valuable changes such as fixes to typographical or other errors, new slides with more detail, improved or additional examples, or other such improvements. The possibility of improving course materials which are shared provides an incentive not found in sites which are read-only.

The second technique we are implementing is automatic attribution. Each slide will automatically have the names and percentage of contribution of anyone who has collaborated on the creation of that slide. This will help instructors build reputation. Further, we will ask users to indicate if they are using a set of slides in a course at their university. This information can be used to automatically generate a numeric index which indicates the total academic contribution of any particular instructor. Having accessible, objective metrics, which measure contributions, will increase people's motivation to contribute. Also, even altruistic contributors will be more likely to contribute if they can know how effective their contributions are.

The third technique is to support the creation of groups centered around topics in computer science. The presentations on the site will be organized into a hierarchy of computer science topics to help users find the presentations they desire. But beyond that, each topic will also be a group of its own with message boards, highlights of recent activity, and other community features. Although creating the tools for community does not guarantee that communities will form, it makes it more likely. Successful communities recognize and encourage the work of contributors through social means. If we are successful, creating

good course materials will lead to recognition in an instructor's professional circles the same way writing good papers does currently.

3 Collaborative Technologies

In the previous section we talked about ways in which we would increase people's incentives to share the material which they have already created. Although those incentives are important to creating communities of sharing, we also wish our communities to be collaborative. As such, the other focus of our site is to encourage on-line collaboration. On-line collaboration can often allow groups of people to create resources which they do not have the time or knowledge to create individually. The centralized model adopted by the most repositories contrasts the web's strongly decentralized nature.

Collaboration requires the use of some set of supporting tools. Although there are existing tools, such as wikis and source control systems, they are not well suited for the development of course materials. Source control tools are limited because they only operate well on text files, and they require some specialized knowledge which many instructors will not want to spend the time to learn. Wikis are easy to pick-up and use and allow for more complex formatting, but they have one large drawback: they require that all participants reach a consensus. When there are disagreements which cannot be settled by the participants, administrators must step in. Instead, our site will use a technology we call "divergent wikis".

A divergent wiki is a tool which combines aspects of a wiki with aspects of distributed source control technology. Specifically, our divergent wiki will allow simple on-line editing like a wiki, but it will also have the ability to maintain different versions of a presentation. Rather than having a single version, which everyone makes changes to, when a user wishes to make a change to a presentation, they will "fork" that presentation, creating their own duplicate. So, if Alice likes Bob's presentation about the semantic web, but thinks that it needs a slide about topic maps, then she can fork her own version. After she has created her own version, she can edit it how she pleases, in this case adding a new slide. Once her changes are complete, she can share that set of changes with the original creator of the presentation. Bob then has the option of applying her changes to his version or not. Whether he does or not, Alice can continue forward with her own version. In the future, other users can fork their own version from either Alice's version or from Bob's version.

If another user makes a change to a part of the presentation found in both Alice's and Bob's version, then both Alice and Bob will have the option of applying those new changes. Thus, the work done on one version of a presentation will also benefit other versions of the presentation. This will address one of the most fundamental problems of traditional wikis: it is difficult to collaborate successfully in the presence of differences in opinion, even when those differences are small. In our system, differences of opinion will simply be reflected in different versions.

We expect that there will be different versions of a presentation which have been specialized not only to differences of opinion, but also to different levels of student background knowledge (or other areas of student need) or even more mundane things like different styles of notation. However, so long as these ver-

sions maintain substantial similarity, changes to one can still be applied to the other. Thus adding a new example to one version of a presentation will allow the maintainers of other versions to use that example as well. This new tool will increase the efficiency of on-line collaboration for the creation of course materials especially when opinions or student needs make it impossible to have a single one-size-fits-all presentation.

In addition to making it easy to collaborate, we must also make it easy for users of the site to find course material which meets their needs. To aid in this we are leveraging semantic annotations, such as the category scheme mentioned previously and also tagging and rating. These will help users quickly find presentations suitable to their needs.

Another thing which we have observed is that many collaborative on-line software projects have a core group of developers who stay connected through mailing lists or other similar tools. In order to help groups like that form and thrive, we also target automatic discovery of implicit communities of interest from which we hope collaborative groups will spring. By examining users, their interests, their contributions, the tags they use, and other factors, we can recognize groups of common interest, even when they are not oriented around a single topic in the hierarchy. When we identify implicit communities we can connect different users together automatically. These communities should be valuable tools for making connections and finding collaborators.

4 Current Status of the Work

We have completed the design phase of the project and are in the implementation phase. We have designed the overall architecture of the software (Figure 1) and developed detailed user interface specifications. We have also designed and implemented the needed database schemas. Currently we are writing the needed code to create a functional web site. Where possible we are utilizing existing products, resources, and code to help us. For example, most of the presentation import/export work will be done by OpenOffice.org through the use of the API which they provide.

When the implementation is complete, we will gather users to test it both to refine our design and to try to collect data.

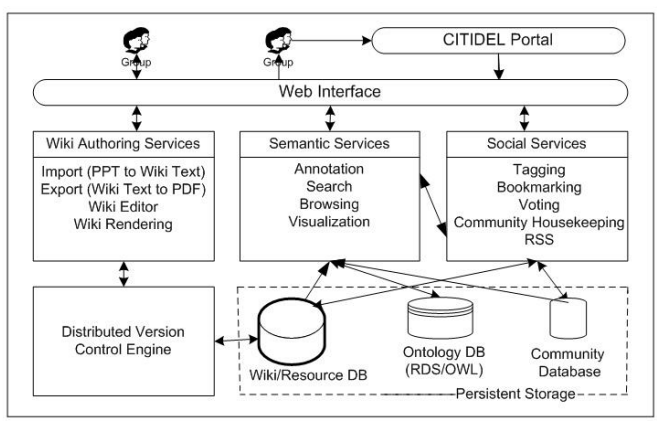

Figure 1

5 Related Work

A recent advance in the field of the traditional version control technology is the distributed version control. Distributed version control systems focus on a decentralized development model, in which each developer can have their own local repository. Distributed version control allows for the maintenance of multiple branched versions of a given resource and for the easy integration of sets of changes across multiple branches [1, 2, 3]. Recently, distributed version control systems such as Git [4], Mercurial [5], and Bazaar [6] have begun to be commonly used to manage large open source projects. From an ontology perspective, [7] offers an environment designed to facilitate the development and maintenance of ontologies among distributed stakeholders. Thus far, these products have only been used for managing source code, but we hope to demonstrate that their ideas can also be applied to collaborative courseware creation.

There is a growing movement towards open educational resources in education and science communities. Examples are MIT OpenCourseware initiative [8], Carnegie Mellon's Open Learning Initiative [9], etc. While attracting a good number of instructors and students, they propose "canned" courses, with specific course objectives and targeting a specific audience. It is known that the "one size fits all" approach does not work in education

Similarly, the Connexions project [10, 11] allows instructors to share learning content in a legal framework using a component approach. Connexions is designed as "The site" for authoring learning modules in any discipline, using proprietary tagging and internal representation of the modules. Our approach seeks to address the development of openly shared courseware adjustable to different course level and objectives, etc., by supporting multiple versions of resources and lawful compilation of content from different sources.

We are not aware of any work on learning repositories focused on online support for collaboration based on distributed version control. The concept "distributed" when used in the context of learning repositories is applied largely to distributed applications focused on search and retrieval of learning resources [12].

Recently the term Library 2.0 [13] has been coined to define how Web 2.0 will affect libraries. Although recent Web 2.0-based projects place focus on collaboration based on grass-root efforts, we are not aware of applications that address course content development and focus on the dynamic nature of knowledge in emerging disciplines. Wikis, including Wikibooks[9], the Wikimedia project with the mission to create a free collection of open-content textbooks that anyone can edit, have a goal very different from ours: they do not support multiple revisions of a resource and do not acknowledge authorship.

The question of what motivates or triggers individuals to join and participate in online communities and how to design the technical features of the community software accordingly can be viewed from a wide range of perspectives [14]. Preece & Maloney-Krichmar [15] identify research in social psychology, sociology, communication studies, computer-supported cooperative work and human-computer interaction as main areas which can help inform designers about how and why people interact in online communities. Social rewarding mechanisms have been introduced to reward people contributing to an online

[9] http://en.wikibooks.org/wiki/Main_Page

community. In the majority of cases, social rewarding is based on accentuation of the most active members. Hoisl et all [16] describe an approach to reward users actively contributing to a wiki. The technique of motivating users to contribute to the community by the means of reputation is described by Resnick [17].

6 Conclusion

Computer-supported collaborative knowledge creation is a continuous process of knowledge assets development, where its successful growth suggests ability for the reused and contributed knowledge to take different evolution paths. A parallel can be drawn between the development of open educational resources and open source software development. Although the requirements of educational resource development are of a different nature to that of software, they are nevertheless comparable in their degree of complexity. The design issues and principles, such as adherence to learning design specifications, levels of granularity, and separation of content from presentation, present challenges that are equally demanding from the development perspective. Since practices for collaborative and open developments are still in its infancy, it is likely that many of the points are not yet fully understood. In this respect, a side effect goal of this work is to contribute to the understanding of some mechanisms related to the open systems technology. For example, to what extend enabling users to reuse, collaborate, extend and re-contextualize for different purposes will improve their level of participation.

From practical perspective we propose a community-oriented course development environment grounded on an integral approach incorporating sharing, lawful content reuse, content evolution supported by distributed version control, and semantic structuring. While most of the characteristics of the model are not individually unique, their *combination* is a novel contribution to this area of research. The proposed distributed wiki model supporting independent branching and "cherry-pick" of single changes fosters sharing and collaboration and promotes lawful reuse of open-content learning resources.

References

1. Morris, R., Yip, E., and Chen, B. Pastwatch: A distributed version control system. 3rd Symposium on Networked Systems Design and Implementation (NSDI) (2006)
2. BitKeeper Website. http://www.bitkeeper.com/
3. Kapitza R., Baumann P. Reiser H.P.: Using Object Replication for Building a Dependable Version Control System. In Proc.of the 8th IFIP International Conference on Distributed Applications and Interoperable Systems - DAIS, Norway (2008) 86-99
4. Git Website. http://git.or.cz/
5. Mercurial Website. http://www.selenic.com/mercurial/wiki/
6. Bazaar Website. http://bazaar-vcs.org/
7. Redmond, T., Smith, M., Drummond, N. and Tudorache, T.: Managing change: An ontology version control system. In OWL: Experiences and Directions, 5th Intl. Workshop, OWLED 2008, Karlsruhe, Germany (2008)
8. MIT Open Courseware http://ocw.mit.edu/OcwWeb/Media-Arts-and-Sciences/MAS-863How-to-Make--Almost--AnythingFall2002/CourseHome/
9. Open Learning Initiative, Available at: http://www.cmu.edu/oli/

10. Baraniuk, R et al. Connexions Project. Avalable at http:Hcnx.rice.edu/
11. Henry G., Baraniuk R. G.: Peer to Peer Collaboration Using Connexions, Proceedings of the 2004 American Society for Engineering Education Annual Conference & Exposition, American Society for Engineering Education, Utah (2004)
12. Hatala, M. and Richards, G.: Making a SPLASH: A Heteregeneous Peer-Learning Object Repository . The Twelfth International World Wide Web Conference (WWW 2003), Association of Computing Machinery (ACM) (2003)
13. Stephens, M.: Web 2.0 and Libraries: Best Practices for Social Software. Library Technology Reports 42(4) (2006)
14. Webster, A.: Visible Relation in Online Communities: Modeling and Using Social Networks. M.Sc. Thesis. University of Saskatchewan (2007). Available at: http://library2.usask.ca/theses/available/etd-09192007-204935/
15. Preece, J. and Maloney-Krichmar, D.: Online Communities: Focusing on Sociability and Usability. Handbook of Human-Computer Interaction. Lawrence Erlbaum Associates, Inc. (2003) 596–620
16. Hoisl, B., Aigner,W. and Miksch, S. Social rewarding in wiki systems - motivating the community (2007) 362–371
17. Resnick, P., Zeckhauser, R., Friedman, E. and Kiwabara, K.: Reputation systems. Comm. ACM, 43(12) (2000) 45–48