Journal of Universal Computer Science, vol. 14, no. 10 (2008), 1717-1738
submitted: 1/10/08, accepted: 6/5/08, appeared: 28/5/08

Exposure and Support of Latent Social Networks among Learning Object Repository Users

Peng Han
(FernUniversität in Hagen, Germany
han.peng@fernuni-hagen.de)

Gerd Kortemeyer
(Michigan State University, USA
kortemey@msu.edu)

Bernd J. Krämer
(FernUniversität in Hagen, Germany
bernd.kraemer@fernuni-hagen.de)

Christine von Prümmer
(FernUniversität in Hagen, Germany
christine.vonpruemmer@fernuni-hagen.de)

Abstract: Although immense efforts have been invested in the construction of hundreds of learning object repositories, the degree of reuse of learning resources maintained in such repositories is still disappointingly low. As the reasons for this observation are not well understood, we carried out an empirical investigation with the objectives to identify recurring patterns in the retrieval and (re-) use of learning resources and to design and test social networking functionality supporting communities of practice. The outcomes of this project, which are reported here, aim to affect the design of a new generation of learning object repositories, like CampusContent, that tries to eliminate deficits of current repositories and involve recent contributions in the area of social software. Object of our investigation was LON-CAPA, a cross-institutional learning content management and assessment system used since 2000. We analyzed hundreds of thousands of log data collected over a period of three years and detected various kinds of latent relationships among LON-CAPA users, such as the co-occurrence of learning resources from independent authors in instructional materials. To understand the rationale behind these findings, we conducted a study with LON-CAPA users. One section of the questionnaire asked for people's opinion about the expected benefit of community support. Nearly 80% of the study participants said that the formation of communities of practice (CoP) would be an asset to LON-CAPA. More than 80% would be ready to provide their profiles for matching up with CoPs and serve the community by spending time on the evaluation of resources they had used. Finally we sketch a faceted search functionality we designed to support CoPs among LON-CAPA users. This functionality is currently tested with two CoPs.

Key Words: metadata, community building, sharing and reuse, structural computing, learning object, repository

Category: H.3.7, H.5.4

1 Introduction

The concept of learning objects (LO) arose in the late nineties driven by the motivation to reduce the development and maintenance cost and increase the quality of digital learning content by means of modularization and reusability. Learning objects promised to offer a new way to create and mediate educational content in terms of smaller units of learning that are self-contained, can be reused in multiple contexts and can be grouped into coherent collections of digital learning content. Although hundreds of LO repositories exist today, including LON-CAPA[1], MERLOT[2] or Connexion[3], reuse and sharing of digital learning materials has not become a reality on a broader scale.

In 2005 the German project CampusContent[4] set out to increase the quality and frequency of reuse of digital learning objects and pedagogical experience. To achieve this, the project proposed a novel solution to overcome an inherent contradiction in the concept of reusable learning objects: context-independence of learning resources versus contextualized learning. The former is needed to enable better reuse, while the latter relates instructional content to the specific conditions of learners' pre-knowledge and interests and aims to increase their motivation to learn. Learning objects are viewed as compound objects consisting of an information object and educational context. Ideally the information object isolates the information content of a learning object from educational guidelines and must not contain implicit references to other objects. The educational context includes a learning objective and an educational scenario [Baumgartner et al. 2007]. The scenario describes a particular learning situation and specifies recommended learning activities and further conditions. Information objects can then be flexibly combined with other information objects and with different scenarios stimulating different levels of cognitive processes [Bobrowski et al. 2005].

A second innovation of the CampusContent project is the inclusion of social software functionality to explicitly support communities of practice through improved search and browsing functions. It maintains a community's shared interest as well as reuse experiences including evaluations, recommendations and annotations of resources.

Keyword-based search is currently the dominant approach employed by most search engines. However, as average web users are usually laymen in information retrieval and are often unable to formulate boolean search expressions, the results they get rarely meet their intentions. For example, when entering the

[1] http://www.lon-capa.org
[2] http://www.merlot.org
[3] http://cnx.org
[4] `http://www.campuscontent.de`; this project is supported under grant no. 44200719 from the Deutsche Forschungsgemeinschaft (DFG, German National Science Foundation; `http://www.dfg.de`.)

term "collaborative filtering", which refers to a small research branch of information retrieval research, into Google Scholar, we may obtain more than 30,000 hits. Inexperienced users will not know how to proceed effectively. On a smaller scale, this problem also exists for learning object repositories with large collections ranging over various disciplines and subject areas. For instance, statistics plays a role in social sciences, economics, engineering, mathematics, medicine and other disciplines but often with a different focus and partly different learning content. In addition, the metadata that repositories associate with learning objects often just address the topic dimension of a learning object. Educational facets such as the intended learning objective, knowledge type or intended cognitive capabilities are typically not considered.

Recently, social software such as Blogs, Wikis, or collaborative tagging has acquired increasing popularity. These solutions either enable users to jointly work on news and articles (such as Wikipedia[5]) or enable them to attach personal remarks (tags) to the URL (e.g., Delicious[6]). No matter what the functionalities are, the concept of community plays a key role in the design and implementation of such systems. By grouping users with common interests or objectives together, these systems provide an effective alternative for people to share and search resources. For learning resources, the most intuitive (and probably the most effective) way to find appropriate resources may be to ask colleagues sharing the same interests.

To verify our hypothesis and meet the needs of repository users, we investigated the significance of the concept "community of practice" for learning object repositories and studied the effectiveness of profile-of-interest-based search processes for learning resources. In this paper we present the results of an empirical study performed to detect communities of practice among LON-CAPA users. We also sketch our design of profile-enhanced social search functions taking into account the conceptual closeness of networked communities of practice. The rest of the paper is organized as follows: after surveying related research activities in Section 2, we describe the design of our experiments and analyze the results by identifying several kinds of social communities existing among LON-CAPA users in Section 3. Section 4 summarizes the findings from survey we conducted with LON-CAPA users to understand their motivation, behavior, and preferences of the current system and obtain their opinion on the expected value of novel social software functionality we plan to provide. Based on these findings we suggest a novel concept of community-aware resource selection in Section 5 and sketch the design of an ongoing prototype implementation, which will then be tested by selected LON-CAPA users. Section 6 concludes and presents future work.

[5] http://www.wikipedia.org
[6] http://del.icio.us

2 Connecting People Through Social Software

The term "social software" refers to a category of applications that support networking and computer-mediated interaction of people with shared interests and preferences. It extends capabilities of communication systems and groupware like email, forum, or group calendar as it helps people with similar interests to find and link with each other. A core element of such software is a computer-readable profile. It may have been defined explicitly by an individual, it may have been derived from information accessible in the web or on personal computers or it may have been mined from log files of relevant systems. Based on the idea of shared interests, researchers have tried to improve the effectiveness of search engines by exploiting information about a user's social environment and his position in a social network of peers [Wasserman et al. 1994]. Collaborative filtering (CF) [Breese et al. 1998] is probably the first attempt to introduce social information into information retrieval. The key idea is to exploit the fact that people often follow the example of others who presumably have similar interests. A CF algorithm will typically calculate the similarities that exist between different users and compute recommendations based on the preferences of those users with high similarities. iTunes uses this technique, for instance, to complement the result of a direct search for a song by suggesting songs that do not directly match the search terms but belong to the same genre because other iTunes users viewed them so.

[Freyne et al. 2004a, Freyne et al. 2004b] present the design and implementation of an experimental search engine that exploits social relationships between users, borrowing the idea from collaborative filtering. By aggregating related judgments from different communities of practice, the authors try to improve the engine's search efficiency by reusing past search results in future searches that use the same keywords. To implement this feature, users are required to join a specific community before executing a query. In [Kautz et al. 1997a, Kautz et al. 1997b], Kautz and others propose a mining system for social relationships that combines social networking and collaborative filtering techniques. This approach focuses on the extraction of a social network from information found on web pages, the identification of experts for a particular topic and the establishment of a link between the searching user and the expert in the social network. The way in which it extracts social links from publicly available information on the Web differs from the approach of other social systems in that it does not require a person to sign up and explicitly name his or her colleagues and collaborators. Similar work has been pursued in [Almeida et al. 2004, Walker et al. 2004]. All these systems have in common that instead of directly searching for a required resource in an unstructured space of resources, they first try to locate people who most likely might have these resources because they share a great deal of interests with the searching person.

Despite these efforts to introduce social network analysis results into information retrieval, current approaches are far from satisfactory. They either require a user to maintain a specific social network for each topic area or they offer their users too few means of influencing a search by selecting the proper social context.

Similar to keywords, hyperlinks or time series, social information is just another facet [Perkio et al. 2005] a user can exploit to search and explore the information space. Therefore, social information should be seamlessly combined with other search dimensions to better fit a user's needs and improve the quality of the search process. In addition, a user should be able to manage and access her social information, to define allowed usage patterns and constrain unwanted accesses from others systems and users. The concept of Community of Practice (CoP) [Wenger 1998] has already been applied as a additional facet to classify information or services in information retrieval [Kirsch 2005] and web service discovery [Perryea et al. 2006], respectively. These approaches exploit the relationship between users with similar characteristics, such as topical preferences and experiences, to enhance the search process.

3 Analysis of LON-CAPA Log Data

LON-CAPA is a learning content management and assessment system originally developed by Michigan State University (MSU). Its predecessor CAPA (Computer-Assisted Personalized Approach) was piloted 1992 in a small physics class of 92 students. Today LON-CAPA exhibits over 16,000 enrollments per semester at MSU and approximately 40,000 enrollments system-wide. The content maintained in the system ranges from middle school courses to graduate level courses in research universities and covers several disciplines including physics, astronomy, biology, business, chemistry, civil engineering and computer science.

Besides educational resources, LON-CAPA also maintains information about people who create, modify, assess, or use these resources in their courseware. In our analysis we investigated a subset of the information stored in the log data collected over a period of three years. The data we looked at are records of the form $(e_1, e_2, \ldots, e_{44})$, where e_1 to e_3 are strings representing the title of a resource, the name of the person who contributed the resource, and a sequence of words denoting the subject areas into which the resource falls, respectively; e_4 is a url acting as a unique identifier for the resource; e_5 is again a string of words denoting characterizing keywords; other elements specify the language, creation and last revision dates, copyright information, resources that are referred to by this resources, targeted grade levels, a list of courses in which the resource is used and other information. The data we investigated refer to 253,972 learning resources developed by 539 authors. These resources have been used in 2,275 courses that were composed by 2,120 instructors.

For the purpose of effective analysis algorithms, these raw data were parsed and the resulting expressions were mapped into a database implementing the data model presented in Fig. 1. The model maintains a table for each major data element, including resources, courses, which are composed of resources, LON-CAPA users, profiles, and subject categories.

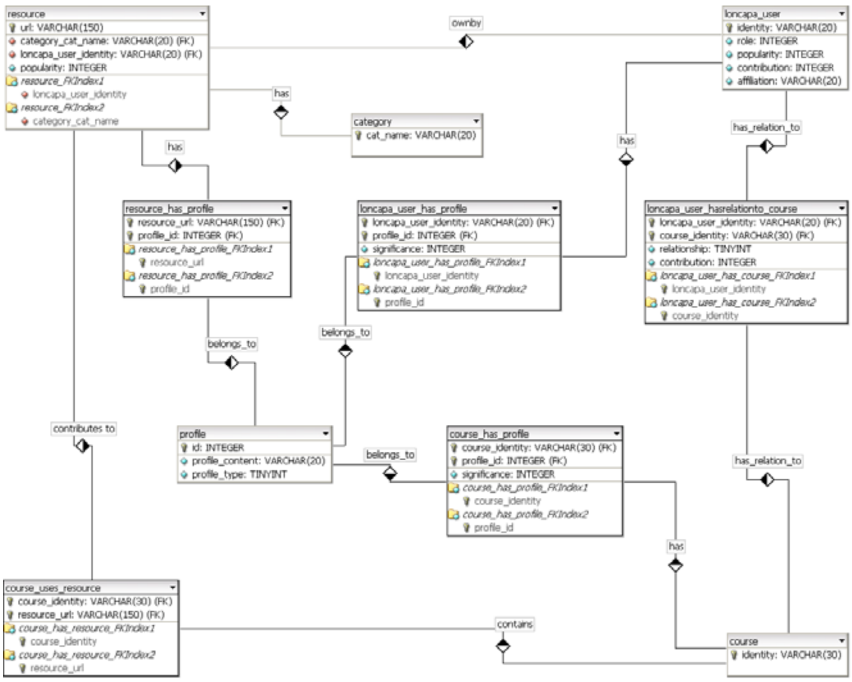


Figure 1: Relational model for LON-CAPA log data

In our data analysis, we were interested in finding answers to different questions including: Who are the most popular authors in a particular subject area? Are there subsets of authors who share some commonality? Can we detect recurring patterns of reuse from these data?

We used concurrence analysis to mine the LON-CAPA log data that were collected before redesign. Particular aims were to find those authors whose resources are reused most and to identify sets of resource authors whose resources are frequently reused in the same courses composed by independent instructors. As resource authors can also be instructors, self-reuse was eliminated from the data analysis. We grouped such resource authors exhibiting significant co-occurrence in courses and took a closer look at common characteristics by constructing a profile for each of author. We used the most popular term frequency method to find out commonalities among the profiles of all authors in a group.

3.1 Popularity of Authors

In the first step of our analysis we detected big differences in the number of learning resources authors had provided. Table 1 lists the top 10 authors in terms of numbers of resources they contributed to LON-CAPA. For privacy reasons, we use numbers to identify authors. Incidentally, these numbers correspond to the contribution ranks of the authors. From Table 1 we can see that the top 10 authors together provided more than 140,000 learning resources, which accounts for 55% of the total number of available resources.

Table 1: Top 10 most contributing authors

Author ID	Number of Contributed Resources
001	89086
002	10831
003	10408
004	6672
005	4643
006	4167
007	3919
008	3816
009	3731
010	2943

As the number of resources an author provides does not tell us anything about the appreciation an author's work receives in the LON-CAPA user community, we studied the popularity of each author. Table 2 lists the top 10 most popular authors based on the frequency an author's resources have been used by instructors in their courseware. Comparing this result with the ranking in Table 1, we can see that the popularity of an author is not directly related to the number of resources he or she contributed, although, at first glance, it seems likely that a high number of contributions may raise the number of reuses. Among the top 10 most contributing authors, only two appear in the popularity table.

To obtain a more objective popularity measure, we define the *Normalized Contribution Popularity (NCP)* by Equation 1.

$$\text{NCP} = \frac{\text{Num}_{Used}}{\text{Num}_{Contributed}} \tag{1}$$

It represents the average usage frequency of the resource an author contributed.

Table 2: Top 10 Most Popular Authors

Author ID	Reused in Courses	Popularity Rank
018	753	1
048	406	2
010	393	3
014	323	4
029	320	5
065	307	6
003	281	7
077	257	8
011	233	9
031	215	10

The top 10 authors with the highest NCP are listed in Table 3. Except for two, all authors from Table 2 occur again in Table 3, but with different ranks. Author 003 disappeared because his or her high number of contributions is not matched by a corresponding high number of reuses.

Another interesting observation read from Table 3 is that the resources contributed by the top 10 authors with the highest NCP value account for more than 40% of the total resource usage in the system, while they add up to only approximately 6% of the total number of available resources.

3.2 Author Communities

In the previous subsection, we analyzed some characteristic reuse figures for the learning resources maintained in LON-CAPA. We could reveal a clear imbalance between the number of resources contributed by an author and reuse frequencies. The results show that there are authors who contribute a lot but are not reused much, for whatever reason. The analysis results also suggest that there exists a small number of authors who own the majority of popular resources in the system. In the sequel we look further into the data to investigate whether we can detect relationships between popular authors.

Table 3: Top 10 Authors with the Highest NCP

Author ID	Contributed Resources	Reuse Instances	NCP
018	1930	52586	27
010	2943	60212	20
065	646	10902	16
077	513	7752	15
014	2413	25837	10
054	835	7726	9
029	1317	12266	9
091	404	3128	7
048	919	6726	7
011	2937	18093	6

Here we consider two authors to have some kind of latent relationship if their resources are used simultaneously in the same course. More formally, we define the Co-Contribution Association (CCA) between two authors a_i and a_j by Equation 2:

$$\mathrm{CCA}(a_i, a_j) = \mid \{C \mid \exists a, b \in C, a \in R_{a_i} \text{ and } b \in R_{a_j}\} \mid \tag{2}$$

where C is a course and R_{a_i} and R_{a_j} denote the sets of resources contributed by user a_i and a_j, respectively. Table 4 lists the top 20 pairs of authors with the highest CCA.

Table 4: Top 10 Couple of Authors with the Highest CCA

CCA Rank	Author_1	Author_2	CCA	CCA Rank	Author_1	Author_2	CCA
1	018	010	274	11	077	014	135
2	018	048	261	12	029	014	130
3	018	065	225	13	077	029	128
4	018	011	182	14	018	066	121
5	048	010	173	15	014	007	121
6	018	029	170	16	029	048	119
7	029	065	154	17	048	011	111
8	065	010	146	18	031	018	108
9	029	010	138	19	065	011	107
10	065	048	137	20	010	011	105

An interesting observation to be made from Table 4 is that a high Co-Contribution Association also exists between the top 10 authors with the highest NCP. There is a total of 12 distinct authors within the top 20 pairs, 9 of which are among the 10 most popular users. To describe this relationship more accurately, we define the concepts of Strong Author Community (SAC) and Weak Author Community (WAC) with a connectivity level n by Equations 3 and 4, respectively:

$$\mathrm{SAC}_n = \{A | \forall a_i, a_j \in A, \mathrm{CCA}(a_i, a_j) \geq n\} \tag{3}$$

$$\mathrm{WAC}_n = \{A | \forall a_i \in A, \exists a_j \in A, a_i \neq a_j \wedge \mathrm{CCA}(a_i, a_j) \geq n\} \tag{4}$$

Table 5 lists some of the SACs and WACs we found among LON-CAPA authors. From this table we can see that even for a very high connectivity such as 90, there exist such WACs or even SACs that have quite a few members. A finite SAC can be represented as a fully connected graph, while the graph of a finite WAC must not have unconnected users.

Table 5: Sample Communities in LON-CAP

Id	Community Type	Connectivity	Community Members
1	SAC	170	010, 018, 048
2	SAC	150	018, 029, 065
3	SAC	90	010, 011, 018, 029, 048
4	WAC	200	010, 018, 048, 065
5	WAC	170	011, 018, 029
6	WAC	100	010, 018, 029, 031, 048, 065, 066

We can summarize the main results of our data analysis of LON-CAPA log data as follows, remembering that self-reuse was eliminated in the analysis:

Discovery 1. Most reuse instances observable in the LON-CAPA system involve a relatively small portion of learning resources and these resources are contributed by a small number of authors only.

Discovery 2. The resources contributed by the popular authors are not only frequently used individually but also frequently used together. In other words, by co-contributing to the same courses, popular authors form tightly connected communities of practice.

3.3 Learning Resource Selection

In the LON-CAPA learning resource repository each learning resource is associated with two metadata fields: subject and keyword. Both consist of one or more terms describing the topic addressed by a resource and its content, respectively.

Let $S = \{s_1, s_2, \ldots, s_n\}$ and $K = \{k_1, k_2, \ldots, k_m\}$ be countable sets of subject and keyword terms in LON-CAPA. Then we can define the subject and keyword profile for a particular course C_i as follows:

$$S(\mathrm{C}_i) = \langle C_{i\_s1}, C_{i\_s2}, \ldots, C_{i\_sn} \rangle \tag{5}$$

$$K(\mathrm{C}_i) = \langle C_{i\_k1}, C_{i\_k2}, \ldots, C_{i\_km} \rangle \tag{6}$$

where C_{i_sj} and C_{i_kj} denote the frequency of subject term s_j and keyword term k_j, respectively, related to the resources contained in course C_i. Similarly, we can define the subject and keyword profile of a particular resource provider P_j as follows:

$$S(\mathrm{P}_j) = \langle P_{j\_s1}, P_{j\_s2}, \ldots, P_{j\_sn} \rangle \tag{7}$$

$$K(\mathrm{P}_j) = \langle P_{j\_k1}, P_{j\_k2}, \ldots, P_{j\_km} \rangle \tag{8}$$

When considering the course composition process from the perspective of an instructor, it seems likely that he or she searches for a learning resource whose content best matches the subject area of the course under development. In practice this would mean that instructors search for resources using keywords characterizing the topic areas of a course. If this were the prevailing reuse pattern and if we assume that no author has contributed a large number of resources covering a hardly asked topic area, authors with a high number of contributed resources would exhibit a high NPC because their resources should be hit more frequently by keyword searches due to their sheer numbers.

However, our data analysis does not support this assumption. One reason why the resources of some authors are less reused could be the "vocabulary problem" [Furnas et al. 1997], which says that the keywords assigned by the authors are often different from the terminology used by searchers. Another explanation suggests that the course composition process follows other behavioral patterns than just keyword search. To support this argument, we selected courses containing resources from more than 5 authors, which count up to 724 courses. Then we analyzed the composition of these courses to find out that the resources are not always drawn from authors whose resources most closely match the subject area and keywords characterizing the actual course. The result is summarized in Table 6. In the first column it lists the ratio of resources contributed by authors with the best subject or keyword match to the course content. We can see that only 495 (or less than 70%) of the courses consist of more than 70% of the resources contributed by the top 10 subject or keyword related authors. For more than half of the courses we found that less than 50% of their resources come from the top 10 subject or keyword related providers.

Table 6: Course Composition Analysis

Contribution Ratio	Keyword Top 5	Keyword Top 10	Subject Top 5	Subject Top 10
0.7	319	445	310	495
0.5	214	302	184	343
0.2	79	124	80	154

The lesson learned from this consideration is that content-related metadata alone are not carrying sufficient information for the learning resource search and selection process or that metadata of often inappropriately chosen.

3.4 Community Analysis

Based on the profile generation approach discussed in the Section 3.3, we can now try to gain further insight into the latent communities detected. Starting with the most populated WAC (WAC 6 in Table 5), we can easily see that it includes SAC 1 and SAC 2, which are connected to each other through author 018. Figure 2 depicts the structure of WAC 6. It also visualizes the overlap of the two strongly connected communities SAC 1 and SAC 2, which both address the subject area physics. A look at the profiles of the members in each SAC reveals that they share some keywords. For example, in SAC 1 the keyword profiles of all its authors share the terms "acceleration", "length" and "velocity".

Based on these results we may want to impose a social structure on the individuals using LON-CAPA. Shared sets of topic keywords occurring in the individuals' profiles would express a common relation of interest and would allow us to modify the search functionality of LON-CAPA's by including social information. For example, if an individual is a member of SAC 1 and SAC 2, these communities could be searched first. If the resulting match is unsatisfactory, the search focus can be widened to the WACs the user belongs to. Alternatively the user may choose to spread the search process into communities (SACs or WACs) in which other members of SAC 1 and SAC 2 are involved and so forth. Relying on the SAC and WAC relationships between users and the weight parameter n, the shared keyword set of a community of practice can be used as a controlled

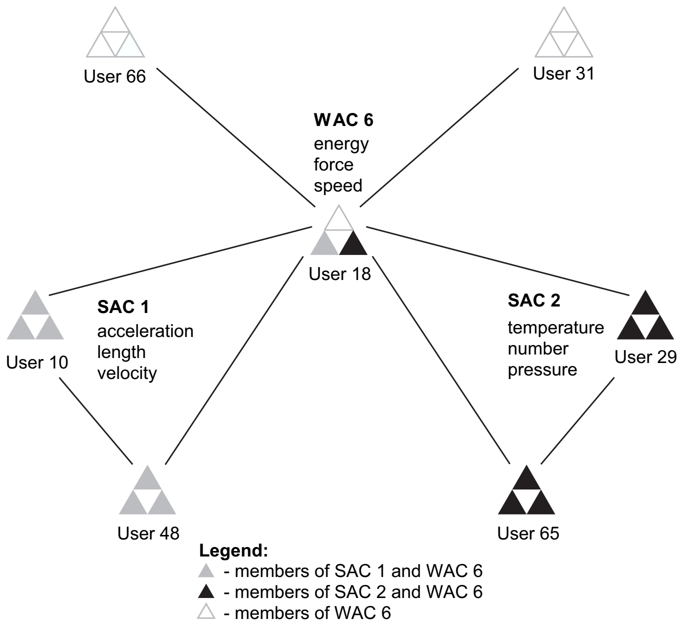


Figure 2: WAC 6 includes SAC 1 and 2

vocabulary in a search process. In addition, search patterns can be stored and visited as a new search process is started. If the new search process turns out to be a prefix of stored search patterns, it could be expanded automatically. The advantage of this kind of exploration is that users do not necessarily need to guess proper search keywords. Instead, they simply reuse community information and keep control over the communities they want to consider.

4 A Survey of LON-CAPA Users

Our empirical research about learning object reuse based on log data mining produced some valuable insights into social aspects of sharing and reusing learning resources. But this research also produced new questions, which we could not answer from a further analysis of the log data. Such questions include: Why are some resources used more frequently than others? Why are particular resources from different authors used frequently in the same course? Can social software facilitate the process of community building?

To learn more about the rationale behind the behavior of resource contributors and course composers and to understand perceived benefits and disadvantages of reusing third party learning resources, we conducted a web-based survey of LON-CAPA users. We received 49 responses. The main subject area in teaching of about 44% of the respondents was physics or a combination of physics and astronomy or chemistry. In addition, we interviewed selected individuals to further examine the trends resulting from a detailed statistical analysis of the responses received.

The questionnaire included 24 questions grouped into four sections: Section 1 addressed the personal background of the respondents, including their main teaching subject area, their role as resource contributors and re-users, the type of institution they are affiliated with and their job title. Section 2 dealt with personal experiences with learning resource repositories in general, and LON-CAPA in particular; it asked for use frequencies, the users' motivation to publish or reuse resources, the adequacy of matches and quality of resources found, the preferred granularity and media type of resources, and their most frequently used way of searching for resources. Section 3 investigated the impact of learning resource repositories on the process of course preparation and learning in general, while Section 4 analyzed the appreciation of future community support for LON-CAPA. Occasionally open questions were used but the majority of questions was closed, often using a five-value scale ranging, e.g., from "strongly agree" to "strongly disagree".

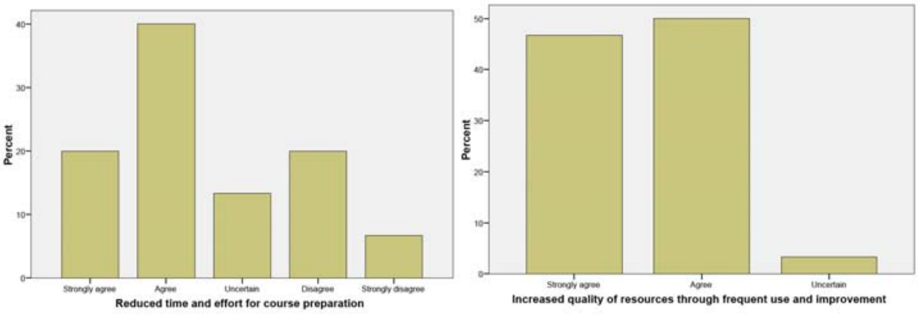
The web-based questionnaire is still accessible[7] but responses are no longer stored and evaluated.

4.1 Subject Characteristics and Motivation

We found a relatively equal distribution among resource contributors and users of resources with 49.3% and 50.7%, respectively, and 87.8% declared to use LON-CAPA a lot to quite a bit. Motivations to use other authors' resources include timesaving and comfort (33%), quality of the resources found (29%), trust in the resource author (17%), and following the example of colleagues (12%). More than 30% claim to have used more than 250 learning resources. More than 40% of the respondents are believers in the share-and-reuse idea and nearly 30% are convinced that their resources are generic enough to be of use for other instructors. Those who hesitate to publish their learning resources consider their work too course-specific to be of interest to others (30%), want to test and improve their work more thoroughly (27.5%) or have copyright concerns (25%).

Answers to two reasons that might motivate people to reuse learning resources are graphically depicted in Fig. 3(a) and 3(b). Other options to the motivation question included: "Lack of personnel to develop resources", "lack of technical competences", "trust in the resource provider", "following the example

[7] `https://eva.fernuni-hagen.de/mrIWeb/mrIWeb.dll?I.Project=LONCAPA1&i.test=1`



(a) Reduced time and effort (b) Improved quality

Figure 3: Motivations for reuse

of a colleague who reused the resource in a course addressing similar topics as my course", and "did never reuse third party learning resources".

4.2 Reuse Experience

A majority of users of resources maintained in the LON-CAPA system stated that they found appropriate resources often (61%) and very often (10%). This positive vision is, however, blurred by the experience that only 16.3% are always happy with the resources they found. This correspond to the response that 41% invested more than 15 minutes to adapt a resource to their needs and still 37% spent between 6 and 15 minutes on reworking it, on the average. This was necessary because the terminology or notation did not fit in the context of use, the technical format did not blend well, the design was considered inappropriate, the resource was pedagogically undemanding or it contained errors. Reasons given by a few respondents why a reuse attempt was not successful at all include: lack of resources in a particular subject area, wrong target audience ("too complex for high school") and the specificity of resources found.

When asked how to overcome these deficits of learning resources, the respondents basically support the approach chosen in CampusContent because the inclusion of pedagogical information such as recommended learner activities and learning objective ranks highest on the wish list with 32%, but also special design with preconceived adaptation mechanisms and more flexible media formats are desired by 27% and 23%, respectively. Motivations "trust" and "following others" were chosen by 16.5% and 11.7%, respectively.

4.3 Search and Browsing Behavior

To better understand the findings about reuse frequencies we reported for the log data analysis, we asked for preferred ways of searching and browsing the LON-CAPA repository. Table 7 summarizes the answers we received.

Table 7: Most Preferred Search Criteria

Approach	**Frequency**
One or two subject keywords	29%
More than two subject keywords	9%
Author names	15%
Browse in the domain of particular institutions	24%
Follow links (or recommendations) received from peers	18%
Browse hit lists	2%

Hence, with 38% of the entries keyword-based search falls behind socially motivated search and browsing activities (rows 3-6), which sum up to 59%.

As one of our hypotheses in this research was that social relationships have an impact on the behavior of repository users, the questionnaire included a section of items that investigated the interest of LON-CAPA users in additional functionality providing community awareness and support. The rounded figures for the responses are summarized in Table 8.

An interesting observation in this section of the questionnaire is that the options "disagree" and "strongly disagree" were not selected at all.

5 Community-Aware Learning Resource Selection

The analysis of massive amounts of LON-CAPA log data revealed the existence of latent communities of practice, and the survey presented in the Section 4 encouraged us to design and implement community-related functionality that aims to enhance the current keyword-based search paradigm for learning resources. We suggest a kind of faceted search. The facets we propose represent community-related information such as the community profile, popular authors or trusted institutions.

Existing e-commerce systems such as eBay or Amazon have demonstrated that the exploration of relationships between different items such as frequent

Table 8: Attitude Towards the Added-Value Functionality

	Strongly Agree	Agree	Uncertain
The possibility of forming communities of practice would be a major improvement	21%	63%	16%
I am willing to invest a few minutes of my time to provide my profile for matching up with communities of practice	28%	58%	14%
I want to know who the authors of the most frequently used resources in my subject area are	52%	37%	11%
I am willing to evaluate and annotate a resource I looked at seriously	37%	58%	5%

occurrences of simultaneous purchases of different stock items is an effective method to help users find products. In existing learning resource repositories, however, resources contributed by different authors are normally considered independent, and latent interest-based relationships as have been revealed in our analyses are completely ignored.

5.1 Implementation Design

Inspired by these findings, we propose a community-aware learning resource discovery approach. In a first step, two or three initial communities composed from the members of selected SACs and WACs, most likely in physics, will be established explicitly. The members of these communities will be mapped into the nodes of a social network implemented by means of an open source networking software that organizes a person's relationships with other people and provides open service interfaces to third party applications. Those third party applications can access the community's and a community member's relationship information to provide added-value functionality. In our case, relationship information is defined by the subject and topic keywords characterizing the interest profiles of resource authors and users.

When a new user wants to join LON-CAPA, the system will first build an initial profile for this user by asking her to select one or several subject areas she teaches. For each area she is prompted a set of characteristic keywords, e.g., by means of a tag cloud, from which a suitable subset can be selected. Each keyword set acts like a controlled vocabulary and is defined by the sum of resources available in LON-CAPA for the selected subject area. Based on the constructed profile, the user will then be assigned to communities exhibiting similar profiles as this user. As the user then contributes or reuses resources, her profile will be updated automatically as described in Section 3 and her membership to certain communities will be adjusted accordingly.

The intersection of the profiles of users within a community defines the profile of the community. Through the continuous update and use of resource repositories, new communities with new profiles may come into existence and existing communities are evolving as new members join a community.

Our approach differs from other community construction methods as discussed in [Witschel et al. 2005, Yang et al. 2007] in that we do not directly use the profile of users to construct communities. Rather, as discussed in Section 3, profiles just serve as one criterion for community construction because profiles only reflect content-related aspects of resources. Other aspects such as preferred difficulty level, quality or media type are not addressed by topic keywords (from which profiles are constructed).

5.2 Community-Based Resource Search

A first prototype following this design is illustrated in Fig. 4. It shows the initial hit list resulting from a direct search using the keywords "power" and "exponent". Now the user can choose to narrow the search result either by selecting the option "Filter" or "Exclude" in the top of the right pane and simultaneously applying one of the social facets "Community profile", "Author", "Keyword" or "Institution" as a modifier. In the situation shown in the right pane, the user has selected "Filter" and activated facet "Community profile", which thereby unfolds into a tag cloud listing all terms defining the community profile. Other than usual, however, the size of a term in the cloud is not a function of its occurrence frequency in resources associated with the selected community but rather an indication of its popularity. As a result of filtering the current hit list, e.g., by the profile topic "mass", only those resources will be listed that have the keyword "mass" in their metadata. If the user had selected "Exclude" instead and clicked on the term "mass", only those resources that do not include the term "mass" in their metadata would be listed. Alternatively the current hit list can be extended by resources from other communities that satisfy the selected facet, say "mass" in "Community profile". When applying these modifiers repeatedly, users can form very complex search expressions, while obviating the need to build logic expressions explicitly. In this way, we integrate social information in the currently dominating keyword-search based paradigm.

The tag cloud "Community profile" in Fig. 4 includes topic terms but also author and institution names. The author and institution names will also show up when the social facets "Author" or "Institution" were selected, respectively. There is a difference in behavior, however, between selecting author Michael from

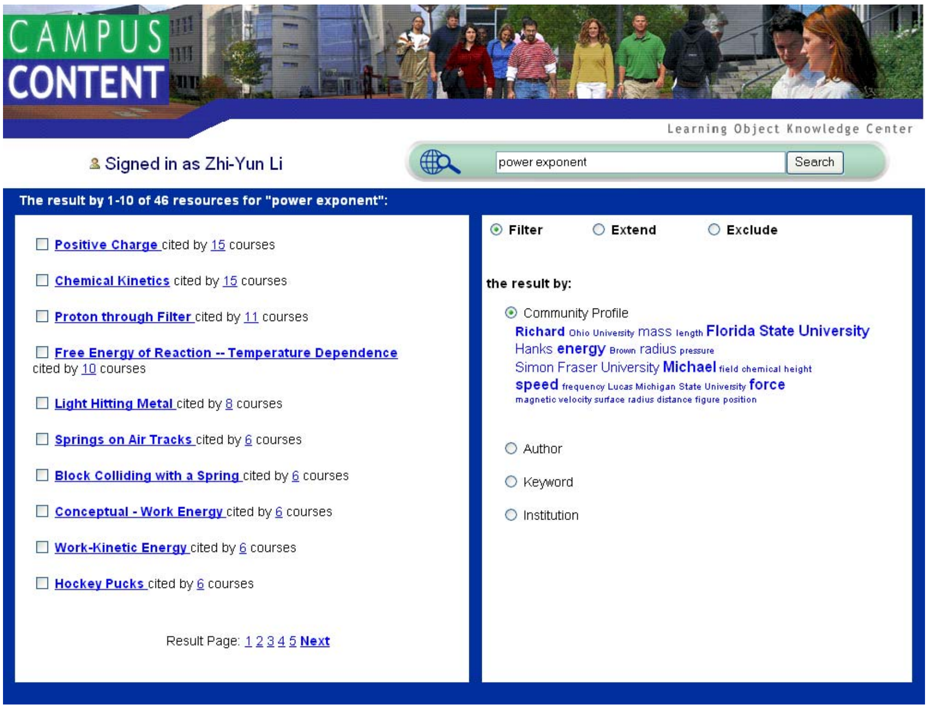


Figure 4: Community-aware search interface

the "Community profile" facet and the "Author" facet. In the first case the result will include the resources contributed by the community to which author Michael has the closest relation. In the second case all resources of author Michael would be listed.

5.3 Expert Evaluation and Rating Services

Expert evaluation applied to learning resources is an instrument of community learning and accountability to the community of resource users. Together with resource rating it forms part of a new set of community services aiming at sharing and exchange of expertise about content quality and pedagogical experience. The social software enhanced prototype will associate each resource with links to an expert evaluation form and a simple rating function.

The expert evaluation form includes a series of questions addressing different aspect of the information quality of a resource, its pedagogical aptness, its design quality and its potential for adaptation. This evaluation form is meant for domain experts who are willing to provide a more elaborate opinion on other authors' resources. Information quality aspects, for example, will question to what degree the content of a resource achieves its intended informational and educational purpose and whether it appears to be complete and accurate, is well organized, and uses standard concepts and notations. The form also asks whether relevant metadata are provided and the intended audience is properly addressed. Design issues refer to the resource's look-and-feel, degree of interactivity, and conformance to browsers and standard plug-ins. The expert can determine the visibility of his or her evaluation. In any case the system will provide anonymous summary reports per evaluated resource. Whether or not the individual contributions of a summary report should be weighted by the expert's recognition in the community and his or her average evaluation will be subject to discussion among LON-CAPA users.

The rating function relies on a simple metric with multiple weighted dimensions including information quality (50%), design and presentation (25%), adaptability (15%), and difficulty level (10%). It allows users of LON-CAPA to assign zero to five stars with each dimension. An average value will be computed and assigned to each resource. An average will also be calculated across the average ratings of all resources an author has provided. This value will be assigned to the author's profile. Again, the visibility of these ratings will be discussed within the LON-CAPA community.

6 Conclusion and Future Work

In this paper we presented an empirical study investigating the behavior of the users of a large learning resource repository, LON-CAPA. Our main objective was to understand: a) why only a small portion of the hundreds of thousands or even millions of learning resources maintained in a plethora of repositories around the world are actively used; b) why the resources being reused frequently are contributed by a small number of authors only; and c) what mechanism have the potential to facilitate the detection of appropriate learning resources.

Through a detailed analysis of masses of log data collected by the operator of LON-CAPA, we identified two kinds of latent author communities that are formed by frequently co-used resources. In a second step we performed an opinion poll among LON-CAPA users to learn more about the motivation to use such a system, perceived benefits or obstacles, and the users' appreciation of planned community support.

Finally we presented the design of a first prototype to provide rudimentary community support based on social networking software and a topic-related notion of shared interest. This prototype will allow new LON-CAPA users to find communities that best match the user's interest profile. It will also enable a user searching for resources to select the search space flexibly by locating the communities whose profile is close to the user's profile.

Future work includes the full implementation of the prototype, its seamless integration with a new version of the LON-CAPA user interface and its test with pilot users. The evaluation of these tests and further functional extensions in planning, such as an annotation and recommendation feature for resources, will then provide the grounds for a qualified revision of the prototype and the design of such functionality as integral part of the second version of the CampusContent learning object repository.

Acknowledgments

We gratefully acknowledge the various improvements Dr. E. Michael Maximilien suggested after a careful review of an earlier version of this contribution.

References

[Almeida et al. 2004] Almeida, R. B., Almeida, V. A. F.: A community aware search engine. In *Proceedings of the 13th International World Wide Web Conference*, pages 413–421, ACM 2004.

[Baumgartner et al. 2007] Baumgartner, P., Heyer, S.: The Trinity: Towards a Model of Learning Objects. Submitted to: Technology, Instruction, Cognition and Learning (TICL), Special Issue: Learning Objects and Pedagogy, 2007.

[Breese et al. 1998] Breese, J. S., Heckerman, D., Kadie, C. M.: Empirical Analysis of Predictive Algorithms for Collaborative Filtering. In *Proceedings of 14th Conference on Uncertainty in Artificial Intelligence*, pages 43– 52, Morgan Kaufmann 1998.

[Bobrowski et al. 2005] Bobrowski, S., Nowaczyk, O.: Methoden des Software-Engineering als Leitbild für den Entwurf wieder verwendbarer Lernobjekte (Using software engineering methods as a guideline for the design of reusable learning objects), Forschungsberichte des Fachbereichs Elektrotechnik und Informationstechnik, FernUniversität in Hagen, No. 8, 2005 (in German) `http://www.fernuni-hagen.de/imperia/md/content/fakultaetfuermathematikundinformatik/forschung/berichteetit/forschungsbericht_8_2005.pdf`

[Freyne et al. 2004a] Freyne, J., Smyth, B.: An Experiment in Social Search. In *Proceeding of 3rd International Conference on Adaptive Hypermedia and Adaptive Web-Based Systems*, pages 95–103, Lecture Notes in Computer Science 3137, Springer 2004.

[Freyne et al. 2004b] Freyne, J. Smyth, B., Coyle, M., Balfe, E., Briggs, P.: Further Experiments on Collaborative Ranking in Community-Based Web Search. *Artificial Intelligence Review*, 21(3-4):229–252, 2004.

[Furnas et al. 1997] Furnas, G. W., Landauer, T. K., Gomez, L. M., Dumais, S. T.: The vocabulary problem in human-system communication. *Communications of the ACM*, 30(11):964–971, 1987.

[Kautz et al. 1997a] Kautz, H., Selman, B., Shah, M.: The hidden web. *AI Magazine*, 18(2):27–36, 1997.

[Kautz et al. 1997b] Kautz, H., Selman, B., Shah, M.: Referral Web: combining social networks and collaborative filtering. *Communications of the ACM*, 40(3):63–65, 1997.

[Kirsch 2005] Kirsch, S. M.: Social Information Retrieval. Diploma Thesis in Computer Science, Rheinische Friedrich-Wilhelms-Universität Bonn, 2005 `http://www.sebastian-kirsch.org/moebius/docs/socialir-thesis.pdf`

[Perkio et al. 2005] Perkio, J. Tuulos, V., Buntine, W., Tirri, H.: Multi-faceted information retrieval system for large scale email archives. In *Proceeding of 2005 IEEE/WIC/ACM International Conference on Web Intelligence*, pages 557–564, 2005.

[Perryea et al. 2006] Perryea, C. A., Chung, S.: Community-based service discovery. In *Proceedings of the IEEE International Conference on Web Services (ICWS'06)*, pages 903–906, 2006.

[Walker et al. 2004] Walker, A., Recker, M., Lawless, K., Wiley, D.: Collaborative information filtering: A review and an educational application. *International Journal of Artificial Intelligence and Education*, 14(1):3–28, 2004.

[Wasserman et al. 1994] Wasserman, S., Faust, K.: *Social Networks Analysis: Methods and Applications*. Cambridge University Press, 1994.

[Wenger 1998] Wenger, E.: *Communities of practice: Learning, meaning, and identity.* Cambridge University Press, 1998.

[Witschel et al. 2005] Witschel, H.-F., Böhme, T.: Evaluating profiling and query expansion methods for P2P information retrieval. In *Proceedings of the 2005 ACM workshop on Information retrieval in peer-to-peer networks*, pages 1–8, ACM Press 2005.

[Yang et al. 2007] Yang, F., Krämer, B. J., Han, P.: Common-interest Based Self-organising E-Learner Communities, Chapter 4 in *Contributions to Ubiquitous Computing*, pages 85-108, Springer 2007.