Educational Material Repository

Jeff Trawick
trawick@gmail.com
Raleigh, NC

January 30, 2014

A vast amount of educational materials are available currently, both from educational vendors as well as from a large number of organizations and individuals. Much of the material hard to find, is available under varying terms, and is not easily consumable by teachers, students, and parents. Concentrating efforts and standardizing the format of and access to these materials would be of great benefit to all parties.

This is a brief description of a (theoretical) common repository of educational materials which are available for use by teachers, students, and parents in different contexts, and consumable using different software. The repository contains the material itself rather than serve as an index to materials available elsewhere.

Key characteristics

The key characteristics of the repository are

- Materials are machine readable, with a structure and open access mechanism that allows them to be consumed in varied and unpredictable ways, allowing maximum reuse of materials using a variety of software.

- The repository provides some means to consume the materials directly, without the use of additional software.

- Materials are created and modifiable in a manner that supports reviewed collaboration over the Internet with minimal expertise required in the software which facilitates collaboration.

- Materials in the repository are all available under a common, permissive license.

- The repository contains enough materials and is usable enough to attract the critical mass of users, both individual and institutional, required to grow and improve the contents.

- The ecosystem has room for and can benefit from commercial involvement.

(There are a number of characteristics of the organization of the materials, including the ability of educators to create custom aggregations of the materials, which are required for the repository to be useful; organizational characteristics are largely ignored in this brief description.)

Accessible, machine readable materials

First, it is useful to give a few examples of different types of materials from different fields of study:

- Glossary items (terms and definitions)

- Examples and counterexamples of sentences with subject-verb agreement

- Mathematical word problems with corresponding equations and solutions

- Essay templates with prompts and structure which remind students of the requirements of a particular type of essay

(There are of course many different types of such materials, some which follow a predictable model, or schema, and some which do not. The repository necessarily contains materials that follow a model which has specifically been designed to represent a particular type of material. Many types of materials are out of scope for the repository.)

Such materials can be consumed in many different ways. Consider the use of glossary terms in a single class:

- A biology teacher enters the terms and definitions corresponding to the current unit of study into a web service such as Quizlet so that

students have access to flash cards and games that help them learn the material.

- A biology teacher prepares a quiz in Microsoft Word by re-typing or cutting and pasting terms and definitions into several different types of questions (e.g., matching, pick the correct statement, fill in the blank).

- A biology teacher prepares a presentation which has a mixture of different types of materials, including glossary items.

(and so on and so forth)

The point is that a particular physical copy of data is typically tied to the specific software that is used to consume and/or format it, leading to a large amount of wasted effort. The inability to reuse the data for different purposes or with different software is one contributor to the duplication of effort not just for a single teacher using the material in different contexts but more importantly among the many thousands of teachers who are teaching the same subject.

The repository must provide APIs for referencing the materials that can be exploited by different software. For example, the biology teacher in the example above should be able to create a customized selection of glossary terms for the current unit (or more likely, review and reuse an existing selection), assign an id to that selection, and plug that id into a service like Quizlet or other software to quickly reuse the definitive material. In short, consume from multiple software types and implementations without duplicate effort. (And even the initial effort should be very small once any teacher using the system has set up the glossary items used for the class.)

Direct consumption

While a key characteristic of the repository is that the materials are consumable from any educational software (once the APIs are exploited by the software provider), it is also important to provide basic capabilities with the repository to make the materials accessible to users via a web browser. A reference view of the materials is required, and other software won't be able to consume the materials immediately anyway. But the expectation for the future is that the best quality end user software that is able to consume the materials will come from other sources, and that the critical mass of raw materials in the repository will be a strong incentive for vendors or individuals

to experiment with many different ways to consume it. The range of possibilities is wide — web sites, tablet applications, Microsoft Office plugins, custom devices, etc.

Collaboration

Existing arrangements exist for collaboration over the Internet from individuals all over the world. Well known examples of this are Wikipedia and the various (major) open source software projects. Each group defines the manner in which untrusted people can contribute additions or improvements to the work, as well as the manner in which some contributors are empowered to make larger changes and help manage the information (i.e., become trusted). Collaboration has key technical considerations, particularly the software and procedures which are used by contributors in order to share information. Technical considerations which might be appropriate for developers of software source code would not be so for educational materials, as low technical barriers are required.

Collaboration also has social considerations. In general, a lot can be said about the necessary technical and social aspects of collaboration, but for now it is sufficient to trust that one of the well known existing arrangements will be workable.

Licensing

The exact licensing terms, including attribution requirements, are key. All materials in the repository must be available under the same terms, and those terms should be as simple and permissive as possible. Individuals who contribute material or corrections or other improvements must acknowledge agreement with the common licensing and that they are legally entitled to make the contribution (i.e., the material is their own work and the individual is not covered by an employer agreement which forbids publishing work in that manner).

One of the unavoidable outcomes of simple attribution requirements is that the names of individual contributors will not be associated directly with the material at the point of consumption. However, tracking of creation and modifications to the materials should allow authorship to be obtained from the repository and the contributions of any particular individual to be reviewed.

An important choice in licensing is whether or not the materials can be reused for profit. Some individual contributors may not want their effort to be reused by others for profit. At the same time, the ability to use the repository as part of a for-profit product will motivate some vendors to assist with the maintenance or contents of the repository, and APIs can be designed with constraints that ensure that infrastructure costs of the repository are covered by heavy users of the repository (i.e., not individual accesses).

Bootstrapping

The repository must provide value before it will attract a large number of users, and large number of users must be attracted before volunteers will surface who will add to and improve the repository.

Among the many resources already available on the Internet, some will contain large amounts of materials in reasonable formats, and it may be possible technically and legally to import such materials into the respository. It is expected that such existing resources will have a relatively small scope (e.g., high school mathematics).

Beyond the data itself, existing resources may have already arrived at a good model for certain types of materials that can be reused by this project.

Another potential bootstrapping mechanism is to start with a very limited scope (e.g., only several common core classes) with a specific plan in place to develop a useful set of materials for those classes.

Commercial involvement

An obvious area for commercial involvement is with for-profit end user software that can utilize the repository for educational materials. Software vendors write educational software that can use the material in the repository.

A less obvious area for commercial involvement is with for-profit educational materials that can be accessed in the same manner (API) as the free repository, allowing arbitrary third-party end user software to be used to access educational materials that have been licensed by the user or by their institution. This project should enable that type of involvement by making the software used to manage and host the repository available under an open

source license with no major commercial restrictions (e.g., ASL 2.0) and by welcoming contributions to that software from any third party.

Perhaps the greatest potential benefit from allowing commercial entities to run their own restricted access repositories using the same repository software is that it is an incentive to share the effort to develop good models for the educational materials.

Types of models

Usually a model is a pattern for which there will be many individual data items in the repository. For example, a model for glossary terms would support thousands of possible glossary terms across many subjects, and each glossary term would be individually defined in the repository.

"Generator models" should be supported as well. These describe data items that can be generated automatically by software without being individually defined in the repository. Arithmetic problems are an example of items that can be generated by software. The purpose of having models for these is to define the constraints as they relate to certain learning objectives. For example, a math objective for early primary school might a constraint on appropriate addition problems that limits the numbers being added to the range 1-4. By representing the constraints of such problems in the repository, client software can generate the appropriate type of problem given the learning objective or a reference to the specific model. Contrast that with some current software that requires the teacher to specify the constraints at the point of creating a worksheet.

Usefulness

It is natural to question the usefulness of the repository of materials given the overall scope of teaching and learning. After all, these materials are not the essence of these activities, and teachers and schools have been managing to purchase, create, or otherwise obtain such materials all along. The potential benefits are straightforward:

- Reduce duplicate effort (allow teachers to spend time on more valuable activities).

- Increase the quality of the materials used (no mispellings, grammatical errors, or bad examples in crowd-sourced, crowd-reviewed materials).

- Concentrate efforts on a single open repository instead of thousands of existing web sites with little or no collaboration capabilities (avoiding the tedius search for existing, freely available materials of sufficient quality).

- Provide machine readable access to these quality materials in order to improve the usefulness of existing and future software, and in particular reduce the effort required to experiment with new ideas for consuming the information.

Possible next steps

- Refine the description of the repository so that more people can understand the key points and help assess its potential value.

- Survey existing resources to identify best practices and determine how close existing resources come to meeting requirements. For example, "Learning Objects" and "Sharable Content Objects" are existing ideas that have some intersection with the repository concept.

- Create a simple prototype to illustrate potential mechanisms for collaborating on the creation and maintenance of the materials as well as consumption of the materials by teachers and students. Use the prototype to discuss the potential value of the respository witho a wider audience and identify particular technical and social challenges.