A Markup Language for Building a Data Warehouse for Educational Peer-Assessment Research

Yang Song, Ferry Pramudianto, Edward F Gehringer
Department of Computer Science
North Carolina State University
Raleigh, NC, U.S
[ysong8, fferry, efg]@ncsu.edu

***Abstract*—Peer assessment has proved to be a useful technique in all levels of education. The process of giving and receiving comments can encourage critical thinking and help students learn both from reviewing and being reviewed. Peer assessment generates a large volume of data, especially if done online. Online peer-assessment systems are designed differently and use different schema for their data, which complicates the work of comparing different designs. For example, some systems are based on ranking – reviewers rank the artifacts they are asked to assess, while other systems use rating – reviewers assess a single artifact at a time and score it on various criteria. Comparing these two types of systems, e.g. on rating accuracy, or usefulness of formative feedback, can be challenging because researchers need to learn the design and terminology of each system before analyzing the data. We introduce a Peer-Review Markup Language to provide a common definition of terminology across multiple systems. We are using this markup language to build a data warehouse for data from different systems. We discuss issues raised during this process and our approach to solving them.**

Keywords—educational peer-review; peer assessment; data sharing; data warehouse

I. INTRODUCTION

Despite the fact that dozens of online peer assessment systems are in use, virtually all research studies on online peer assessment derives their conclusions from data taken from a single system [1–6]. A few surveys have tried to compare different systems, but they have only scratched the surface, because they compare the functionalities of the different systems, but contain no results based on comparison of data [7, 8] from different systems. This is symptomatic of the lack of protocols for sharing data generated by different peer assessment systems.

Different online peer assessment systems may use different terms for the same concept, or use same terms for different concepts. For example, in peer assessment activity, the student authors need to submit their work first. Different systems refer to this as "artifact", "work" or "submission". Even the term "peer review" has different meanings in different contexts. It can mean an assessment of work submitted by a different individual or team, or it can mean a review of an individual's *contribution* to the work of the team.

Moreover, different systems implement the process of peer review in different way. For example, some systems give each reviewer several submissions and ask them to rank them; whereas other systems may give each reviewer one submission at one time and ask the reviewer to rate it based on different aspects. There is no easy way to compare rankings from one system with ratings for another. To make sense of the results, researchers need to thoroughly understand the design of each system, and this will take a long time to achieve.

These differences make it hard to compare results from different systems [7]. However, many research questions in this research area, e.g., comparing ranking and rating, or comparing reviews done by individuals with reviews performed by teams, cannot be addressed without first obtaining a sufficiently general data set. To surmount this hurdle and facilitate data sharing between systems, this paper presents our effort to design a Peer-Review Markup Language (PRML).

II. PRML STRUCTURE

Our Peer-Review Markup Language (PRML) was designed to be a generic data model/schema for modeling most of the data – both numerical and textual – generated by online peer assessment systems used by students. PRML was originally designed jointly by the owners/designers of four online educational peer assessment system: Expertiza [3, 9], Mobius SLIP [4], CritViz [5], and Crowdgrader [6]. All those four systems are in active use and development, with new users, new functionalities and new patches.

PRML defines a common terminology for concepts that are pervasively used in educational peer assessment. This common terminology, allows researchers, system designers and owners to discuss common concepts related to their research and their systems. In addition, PRML also defines a generic schema to model the relationship of data generated by peer-assessment activities. Thus, PRML can help build a data-sharing protocol and moreover, a data warehouse, which can hold data produced by different peer-assessment systems. The number of users and reviews, plus the variety of different peer-assessment settings can help researchers to study and compare hypotheses and come up with general conclusions.

PRML defines entities in educational peer assessment and the relationship between them, as shown in Figure 1. Peer

assessment is usually performed within a Course, which may have one or several Assignments. The Course will have many Participants enrolled. Those individual participants may need to form teams or act alone in some cases; e.g., some system support Students to work in a group and submit a piece of an Artifact. Therefore PRML uses the concept of Actor to model an individual participant or team that undertakes any peer assessment activity. An Actor can either be a reviewer or a reviewee. An Actor may consist either of one participant, or multiple participants acting as a team. This design allows PRML to model the data generated by individual reviews of an individual's work, individual reviews of a team's work, team reviews of an individual's work, or a team's review of another team's work. Actors (individuals or teams) are also the authors of Artifacts. Those artifacts may relate to a common task (e.g., all students are doing the same homework assignment) or different tasks (e.g., each student or team is working on a different topic).

To facilitate peer assessment, Instructors may also create their own review Rubrics. A Rubric can either be holistic – containing only one Criterion asking the reviewers to give overall comment and score an overall artifact – or detailed – containing multiple Criteria and scored on different aspects. Each Criterion is a prompt asking reviewers to evaluate some aspect/dimension of the artifact. Criteria may be of different types, such as textbox questions, rating questions, checkboxes or ranking questions. The response that a reviewer filled in for the criterion is called a Critique. A Critique can be textual or numerical or even both (to accommodate systems that ask for both numerical and textual feedback in one question).

In the designing phase of PRML, we tried to make the PRML as generic as possible so that we can model the data generated by different systems. For example, we are aware of both systems designed to support individual's work and systems designed to support team's work. There are even systems can handle both, too. Therefore we introduced the concept to Actor to support this. In addition, we also considered and made sure that the PRML can support both holistic rubrics and details rubrics.

In addition, some data related to class settings are important for researchers. Those data are not necessarily represented in the database of the peer assessment system – they could be some rule enforced by the system, or set up by the instructor. PRML can also handle those data. Below are some examples:

Reviewer assignment strategy. Most systems start the peer-review process after the submission is due. Reviewers may be assigned either statically or dynamically. If they are assigned statically, either the instructor or some algorithm decides who will review whom. However, dynamic assignment is often preferred because inevitably, some students will fail to submit their reviews, and thus, some submissions will be short of feedback. Dynamic assignment assigns submitted artifacts to reviewers on the fly. This allows a reviewer to be assigned, for example, to the submission that has received the fewest reviews so far. In reality, many more factors are at play when assigning reviewers. For example, if students are working on different topics, the system may allow reviewers to choose the topics they want to review. Another strategy is to use reviewer's reputation (determined, for example, by how closely a reviewer's ratings matched that of the teaching staff on a previous assignment) [10] to make sure that each artifact is reviewed by at least one competent reviewer.

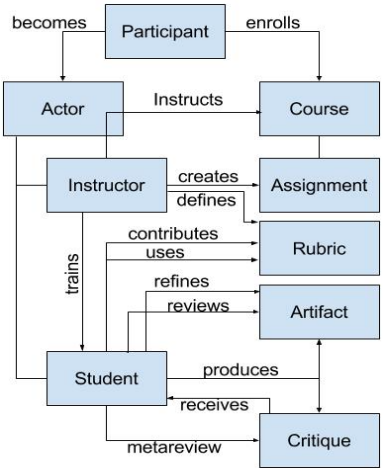

Fig. 2. PRML main concepts and their relations

Degree of anonymity. Most systems use some degree of anonymity in the review process. Anonymity may be applied both to the reviewers and authors (double blind) or only to one side (single blind). Some systems adopt anonymity in the review process, but then publish the artifacts and reviews to provide a sense of accountability for the authors and reviewers.

Rejoinders/back-reviews policy. Some systems allow authors to give back-reviews, or "rejoinders" to the reviewers. This feedback can be qualitative (prose, checkboxes) or quantitative (e.g., Likert scale), depending on the rubric. The rejoinder may or may not be included in the computation of the reviewer's grade.

III. Data Warehouse

Based on the entities and relations defined in PRML, we have also created a data-warehouse schema to facilitate data sharing across different online educational peer-assessment systems. This data-warehouse schema was designed based on the star schema [11], which separates the data into facts that hold measurable, quantitative data about the peer assessment and dimensions, which are descriptive attributes related to the fact data.

In our data-warehouse schema (Figure 2), the fact table in the middle is the Critique table. It contains information from textual and/or quantitative feedback based on a single criterion. This is also the finest grained data which can be created in educational peer assessment process. Around the Critique table, there are many dimension tables which include information on additional attributes of this Critique. The Actor table contains references to both reviewer and reviewee for each critique. The actors can either be individuals or a group. The Participant table stores information on users who participate in an assignment. Between the Actor table and the Participant table, there is another joint table called Actor_participant, which is

designed to maintain the *n*-to-*n* relationship. However, since this data warehouse is designed to be anonymized, we barely need to store any data in Participant table. The Eval_Mode records the evaluation mode of this critique, which can be ranking-based, or rating-based, or even both.

The Criterion table holds information on the criteria in the review rubrics, including scales, weight and type (textbox, rating, checkbox, etc.). A Criterion can either be holistic or detailed. The Task table contains information on when the review period starts and ends, the CIP (Classification of Instructional Programs) code of the course, the type of the task (submission, review, etc.) and further information for the assignment and course. Note that, we do not use a separate assignment or course table, instead, we "flatten" the schema and put all the information related to assignments and courses into Task table. This is to reduce the number of join-table queries and thereby speed up querying of the data warehouse. The Artifact table contains information about the student's work for each task. Considering that the artifact may contain many items (e.g., a .pdf file for project report plus a video for demonstration), we also created an Item table. The Item table is also associated with the Critique table, since reviewers may also create some items (e.g., a reviewer may be able to add comments to the original .pdf file and submit it to the online peer assessment system).

In addition to the information above, the data warehouse also collects the meta-information about the assignment settings in the Course_setting table. The information includes the anonymity setting (none, single blind, double blind, etc.), the workflow (in term of tasks or assignments) and the rubric mode (holistic or detailed). This table facilitates further comparisons on different assignment settings after enough data has been collected.

In the star data-warehouse schema, all the entities except for Actor_Participant and Participant are directly associated with the fact table Critique. This can make the query easy to implement and fast enough even when the data volume is huge. Any comparison based on this data warehouse can be considered as "slicing" the data using one or more dimensions as constraints. For example, we can create a query to retrieve all the reviews generated by ranking-based assessments and another query to get all the reviews generated by rating-based assessments. Then, a comparison can be done using metrics such as the average length of comments or the number of suggestions offered. In this way, we can come up with a more general conclusion about which type of assessment is better.

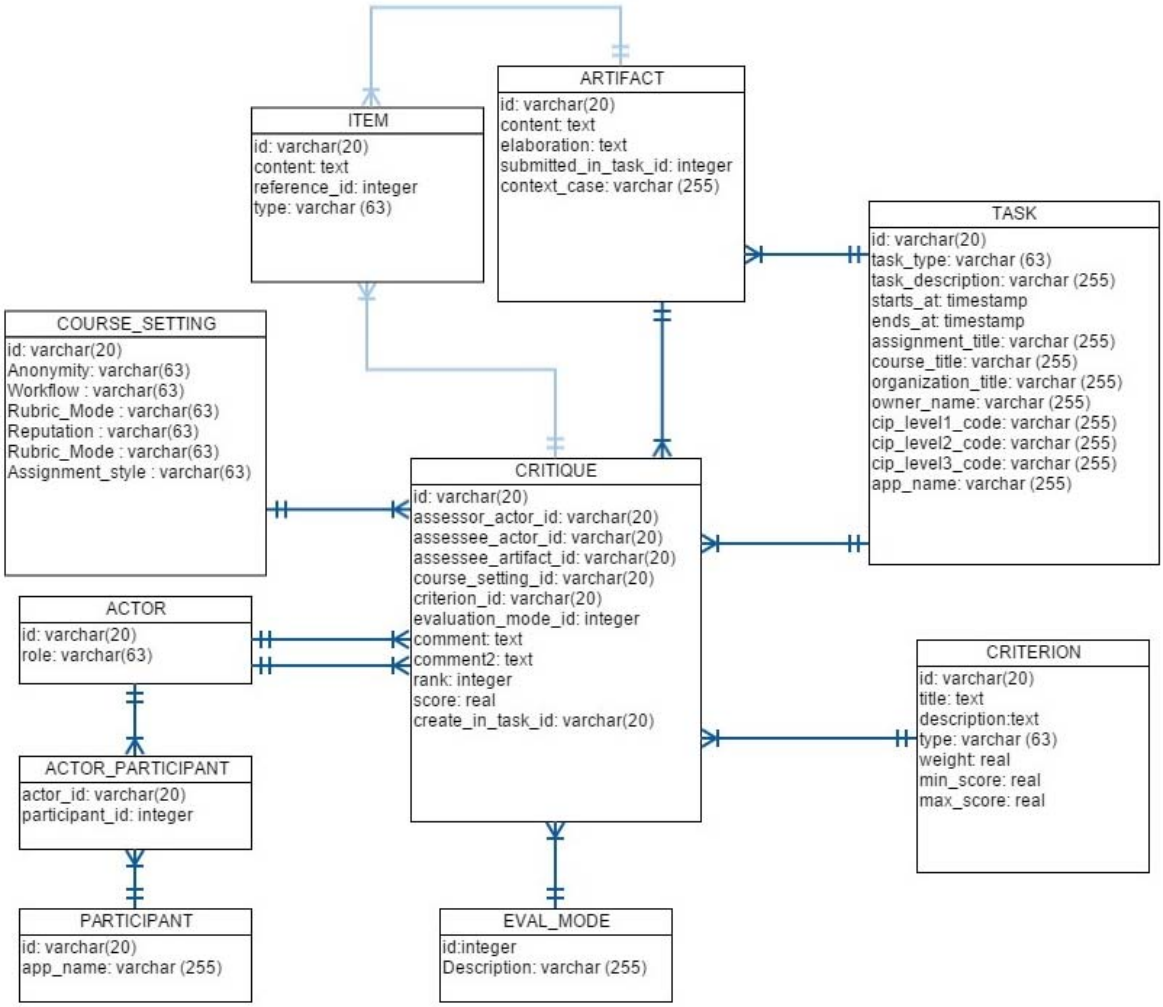

Fig. 2. Data-warehouse schema

The data sharing between an online peer assessment system and the data warehouse can be considered to be an Extract, Transform and Load (ETL) process [12]. Figure 3 shows the process for data sharing process from two systems, Expertiza and CritViz. The data is extracted from the source database, then transformed into the data-warehouse schema and loaded to the staging data warehouse first. (The staging data warehouses serve as temporary repositories for data-verification purposes.) This process is done using Pentaho [13]. The Central data warehouse can pull the data from the staging data warehouses after the data is validated.

The ETL processes for going from the original databases to the staging data warehouses differ for each system. Depending on the design of the peer-assessment system and its own database schema, it could be challenging to code the ETL. We may need to combine data from different tables into a single table, split original data from one table into several tables, or add new data or relationships in this ETL process. For example, in our data-warehouse schema, both reviewer and reviewee are actors, which can either be individuals or teams. The actors are the owners of artifacts for each assignment. This design is to make the data-warehouse schema more generic. However, not all systems are designed like this. CrowdGrader, for example, also supports team authorship, but it is handled differently: there is a lead user of the team, and this user is the owner of the artifact instead of the whole team. Figure 4 shows the difference between CrowdGrader schema and data-warehouse schema. To transform the CrowdGrader data into the data-warehouse schema, we need to create actor entities for the teams and associate the artifacts to the newly created actors. This ETL process should be performed by collaboration between the owners/designers of a system and the designers of the data warehouse.

Besides transforming the data into the data-warehouse schema, the ETL process also anonymizes the original data by skipping all the personal information and not including it into the staging data warehouse. The central data warehouse, thereby, will exclude any personal information of the users'. The data warehouse is purely for research uses and need not include any personal data such as names, emails or campus IDs.

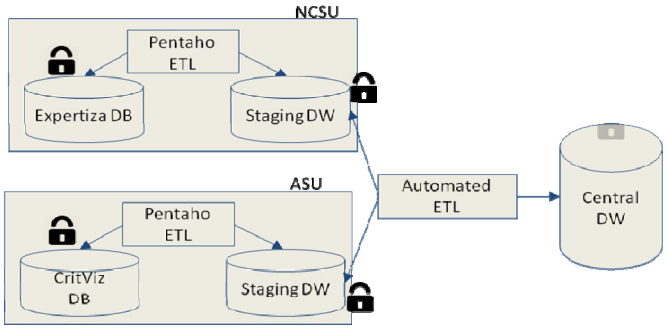

Fig. 3. An example of Extract, Transform and Load (ETL) process

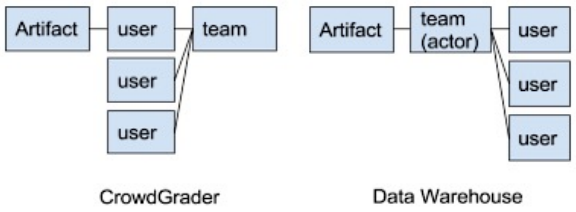

Fig. 4. An example of changing entities relation

We have finished the ETL process for two systems, Expertiza and Critviz. The data warehouse can be accessed by querying the MySQL database at peerlogic.csc.ncsu.edu (note that the db is not accessible via a web browser; however, you can read about querying the data warehouse on our website peerlogic.org). We provide a read-only credential upon request.

IV. Conclusion and Future Work

In this paper, we have presented our work on a Peer-Review Markup Language (PRML), which defines the common entities and relations in the educational peer-assessment process. This common terminology makes it easy for researchers to talk about the process, and more importantly, share peer-assessment data.

Based on the PRML, we further designed a data warehouse which makes data sharing across different systems much easier. Our project is comprised of four systems, of which two have already shared their data with us. The courses whose data is in the data warehouse cover a wide range of majors such as computer science, mechanical engineering, education, and art.

In the future, we will try to involve more systems and researchers in this project. The data warehouse will hold a more extensive set of data in more dimensions, such as number of users, range of educational levels, number of majors, etc. Researchers will be able to apply data mining and data analysis on this data warehouse and derive more general conclusions based on data collected by multiple systems with a large number of users.

Currently, we only provide data access via database queries. We are also building a web-based data visualization tool to facilitate data access for researchers seeking to test their hypotheses and do comparisons.

Acknowledgement

This research is part of the PeerLogic project, which is funded by the National Science Foundation under grants 1432347, 1431856, 1432580, 1432690, and 1431975.

References

[1] Y. Song, Z. Hu, and E. F. Gehringer, "Closing the Circle: Use of Students' Responses for Peer-Assessment Rubric Improvement," in *Advances in Web-Based Learning -- ICWL 2015*, F. W. B. Li, R. Klamma, M. Laanpere, J. Zhang, B. F. Manjón, and R. W. H. Lau, Eds. Springer International Publishing, 2015, pp. 27–36.

[2] Y. Song, Z. Hu, Y. Guo, and E. Gehringer, "An Experiment with Separate Formative and Summative Rubrics in Educational Peer Assessment," in *Submitted to IEEE Frontiers in Education Conference (FIE), 2016*, 2016.

[3] E. Gehringer, L. Ehresman, S. G. Conger, and P. Wagle, *Reusable Learning Objects Through Peer Review: The Expertiza Approach*. .

[4] "Mobius SLIP: UNCG develops a new online learning tool," *Research & Economic Development*, 06-Nov-2013. [Online]. Available: http://research.uncg.edu/spotlight/mobius-slip-uncg-develops-a-new-online-learning-tool/. [Accessed: 08-Jul-2016].

[5] D. Tinapple, L. Olson, and J. Sadauskas, "CritViz: Web-based software supporting peer critique in large creative classrooms," *Bull. Tech. Comm. Learn. Technol.*, vol. 15, no. 1, 2013.

[6] L. de Alfaro and M. Shavlovsky, "CrowdGrader: A Tool for Crowdsourcing the Evaluation of Homework Assignments," in *Proceedings of the 45th ACM Technical Symposium on Computer Science Education*, New York, NY, USA, 2014, pp. 415–420.

[7] A. Luxton-Reilly and A. Luxton-Reilly, "A Systematic Review of Tools that Support Peer Assessment," *Comput. Sci. Educ.*, vol. 19, no. 4, pp. 209–232, 2009.

[8] D. Babik, E. F. Gehringer, J. Kidd, P. Ferry, and T. David, "Probing the Landscape: A Systematic Meta-review of Online Peer Assessment Systems in Education," in *In Proceedings of the CSPRED in EDM 2016*, Raleigh, N.C, 2016.

[9] E. Gehringer, "Expertiza: information management for collaborative learning," *Monit. Assess. Online Collab. Environ. Emergent Comput. Technol. E-Learn. Support*, pp. 143–159, 2009.

[10] Y. Song, Z. Hu, and E. F. Gehringer, "Pluggable reputation systems for peer review: A web-service approach," in *IEEE Frontiers in Education Conference (FIE), 2015. 32614 2015*, 2015, pp. 1–5.

[11] R. Kimball and M. Ross, *The Data Warehouse Toolkit: The Complete Guide to Dimensional Modeling*, 2 edition. New York: Wiley, 2002.

[12] P. Vassiliadis, "A Survey of Extract-Transform-Load Technology," Nov. 2015.

[13] A. S. Pulvirenti and M. C. Roldán, *Pentaho Data Integration 4 Cookbook*. Birmingham, U.K.: Packt Publishing, 2011.