

Transaction-Based Analysis of Piazza and Other Learning Platforms

Collin F. Lynch

cflynch@ncsu.edu

North Carolina State University

Raleigh, North Carolina

ABSTRACT

With the rise of blended courses clickstream data has become increasingly prevalent in education, as has the need to exchange this data across diverse platforms and to evaluate it for research. These twin goals of interoperability and integration make it ever more important for us to develop appropriate standards for data storage and data exchange. In this talk I will briefly describe our ongoing work on the analysis of user activity in Piazza data and the new working group for standards analysis in clickstream data.

ACM Reference Format:

Collin F. Lynch. 2018. Transaction-Based Analysis of Piazza and Other Learning Platforms. In *Proceedings of 2019 CSSPLICE Workshop (CSSPLICE '19)*. ACM, New York, NY, USA, 2 pages. <https://doi.org/10.1145/nnnnnnn.nnnnnn>

1 OVERVIEW

Modern classes are defined not by face-to-face interactions in lectures or office hours but by suites of tools such as online textbooks, student forums, intelligent tutors, and automatic graders, many of which may be class-specific. As a consequence clickstream data has become increasingly important in education, particularly CS Education. Students in introductory CS courses are now expected to work with learning management systems to access course materials, electronic textbooks to and tutoring platforms to work through practice problems, and online forums and ticketing systems to obtain help. This provides new opportunities for intervention and adaptation as well as new challenges for class management, and it raises new questions of privacy, security, and ethics.

On the one hand, this diversity of platforms allows instructors to tailor their course infrastructure to students' needs including the provision of automated feedback and tutoring. And the widespread use of online tools provides us with a wealth of data to better understand how students learn, how they engage in help-seeking [8] and collaborative problem-solving, and how best to adapt to their personalized needs [1, 4]. This data can also support research on new educational interventions such as personalized learning and adaptive collaboration.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

CSSPLICE '19, Feb 27th, 2019, Minneapolis Minnesota

© 2018 Copyright held by the owner/author(s). Publication rights licensed to ACM.

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM... \$15.00

<https://doi.org/10.1145/nnnnnnn.nnnnnn>

Yet, the requirement that students manage multiple learning platforms can serve to add complexity to otherwise difficult courses and to increase student confusion, reduce engagement, and even inhibit learning. Students often face difficulty navigating multiple interaction modes, or even knowing when to switch from one platform to another [5]. As a consequence, students often work in 'silos,' working first on one tool and then on another [7], or go to office hours without first checking for similar questions online, or even skip using some platforms entirely. Moreover the fact that many tools operate with little to no connection between them can add substantial logistical complexity and make it difficult for students and instructors to effectively integrate learning.

And finally, the fact is that all of these opportunities and challenges are predicated on ongoing data collection, data sharing, and data mining. Students' are no longer working in the privacy of their dorm rooms or their personal machines. Their every educational act is now shared with instructors, institutions, researchers, and platform developers. As we collect this data and depend upon it for classroom management and personalized learning it becomes ever more essential for us to implement robust standards for privacy and security, and to come to some agreement on basic questions of ownership, transparency, and even anonymity all of which are central to education.

2 WORKING GROUP

Our goal in this working group is to coordinate discussion on these issues and to help develop appropriate technical and policy standards in three key areas: access methods, metadata, and representations. In particular we will focus on the following questions:

- (1) How can platform providers offer efficient access to clickstream data? And can those models be unified across platforms?
- (2) What level of granularity should we use when recording and reporting transaction data?
- (3) What metadata should be recorded for individual users or actions to support class- and student-level integration?
- (4) What standards of encryption or access can be maintained to prevent intrusion?
- (5) Who should be provided with access to raw clickstream data?
- (6) What standards of retention, deletion, and anonymity should be offered by platform developers and institutions to protect student and instructor privacy and to comply with existing preservation standards.

We will begin by coordinating discussion among working group participants to identify the existing de-facto standards for these questions. We will also work to share open-source code or other analysis tools that can be shared by working group members. Where

possible we will then work to formalize these standards into technical and policy recommendations for the CSSPLICE community.

3 PIAZZA DATA

One of the essential skills for any student is *help seeking* that is knowing when to seek help, where to get help from, and how to ask. Skills of this type are essential in any domain, none more so than in CS Education where students face a wide range of initial hurdles from higher level conceptual issues to basic language skills (both in human and machine languages). For online courses and for many blended courses this help seeking occurs online with students turning to forums to pose questions as a supplement, or even an alternative to posing questions in class, working with their peers, or reading course materials. In order to understand some of what is happening in our courses we need to understand how students are using these online forums to communicate, to exchange information, and to learn over time.

One of the more popular online forums is Piazza¹ which provides an online question and answer forum that supports student questions and instructor or peer replies. It also supports up-voting and down-voting of questions, boxing in folders, and other familiar structures. Piazza is popular at a number of universities including NCSU and CMU. In prior research at NCSU we have analyzed data from the Piazza site to evaluate how students work with online course materials and how they engage in study and help-seeking [6]. And we have performed social network analysis to evaluate how students engage with one another and with instructors in blended courses [2, 3].

4 PIAZZA DATA ANALYSIS TOOL

Piazza data has typically been made available to researchers as shared database dumps with a set MongoDB format. This database format records the complete history for a Piazza course including: student and staff membership; Posts and the history of post edits; replies and edits to replies; as well as likes, folder structure, and other information. The database does not record student page views, nor does it record all anonymized student data.

We previously developed private code for analysis of this data at NCSU with the support of the National Science Foundation (NSF grant #1418269: “Modeling Social Interaction & Performance in STEM Learning” Yoav Bergner, Ryan Baker, Danielle S. McNamara, & Tiffany Barnes Co-PIs). With the support of the CSSPLICE initiative we at NCSU and CMU have jointly collaborated to conduct research on Student-TA communication in Piazza, and to develop a public codebase for the anonymization and analysis of student data. The code is now publicly available² under a GPL License. The code

itself is in Python supported by MongoDB an open-source object database. When given a standard dump this code will clean and anonymize the data and refactor the raw combined format into a series of linked objects. It then provides a standard object-relational interface for accessing the data.

Potential collaborators are invited to access the repository for their own analysis and to contact Dr. Collin F. Lynch (cflynch@ncsu.edu) or Niki Gitinabard (ngitina@ncsu.edu) for additional information

¹<https://www.piazza.com>

²<https://github.com/ArgLab/PiazzaDataAnalysisTool>

on the tool. Users of the tool should cite [2] to reference the tool itself.

REFERENCES

- [1] Peter Brusilovsky, Stephen Edwards, Amruth Kumar, Lauri Malmi, Luciana Benotti, Duane Buck, Petri Ihantola, Rikki Prince, Teemu Sirkiä, Sergey Sosnovsky, Jaime Urquiza, Arto Vihavainen, and Michael Wollowski. 2014. Increasing Adoption of Smart Learning Content for Computer Science Education. In *Proceedings of the Working Group Reports of the 2014 on Innovation #38; Technology in Computer Science Education Conference (ITiCSE-WGR '14)*. ACM, New York, NY, USA, 31–57. <https://doi.org/10.1145/27113609.27113611>
- [2] Niki Gitinabard, Collin F. Lynch, Sarah Heckman, and Tiffany Barnes. 2017. Identifying Student Communities in Blended Courses. In *Proceedings of the 10th International Conference on Educational Data Mining (EDM 2017) Wuhan China*, Xiangen Hu, Tiffany Barnes, Arnon Hershkovitz, and Luc Paquette (Eds.). International Educational Data Mining Society, 402 – 403.
- [3] Niki Gitinabard, Linting Xue, Collin F. Lynch, Sarah Heckman, and Tiffany Barnes. 2017. A Social Network Analysis on Blended Courses. *CoRR* abs/1709.10215 (2017). <http://arxiv.org/abs/1709.10215> Presented at the 2017 Workshop on Graph-Based Educational Data Mining held in conjunction with EDM 2017 Wuhan PRC.
- [4] Petri Ihantola, Arto Vihavainen, Alireza Ahadi, Matthew Butler, Jürgen Börstler, Stephen H. Edwards, Essi Isohanni, Ari Korhonen, Andrew Petersen, Kelly Rivers, Miguel Ángel Rubio, Judy Sheard, Bronius Skupas, Jaime Spacco, Claudia Szabo, and Daniel Toll. 2015. Educational Data Mining and Learning Analytics in Programming: Literature Review and Case Studies. In *Proceedings of the 2015 ITiCSE on Working Group Reports (ITiCSE-WGR '15)*. ACM, New York, NY, USA, 41–63. <https://doi.org/10.1145/2858796.2858798>
- [5] Paul Salvador Inventado, Roberto S. Legaspi, Rafael Cabredo, Koichi Moriyama, Ken-ichi Fukui, Satoshi Kurihara, and Masayuki Numao. 2013. Identification of Effective Learning Behaviors. In *Artificial Intelligence in Education - 16th International Conference, AIED 2013, Memphis, TN, USA, July 9-13, 2013. Proceedings (Lecture Notes in Computer Science)*, H. Chad Lane, Kalina Yacef, Jack Mostow, and Philip I. Pavlik (Eds.), Vol. 7926. Springer, 670–673. https://doi.org/10.1007/978-3-642-39112-5_85
- [6] Adithya Seshadri, Niki Gitinabard, Collin F. Lynch, Tiffany Barnes, and Sarah Heckman. 2018. Predicting Student Performance Based on Online Study Habits: A Study of Blended Courses. In *Proceedings of the 11th International Conference on Educational Data Mining, EDM 2018, Buffalo, NY, USA, July 15-18, 2018*, Kristy Elizabeth Boyer and Michael Yudelson (Eds.). International Educational Data Mining Society (IEDMS), 411 – 417. <http://educationaldatamining.org/EDM2018/proceedings/>
- [7] Adithya Seshadri. 2016. *Sessions Analysis: Mining Online Tool Use to Predict Performance in Blended Courses*. Master’s thesis. North Carolina State University, Raleigh, North Carolina, United States of America.
- [8] Aaron J. Smith, Kristy Elizabeth Boyer, Jeffrey Forbes, Sarah Heckman, and Ketan Mayer-Patel. 2017. My Digital Hand: A Tool for Scaling Up One-to-One Peer Teaching in Support of Computer Science Learning. In *Proceedings of the 2017 ACM SIGCSE Technical Symposium on Computer Science Education, Seattle, WA, USA, March 8-11, 2017*, Michael E. Caspersen, Stephen H. Edwards, Tiffany Barnes, and Daniel D. Garcia (Eds.). ACM, 549–554. <https://doi.org/10.1145/3017680.3017800>