

Building an Infrastructure for Computer Science Education Research and Practice at Scale

Peter Brusilovsky
University of Pittsburgh
Pittsburgh, PA, USA
peterb@pitt.edu

Ken Koedinger
Carnegie Mellon University
Pittsburgh, USA
koedinger@cs.cmu.edu

David A. Joyner
Georgia Tech
Atlanta, USA
david.joyner@gatech.edu

Thomas W. Price
North Carolina State
University
Raleigh, USA
twprice@ncsu.edu

ABSTRACT

The goal of this workshop is to bring together the existing community of researchers working on Infrastructure Design for Data-Intensive Research in Computer Science Education and a community of Learning at Scale researchers focused on Computer Science Education. While both communities share many similar goals and could greatly benefit from each other work, the interaction between the communities is small. We hope that the proposed workshop will be instrumental in bringing together like-minded researchers from different communities, establishing collaboration, and expanding the scope of infrastructure project to address critical scaling issues.

BACKGROUND

Computer Science Education makes heavy use of online educational tools like IDEs, Learning Management Systems (LMS), eTextbooks, and interactive programming environments. Many interactive examples, animations, and auto-assessed exercises, referred to as “smart content”, are now available. Instructors and students would benefit from greater inter-operability between these tools, which is especially critical in a large-scale education context. Computer Science Education (CSEd) researchers increasingly make use of large collections of data generated by click streams coming from such tools. However, students, instructors, and researchers all face barriers that slow progress: (1) While the standards are improving, educational tools still do not integrate well. (2) CSEd process and outcome data from generated by one system are not compatible with that from other systems. (3) Computer science problem solving and learning tasks (e.g., open-ended coding solutions to complex problems) are quite different from

the type of learning tasks that current educational data mining approaches focus on (e.g., discrete answers to questions or verbal responses).

SPLICE is an NSF-supported project and a community of researchers focused on building an Infrastructure for Computer Science Education Research. The infrastructure should support (1) broader re-use of innovative learning content instrumented for rich data collection, (2) extensive data collection from multiple learning tools in inter-operable formats, (3) repositories to share learner data and tools for data analysis, and (4) development of best practices in collecting, sharing, and processing learner data. The goal of this workshop is to bring together the existing community of researchers working on the Infrastructure and a community of Learning at Scale (L@S) researchers focused on CSEd. While both communities share many similar goals and could greatly benefit from each other’s work, the interaction between the communities is small.

Many L@S attendees are either developing or studying online educational tools and tool-enriched learning processes, and all will benefit from better inter-operability among these tools and data as well as better support for analysis of data coming from those tools. While L@S is not focused explicitly on CSEd, it is one of the most popular topics at L@S, making the proposed workshop mutually beneficial for both communities. Past L@S conferences featured a range of research projects focused on tools, approaches, and data in CSEd - such as new tools for smart content creation [6, 8], approaches and tools to teach programming at scale [3, 9, 1, 10], automated assessment [2, 5], and data-driven learning technologies - which could benefit from and contribute to the common infrastructure. Our immediate goal is to focus attention within the CSEd community on current barriers to collaboration, and building community structures to address infrastructure problems.

This workshop follows a series of workshops organized at other conferences relevant to the SPLICE goals - AIED, EDM, SIGCSE, ICER, LAK (see cssplice.github.io). Each of this workshop was successful in bringing like-minded researchers

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.

L@S '20, August 12–14, 2020, Virtual Event, USA.

© 2020 Association for Computing Machinery.

ACM ISBN 978-1-4503-7951-9/20/08 ...\$15.00.

<http://dx.doi.org/10.1145/3386527.3405936>

from different subcommunities into conversation, establishing collaboration, and expanding the scope of the project. We hope that the workshop at L@S will continue this tradition.

WORKSHOP CONTENT AND STRUCTURE

The workshop focus will be on how novel online learning tools inter-operate in the context of a large-scale educational process, how learner data could be collected, integrated and processed at scale, and how data analysis approaches could be accumulated, shared, and reused to improve education research efforts across the discipline and beyond. The organizers will introduce the goals of SPLICE Community, report progress to date, present a range of relevant activities, and plan future collaboration. The attendees will be able to share their relevant research, establish collaboration, get engaged in the SPLICE Community work, and obtain SPLICE funding for collaborative research in the area.

The overarching goal of SPLICE project is to support and better coordinate efforts to build community and capacity among CSED researchers, data scientists, and learning scientists toward reducing barriers. In this context, the primary goal of the workshop itself is to report progress, share best practices, develop an action plan for the coming year, and to build upon our existing collaborations developed over the course of the project to engage more members of the community in tasks that will advance the project agenda. In particular, we will report out on existing collaborations, working groups, and progress on content and data re-use [4, 7].

POST-WORKSHOP PLANS

The workshop structure maximizing attendees' chance to find more about matching projects and expertise of other organizers. As a part of the community building activities, we will encourage forming small groups for collaborative work on the infrastructure issues. Following the workshop, we will request proposals for collaboration project between participants. Most promising projects will be funded by the our NSF-supported project as a part of community-building efforts.

ORGANIZERS

The workshop will be organized by a team of SC-SPLICE project members who have an extensive experience in both research and past workshop organization.

Peter Brusilovsky is a Professor of Information Science and Intelligent Systems at the University of Pittsburgh, where he also directs Personalized Adaptive Web Systems (PAWS) lab. He has been working in the field of adaptive educational systems, user modeling, and intelligent user interfaces for more than 30 years. He published numerous papers and edited several books on adaptive hypermedia and the adaptive Web. He is a founder of CS-SPLICE, an organizer of several SPLICE workshops, and a frequent attendee of L@S conferences.

Ken Koedinger is Professor of Human-Computer Interaction and Psychology at Carnegie Mellon. He explores how people think and learn by developing and studying technology-enhanced learning. He leads the LearnSphere effort (learnsphere.org), which integrates learning data and

analytics across multiple resources. And he directs LearnLab (learnlab.org), which started with 10 years of National Science Foundation funding and is now the scientific arm of CMU's Simon Initiative (cmu.edu/simon). He is also a founder of CS-SPLICE and a frequent attendee and presenter at L@S conferences.

David Joiner is Associate Director of Student Experience at Georgia Tech's College of Computing and a director of LucyLabs, an online research lab. David's work focuses on online learning not through MOOCs, but through large online classrooms. He is interested in the unique opportunities these classes have for personalizing student learning and granting students greater ownership and autonomy over their education. David is one of the few researchers who contributed to both L@S and SPLICE communities attending and presenting at several L@S conferences and SPLICE workshops.

Thomas Price is an Assistant Professor of Computer Science at North Carolina State University. His primary research goal is to develop learning environments that automatically support students through AI and data-driven help features. His work has focused on the domain of computing education, where he has developed techniques for automatically generating programming hints and feedback for students in real-time by leveraging student data. He has helped organized a number of efforts at the intersection of AIED, Data Mining and CS Education, including the CS-SPLICE working group on programming snapshot representation and prior CSED and CS-SPLICE workshops.

ACKNOWLEDGEMENTS

This material is based upon work supported by the National Science Foundation under Grant No. 1740775. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

REFERENCES

- [1] Charles Chen and Philip Guo. 2018. Improv: Teaching Programming at Scale via Live Coding. In *Proceedings of the 6th Annual ACM Conference on Learning at Scale*. ACM, Article 9.
- [2] Andrew Head, Elena Glassman, Gustavo Soares, Ryo Suzuki, Lucas Figueredo, Loris D'Antoni, and Björn Hartmann. 2017. Writing Reusable Code Feedback at Scale with Mixed-Initiative Program Synthesis. In *Proceedings of the Fourth (2017) ACM Conference on Learning at Scale*. ACM, 89–98.
- [3] David A. Joyner. 2018. Toward CS1 at scale: building and testing a MOOC-for-credit candidate. In *Proceedings of the Fifth Annual ACM Conference on Learning at Scale*. ACM, 59:1–59:10.
- [4] Hamza Manzoor, Kamil Akhuseyinoglu, Jackson Wonderly, Peter Brusilovsky, and Clifford A. Shaffer. 2019. Crossing the Borders: Re-Use of Smart Learning Objects in Advanced Content Access Systems. *Future Internet* 11, 7 (2019), 160.

- [5] Joseph Moghadam, Rohan Choudhury, HeZheng Yin, and Armando Fox. 2015. AutoStyle: Toward Coding Style Feedback at Scale. In *Proceedings of the Second (2015) ACM Conference on Learning at Scale*. ACM, 261–266.
- [6] Jungkook Park, Yeong Hoon Park, Jinhan Kim, Jeongmin Cha, Suin Kim, and Alice Oh. 2018. Elicast: Embedding Interactive Exercises in Instructional Programming Screencasts. In *Proceedings of the Fifth Annual ACM Conference on Learning at Scale*. ACM, 58:1–58:10.
- [7] Thomas W. Price, David Hovemeyer, Kelly Rivers, Ge Gao, Austin Cory Bart, Ayaan M. Kazerouni, Brett A. Becker, Andrew Petersen, Luke Gusukuma, Stephen H. Edwards, and David Babcock. 2020. ProgSnap2: A Flexible Format for Programming Process Data. In *Proceedings of the 2020 ACM Conference on Innovation and Technology in Computer Science Education*. ACM, 356–362.
- [8] Rémi Sharrock, Ella Hamonic, Mathias Hiron, and Sebastien Carlier. 2017. CODECAST: An Innovative Technology to Facilitate Teaching and Learning Computer Programming in a C Language Online Course. In *Proceedings of the Fourth ACM Conference on Learning at Scale*. ACM, 147–148.
- [9] Ralf Teusner, Thomas Hille, and Thomas Staubitz. 2018. Effects of automated interventions in programming assignments: evidence from a field experiment. In *Proceedings of the Fifth Annual ACM Conference on Learning at Scale*. ACM, 59:1–59:10.
- [10] Eliane Wiese, Michael Yen, Antares Chen, Lucas Santos, and Armando Fox. 2017. Teaching Students to Recognize and Implement Good Coding Style. In *Proceedings of the Fourth (2017) ACM Conference on Learning @ Scale*. ACM, 41–50.