An Infrastructure for Sustainable Innovation and Research in Computer Science Education (SPLICE)

Peter Brusilovsky, Clifford Shaffer, Kenneth Koedinger, Tiffany Barnes, Thomas Price

Introduction and Objectives

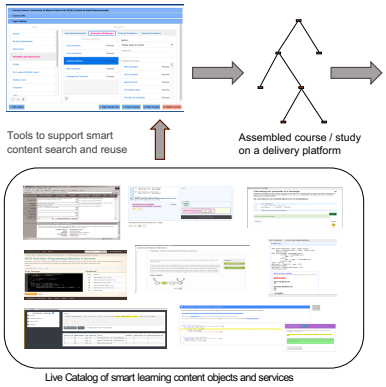
Researchers from the University of Pittsburgh, Carnegie Mellon University, North Carolina State University, and Virginia Tech collaborate to support the SPLICE Portal, a social and technical Infrastructure for Sustainable Innovation and Research in Computing Education. Our goal is to accelerate and scale research on teaching and learning of computing disciplines by leveraging the power of data-driven AI and ML. infrastructure. The Portal facilitates creation and adoption of new tools to support computing educators and students, and new standards to support data collection and data-enabled research. Cross-disciplinary scientific advances are disseminated through the SPLICE infrastructure community and through scientific publications.

The project aims to expand re-use of educationally efficient *smart learning content and services* (SLC). SLC is different from traditional learning content: every SLC item is an interactive service that communicates with the learner working with a specific learning activity (a problem or a worked example), collects data, and provides feedback. Rich data collected during learner interaction with SLC open opportunities for personalization and provide important insights about the learning process. The re-usability of SLC components is supported by their adherence to several standards such as LTI and Caliper, ensuring the connection of SLC to learning management systems and allowing centralized collection of learner data.

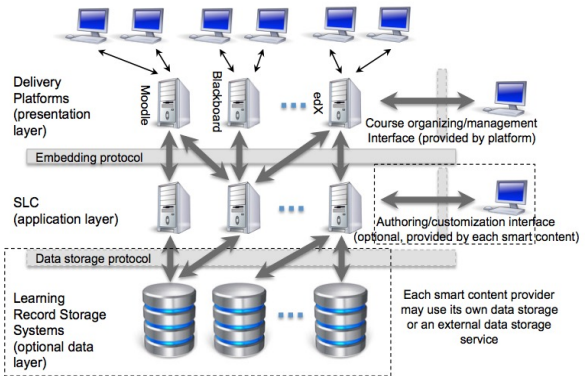
Our development efforts are focused on two hubs: the content hub and the data hub. The content hub is centered around a repository of reusable SLC developed by the project partners and collaborators worldwide. It provides a range of services to contribute and find SLC as well as to assemble a collection of SLC for a course or a research study. The data collected during student work with SLC in a course or a study are archived in the data hub, which provides support for contributing and finding datasets. The data hub also offer an extensive collection of tools to analyze collected data. Adherence to data representation standards and re-usability of analysis components facilitate data analysis and reduce the entry barriers for instructors and other users without data mining background.

The SPLICE Infrastructure

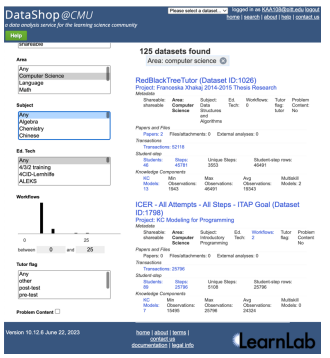
The Content Hub: Course / Study Assembly



The Infrastructure in Use: Class / Study Support



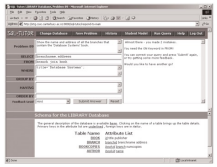
The Data Hub: Data Collection and Analysis



DataShop learning data archive with search and analysis tools

Smart Learning Content (SLC)

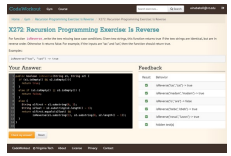
The content hub brings together a broad variety of SLC developed by project partners and collaborators. Here are some examples of Smart Learning Content interfaces developed for courses on Java, Python, Algorithms, and Databases, which reside on the application layer in the SPLICE infrastructure (see diagram).



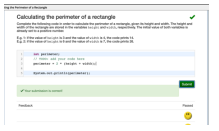
SQL Tutor, University of of Canterbury, New Zealand

Jsvee: Python code visualization, Aalto University, Finland

OpenDSA algorithm exercises, Virginia Tech

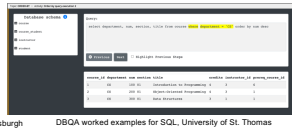


CodeWorkout coding problems, Virginia Tech



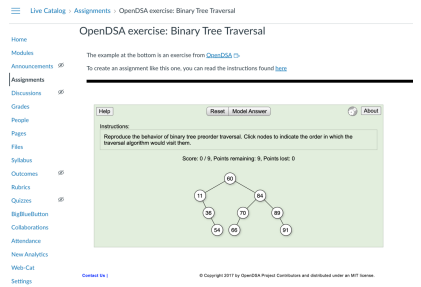
PCRS coding problems, University of Toronto, Canada

QuizJet tracing problems, University of Pittsburgh

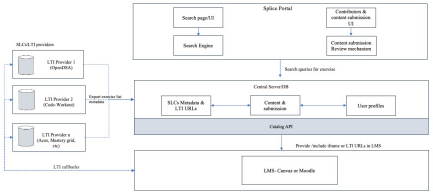


DBQA worked examples for SQL, University of St. Thomas

Live Catalog

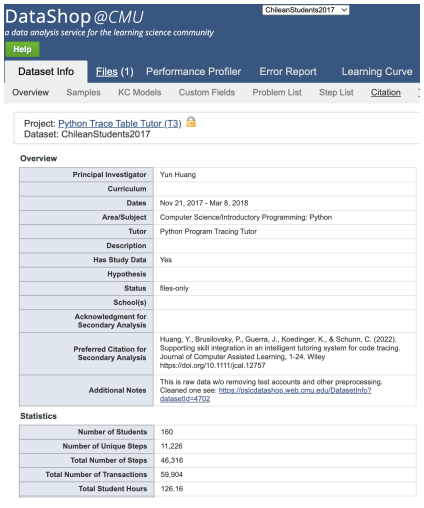


Live Catalog, the main component of the content hub, is a centralized repository designed to streamline the discovery and use of LTI-compatible SLCs. The catalog aggregates a range of learning content from various LTI tool providers. The portal includes a "live" display functionality, where users can choose relevant SLC items by interacting with SLCs directly. The first version of the Live Catalog is now available on Canvas Instructure. An enhanced version of the Live Catalog is currently under development (see below).



Learn Sphere's DataShop

The data hub is centered around a repository of reusable learning datasets hosted on LearnSphere's DataShop, the world's largest open repository for educational technology data. The use of LearnSphere Workflow mechanism enables the users to reuse not just datasets, but also data analytics. Workflows offer an extendable collection of data import, processing, and data mining components, which could be assembled into structured pipelines.



Overview

Principal Investigator	Yun Huang
Curriculum	
Dates	Nov 21, 2017 - Mar 8, 2018
Area/Subject	Computer Science/Introductory Programming: Python
Tutor	Python Program Tracing Tutor
Description	
Has Study Data	Yes
Hypothesis	
Status	files-only
School(s)	
Acknowledgment for Secondary Analysis	
Preferred Citation for Secondary Analysis	Huang, Y., Brusilovsky, P., Guerra, J., Koedinger, K., & Schunn, C. (2022). Supporting skill integration in an intelligent tutoring system for code tracing. Journal of Computer Assisted Learning, 1-24. Wiley https://doi.org/10.1111/jcal.12757
Additional Notes	This is raw data w/o removing test accounts and other preprocessing. Cleaned one see: https://pslcdatashop.web.cmu.edu/DatasetInfo?datasetId=4702

Statistics

Number of Students	160
Number of Unique Steps	11,226
Total Number of Steps	46,316
Total Number of Transactions	59,904
Total Student Hours	126.16

Recent Community Events

- CS Education Research Track at Simon Initiative LearnLab Summer School, Pittsburgh, PA, July 24-28, 2023
- 7th Educational Data Mining in Computer Science Education (CSEDM) Workshop In conjunction with LAK 2023, Arlington, TX, USA - March 13, 2023
- 8th SPLICE Workshop on Technology and Data Infrastructure for CS Education Research in conjunction with SIGCSE 2023, Toronto, March 15, 2023.

Community portal at http://cssplice.org/

We maintain the website at http://cssplice.org/ Here we share all project information including standards, best practices, and resources. It informs the community about events and opportunities to contribute. The site provides links to project publications, code, data, and learning content hosted in archival repositories such as GitHub and DataShop. It also provides space to share information about activities and products of community working groups (below).

SPLICE Working Groups

- Smart Learning Content Protocols. Leaders: Cay Horstmann, Steve Edwards, Cliff Shaffer
- Parsons Problems Interoperability Standards. Leader: Cliff Shaffer and [need some more!]. Google Group.
- Reusable Code Examples. Leaders: Peter Brusilovsky, Vasile Rus
- Large Language Models. Leader: Juho Leinonen. Google Group.
- Small Code Snapshots (ProgSnap). Leaders: Thomas Price and Ayaan Kazerouni.
- Programming Exercise Markup Language (PEML). Leaders: Phill Conrad, Cory Bart, and Stephen Edwards.
- Packaging Curricular Materials. Leaders: Cory Bart, Phill Conrad, Michael Hilton, Bob Edmison.
- Interoperable Data Collection. Leaders: Michael Yudelson.