

Invited Talk: Mathematics in the Scholarly Literature

Lucy Lu Wang

Allen Institute for AI

Abstract

Formulas, equations, proofs, and symbols are used throughout the scholarly literature, in mathematics papers, but also throughout engineering, the sciences, and beyond. Modeling the interplay between scientific text and mathematics is an important part of document understanding. In this talk, I will introduce corpora and tools released by Semantic Scholar that can be used to study mathematics in context. Specifically, I will discuss S2ORC, a structured corpus of scientific papers created to support scientific NLP and text mining. S2ORC contains over 700K mathematics papers, along with many other papers in math-heavy domains such as Computer Science and Physics, a treasure trove of mathematics content. I will also briefly discuss several ongoing initiatives at Semantic Scholar that aim to understand and explain math equations using the paper context around these equations.