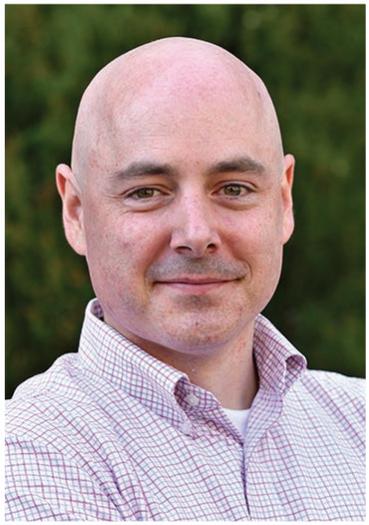


Distinguished Lecture Series

EECS COLLOQUIUM

Spring 2016



DeepDive: A Dark Data System

Chris Ré

Professor, InfoLab, Department of Computer Science
Stanford University

Abstract

Many pressing questions in science are macroscopic, as they require scientists to integrate information from numerous data sources, often expressed in natural languages or in graphics; these forms of media are fraught with imprecision and ambiguity and so are difficult for machines to understand. Here I describe DeepDive, which is a new type of system designed to cope with these problems. It combines extraction, integration and prediction into one system. For some paleobiology and materials science tasks, DeepDive-based systems have surpassed human volunteers in data quantity and quality (recall and precision). DeepDive is also used by scientists in areas including genomics and drug repurposing, by a number of companies involved in various forms of search, and by law enforcement in the fight against human trafficking. DeepDive does not allow users to write algorithms; instead, it asks them to write only features. A key technical challenge is scaling up the resulting inference and learning engine, and I will describe our line of work in computing without using traditional synchronization methods including Hogwild! and DimmWitted.

DeepDive is open source on github and available from DeepDive.Stanford.Edu.

Biography

Christopher (Chris) Ré is an assistant professor in the Department of Computer Science at Stanford University and a Robert N. Noyce Family Faculty Scholar. His work's goal is to enable users and developers to build applications that more deeply understand and exploit data. Chris received his PhD from the University of Washington in Seattle under the supervision of Dan Suciu. For his PhD work in probabilistic data management, Chris received the SIGMOD 2010 Jim Gray Dissertation Award. He then spent four wonderful years on the faculty of the University of Wisconsin, Madison, before moving to Stanford in 2013.

Wednesday

February 3

4:00 - 5:00 pm

Refreshments

served at 3:30

Hewlett-Packard

Auditorium

306 Soda Hall

For more information please see:

<http://www.eecs.berkeley.edu/Colloquium/>