ANANT BHARDWAJ

MIT Computer Science & Artificial Intelligence Laboratory
http://people.csail.mit.edu/anantb/   anantb@csail.mit.edu

Education

Massachusetts Institute of Technology	09/2012 - present
Ph.D. Student, MIT CSAIL	
Advisors: David Karger, and Samuel Madden	
Stanford University	09/2010 - 06/2012
M.S., Computer Science	
Advisor: Scott Klemmer	
Pune University	07/2003 - 06/2007
B.E., Computer Engineering	
First Class with Distinction	

Research Interests

data-management platforms/tools, databases, distributed systems, human-computer interaction

Employment

Massachusetts Institute of Technology	09/2012 - present
Research Assistant: MIT CSAIL (Haystack Group, and Database Group)	
Google	06/2014 - 09/2014
Research Intern: Search Team	
Massachusetts Institute of Technology	01/2014 - 05/2014
Teaching Assistant: User Interface Design and Implementation	
Google	05/2013 - 08/2013
Research Intern: Search Team	
Sookasa	06/2012 - 08/2012
Engineering Consultant: Server Infrastructure	
Stanford University	04/2012 - 06/2012
Teaching Assistant: Web Applications	
Stanford University	01/2012 - 03/2012
Teaching Assistant: HCI - Interaction Design Studio	
Stanford University	09/2011 - 12/2011
Teaching Assistant: OOP from Modeling and Simulation Perspective	
Oracle	06/2011 - 09/2011
S/W Engineering Intern: Identity Management Team	
Stanford University	01/2011 - 06/2011
Research Assistant: Elias Lab	
BMC Software	07/2007 - 08/2010
Product Developer: Marimba	

Projects

DataHub [*Anant Bhardwaj, David Karger, Sam Madden*]
DataHub is a unified, managed, collaborative platform for making data-processing easy. It consists of: (1) *a flexible data store* (files, relational databases, extensible to other data-storage backends) with sharing/collaboration capabilities, managed on behalf of different users/groups, and (2) *an app ecosystem* that hosts apps for various data-processing activities such as ingestion, curation, integration, discovery, query, analytics, visualization, and machine learning. The DataHub users can use any of the apps from the App Center for processing their data as it fits their need.

Confer [*Anant Bhardwaj, Juho Kim, David Karger, Sam Madden, Rob Miller, Haoqi Zhang*]
Confer is a tool for conference planning. It helps a) conference attendees find interesting papers/talks,

organize schedule, and discover people with similar interests, and b) conference organizers schedule sessions, plan community interactions, understand community structure, and discover new areas of research, practice, methodologies, and emerging application areas. It has been deployed at 15 academic conferences including CHI, CSCW, KDD, ACM MM, SIGMOD, SIGIR, and WSDM, and has more than 20,000 unique users.

Distill [*Anant Bhardwaj, Eugene Wu, Aditya Parameswaran, Sam Madden, David Karger*]
Distill is a general purpose example-based data cleaning/extraction tool for converting semi-structured text into a structured table. A user provides a few examples (2 or 3 examples) by specifying the desired tabular output for a given sample input text. The system uses the sample examples to automatically infer a model which can be used to extract the complete table from the raw file.

Barista [*Anant Bhardwaj, Rebecca Taft, Manasi Vartak, David Goehring*]
Barista is a distributed, synchronously replicated, fault tolerant relational data store. It runs as a middleware service over database instances to provide an abstraction for a distributed relational store. It ensures that the data is replicated across many sets of Paxos state machines in replica groups to provide fault-tolerance and recovery. The replication enables load balancing and availability; clients automatically failover between replicas. Barista exposes SQL for data management. Client applications can use Barista with the same SQL code they used before, and under the hood it guarantees replication, consistency, and fault-tolerance seamlessly.

Mail-X [*Anant Bhardwaj, Amy Zhang, David Karger*]
The problem with current mailing lists is that they are all-or-nothing; when one sends an email, everyone gets it. People need to feel like their email is very important before they use the list, which makes polite people reluctant to do it at all. Many not-so-polite people do reply all and send to entire list which causes inbox flooding. Many unsubscribe from the list and miss useful information. Mail-X is a mailing list manager that allows more controlled, yet effective information dissemination.

Semantic Index [*Anant Bhardwaj, Ekaterina Ob'yedkova, Ravi Netravali*]
Current search engines use inverted index to encode natural language text. Unfortunately, it fails to capture the semantic understanding and thus can only go as far as giving relevant document hits. This project introduces an indexing scheme which can capture semantic understanding.

Voice-X [*Anant Bhardwaj, Trisha Kothari, Sakshi Agarwal, John Butler, Chad McClymonds, Joshua Cohen, David Karger, Terry Winograd*]
Voice-X is an infrastructure designed for rural areas/slums where people do not have computers/internet access. The goal is to enable people to search for information, generate and share information, exchange knowledge, be social, and much more, all through their feature phones, over SMS.

Redprint [*Anant Bhardwaj, Dave Luciano, Scott Klemmer*]
Redprint is a browser based IDE for PHP that integrates web search into the development environment to assist programmers with context specific example code in real time, thus reducing the cost of searching and the time to develop software.

Sookasa [*Anant Bhardwaj, Abhinay Nagpal*]
Sookasa is a security and data management layer on top of cloud services such as Dropbox. Wrote the initial system which later got converted into a commercial company (http://sookasa.com).

Data Wrangler [*Sean Kandel, Anant Bhardwaj, Jeff Heer*]
Data Wrangler is an interactive system for creating data transformations. It combines direct manipulation of visualized data with automatic inference of relevant transforms, enabling analysts to iteratively explore the space of applicable operations and preview their effects.

Publications

Anant Bhardwaj, Amol Deshpande, Aaron Elmore, David Karger, Sam Madden, Aditya Parameswaran, Harihar Subramanyam, Eugene Wu, Rebecca Zhang. *Collaborative Data Analytics with DataHub.* VLDB 2015.

Anant Bhardwaj, Souvik Bhattacherjee, Amit Chavan, Amol Deshpande, Aaron J Elmore, Samuel Madden, and Aditya G Parameswaran. *DataHub: Collaborative Data Science & Dataset Version Management at Scale.* CIDR 2015.

Anant Bhardwaj, Juho Kim, Steven P. Dow, David Karger, Sam Madden, Robert C. Miller, and Haoqi Zhang. *Attendee-Sourcing: Exploring the Design Space of Community-Informed Conference Scheduling.*

HCOMP 2014.

Juho Kim, Haoqi Zhang, Paul Andre, Lydia B. Chilton, Anant Bhardwaj, David Karger, Steven P. Dow, and Robert C. Miller. *Cobi: Community-Informed Conference Scheduling.* HCOMP 2013.

Anant P. Bhardwaj, Dave Luciano, and Scott R. Klemmer. *Redprint: Integrating API specific "instant example" and "instant documentation" display interface in IDEs.* UIST 2011.

Invited Talks

DataHub: A Hosted Platform for Organizing, Managing, Sharing, Collaborating, and Processing Data

University of Washington – Seattle	Jan 19, 2015
Microsoft Research – Redmond	Jan 15, 2015
Indian Institute of Technology (IIT) Bombay – Mumbai, India	Jan 07, 2015

Relevant Coursework

Massachusetts Institute of Technology
Database Systems, Distributed Systems, Theory of Computation, Natural Language Processing, and User Interface Design and Implementation.

Stanford University
Operating Systems, Computer Networks, Web Applications, Information Retrieval and Web Search, Machine Learning, Object Oriented Programming from Modeling and Simulation Perspective, Introduction to HCI, HCI Design Studio, Research Topics in HCI, Designing Computer Science Experiments, and Designing Liberation Technology.

Pune University
Artificial Intelligence, Design and Analysis of Algorithms, Programing Languages, and Software Engineering.

Teaching

User Interface Design and Implementation	Spring 2014

Teaching Assistant for Prof. Phil Guo, Prof. Daniel Jackson, Prof. David Karger, and Prof. Rob Miller, Massachusetts Institute of Technology

Introduction to Programming in Java	IAP 2013

Instructor, Massachusetts Institute of Technology

Web Applications	Spring 2012

Teaching Assistant for Prof. John Ousterhout, Stanford University

HCI: Interaction Design Studio	Winter 2012

Teaching Assistant for Prof. Jeff Heer, Stanford University

Object Oriented Programming from Modeling and Simulation Perspective	Fall 2011

Teaching Assistant for Prof. David Cheriton, Stanford University

Service

Program Committee: CHI (2014, 2013)
Reviewer: CHI (2015, 2014), ICDE (2014), SIGMOD (2015, 2014), VLDB (2014)
Chair: MIT CSAIL Student Workshop 2013
M.S. Faculty Liaison: Stanford University (2010-2011)
Student Volunteer: CHI (2015)

Awards

Finalist, Facebook Research Fellowship	2015
American Alumni Association Student Grant	2010
K.C. Mahindra Scholarship	2010