

Reading class

Serge Abiteboul

Introduction

General presentations by Serge (2 classes)

- How to write a scientific article
- The life of a scientist (professor or researcher)
- How to choose a research topic
- Scientific practical ethics

Student presentations (one per student)

- Presentation of an article (1 hour) or a book chapter, possibly several, in one of the topics described further
 - Presentation of the general area
 - Presentation of the specific article/chapter (s)
- Questions (30mn to 1 hour)

Deliverable

- Slides
- Short scientific mediation style article (one or two pages)
 - For the style see « Il était une fois ma thèse » in <http://binaire.blog.lemonde.fr>
- Annotated bibliography (up to 5 pages)

Topics

- database theory
- database systems
- ethics in data management

The articles can be chosen in the following list. Other articles can be chosen after validation by Serge.

The idea is to present (i) the context, (ii) specific contributions, (iii) further work and/or criticism of the contributions. Please use extensively:

- <https://scholar.google.fr/>
- <http://dblp.uni-trier.de/>
- <https://en.wikipedia.org>

Some material taken from:

- Alice: Foundations of Databases (Addison Wesley), Serge Abiteboul, Rick Hull and Victor Vianu.
 - <http://webdam.inria.fr/Alice/>
- Jorge: Web data management (Cambridge University Press 2011), Serge Abiteboul, Ioana Manolescu, Philippe Rigaux, Marie-Christine Rousset, and Pierre Senellart
 - <http://webdam.inria.fr/Jorge/>

Database theory

1. Incomplete information
 - a. Alice's chapter 19
 - b. T. Imielinski et al. Incomplete Information in Relational Databases, JACM 1994
2. Probabilistic databases
 - a. Suciu et al. Probabilistic Databases, Synthesis Lectures, 2011.
 - b. Google scholar
 - c. One doesn't have to consider the entire book
3. Provenance
 - a. Cheney, Chiticariu, Tan, Provenance in Databases: Why, How and Where, Foundations and Trends in Databases, 2009.
 - b. Todd J. Green, Grigoris Karvounarakis, and Val Tannen. 2007. Provenance semirings. In *Proceedings of the twenty-sixth ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems (PODS '07)*. ACM, New York, NY, USA, 31-40. DOI: <https://doi.org/10.1145/1265530.1265535>
 - c.
 - d. <http://pierre.senellart.com/publications/senellart2017provenance.pdf>
4. Datalog
 - a. Alice's Chapter 12,13

- b. T. J. Green, S. Huang, B. T. Loo, W. Zhou, Datalog and Recursive Query Processing, Foundations and Trends in Databases, Vol. 5, 2012.

Web data management

1. Distributed hash tables
 - a. In Jorge
2. Web search engine
 - a. In Jorge
 - b. The Anatomy of a Large-Scale Hypertextual Web Search Engine, Sergey Brin and Lawrence Page

Ethics in data management

1. Private data analysis
 - a. C. Dwork, A Firm Foundation for Private Data Analysis, Communications of the ACM, 2011.
2. Ethical issues in data analysis
 - a. Weapons of math destruction, Cathy O'Neil
3. Personal information systems
 - a. Managing your digital life with a Personal information management system, Serge Abiteboul, Benjamin André, Daniel Kaplan, Communications of the ACM, ACM, 2015, 58 (5), pp.32-35
 - b. PIMS, S. Abiteboul, Amélie Marian, <https://www.slideshare.net/ameliemarian/personal-information-management-systems-edbticdt15-tutorial>

More articles may be found from

U. Wisconsin:

https://www.cs.wisc.edu/sites/default/files/pictures/Database%20systems%20qual_Summer%202014.pdf.

Stanford: <http://cs.stanford.edu/people/chrimre/cs345/cs345.html>

CMU: <http://www.cs.cmu.edu/~pavlo/courses/fall2013/reading-list.html>

Utah: <https://utah.instructure.com/courses/462398/assignments/syllabus>

Possible ideas

- Crowdsourcing
- Distributed data store
- NoSQL
- Blockchain
- XML query language