



La lettre du Collège de France

7 | 2015

Collège de France Newsletter n°7

Data Sciences: From First Order Logic to the Web

Serge Abiteboul



Electronic version

URL: <https://journals.openedition.org/lettre-cdf/2669>

DOI: 10.4000/lettre-cdf.2669

ISSN: 2109-9219

Publisher

Collège de France

Electronic reference

Serge Abiteboul, "Data Sciences: From First Order Logic to the Web", *La lettre du Collège de France* [Online], 7 | 2015, Online since 02 November 2015, connection on 17 August 2022. URL: <http://journals.openedition.org/lettre-cdf/2669> ; DOI: <https://doi.org/10.4000/lettre-cdf.2669>

This text was automatically generated on 17 August 2022.

All rights reserved

Data Sciences: From First Order Logic to the Web

Serge Abiteboul

EDITOR'S NOTE

Excerpts from the Inaugural Lecture 8 March 2012

Source: *La lettre*, no. 33, May 2012

- 1 **To obtain information, we can query a *database management system*.**
- 2 To do so, we express our queries in a simple computer language, perhaps using graphics, perhaps even in our natural language¹. The system translates this request into a formal language. This consists of a syntax, which allows the user's query to be specified, and a formal semantics that gives this syntax an exact meaning. Mathematical logic allows for this kind of formal language. In this lecture, I will discuss the profound ties between what I here call *data sciences* and mathematical logic or, to be more precise, *first order logic*.
- 3 We will consider the *Web's information systems* that serve as entry points to information of all sorts. The most widespread examples of this kind of system are search engines like Google, which provides an index to billions of documents on the Web, and in a way allows us to think of the Web as a gigantic database. As for social network systems like Facebook, they serve as entry points to hundreds of millions of users' personal data.
- 4 The Web's information systems, just like centralized data management systems, are mediators between intelligent individuals with little desire to trouble themselves with programming details, and physical objects like discs or USB keys. We are interested in intelligent systems that can manage information, understand it and make it available to human users. This last sentence deliberately echoes an anthropomorphic view of computer systems.
- 5 We interact with machines that are becoming ever more autonomous, more difficult to distinguish from human beings. While the intelligence of a database management

system is a small step toward artificial intelligence as defined by Alan Turing, the intelligence of the Web is a recent consideration, from both a philosophical and a scientific point of view. In this lecture I will discuss the emergence of collective knowledge fuelled by the sharing of large volumes of information. We will try to imagine what tomorrow's Web might be like with millions, perhaps even billions of interconnected machines reasoning collectively.

- 6 The World Wide Web, introduced by Tim Berners-Lee and Robert Cailliau around 1990, is based on hypermedia documents. This is the Web to which we so quickly became accustomed. The information is in a natural language and the texts are loosely structured with HTML tags, for example for titles or enumerations. The anchors on which Internet users can click lead not only to other HTML pages but also to pictures, music and videos. In this part, I will talk about one of the Web's greatest success stories, the *search engine*. Web search engines save us from having to tediously browse through the multitude of pages, and instead plunge us into a global digital Library.
- 7 The search engine sees the Web as a global library. The Internet user searches for information, and even though the Web most likely cannot answer all of his or her questions, this information may be somewhere amongst the truly extraordinary masses of information and knowledge it holds.
- 8 Like children, we marvel at the tens of billions of documents on the Web. But from the youngest age a child learns to evaluate, classify, and filter the considerable mass of information it encounters. What about us? If the search engine did not help us to focus on a small number of pages, what would we do? The technical exploit lies in finding, within an instant and thanks to its index, the Web pages that contain the words of the query. The magic part is that, from among tens or even hundreds of millions of possible pages, the search engine comes up with the few pages that so often contain what the user is looking for.

- Videos of lectures available online at www.college-de-france.fr

- The Collège de France Chair of Information Technology and Digital Sciences is supported by the Institut national de recherche en informatique et en automatique (INRIA – French National Research Institute for Computer Science and Applied Mathematics).

NOTES

1. “Natural languages” refer to languages elaborated over time by groups of speakers, like French or English. This is not so much in opposition with “constructed” languages like Esperanto, as with formal languages like first order logic, SQL or Java.

AUTHOR

SERGE ABITEBOUL

Researcher at the French National Research Institute for Computer Science and Applied Mathematics (INRIA), member of the French Academy of Sciences, and of the Laboratoire Spécification et Vérification (LSV) at Cachan's École normale supérieure.