

# BIG DATA: TOWARDS BUILDING CUSTOM IN-HOUSE SOLUTIONS

Alexander Trousov

*RANEPA, Russia*

## **Abstract**

Off-the-shelf business intelligence tools often do not meet the needs of clients who want to derive custom insights from their data. The lecturer, based on his personal experience in building solutions for IBM and papers by several experts in the domain, concurs with this opinion: medium-to-large organizations with access to strong technical talent usually prefer to build custom, in-house solutions. If you work for such organization or you are a student or a researcher applying for a position in such organization, you need not only to have engineering skills in handling Big Data, you must have certain knowledge about what works and what doesn't work. Although the demand for big data analytics is high, there is currently a shortage of data scientists and other analysts who have experience working with Big Data. This talk is a short guide for choosing methods suitable for mining Big Data.

We present the best practices for Big Data paradigm (if you prefer to think in terms of Thomas Kuhn's approach to the Structure of Scientific Revolutions) or ideologeme (when we consider Big Data as social phenomena, where the term has no precise definition and is used pragmatically). We trace the origins of Big Data paradigm as the modern stage of databases and knowledge management systems, as well as enterprises software tools for social collaboration and team collaboration (such as Microsoft SharePoint, IBM Lotus Notes and others). We overview various aspects of Big Data, including ontological (the dichotomy between top-down and bottom-up approaches to knowledge creation and use), artificial Intelligence, and cultural aspects (including utopian and dystopian views related to neural realism).

We critically review misuse of Big Data & Artificial Intelligence terms by journalists, big vendors, and groups of researchers who have vested interest in promoting particular methods. We provide our take on most popular ideas in this area, explaining why some of them are blatantly false (noisily opinionated, fall into the category of propaganda

or marketing tricks, etc.), while some of this popular opinions do have some currency among serious researchers and business decision makers. For instance, we explain what is wrong and what is right in the slogan Big data is about what, not why. We dont always need to know the cause of a phenomenon; rather, we can let data speak for itself., because your position regarding statements like this will significantly affect your choice of methods suitable for solving particular problems. This rather high-level and somewhat abstract exposition will be illustrated by very specific comments (and these comments will probably take more time than the exposition itself), allowing to grasp the implications for concrete practical tasks. For instance, in the talk about approaches to knowledge management, we will show limitations of artificial neural networks for several tasks of natural language processing.

Recommended reading

=====

Big Data: A Revolution That Will Transform How We Live, Oct 2013 by Viktor Mayer-Schonberger and Kenneth Cukier (Russian translation is available)

You will miss the point of this talk if your mentality equates academic science and industrial research, if you underestimate the importance of interpreting well established formal methods in terms of the application domain. To love industrial research and grasp the difference (but not dichotomy) between industrial research and science, between publishing and building solutions, between building complex solutions out of dozens of technologies and working with elegant and captivating formulas, it might be useful to quickly browse following resources:

Ferrucci, D, et al. (2010), "Building Watson: An Overview of the DeepQA Project", AI Magazine (AI Magazine) 31 (3), <http://www.aaai.org/ojs/index.php/aimagazine/article/view/2303/2165> retrieved 2016-11-08

[https://en.wikipedia.org/wiki/Cognitive\\_computing](https://en.wikipedia.org/wiki/Cognitive_computing)

A Theory on Power in Networks. By Enrico Bozzo, Massimo Franceschet. Communications of the ACM CACM. Volume 59 Issue 11, November 2016