Dataology and DataScience：Up to Now

Yangyong Zhu and Yun Xiong

Research Center for Dataology and DataScience
School of Computer Science, Fudan University, Shanghai 201203, China
{yyzhu, yunx}@fudan.edu.cn

EXTENDED ABSTRACT

The data explosion is the rapid increase in the amount of data in cyberspace, which brings human into a big data era. Big data means more challenges and new opportunities. In big data era, fundamental theories and methods dealing with big data are demanded urgently. It is necessary to form a new discipline called "Dataology and DataScience" which takes data as research object.

"Dataology and DataScience" has been attracting a great deal of attention. The term "Dataology" was firstly used by Peter Naur to suggest that "Computer Science" should be called "Dataology" who was a Danish pioneer in computer science and Turing award winner[1]. It is very interesting suggestion because he argued that it was not reasonable to regard "Computer" as "Science". However, the idea has not been developed yet until "Dataology" was defined as a new academic discipline (also called "Data Science" or "Science of Data") which is an umbrella of theories, methods and technologies for studying data nature by Yangyong Zhu *et al* in 2009[2-3]. On the other hand, the term "Data Science" began to be used until 1990's[4]. Unfortunately, the definition on "Data Science" has not been formally given except that some research content, scope and topics are pointed out[4-6] (e.g., [4] gives a relatively comprehensive outline of "Data Science"). In [2] and [3], "Dataology and DataScience" is unified and defined as a new science (or academic discipline) whose research object is data. Nowadays, "Dataology and DataScience" is entering into a brand-new stage. Recently, a lot of research organizations have been established around the world including the USA, Canada, Australia, Japan, Korea and China, as well as some journals and proceedings have been published. We believe that

"Dataology and DataScience" would be a new kind of Science exactly as "Natural Science" and "Behavioral Science (or Social Science)".

In this paper, an overview of the progress of "Dataology and DataScience" is given. We investigate various terms about data science and their definitions and concepts proposed in these years and analyze the background and context of existing research on data science. Furthermore, we discuss some key issues (e.g., fundamental theories, new methods and research topics) faced by Dataology and DataScience when it is expected to essentially and entirely become an academic discipline taking data as research object. In addition, we also review as comprehensive current proceedings, journals, research organizations and related data scientists as possible.

Reference

[1] Wikipedia on Peter Naur. Available at http://en.wikipedia.org/wiki/Peter_Naur. 2011-05-09.

[2] Yangyong Zhu, Ning Zhong, Yun Xiong. Data Explosion, Data Nature and Dataology. In *Proceedings of International Conference on Brain Informatics (BI'09)*. 2009.

[3] Yangyong Zhu, Yun Xiong. *Dataology and DataScience*. (in Chinese with English abstract). Fudan University Press. 2009. ISBN 978-7-309-06956-3 /T.350.

[4] F. Jack Smith. Data Science as an academic discipline. *Data Science Journal*. 2006, 5(19).

[5] Lei Liu, Hui Zhang, Jianhui Li, *et al*. Building a Community of Data Scientists: an Explorative Analysis. *Data Science Journal*. 2009, 8(24).

[6] Honglei Sun, Chuang Liu. Research on Progress of International Scientific and Technological Data Frontier Fields. *Advance in Earth Sciences*. 2003, 18(3). (in Chinese)