

Big Data: Issues in organizing and retrieval

Prof. Dr. A.R.D. Prasad
DRTC, Bangalore, India

Abstract

As data grows in volume and complexity, the research world is now gearing up to the era of Big data. Data maybe domain specific mostly but its exploitation and applications are cross-disciplinary. Metadata so far has addressed the issues of representation of resources such as monographs, books articles and journals while some others have dealt with representation of artifacts. Other diverse standards are deployed for representation of processes, organizations and persons. With the collections of data that are growing rapidly into Big data the question of standards and methods for representation of data arises. The talk aims to discuss approaches to representation of Big Data in order to facilitate its organization and retrieval. The architecture of the metadata for information resources and physical objects is already complex but representation of data brings an entirely new dimension to the challenge. The discussion argues for design of purpose oriented metadata for harnessing domain based data. The discussion is emphasized using examples of actual agricultural data sets from the data.gov.in, the Government of India initiative of open data. The talk aims to elaborate on the experience working with data sets of apex research labs in India and discusses issues and challenges to Big Data while offering possible leads of addressing them using a hybrid approach of knowledge organization along with automated techniques.

Keywords: Big Data, Big Data Organization, Metadata and Semantics for Big Data