Intelligent Software Defined Storage

Simon Berkovich
Department of Computer Science
The George Washington University

The Big Data situation requires a qualitatively different type of information processing. This problem brings in a new type of a computational model that explicitly works only with a relatively small portion of the available data, while the rest of the data just implicitly affects selection of the given working portion [1]. The unavoidable restrictions on the operations with overabundant data translate into the design of the brain in accordance with the fundamental Freud’s idea of unconsciousness. This design is contemplated in our paper [2]. Diversified information in overwhelming amounts appears ambiguous, volatile, and unreliable. So, the contents of Big Data systems cannot be treated with confidence as in traditional searching and data mining. Instead, Big Data should be utilized essentially through what can be seen as “knowledge formation”. In other words, processing of Big Data must be performed by what can be considered as “scientific method”. Namely, besides simple extraction of references as from regular information systems the full exploitation of Big Data necessitates formulating testable hypotheses and creating prediction models. A classical illustration presents usage of the observational data of Tycho Brahe through transformation of Kepler’s laws into Newton’s model of “Universal Gravitation”. Thus, employing Big Data falls into the realm of Artificial Intelligence. As a matter of fact, the intelligence facilities of the brain can be considered as a necessary condition to deal with the Big Data challenge. A special type of holographic memory is a pivot point in the realization of these facilities [2]. To implement such kind of Big Data processing facilities in practice we introduce a particular construction of Software Defined Storage. This construction emulates the basic features of the suggested memory organization of the brain: multi-attribute cortical map, content-addressable access, and stream resolution of multiple responses. The envisioned Software Defined Storage incorporates two developments: memory device for multi-attribute items that can be accessed by any combinations of attributes using Fuzzyfind procedures [3] and massive distributed streaming for resolution of multiple responses [4].

References


