Discovering Contexts and Structured Data Structures from Hazy Data

ABSTRACT

A method for storing data of an unknown data context and structure in a structured storage system is described. The method includes extracting a list of key-value pairs from one or more inbound data blocks from one or more external data sources; calculating similarities for the list of key-value pairs from the one or more inbound data blocks based on one or more current existing data structures; determining a highest similarity from the similarities for the list among the one or more current existing data structures; determining that the highest similarity for the list does not meet a predetermined similarity threshold; creating a new data structure to store the list of key-value pairs for a first inbound data block from the one or more inbound data blocks; and storing the list of key-value pairs in the new data structure in the structured storage system.

Fig. 2