REVIEW ARTICLE

Literature Review on Information Extraction by Partitioning

Sonam S.Chauhan[1], Prashant R. Deshmukh
[1]Department of Information Technology
Sipna College of Engineering and Technology,
SANT GADGE BABA AMRAVATI UNIVERSITY, AMRAVATI
[1]Email:sonams.chauhan@yahoo.in

Abstract— **Information extraction systems are implemented traditionally as a pipeline of special-purpose processing modules that target at the extraction of a specific type of information. Such an approach has major drawback that whenever a module is improved or a new extraction goal emerges, extraction need to be applied to all parts even if only a small part is affected.**

Here we describe an approach for information extraction in which we partition the dimensions (attributes), i.e., a higher dimension of large data set can be transformed into relatively smaller subsets of data in certain numbers that might be processed easily. Here we use hetrogeneous database as large data set. Thereafter, based on the separation of dimensions, the discernible dataset of all data are computed so as to get their core attribute sets. Furthermore, the attribute reduction and data redundancy methods are used to obtain the partition results.

In our approach towards information extraction we tried to partition are data sets so that information can be extracted easily and correctly with query processing which generate results in less time. Feasibility of our incremental extraction approach can be demonstrated by performing experiments to spotlight two important aspects of an information extraction system: quality of extraction results and efficiency.

Keywords— **Hetrogeneous databases; multidimensional data; partitioning; association rules; algorithm; itemsets**

Full Text: http://www.ijcsmc.com/docs/papers/December2013/V2I12201376.pdf