

INFOBRIGHT CONTEST PReMI & RSFDGrC 2009

Infobright and the PReMI & RSFDGrC 2009 organizers are pleased to announce a contest for data mining algorithms for handling rough tables with rough attributes. The main goal is to investigate whether the Infobright Knowledge Grid statistics currently applied in the Infobright Community Edition (ICE) analytic database may also be used to speed up the data mining processes over large data volumes.

Background: ICE is an open source, free of charge database designed for SQL-based analytics and Business Intelligence. It is easy to use and ideal for data volumes up to 30 TB. It combines the principles of a column-oriented data store and data compression with a unique Knowledge Grid architecture to eliminate the complexity of data warehousing. For further information on how to download and use ICE, visit Infobright at:

<http://www.infobright.org/Open-Source/Home>

The **Knowledge Grid** is based on data granulation. For example, consider a data table with one billion rows. In ICE, it corresponds to 15259 **rough rows**. Each rough row groups up to 65536 consecutive rows together. Physically, each rough row is split into data packs corresponding to particular attributes, assuring both horizontal and vertical data decomposition. Logically, each rough row can be treated as a row in a new data table, where the attributes' values correspond to statistics calculated for particular data packs. Compared to the original data, a **rough table** has 65536 times less rows, the same number of attributes, but more compound values. For example, given 10 attributes in the original data, we'll now have 10 **rough attributes**, where every rough attribute labels every rough row with – in the simplest scenario – the minimum and maximum values within the corresponding data pack. For information about the other statistics and the usage of the Knowledge Grid in ICE please refer to the following paper:

<http://www.vldb.org/pvldb/1/1454174.pdf>

The papers describing the analysis of rough tables should be uploaded to the RSFDGrC online system in the form of regular paper submissions by July 20, 2009. All the papers with positive reviews will be accepted for publication in RSFDGrC 2009 proceedings.

Infobright provides the **15259 rough rows sample of rough data** that is recommended to be used by the contest participants. For further information visit Infobright forums at:

<http://www.infobright.org/index.php/Forums/viewthread/538/>

With best regards,

Dominik Ślęzak
Chief Scientist
Infobright Inc.

Santanu Chaudhury
PReMI&RSFDGrC
I.I.T. Delhi, India

INFOBRIGHT
Open Source Data Warehousing

