

Division of Computational Informatics: Database Group

Group Leader: AMAGASA Toshiyuki, Professor, Ph.D.

The management and utilization of big data are major issues in computational science. The Database Group in the Division of Computational Informatics is in charge of research and development in the field of data engineering for the utilization of big data. Specifically, we work on data integration infrastructure technology to handle diverse information sources and real-time data in an integrated manner, high-performance large-scale data analysis technology (Fig. 1), data mining and knowledge discovery technology for knowledge and patterns found in large-scale scientific data and social media, open data-related technologies for handling various data and knowledge on the Internet in a unified manner, and other such fundamental technologies. We also promote applied research in various fields of computational science in collaboration with the Division of Global Environmental Science, Division of Particle Physics, and Division of Life Sciences at the Center for Computational Science, International Institute for Integrated Sleep Medicine (IIIS), Center for Artificial Intelligence Research (C-AIR), etc.

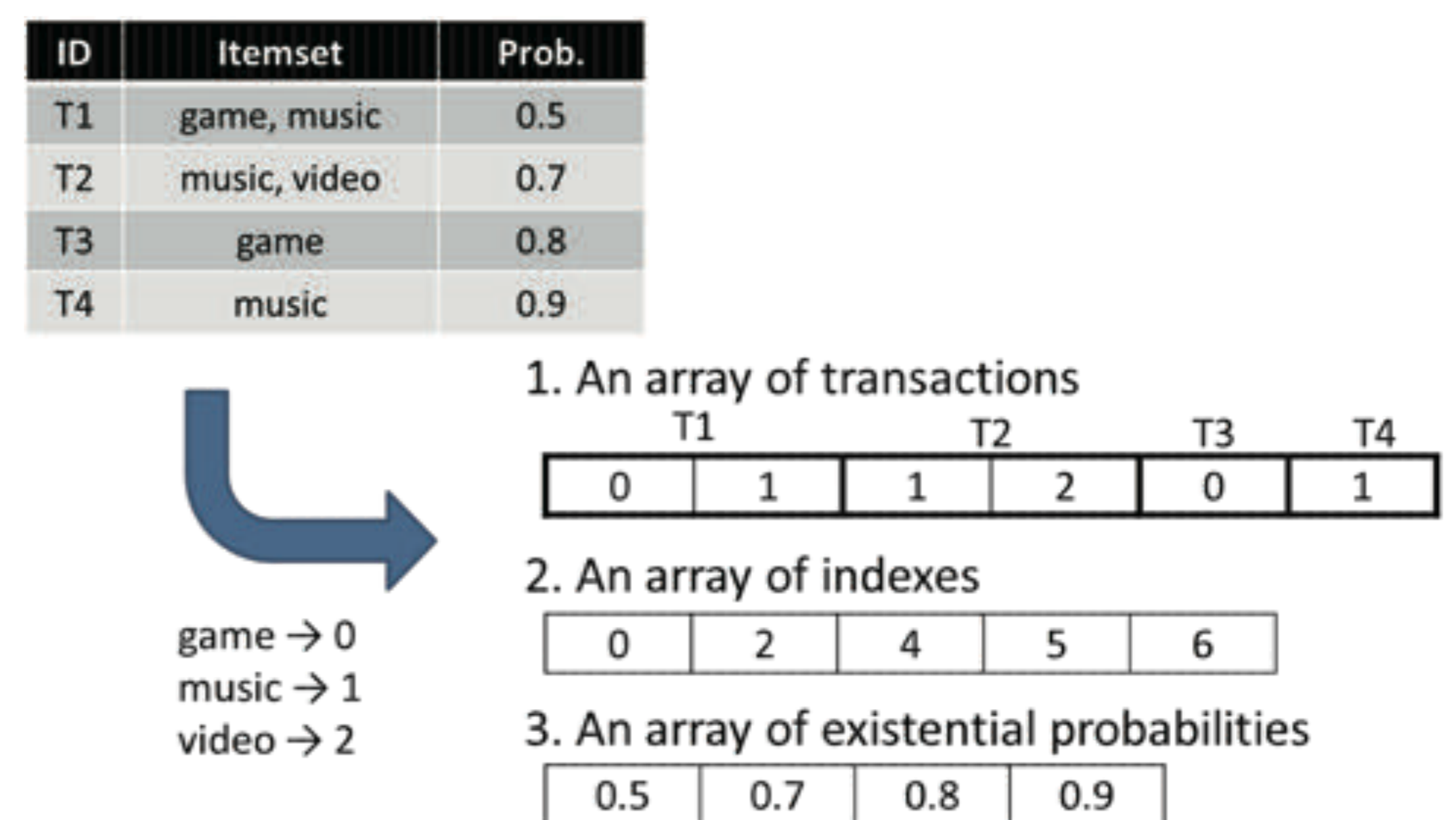


Fig.1 GPU probabilistic frequent item set mining

Big data infrastructure technology

Research on fundamental technologies for processing and analyzing big data characterized by the 3Vs (volume, variety, and velocity): 1) development of fundamental systems for linking streams such as sensor data in addition to databases and the Web, etc.; 2) high-performance big data analysis technology using massively parallel processing with GPUs, FPGAs, etc.; 3) privacy and security in big data processing; and 4) systems for processing semi-structured data such as RDF and LOD.

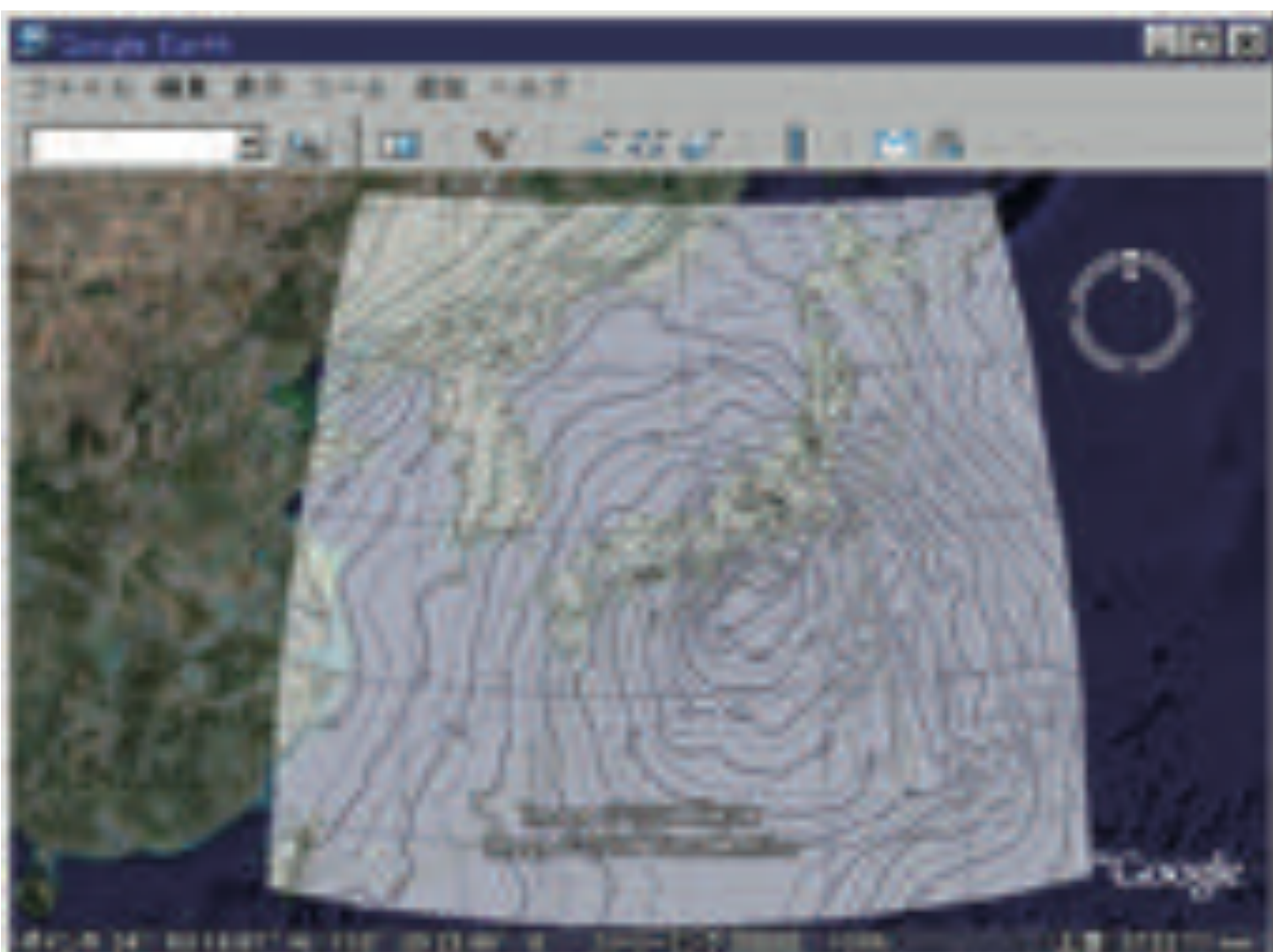
We are working on research and development of heterogeneous stream integration infrastructure systems, stream/batch integrated big data-processing infrastructure, and ultrahigh-performance data analysis methods using GPUs.

Data mining and knowledge discovery

Research on data mining and knowledge discovery methods for various types of data: 1) various analysis algorithms for text, images, graphs, etc.; 2) social media analysis and mining; and 3) biological data analysis using machine learning.

We are working on research and development of high-speed analysis algorithms for large-scale graphs and images, advanced metadata extraction algorithms that significantly improve the scope of social-media use, etc.

Utilization of scientific data



Research aimed at the management and utilization of explosively growing scientific data: 1) operation and development of the GPV/JMA large-scale meteorological database and the JRA-55 archive; 2) operation and development of the JLDG/ILDG lattice QCD data grid; and 3) advancement of genomic and biological data utilization using machine learning.

We are developing and operating the GPU/JMA archive and the JRA-55 archive, which are databases for archiving numerical weather data (GPV) published by the Japan Meteorological Agency (JMA), and making them available to researchers (Fig. 2).

Fig.2 Viewing weather maps using Google Earth