

Kitsuregawa, Toyoda, Nemoto, Yoshinaga,  
Goda, Ikoma LAB.

## [Platform for Mining, Learning, and Visualizing Real World and Social Big Data Fusion]

Center for Socio-Global Informatics (CSGI)

Data Science, Data Integration and Analysis, Web  
Mining, and Natural Language Processing

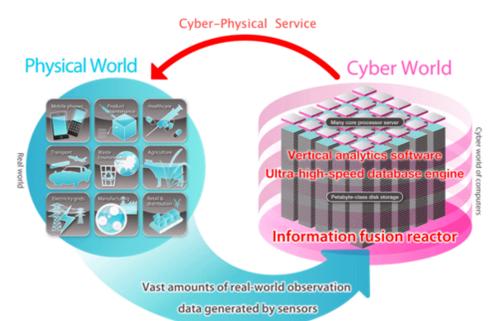
Information and Communication Engineering

<http://www.tkl.iis.u-tokyo.ac.jp>

Our Lab has focus on completely novel and surprising research on system software, advanced applications, hardware and algorithms, which is the basis of the technology for handling large quantities of data. We have been developing high performance database engine based on a novel out-of-order execution principle, ultra-large-scale cyber space mining systems, and 25 petabyte-scale global environment information system based on database engineering.

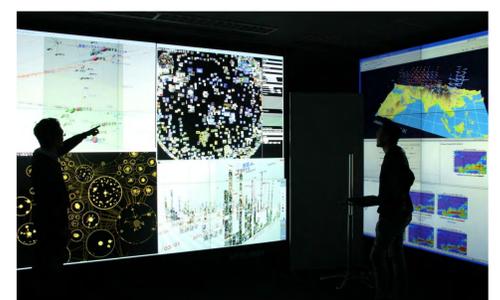
### Development of High Performance Database Engine for Big Data Era

Our research group has been developing Ultrafast Database Engine based on a novel out-of-order execution principle. This new database engine has the capability to achieve significant performance boosting for analytical queries in the Big Data era. The group has been also constructing an experimental system of next-generation strategic social services (cyber physical services) to clarify effectiveness of the new database engine.



### Ultra-large-scale Cyber-Physical Space Analysis System

Our lab has continuously collected Japanese Web pages since 1999 and has constructed a Web archive system including about 30 billion pages, 2 billion blog articles, and 25 billion tweets. Based on this peta scale archive, we are developing structural, contents, and temporal analysis systems including information diffusion extraction, inter-media comparison, and real-time deep text analysis. The cyber space information is integrated with physical space information such as mobility data for traffic analysis. Results of analysis can be interactively visualized on a large-scale high-resolution display wall.



Huge-scale spatio-temporal  
visualization system  
on the display wall

### Over 25 Petabyte Global Environmental Information Fusion

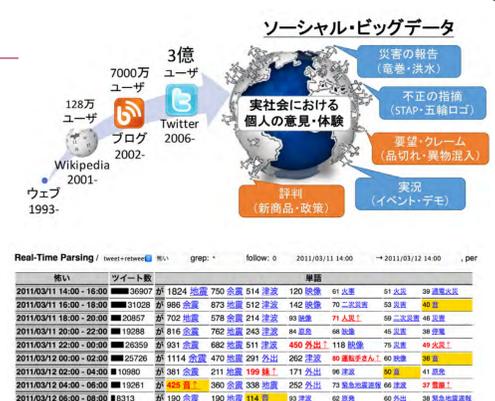
We have been developing a large scale global environmental information fusion system for data integration and analysis that includes the supporting functions of life cycle data management, data search, information exploration, scientific analysis, and partial data down-loading.



Global environmental information storage

### Natural Language Processing for Social Bigdata Mining

We have been developing a natural language processing platform that utilizes massive bigdata on social web services such as microblogs on the world wide web. One of the goals of our project is to mine knowledge that benefits our society. We develop fast and accurate natural language software and related web mining methodology that utilize information available from verbal and non-verbal data. We also deal with core natural language applications such as machine translation, dialog generation, and sentiment analyses.



Stream analysis during the 2011  
Tohoku earthquake and tsunami