Challenge for Info-plosion

Masaru Kitsuregawa

Institute of Industrial Science, The University of Tokyo
4-6-1 Komaba, Meguro-ku, Tokyo 153-8505, Japan
kitsure@tkl.iis.u-tokyo.ac.jp

Abstract. Information created by people has increased rapidly since the year 2000, and now we are in a time which we could call the "information-explosion era." The project "Cyber Infrastructure for the Information-explosion Era" is a six-year project from 2005 to 2010 supported by Grant-in-Aid for Scientific Research on Priority Areas from the Ministry of Education, Culture, Sports, Science and Technology (MEXT) of Japan. The project aims to establish the following fundamental technologies in this information-explosion era: novel technologies for efficient and trustable information retrieval from explosively growing and heterogeneous information resources; stable, secure, and scalable information systems for managing rapid information growth; and information utilization by harmonized human-system interaction. It also aims to design a social system that cooperates with these technologies. Moreover, it maintains the synergy of cutting-edge technologies in informatics.

1 New IT Infrastructure for the Information Explosion Era

The volume of information generated by mankind has increased exponentially, i.e., "exploded" since 2000. The purpose of our research project, "Cyber Infrastructure for the Info-plosion Era" in the Ministry of Education, Culture, Sports, Science and Technology (MEXT) Grand-in-Aid for Scientific Research on Priority Area, is to build advanced IT infrastructure technologies for this information explosion era [1]. According to the research by the University of California at Berkeley, the volume of information created by human is explosively increasing [2, 3]. Huge volume of data is also created by sensors and machines. We have considered that the most important theme for researchers in the field of computer science is the research on new IT infrastructure for the Information-explosion Era.

The project has three major research components (Research Groups) to achieve this goal: build technologies to search for needed information efficiently, without bias and without being at risk from the rapidly growing volume of information (A01); build new and sustainable technologies that can operate large-scale information systems managing enormous amounts of information safely and securely (A02), and build human-friendly technologies to enable flexible dialogue between men and machines and enable everyone to utilize information (A03).

V. Corruble, M. Takeda, and E. Suzuki (Eds.): DS 2007, LNAI 4755, pp. 1–8, 2007.

Underlying these three components is the research into new social systems that can facilitate the use of advanced, information-based IT services (B01). In addition, Large-scale Info-plosion Platform (LIP) is implemented as the shared platforms used by all Research Groups. The project is interdisciplinary in its structure, bringing together advanced research methods in information-related areas. (Project leader: Masaru Kitsuregawa)

2 Infrastructure for Information Management, Fusion and Utilization in the Information Explosion Era (A01)

This Group focuses on the shortfalls of present internet searches and looks into new search methods, including better ranking systems (where minority opinions are not overlooked), interactive searches, reliability assessments and time-space searches. Currently, for knowledge workers, about 30% of their time on intellectual activities is spent just for retrieval information [4]. The Group will attempt to create a system of search platforms to search a massive volume of web page contents.

Another important issue for information retrieval is the dangers associated with information rankings. Search engines are extensively used in the web world. When a general word is given, it may hit millions of pages, and only about 10

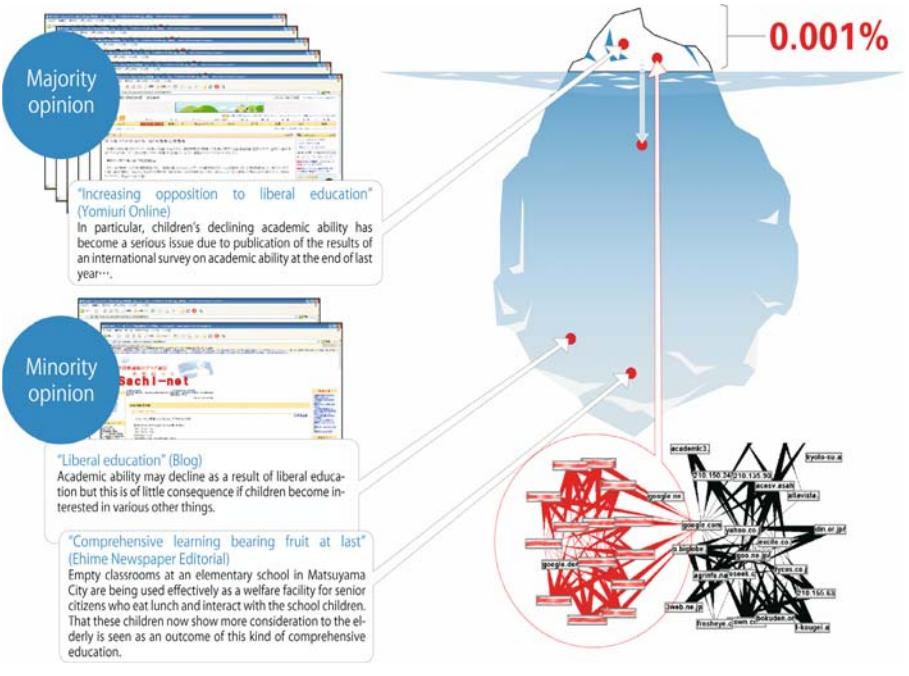

Fig. 1. Dangers Associated with Information Rankings

probable candidates are listed as top rankers on the first page of the search result. Now, is this ranking really reasonable? Who guarantee the correctness of it? Actually, this is controlled by just a private company. There should be a possibility that the ranking is sold and bought. In addition, it is possible to control the ranking deliberately. We have found, as shown in the Fig.1, that the red island of links suddenly appear on February 2004, while only the black island of links is found in 2003. This is an example of a trick to raise their own rank by linking them to a well-known site.

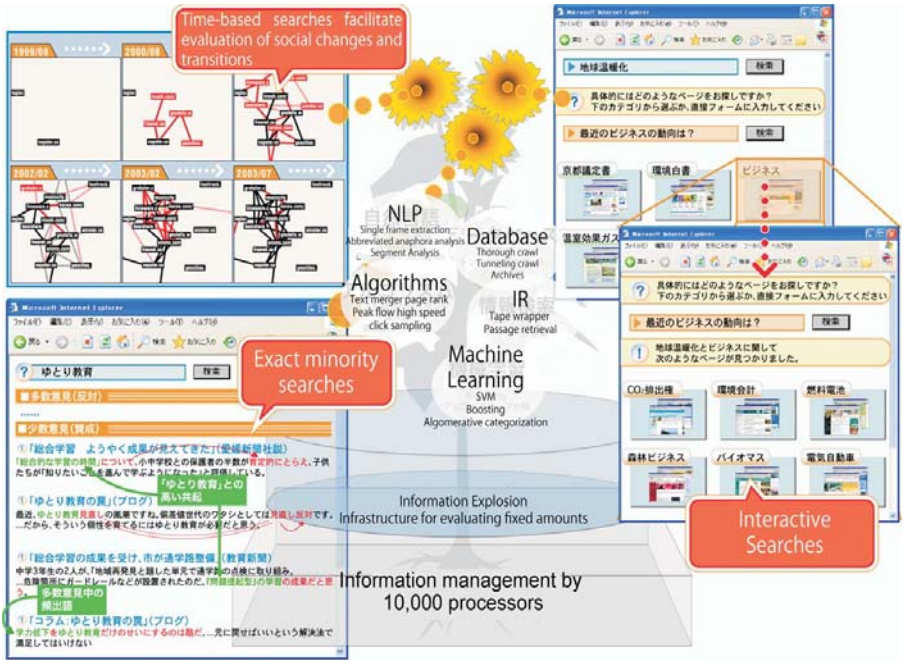

Fig. 2. Next generation searches

Also, in the current ranking system, majority opinions are highlighted while minority opinions are buried. For example, when "Yutori Kyouiku" (liberal education) is searched, majority opinions are found easily like "increasing opposition to liberal education". Although there are minority opinions such as "comprehensive learning bearing fruit at last", they could be completely neglected.

Various technologies, including information searching, natural language processing, machine learning, artificial intelligence (AI) and database technology, are integrated into this system to enable quantitative analysis to be performed (Fig.2). We hope to enable a remarkable level of interdisciplinary synergy among different fields. It is better for human to understand by presenting relational information also than information itself. Several innovative researches are expected in this area, for example, presenting comparative information [5] and information

on time sequence [6]. Research on the methods of management, integration and processing of "exploding" real-world information, including cyber information and information obtained from remote sensors, will also be conducted. (Group leader: Masaru Kitsuregawa, University of Tokyo)

3 Infrastructure for Information-Explosion-Proof IT Systems (A02)

The exponential increase in the amount of information requires far larger IT systems to handle the volume. According to [7], Google has now over 450,000 servers over 25 locations around the world. The ratio of computers used for search engines among shipped computers is 5% in total, according to MSRA (Microsoft Research Asia) Summit in 2006. Data on the Internet is dispersed over millions of nodes, making the overall system unstable and vulnerable to information overload.

In order to keep stable operation of such huge systems, real-time monitoring of behavior of software is inevitable. For such a purpose, explosive volume of information extracted by software sensors should be analyzed so as to point out anomaly behavior of the system and stabilize it. Researches on mining huge volume of data yielded from monitoring very-large-scale systems are particularly important for the age in which a nation-wide cyber attack becomes reality like Estonian case.

This Group aims to establish a new "resilient grid" infrastructure which can automatically allocate computer resources, handle large-scale system faults over

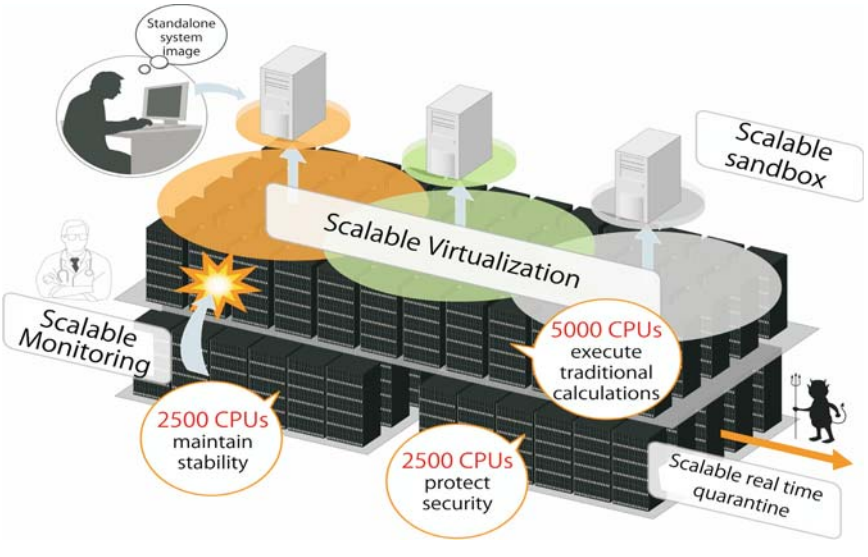

Fig. 3. Infrastructure for large scale system

the network without human intervention and without the modern-day concerns of security breaches and intrusion (Fig.3). This is partially based on Autonomic Computing [8] and interesting research works performed by UC Berkeley and Stanford University [9]. The resilient grid will allow a high-performance virtual computing environment to be configured autonomously, on which applications can be deployed safely and securely. (Group leader: Satoshi Matsuoka, Tokyo Institute of Technology)

4 Infrastructure for Human Communication in the Information Explosion Era (A03)

The information explosion has two aspects: qualitative (volume) and quantitative (complexity). This Group proposes studies of the advancement of human communication to address the issues related to complexity. The underlying concept is a mutually adaptable multi-modal interaction that can fill the communication gaps between people and information systems. This is the key to overcoming the complexity resulting from highly functional and multi-functional information systems, and establishing a secure and user-friendly interactive environment (Fig.4).

Searching information from explosively huge size of information space still requires advanced skills, since existing tools for such a purpose is not necessarily easy to use for naive users. Human-friendly interfaces as well as communications

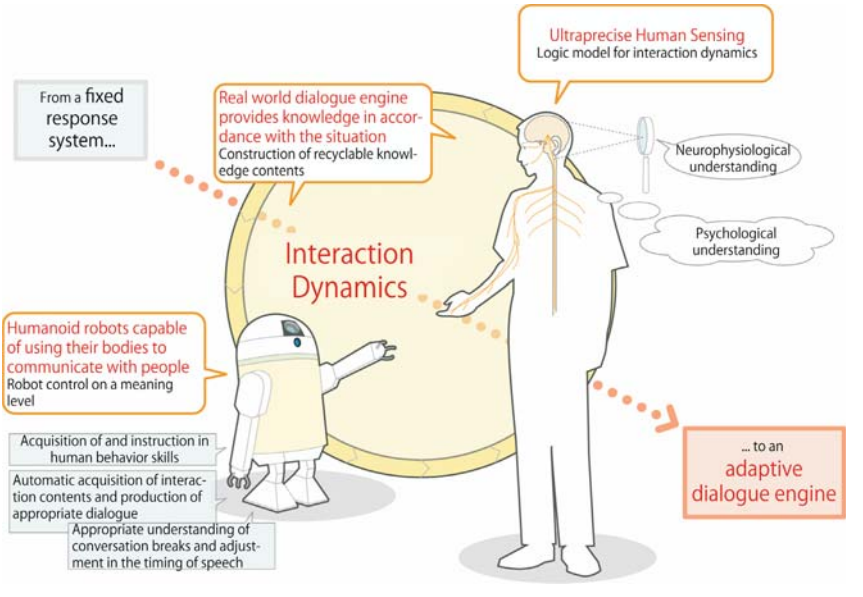

Fig. 4. Infrastructure for human

with a robot in the near future are extremely important, which is covered by this Group. Monitoring human behavior also produces explosive volume of data, which plays essential role to analyze nonverbal communication dynamics.

Recently, new ideas have been demonstrated such as scientific formulation of "knacks", which establishes a firm theoretical ground for connecting highly complex physical dynamics and symbolic information of skillful actions [10]. The effect of the theory is demonstrated by the world's first experiment on highly dynamic whole body motion of a humanoid robot. The ultimate goal of the research is to establish a new framework fusing the robot and information technology that solves the above stated bidirectional connection problem. (Group leader: Takashi Matsuyama, Kyoto University)

5 Governing the Development of a Knowledge-Based Society in the Information Explosion Era (B01)

Advances in technology often race far ahead of consideration of how people will actually use the accompanying new benefits. Engineers focus on the technology rather than the laws and regulations that will be required for the technologies they develop. Support for engineers and keeping them aware of the legal and social implications of technological advances is indispensable. Practical experiments require cooperation and coordination with a number of partners in the

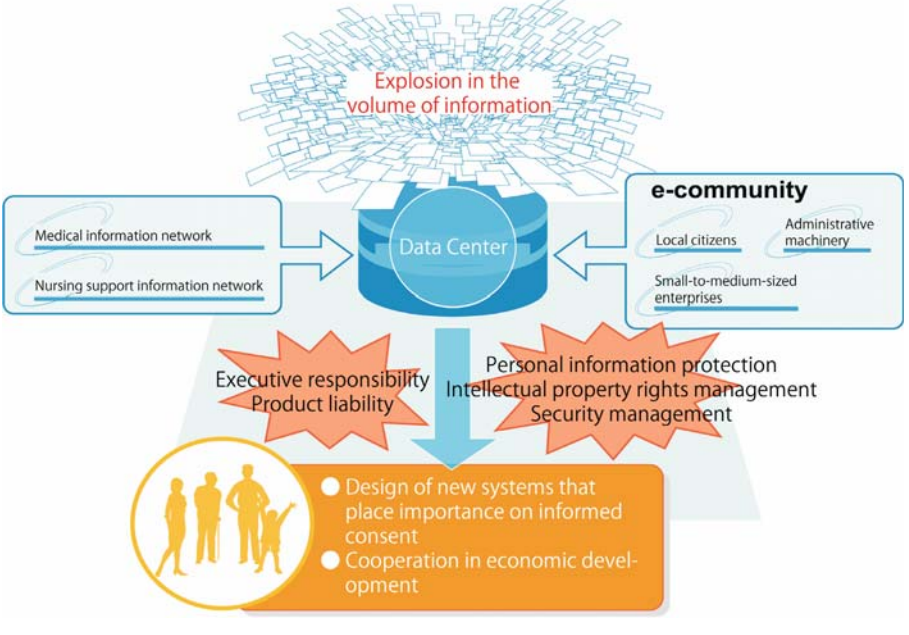

Fig. 5. Promotion of Society-based Research

real world. Advances in IT have both a "bright" side and a "dark" side. Social problems arising from the abuse of the dark side need to be addressed.

The fourth feature of our project is promotion of society-based research. For investigating IT infrastructure for the Information-explosion era, society-based research is inevitable, as shown in Fig.5. In order to realize universal social infrastructure including medical information network and nursing support information network, various kind of problems should be resolved, for example, product liability, personal information protection, intellectual property rights management, and security management.

In such a situation, cooperation in economic development between technology providers and receivers is required. Design of new systems that place importance on informed consent is important. We are making researches into information governance that cover the explosion in the volume of information. This Group focuses on social issues and will study the governance of information technology development, its relationship with the legislative process, and its implications for new legal and social systems [11]. (Group leader: Osamu Sudoh, University of Tokyo)

6 The Aims and Roles of the Large-Scale Info-plosion Platform in the Research Project

One of the most remarkable distinctions of this research project from other national scientific projects is that we will inject a quarter of the total research fund into the Large-scale Info-plosion Platform (LIP). This fund will be spent to implement shared platforms which informatics researchers intriguingly and collaboratively construct for innovative researches. These platforms enable each research team to conduct novel and original research activities that cannot be realized only within a Group. The resources integrated in the platforms will be expected to be also used by researchers outside the research project.

7 Concluding Remarks

In this article, an overview of our Info-plosion project is introduced. The relation of all Research Groups is shown in Fig.6. This research project is approved in 2005 and started full part since April 2006. The budget of the project is about 600 million Yen/yr and over 200 researchers from various fields are participating currently. They are consisting of Planning Researches and Proposal Researches. Over 300 research themes are applied, and more than 60 Proposal Researches are accepted and started their innovative research works. Info-plosion project is in coalition with the Consortium for New Project on "Intellectual Access to Information" (jouhou-daikoukai) sponsored by the Ministry of Economy, Trade and Industry (METI) started in 2006 [12].

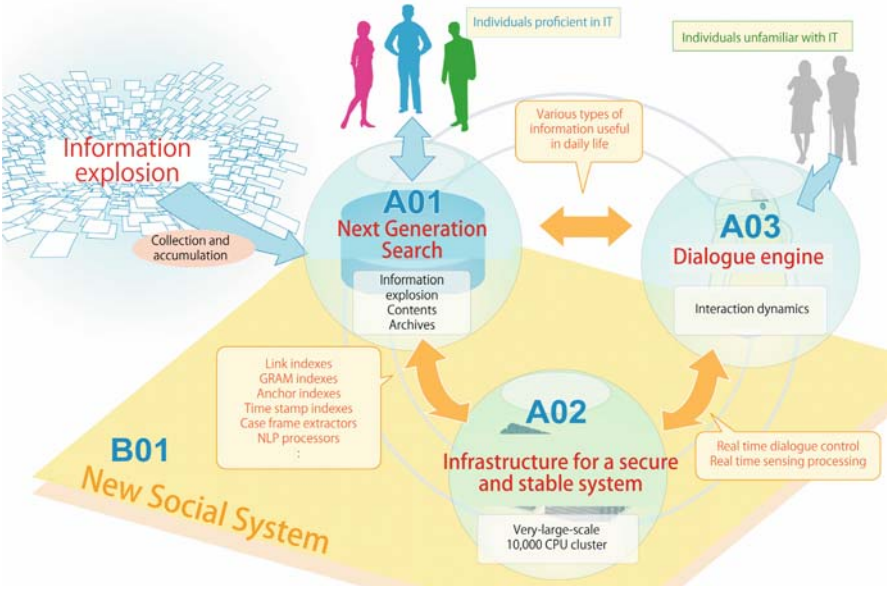

Fig. 6. Overall image of research project

References

[1] http://itkaken.ex.nii.ac.jp/i-explosion/
[2] http://www2.sims.berkeley.edu/research/projects/how-much-info-2003/
[3] http://www.horison.com/
[4] http://www.delphiweb.com/knowledgebase/documents/upload/pdf/1802.pdf
[5] Nadamoto, A., Tanaka, K.: A Comparative Web Browser (CWB) for Browsing and Comparing Web Pages. In: Proc of the 12th Int. World Wide Web Conf. (WWW2003) (2003)
[6] Toyoda, M., Kitsuregawa, M.: What's Really New on the Web? Identifying New Pages from a Series of Unstable Web Snapshots. In: Proc of the 15th Int. World Wide Web Conf. (WWW2006), pp. 233–241 (2006)
[7] http://www.nytimes.com/2006/06/14/technology/14search.html ?pagewanted=2&ei=5090&en=d96a72b3c5f91c47&ex=1307937600
[8] http://www.research.ibm.com/autonomic/
[9] Bodk, P., Friedman, G., et al.: Combining Visualization and Statistical Analysis to Improve Operator Confidence and Efficiency for Failure Detection and Localization. In: Proc. 2nd International Confidence on Autonomic Computing (ICAC'05) (2005)
[10] Kuniyoshi, Y., Ohmura, Y., Terada, K., Nagakubo, A., Eitoku, S., Yamamoto, T.: Embodied Basis of Invariant Features in Execution and Perception of Whole Body Dynamic Actions — Knacks and Focuses of Roll-and-Rise Motion. Robotics and Autonomous Systems 48(4), 189–201 (2004)
[11] Sudoh, O. (ed.): Digital Economy and Social Design, pp. 1–236. Springer, Heidelberg (2005)
[12] http://www.jyouhoudaikoukai-consortium.jp/