Medicare-Grid: New Trends on the Development of E-Health System Based on Grid Technology

Yeh-Ching Chung, Po-Chi Shih, Kuan-Ching Li, Chao-Tung Yang, Ching-Hsien Hsu, Fang-Rong Hsu, Don-Lin Yang, Chia-Hsien Wen, Chuang-Chien Chiu

Medicare-Grid: new trends on the development of E-Health System based on Grid Technology

Yeh-Ching Chung[1], Po-Chi Shih[1], Kuan-Ching Li[2], Chao-Tung Yang[3], Ching-Hsien Hsu[4], Fang-Rong Hsu[5], Don-Lin Yang[5], Chia-Hsien Wen[2], Chuang-Chien Chiu[5]

[1]National Tsing Hua University, [2]Providence University, [3]Tunghai University, [4]Chung Hua University, [5]Feng Chia University
ychung@cs.nthu.edu.tw[1]

Abstract. The evolution of information technology over the last decade has brought opportunities to improvements in the state of art of medical services. One scenario is that a patient's digital health record can easily be shared among hospitals and medical centers via internet, enabling the examination performed in one location while clinical diagnosis be done by physicians in another. In this paper, we propose a Medicare-Grid — a novel grid-based E-health system proposed to ease the process of retrieving and exchanging personal health data among hospitals and medical centers. Grid and peer-to-peer technologies have been used to integrate computing and storage resources provided by hospitals, as also to develop an Electronic Health Record (EHR) center to store and share EHRs among these locations. We have also developed a system prototype using ultimate hardware resources and open source software systems to simulate a real scenario as described above. We demonstrate that the idea proposed in the project is feasible, possible to be implemented and applicable to real world.

Keywords: Grid, E-health, P2P, data mining, RFID, wearable measuring system.

1 Introduction

During the last century, the development of medical services has greatly improved regarding to the quality of medical treatments, results that successfully prolong human lives. One of major evolutions on software is the digitalization process of personal health record. The digitalized record, which is formally named Electronic Health Record (EHR), can thereby be shared easier among hospitals via internet. In order to make the EHR sharing mechanism feasible, major issues to be considered are twofold. The former one is how to digitalize personal health record to a standard form which must be recognizable by all hospitals, while the latter is regarding on how to share them among hospitals. For such a goal, we propose Medicare-Grid — a grid based E-health system addressing the above issues, facilitating the sharing and exchange of digitalized personal health record among hospitals.

As for the first issue, we adopted Taiwan electronic Medical record Template (TMT) [1], which is a standard EHR format established by Department of Health, Executive Yuan, Taiwan. Based on this format, we developed an application that translates specific EHR format used in each hospital or medical center into a standard TMT format. As part in this project, we developed an application for Taichung Veterans General Hospital, Taiwan which translated more than 500,000 EHR samples as testing data.

As for the second issue, we developed a Medicare-Grid platform to address the issue of exchanging EHRs. First, grid and peer-to-peer technologies were used to develop an Electronic Health Record (EHR) center as a decentralized database to store and share EHRs among participating hospitals and medical centers. For each site, we developed a client application that permits them to connect to the EHR center, to upload or download EHRs. Although this mechanism is actually a centralized approach, which has potential drawback on scalability and single point of failure, peer-to-peer is considered to decentralize this “single” server and makes it scalable, fault-tolerant and robust.

In addition to the data sharing mechanism mentioned, we also integrate computing resources provided by hospitals, to form a computational grid for medical related applications. We use the *de facto* standard grid middleware Globus [2] to build up a computing grid platform and implement a workflow-based resource broker to efficiently match and select available resources in reply to user’s requests. Additionally, a web portal is also developed which supports users to utilize underlying grid resources with ease.

Based on our computing and data grid platform, we developed medical related applications to improve the in-hospital medical services. Applications include (1) a data warehouse for medical decision support system, (2) a RFID-based mobile monitoring system to precisely identify people or items, and (3) a wearable physiological signal measurement system that monitors the health condition of a patient.

The remainder of this paper is organized as follows. Section 2 describes similar projects with respect to domestic and international perspectives, while in Section 3 issues regarding to the development of grid platform and EHR sharing mechanism are discussed. Later, three medical related applications are presented in Section 4, and finally, conclusion remarks are presented in Section 5.

2 Related Works

As far as we know, the insights presented in this project are novel not only in Taiwan, since it is focused on the use of grid technology to enable EHR sharing among hospitals and to integrate various Medicare applications. Nevertheless, there are some medical related projects that utilize the computing grid platforms as underlying platform such as Knowledge Innovation National Grid (KING) project [3] and BioGrid related project [4].

There exists a similar project called National Grid [5], which is focused on enabling medical related documents such as EHR or X-ray image to be shared by using grid technology; currently, over one thousand hospitals have joined to this project. IBM cooperates with University of Pennsylvania to integrate computing resources provided by each hospital and form a computational grid environment, so that all participating hospitals can easily utilize remote resources and share medical data.

3 Medicare-Grid Platform

Medicare-Grid platform can be divided by three components, namely computing grid, data grid and EHR management system. Computing grid platform provide the computational cycles needed to the execution of Medicare related applications, while Data grid provides a virtual data storage system to support EHR sharing, which management system handles the EHR format translation among specialized HIS and standard TMT format and to provide user friendly web interface for users to operate the system. In subsections that follow next, we will describe detailed system design of these components.

3.1 Computing Grid

A grid platform is an aggregation of geographically distributed resources that working all together over the Internet as a vast virtual computing environment [6, 7, 8]. The main task of computing grid is resource brokering to optimize a global schedule for requested grid jobs matching and selecting suitable and available requested resources. With a resource broker, users are insulated from the grid middleware, thus avoiding communicative burdens between users and resources.

A workflow-based grid resource broker is presented whose main function is to match available resources with user requests. Also, the broker solves the job dependency problem by sorting topologically and then execution of workflows. In order to deal with communication-intensive applications, the broker considers network information statuses during matchmaking and allocates the appropriate resources, thus speeding up execution and raising throughputs. In Fig. 1, the architecture of Resource Broker system and component relationships are presented, including also functions of each component are listed in the relation link. As Grid Portals allow easy access to the system [9, 10], a schematic diagram of the complete Workflow System is shown in Fig. 2.

The achievements are described as follows. First, we construct a computational grid platform using Globus toolkit, distributed in 5 different locations (universities). Second, we have designed and implemented a resource broker which main function is to match available resources according to users' needs. Finally, we provide a uniform graphic user interface (GUI) to use Medicare-Grid platform, to achieve automatic resource discovery and efficient available resource usage. Indeed, this supports grid users to submit their jobs to the suitable grid resources without knowing in advance any information on available resources.

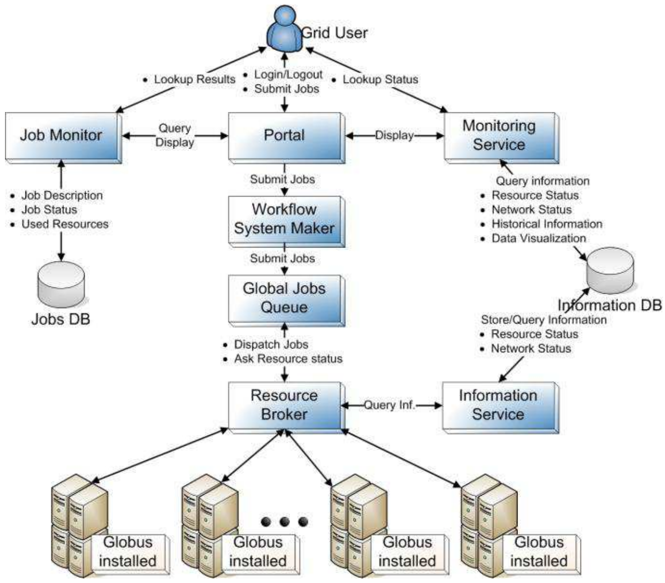

Fig. 1. The architecture of Resource Broker system and the relationships of each component.

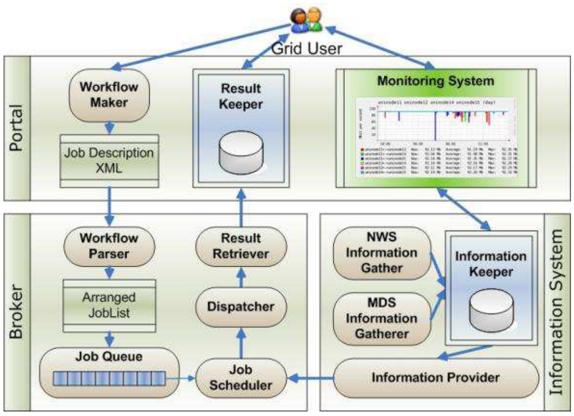

Fig. 2. A schematic diagram of the complete Workflow System.

3.2 Data Grid

Data grid system is designed as distributed two-layer hierarchical peer-to-peer architecture based on the principle of locality. The bottom layer called intra-group overlay is constructed with Chord system, which the overlay clusters neighboring peers provide services within local regions; the upper layer called inter-group overlay is constructed to connect local groups together with consideration to locality. Specifically, the data grid system consists of three modules:

1. File management module: responsible for file operations, such as file insertion, file retrieve, file recovery, replicate and cache from a storage peer,
2. Intra-overlay module: to provide functions to locate peer in local group using Chord architecture,
3. Inter-overlay module: to provide API for communication need among groups.

A novel file replication mechanism is proposed, different from existing replication mechanisms such as PAST [11], OceanStore [12], and Freenet [13] that rely on global information of system. We made use of levels to control the degree of replication in our system. Peers that originally hold a file is skipped by replication level will hold a simple indicator to the peer which really holds such file.

In order to evaluate performance of the proposed system and its potential, we have implemented and deployed such proposal on Taiwan Unigrid [14] to perform large scale experiments. These experiments have been executed on the storage system located in 9 geographically distributed sites with total of 42 servers, as shown in Fig. 3(a). During the experimentation process, we selected top 10 download files from SourceForge.net as the source of testing data, and three different group bound network bandwidth {1000kbps, 100kbps, 10kbps} to cluster storage nodes.

Fig. 3(b), 3(c) and 3(d) show that we have successfully cluster some closer located peers under specific group bound. In the experiment, we noticed that the average measured bandwidth time between the newly coming peer and the measured target is less than 10 ms, and the average locate operation is less than one second.

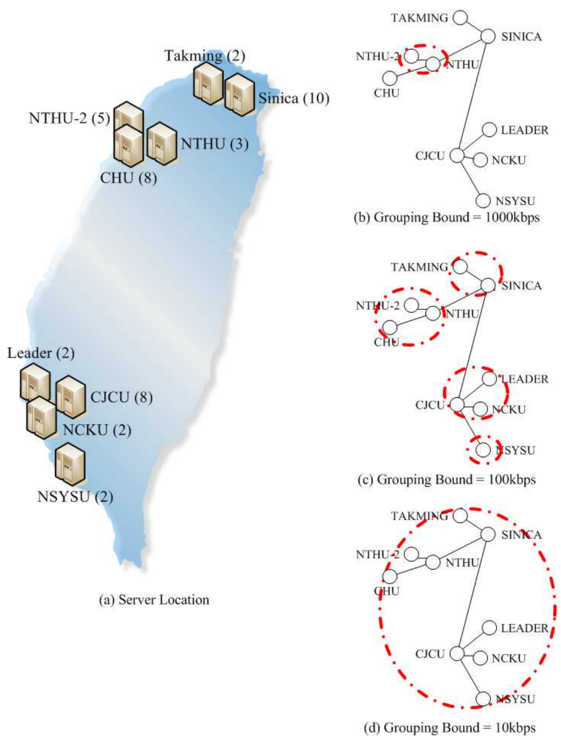

Fig. 3. (a) Server location in Taiwan Unigrid testbed, (b), (c) and (d) Server group result with different group bound 1000, 100, 10kbps.

2.3 EHR Management System

EHR management system is the core service in this paper, addressing the issues of sharing standardized EHR among each participating hospital, and they are threefold. The former one is how to standardize EHR, followed by how to share them, and the latter is regarded to the user interface. For the first issue, we adopted TMT, which is a standard EHR format established by Department of Health, Executive Yuan, Taiwan, and developed a translation application to translate the specific EHR format used in each hospital into standard TMT format. Since the HIS (Hospital Information System) used in each hospital is developed by difference software company, each of them have a specific format and database schema. To make standard TMT format practicable, it requires understanding both TMT and specific hospital format in order to develop translation application for that hospital. Throughout the development of this paper, we focused our attention to the development on the transformation application between TMT and the format used in Taichung Veterans General Hospital, organization which made available more than 500,000 EHR sampling data, and translated with success to standard TMT format and stored next on data grid.

As for the second issue, we exploit the data grid as the fundamental EHR storage space to store and share EHR among hospitals, as illustrated in Fig. 4 the proposed EHR sharing mechanism. The text in black color represents the HIS used in each hospital. Physicians in each hospital make use a desktop computer to read and record patient's health record in the local database. The text in red color is the server and application developed in this paper to facilitate EHR sharing. The EHR Center is constructed using data grid as described in subsection 3.2. A client application is then implemented with the functionality of search, upload, and download data from EHR Center. In order to connect hospital with data grid, the TMT Translator is responsible for the translation of specific EHR format to stand TMT format.

A case study is used to demonstrate the entire sharing operation. For instance, a patient X has appointment with a doctor in hospital A, his/her health record is then stored in local HIS database according to operation procedure of hospital A. Patient X's EHR is then translated to TMT format (by TMT Translator) and uploaded to EHR Center (by P2P Storage Client). As patient X register in another hospital for a diagnosing session, say hospital C, the EHR of X will be downloaded partial or entirely (depending on the purpose of the appointment patient X has in this medical center) from EHR center and translated to the format used in hospital C. This mechanism enables the inspection done by one hospital to be diagnosed by doctor served in another hospital.

Advantages of this sharing mechanism are twofold. First, it is easy to deploy in hospital. Only two applications (TMT Translator and P2P Storage Client) are required to install in each hospital. P2P Storage Client is a universal application that is developed only once. Although each hospital must develop and own a personalized version of TMT Translator, this application may have been developed or under development since Taiwan government is promoting the TMT as standard and requests that all hospitals to follow this standardization process. Second, this sharing mechanism will not interfere with the procedure of taking medical treatment. All EHR translation and exchange are performed in background and on-demand. Moreover, physicians do not needed to learn how to obtain EHR from other hospitals; instead, in their perspective, all the patient's health records can just be read from its own hospital.

As listed in early this subsection, we address the third issue by developing a web portal interface, in order to demonstrate our prototype. This web-based portal integrates all enabled functions provided by computing grid, data grid, and EHR sharing mechanism and able to connect between hospital and Medicare-Grid platform to exchange the EHR data, as shown in Fig. 5.

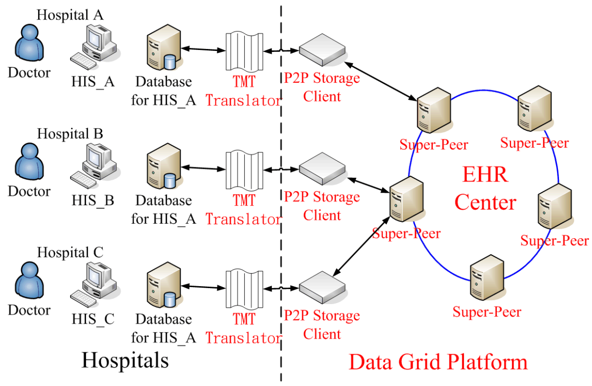

Fig. 4. The EHR sharing mechanism.

Fig. 5. The web portal interface with enabled functions of Medicare-Grid platform.

4 Medicare Applications

In this section, we present medical related applications developed in this project, that include (1) a data warehouse for medical decision support system, (2) a RFID based mobile monitoring system to precisely identify people or items, and (3) a wearable physiological signal measurement system that periodically monitor the health condition of patients. In subsections that follow next, we describe details on the design of these applications.

4.1 Medical Decision Support System

With the rapid growth on the amount of medical data, to find useful information among such a large dataset in an efficient way is desired. The medical decision support system, which is used to help physicians or medical professionals to make better decisions for treatment, is built based on data mining techniques. In this system, we focused on the cardiovascular disease (CVD) and made use of patients' EHR data as the source data. To build an EHR medical decision support system, we collected patients' EHRs on the data grid system and then stored them in a data warehouse, in which data mining techniques are utilized to analyze the EHR information.

There are a number of previous works on mining medical data, such as mining approaches to analyze medical database to find useful patterns, and mining personal health information and providing the results as references for doctors [15]. Data warehousing has been extensively investigated, including data preprocess and data warehouse maintenance over changing information sources [16]. On the technology of data mining, it consists of many approaches such as association rule [17] which discovers hidden and interesting rules in database, clustering which divides a data set into several groups by their characteristics, and decision tree which is used to predict the class of the new input data.

Based on related works, we considered these approaches and then developed a medical decision support system, to analyze the EHR data warehouse and provide useful results. Results obtained relate to the data we collected from some specific hospitals, since they represent the analysis of the heart disease EHR in a particular population, where the results must be confirmed by medicine experts or physicians.

Additionally, we establish a genetic database of cardiovascular diseases. In this system, we integrate 34 cardiovascular diseases and its related gene expression data, SNP, protein-protein interaction, alternative splicing and protein-protein interaction (PPI) information into a web-based interface. Through the analysis of this data model, we obtain significant result of rules for future research and tracking.

The source of cardiovascular disease related gene data is from NCBI OMIM database [18]. OMIM disease data is collected first, and then the text-mining technique is used to generate a dataset of cardiovascular disease. Next, analysis on this dataset is performed to obtain a list of cardiovascular diseases and their related genes. Lastly, we parse these annotations and store them in a database, and then filter out those incorrect results from the dataset, as in Table 1.

We use the list of cardiovascular disease related genes to search the STRING database [19], and get the PPI network graph. Due to the cardiovascular disease is multi-complex disease, these graphs can help to understand the interactions involved in these related genes. The source of alternative splicing data is AVATAR [20], which is a value-added alternative splicing database. Alternative splicing is an important event of gene transcript, and it causes the polymorphism of the gene expression. We link this database and obtain the alternative splicing result of these cardiovascular disease related genes to help us observe the form of the specific gene. The source of SNP data is HAPMAP [21], which provides plentiful SNP information, like the Linkage Disequilibrium (LD) Maps, tagSNPs and the race classification data. We performed analysis on these data and reserved the SNP data that show a high LD value that is related to the cardiovascular disease gene. These SNP data can help us research the relation of a specific SNP in a specific race between the cardiovascular diseases, and these data can also help us design the microarray experiment.

By now, the prototype system has 34 cardiovascular diseases and their genetic data. Each disease has alternative splicing form graph, protein-protein interaction graph and related gene list, and haplotype data. In addition, the number of all CVD related genes is 480 and the number of CVD related tag SNPs is 79621.

Table 1. The list of cardiovascular diseases we provided and their number of related genes.

Disease Name	***Gene Number***
Aortic aneurysm	48
Arrhythmogenic right ventricular cardiomyopathy	22
Arterial thromboembolic disease	13
Ascending aortic disease	28
Atherosclerotic vascular disease	48
Brugada syndrome	6
Cardiac amyloidosis	9
Cardiomyopathy familial restrictive	26
Carney complex	21
Carnitine palmitoyltransferase II deficiency, late-onset form	2
Cerebral amyloid angiopathy	26
Congenital sick sinus syndrome	6
Coronary disease	212
Digeorge syndrome	79
Dilated cardiomyopathy	122
Familial hypercholesterolemia	76
Familial hypertrophic cardiomyopathy	78
Infantile dilated cardiomyopathy	18
Insulin resistance-related hypertension	21
Jervell and Lange-Nielsen syndrome	3
Myocardial infarction	146
Naxos disease	4
Orthostatic hypotension	31
Polymorphic ventricular tachycardia	21
Venous thrombosis	40
Ventricular tachycardia	46
Watson syndrome	156
Williams syndrome	514

4.2 Mobile Intelligence System

The *Mobile Intelligence System* (*MIS*) establishes an active RFID environment comprising various components and approaches for context acquisition of individuals, environment's variables and their associated values. In *MIS*, RFID-based localization, tracking and monitoring techniques were developed for enhancing context acquisition in medical-care environments. Among these functions, localization is the most important component in *MIS* and serves as the key technology for developing mobile intelligence services. The localization system termed as *Real-Time Location System* (*RTLS*), employs active RFID technologies and has three major components as shown in Fig. 6. The monitoring and tracking system can reflect the position of individuals that with active RFID Tag through web interface. In addition, for some areas that are dangerous or private, they can be marked as off limits from the system or restrict the time-duration for stay. For example, 30 minutes is set for a bathroom to avoid accident such as tumble of elder people who is not able to move or unconscious. Once an abnormal event occurred, such as illegal entrance to a limited area or over stay-duration in a specific zone, an alarm can be dispatched and email or short message is sent to the system administrator.

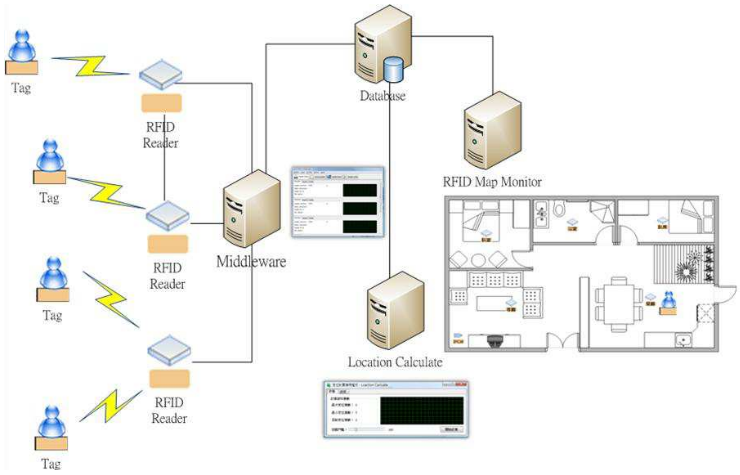

Fig. 6. The Real-Time Location System (RTLS) architecture.

4.3 E-Texcare Health Care System

A number of research topics listed next have been investigated to achieve the e-Texcare health care system requirements for good functionality, portability, comfortable, endurance, and ease of use. The system configuration established and developed is listed as follow:

1. Wearable research (design, fabricate and integrate)
   - Physiological measurement
   - Electronic circuits
2. Physiological measurement research
   - Microprocessor, and micro sensor circuits for physiological measurement
   - Signal processing
3. Mobile and wireless communication research
   - RF and wireless communications
   - Information transmission and reception

A wearable and portable health care system is available for measurements on humans, as shown in Fig. 7. The wearable platform contains sensors that acquire and process vital signals such as ECG, body temperature, etc. Measured bio-signal data are transmitted via Bluetooth technology to the "Mobile Medical Information Processing Module", like PDA or Notebook, for further processing and analysis. All physiological measurement results can be sent to "Remote Mobile Medical Information Processing Module", which are e-health PC workstations in the health care centers through GSM or internet/wireless networks when connection is possible.

The e-Texcare system contains electrocardiogram (ECG/EKG), heart rate (HR), respiration rate, body temperature, and falling detection unit. Therefore, functions as the lethal arrhythmias monitoring, continuous examination of cardiovascular and cardiopulmonary functions, respiration activities, and falling detection are achieved.

Fig. 7. A prototype of e-Texcare® wearable multi-functional physiological measurement system.

5 Conclusion Remarks

In this paper, we integrate grid and peer-to-peer technologies to build up a high-performance computing and storage environment as underlying backbone and proposed an EHR sharing mechanism based on this backbone to form a Medicare-Grid platform. As prototype, we closely collaborated with Taichung Veterans General Hospital, Taiwan who kindly provides us EHR sampling data for experiment purposes. These sampling data are then translated to standard TMT format and stored on Medicare-Grid platform.

In this platform, we have developed three Medicare applications: Medical Decision Support system, which provides analysis of cardiovascular diseases and its genetic data. Each disease has the alternative splicing form graph, protein-protein interaction graph and related gene list and haplotype data; RFID-based localization, tracking and monitoring techniques were developed for enhancing context acquisition in medical-care environments, and finally, the development of e-Texcare health care system and a wearable multi-functional physiological measurement system, to demonstrate that the lethal arrhythmias monitoring, continuous examination of cardiovascular and cardiopulmonary functions, respiration activities, and falling detection are achievable.

Acknowledgment. This paper is based upon work supported by National Science Council (NSC), Taiwan, under grants no. NSC97-3114-E-007-001- and NSC98-2218-E-007-005-.

References

1. Taiwan electronic Medical record Template (TMT), http://emr.doh.gov.tw/old/index.html
2. Foster I., Kesselman C.: Globus: A Metacomputing Infrastructure Toolkit. J. Supercomput. Appl. 11(2), 115--128 (1997)
3. Knowledge Innovation National Grid (KING) Project, http://www.nchc.org.tw/tw/about/publication/king.php
4. National Bioinformatics Applied Grid, http://biogrid.genomics.org.cn/index.jsp
5. NationalGrid, http://www.nationalgrid.com/corporate
6. Czajkowski, K., Fitzgerald, S., Foster, I., Kesselman, C.: Grid Information Services for Distributed Resource Sharing. In: 10th IEEE International Symposium on High Performance Distributed Computing, pp. 181--184. IEEE Press, New York (2001)
7. Foster, I., Kesselman, C.: The Grid: Blueprint for a New Computing Infrastructure. Morgan Kaufmann, San Francisco (1999)
8. Foster I.: The Grid: A New Infrastructure for 21st Century Science. Physics Today, 55(2), 42--47 (2002)
9. Yang C.T., Lai K.C., Shih P.C.: Design and implementation of a workflow-based resource broker with information system on computational grid. J. Supercomput. 1--34 (2009)
10. Yang C.T., Shih P.C., Lin C.F., Chen S.Y.: A resource broker with an efficient network information model on grid environments. J. Supercomput. 40(3), 76--109 (2007)
11. Druschel P., Rowstron A.: PAST: A large-scale, persistent peer-to-peer storage utility. In HotOS VIII, Germany (2001)
12. Kubiatowicz J., Bindel D., Chen Y., Czerwinski S., Eaton P., Geels D., Gummadi R., Rhea S., Weatherspoon H., Weimer W., Wells C., Zhao A.B.: OceanStore: An Architecture for Global-Scale Persistent Storage. In: 9th international Conference on Architectural Support for Programming Languages and Operating Systems (2000)
13. Clarke I., Sandberg O., Wiley B., Hong T.W.: Freenet: A Distributed Anonymous Information Storage and Retrieval System. In: Proceedings of the ICSI Workshop on Design Issues in Anonymity and Unobservability, USA (2000)
14. Taiwan Unigrid, http://www.unigrid.org.tw
15. Lloyd-Williams M.: Case Studies in the Data Mining Approach to Health Information Analysis. In: IEE Colloquium on Knowledge Discovery and Data Mining, vol. 1, pp. 1--4 (1998)
16. Rundensteiner E.A., Koeller A., Zhang X.: Maintaining Data Warehouses over Changing Information Sources. Communications of the ACM, 43(6) (2000)
17. Dai P.T., Chu J.L., Lin S.J., Chang H.B., Yang D.L.: Construction and Analysis of a Data Warehouse System for Customer Relationship Management. In: Proceedings of EC2003, Taiwan (2003)
18. Online Mendelian Inheritance in Man, OMIM. McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University and National Center for Biotechnology Information, National Library of Medicine, http://www.ncbi.nlm.nih.gov/omim/
19. Mering C., Jensen L.J., Kuhn M., Chaffron S., Doerks T., Krüger B., Snel B., Bork P.: STRING 7 -- recent developments in the integration and prediction of protein interactions. Nucleic Acids Res., vol. 35, pp. 358--362 (2007)
20. Chang H.C., Yu P.S., Huang T.W., Hsu F.R., Lin Y.L.: The Application of Alternative Splicing Graphs in Quantitative Analysis of Alternative Splicing Form from EST Database. J. Comput. Appl. Technol., 22(1), 14--22 (2005)
21. The International HapMap Consortium. The International HapMap Project. Nature 426, pp. 789--796 (2003)